Regional Economics

The official recognition of regional economics as a distinct branch of economics prompted a vast amount of theories and models to be developed around the subject. This fully revised second edition provides an authoritative and up-to-date treatment of the evolution of these theories and includes the key theoretical developments of the last ten years.

Regional Economics addresses topics that span from the earliest location theories to the most recent regional growth theories. Complete with the contemporary debate on smart specialization strategies developed by the EU for the design of new cohesion policies, this book provides students with a comprehensive guide to the subject.

Key elements covered in the new edition include:

- proximity and innovation theories;
- the concept of territorial capital;
- the debate on the role of agglomeration economies in urban growth.

This textbook is for undergraduate students in regional and urban economics as well as spatial planning courses.

Roberta Capello is Full Professor of Regional and Urban Economics, Politecnico di Milano, Italy.

19. **Applied Health Economics, Second edition**
    Andrew M. Jones, Nigel Rice, Teresa Bago d'Uva and Silvia Balia

20. **Real Estate Economics**
    A point to point handbook
    Nicholas G. Pirounakis

21. **Finance in Asia**
    Institutions, regulation and policy
    Qiao Liu, Paul Lejot and Douglas Arner

22. **Behavioral Economics, Second edition**
    Edward Cartwright

23. **Understanding Financial Risk Management**
    Angelo Corelli

24. **Empirical Development Economics**
    Måns Söderbom and Francis Teal with Markus Eberhardt, Simon Quinn and Andrew Zeitlin

25. **Strategic Entrepreneurial Finance**
    From value creation to realization
    Darek Klonowski

26. **Computational Economics**
    A concise introduction
    Oscar Afonso and Paulo B. Vasconcelos

27. **Regional Economics, Second edition**
    Roberta Capello

28. **Game Theory and Exercises**
    Gisèle Umbhauer

29. **Innovation and Technology**
    Business and economics approaches
    Nikos Vernardakis

Routledge Advanced Texts in Economics and Finance

1. Financial Econometrics
   Peijie Wang

2. Macroeconomics for Developing Countries, Second edition
   Raghbendra Jha

3. Advanced Mathematical Economics
   Rakesh Vohra

4. Advanced Econometric Theory
   John S. Chipman

5. Understanding Macroeconomic Theory
   John M. Barron, Bradley T. Ewing and Gerald J. Lynch

6. Regional Economics
   Roberta Capello

7. Mathematical Finance
   Core theory, problems and statistical algorithms
   Nikolai Dokuchaev

8. Applied Health Economics
   Andrew M. Jones, Nigel Rice, Teresa Bago d'Uva and Silvia Balia

9. Information Economics
   Urs Birchler and Monika Bütler

10. Financial Econometrics, Second edition
    Peijie Wang

11. Development Finance
    Debates, dogmas and new directions
    Stephen Spratt

12. Culture and Economics
    On values, economics and international business
    Eelke de Jong

13. Modern Public Economics, Second edition
    Raghbendra Jha

14. Introduction to Estimating Economic Models
    Atsushi Maki

15. Advanced Econometric Theory
    John Chipman

16. Behavioral Economics
    Edward Cartwright

17. Essentials of Advanced Macroeconomic Theory
    Ola Olsson

18. Behavioral Economics and Finance
    Michelle Baddeley

Regional Economics

Second edition

Roberta Capello

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

Figures

Tables

Preface to the second edition

Regional economics is the branch of economics that incorporates the dimension 'space' into analysis of the workings of the market. It does so by including space in logical schemes, laws and models that regulate and interpret the formation of prices, demand, productive capacity, levels of output and development, growth rates and the distribution of income in conditions of unequal regional endowments of resources. The path that this discipline has followed since its foundation is remarkable, and the last ten years are no exception. They have registered an interesting evolution in the rich toolbox of theories and models used by the discipline to insert the spatial dimension into the functioning of the market.

Stimulated by the pace of the changes to which the discipline is required to respond, and by colleagues who have kindly asked for an updated version of the textbook, I have decided to work on a second edition of the work published for the first time in Italian in 2004.

At that time, I decided to make an effort to ensure that the textbook was not a collection of the various approaches that exist in regional economics, but rather was organized in a way that expressed the evolution of economic thought in the discipline by concentrating on changes in the concept of space within the various theories: from the traditional location theory to the theory of local development, the latter divided between theories based on a uniform space (and constant returns) and theories based on diversified space (and increasing returns). The latter approaches represent the core of regional economics based on increasing returns and agglomeration economies. Their important feature is that they are able to merge location theories with local development theories. It is the formalization of increasing returns within elegant macroeconomic models that constitute their most advanced element of novelty.

On considering the evolution of the past ten years, I immediately realized that the theoretical advances did not require further parts to be added to the textbook: no theories have been formulated using a new concept of space, and the conception of growth used is still the one that interprets it as competitiveness.

However, especially as regards local and regional development theories, the past decade has seen the development of interesting new theories, which can be synthesized into three main lines of inquiry.

First, the theory of local development has been enriched by a systematization and analysis of the local elements on which regional competitiveness relies. A synthesis concept – labelled 'territorial capital' – has been identified. Its purpose is not the mere classification of the endogenous and exogenous, tangible and intangible, public and private assets highlighted by all theories, but rather to raise awareness of the different economic nature of these elements, which influences their laws of accumulation and depreciation. Moreover, the new reflections engendered by the new concept stress the importance of the co-existence of different elements if a new trajectory of cumulative development is to begin. Thus reiterated in modern terms is the importance of an 'equilibrated development' no longer between tangible elements, but between intangible ones, between knowledge and relational capital, between creativity and cognitive capital. It is the interaction and the synergies among these 'soft' elements that determine increases in the dynamic efficiency of local areas.

The summary of an area's success factors reminds us that local assets have been identified by very different scientific approaches. The traditional approach – defined as functional, positivist and cognitivist – interprets the reality on the bases of a deterministic, cause–effect, logic of individual actions. Another approach has recently been developed. Known as the relational approach, it suggests analysing more complex, inter-subjective relationships based on how economic agents perceive the reality, react to external stimuli, and are capable of synergic and co-operative behaviours. This new approach maintains that local competitiveness arises from trust and a sense of belonging more than pure resource availability; from creativity more than the pure presence of qualified labour; from relationality more than pure accessibility; from local identity more than elements like the quality of the environment and the efficiency of the economic system.

The second direction followed by the new theoretical reflections of the past decade centres on the concept of proximity in the creation of new knowledge. Until ten years ago, scholars were aware that physical proximity should be combined with a relational and social proximity in order to interpret local development patterns. The French proximity school had already addressed the challenge raised by the '*milieu innovateur*' and 'local district' theories, and re-launched the idea of extending geographical proximity to include an 'organized proximity'. However, at the international level (especially in the English-speaking countries) the idea that geographical proximity should be enlarged to other concepts of proximity did not find acceptance. In the past decade, studies on the different types of proximity, from institutional to cognitive, have clearly been taken into consideration by local development theories, and they have obtained broad consensus also thanks to rigorous quantitative empirical analyses supporting the theoretical approaches.

On analysing the current literature, there is an evident dilemma in modern theories in their attempts to supersede pure physical proximity – and therefore move towards an a-spatial paradigm where co-operation among individuals not necessarily located in the same area explains the sources of dynamic advantages of local areas – and the local and geographical anchoring in which the sources of cognitive and institutional proximities are sought.

I believe that the '*milieu innovateur*' theory is still the only theory that finds a way out of this impasse; in fact, it remains the only theory able to merge local advantages – in the form of relational capital and collective learning processes – and the long-distance co-operation able to overcome the decreasing returns that characterize pure local knowledge.

In regard to all these concepts of proximity, no attempt has been made to provide a synthesis of what has been said and what can still be said. Instead, fundamental for understanding the real interpretative capacity of these theories is to consider what relation exists among the different concepts of proximity: do they represent the same phenomenon from a different perspective or are they different elements that complement each other in the explanation of knowledge exchange? Moreover, the question spontaneously arises as to whether it still makes sense to speak about proximity in its traditional meaning: does geographical proximity still play a role in the exchange of knowledge?

This textbook suggests some answers to these questions. In particular, I am convinced that there are complementarities among the different concepts of proximity. As in the balance of development theory, so the balance of different complementary productive resources (transport infrastructure, skilled jobs, productive capital, managerial, entrepreneurial and technological competences) is interpreted as the best strategy for long-run regional development. I am convinced that the presence of, and interaction among, different forms of proximity is the best means to support cumulative innovative processes in the long run.

It is no coincidence that it is precisely in these years that the concept of proximity among actors for the exchange of knowledge has aroused particular interest. As happened for innovation in the 1970s and 1980s, in the past two decades knowledge has been interpreted as the strategic asset upon which regional competitiveness depends; knowledge, and in particular its creation and diffusion mechanisms, has become an increasingly important area of inquiry in regional economics. In particular, the determinants of local learning processes, the role played by space as a source of new knowledge, the effects of proximity as channels for knowledge transmission, are increasingly analysed in an evolutionary approach to regional economics. The latter interprets territory as the source of both uncertainty reduction associated with innovation, and of lock-in mechanisms in specific technological trajectories and innovative paradigms. The recent hermeneutic approach views territory as the source of symbols, emotions and collective identity at the basis of local creativity and knowledge creation.[1]

The third direction followed by theoretical reflections over the past ten years is, from my personal perspective, a rather important one, since it responds to the challenge with which I terminated the first edition of this textbook: the creation of a formalized macroeconomic regional growth model that comprises, together with the traditional macroeconomic elements, modern territorial ones able to interpret regional specificities and regional development trajectories relatively autonomous with respect to those of the nation. Especially these years of crisis, which have reaffirmed with clear evidence the importance of macroeconomic conditions in the explanation of regional growth trajectories – public debt, supranational restrictions on the deficits of single countries, the advantages and disadvantages of a common currency, exchange rates – have raised the challenge of finding a way to merge the two traditionally separated driving forces of growth. In the last chapter of this edition, I present a recent model developed by the Politecnico of Milan that takes up this challenge. The model is labelled MASST, an acronym that contains all the dimensions – macroeconomic, sectoral, social and territorial – that must be considered to interpret regional growth trajectories. The conceptual step forward taken by the model does not reside in a new theory, but rather in the integration of existing theories into a logical framework where macroeconomic and territorial elements find a role. The internal logic of the model, in fact, is an elegant merger of two approaches: the Keynesian theories of effective demand; and the theory of endogenous growth, based on supply elements, as regards the regional differential.

Despite the interesting novelties of the model, there is space for further conceptual reflections. Whilst the MASST model has managed to insert a diversified-relational space into a macroeconomic growth model, still required is a definition of the role of the territorial micro-foundations of growth through study of the location behaviours of firms and individuals. Some young colleague has the task of taking up this challenge.

As regards regional growth models, no particular new idea has come to the fore in the past ten years. The New Economic Geography models of the 1990s and the neoclassical growth models in general have proved to be elegant, and in certain respects useful, restylings of issues already treated. But they are still impenetrable to the evolution of local development theories by which they could instead be inspired to abandon the limiting neoclassical approach of pecuniary externalities, and take technological externalities from the local context into account in their micro-founded macroeconomic growth models.

The revision of a textbook is an opportunity to revise, deepen and rewrite parts of it. This edition is no exception. The neoclassical general spatial equilibrium models have been revised in order to use their mathematical formulations more as an aid to understanding their elegant and refined economic logic than as a barrier to their interpretation. On the basis of remarks and questions by my students over the past ten years, I have revised and rewritten part of Chapter 2, trying to strike the right balance between formal modelling and economic logic.

I have also enlarged on some issues that were too briefly treated in the first edition. Among these issues, I have given more space to the agglomeration economies concept, which has important implications for the normative consequences of regional policy. The presence of cities able to maximize static efficiency, to generate and exploit agglomeration economies, and to grow along cumulative self-reinforcing trajectories is of extreme importance for regional dynamics. The European Union has always been aware of the trade-off between policies to support weak areas (equity principle) and policies able to achieve higher resources to redistribute to weak areas through the exploitation of the efficiency of strong areas (in general large city regions) (efficiency principle). In this version of the textbook, I provide the theoretical tools to reply to such a question.

Also a revision requires, as I have discovered, a huge intellectual effort, which I have managed to face thanks to the co-operation of many colleagues and friends. First of all, my primary source of inspiration has been the regional and urban economics group at the Politecnico of Milan, the scientific school to which I belong. Roberto Camagni – with whom I have shared thirty constructive, enthusiastic and intense years of work – has also on this occasion devoted time to meetings and discussion on the new edition, and read the new parts with interest and dedication. With Roberto, over the past ten years, I have developed scientific ideas and contributions that are included in this edition: for instance, the reflections on dynamic agglomeration economies, the MASST model and regional innovation patterns. Moreover, I have been greatly helped by the young, constructive and dynamic research group that over time has arisen around Roberto and myself: Andrea Caragliu, Ugo Fratesi, Camilla Lenzi and Giovanni Perucca have developed important research fields together with Roberto and myself. They have sometimes had to cope with stressful research schedules, but they have always maintained scientific rigour and enriched common work with creativity, competence and dedication.[2]

The second source of inspiration is the international and national scientific community to which I belong. The international (RSAI), European (ERSA) and national

regional science associations (the Italian one AISRe *in primis*) are a constant source of innovation and inspiration thanks to the presence of exceptional colleagues. After many years, I still have a particular regard for Peter Nijkamp, my mentor together with Roberto Camagni; I am indebted to him for his guidance through the first ten years of my career.

Over the past ten years my scientific activity has been enriched by new experiences. My presidency of the Regional Science Association International enabled me to make a modest contribution to developing regional science in areas, like South America, where it was not yet recognized as a discipline; to meet colleagues all over the world; to discover the richness and variety of problems, issues, themes, methods and, last but not least, persons that exist in every country. My co-ordination of research projects for ESPON and DGRegio obliged me to use theoretical approaches for the solution of practical problems and for the proposal of new policy directions. The editorship of two journals – *Papers in Regional Science* and *Italian Journal of Regional Science* – has provided me with a constant overview on all theoretical and methodological novelties of these years. Last, but not least, the co-ordination of the master's course Management of Built Environment at the Politecnico of Milan has obliged me to pay particular attention to the treatment of issues and themes for academic purposes.

As with the first edition, also in this case I have constantly tried to keep the book's final aim clearly in mind: that of providing students with a tool to understand in depth the economic laws that govern regional growth and development patterns, and the theories with which to forecast these patterns. I hope that I have achieved this goal.

Roberta Capello
Department of Architecture, Built Environment and
Construction Engineering
Politecnico of Milan
14 May 2015

Notes

1 For a hermeneutic approach to knowledge creation, see Cusinato and Philippopoulos-Mihalopoulos (2015).

2 Together with Andrea Caragliu, we developed the reflections on dynamic agglomeration economies. With Andrea Caragliu and Ugo Fratesi we developed the MASST model. Together with Camilla Lenzi, we developed the regional innovation growth patterns. With Giovanni Perucca we developed solid indicators and robust empirical analyses on the concept of territorial capital.

Symbols

An attempt has been made to keep the symbols for variables unchanged throughout all the chapters of this book, and particular effort has been made not to attribute different meanings to the same symbol. However, this has not always been possible when traditional symbols from micro- and macroeconomics are applied: it sometimes happens, in fact, that the same symbol is used with different meanings in the two branches of economics. To avert confusion, there follows a list of the symbols used in the book and their meanings.

τ = Unit transport cost
π = Unit profit/productivity
Π = Total profit
A = Technical progress/intermediate purchases and sales/producers
B = Producer
C = Total costs/consumption/producer
c = Average and marginal propensity to consume/average cost
d = Distance from the centre
D = Demand for a good/cumulated number of adopters
e = Net migration balance/export growth rate
E = Employment
G = Public expenditure
h = Growth rate of human capital
H = Human capital
I = Investments
i = Interest rates
i,j = Sectors or industries
k = Growth rate of physical capital/nesting coefficients
K = Physical capital
l = Growth rate of labour
L = Labour
m = Propensity to import/growth rate of imports
M = Import/raw materials
n = Natural growth rate of population/number of firms/nation or country
p = Unit price of a good
P = Population/prices

q = Size of the house
R = Total revenue/public transfers
r = Land rent/region
s = Average and marginal propensity to save/share of urban land occupied by productive activities/share of patents in disaggregated technological classes (e.g. five-digit classes)
S = Supply of a good/savings/share of patents in aggregated technological classes (e.g. two-digit classes)
t = Time/income tax rate
T = Land/fiscal revenues
u = Utility of a good for a consumer
v = Investment accelerator coefficient
w = Unit wage
x = Quantity of a good
X = Exports
Y = Total income/total production
y = Income growth rate
z = Set of goods/share of employees in a firm

Introduction

1 Economics and space

Economic activity arises, grows and develops in space. Firms, and economic actors in general, choose their locations in the same way as they choose their production factors and their technology. Productive resources are distributed unevenly in space: they are frequently concentrated in specific places (regions or cities) while they are entirely or partly non-existent in others. Quantitative and qualitative imbalances in the geographical distribution of resources and economic activities generate different factor remunerations, different levels of wealth and well-being, and different degrees of control over local development. The problem of factor allocation – which economists have conventionally treated as being the efficient allocation of the factors among various types of production – is more complex than this, in fact; and it is so because the spatial dimension is of crucial importance.

Space influences the workings of an economic system. It is a source of economic advantages (or disadvantages) such as high (or low) endowments of production factors. It also generates geographical advantages, like the easy (or difficult) accessibility of an area and a high (or low) endowment of raw materials. Space is also the source of advantages springing from the cumulative nature of productive processes in space: in particular, spatial proximity generates economies that reduce production costs (e.g. the transportation costs of activities operating in closely concentrated *filières*) and, in more modern terms, transaction costs (e.g. the costs of market transactions due to information gathering). These considerations highlight the need to supersede the purely allocative approach typical of a static interpretation of economic phenomena with a dynamic, indeed evolutionary, approach which ties allocative decisions to processes of development. The geographic distribution of resources and potentials for development is only minimally determined by exogenous factors (raw materials, natural advantages). To a much larger extent, it results from past and recent historical factors: human capital, social fixed capital, the fertility of the land (due to the work of man) and accessibility (measured as the weighted distance from the main centres of production and consumption).

Already evident is an aspect that informs the entire treatment of this book: regional economics is *not* the study of the economy at the level of administrative regions, as is often superficially and erroneously believed. Regional economics is the branch of economics that incorporates the dimension 'space' into analysis of the workings of the market. It does so by including space in logical schemes, laws and models that regulate and interpret the formation of prices, demand, productive capacity, levels of output and development, growth rates and the distribution of income in conditions of unequal regional endowments of resources. Furthermore, regional economics moves from 'space' to 'territory' as the main focus of analysis when local growth models include space as an economic resource and as an independent production factor, a generator of static and dynamic advantages for the firms situated within it – or, in other words, an element of fundamental importance in determining the competitiveness of a local production system.

It may seem somewhat banal to emphasise the importance of space for economic activity. And yet, only recently has it been given due consideration by economic theory. Indeed, in the history of economics, analysts have devoted most of their attention and effort to determining the quantities of resources to be used for various purposes; they have concerned themselves with where those resources and activities are located or where they will be located only in the recent past. Analytical precedence and priority has thus been given to the temporal dimension over the spatial one.

There are several reasons for this belated consideration of space by economists. First, as often pointed out by the founder himself of regional economics, Walter Isard,[1] it has been due to the decisive influence of the neoclassical school, which has conceived the temporal analysis of economic development as crucial and neglected the variable 'space' as a consequence – often in order to simplify the treatment. As Alfred Marshall wrote: 'The difficulties of the problem depend chiefly on variations in the area of space, and the period of time over which the market in question extends; the influence of time being more fundamental than that of space' (*Principles of Economics*, 1920, 8th edition, vol. V, chap. 15, section 1).

Second, the treatment of the variable 'space' in economic analysis – especially if it is included in a dynamic approach – complicates the logical framework. The analytical tools until recently available to economists could not handle temporal and spatial dynamics simultaneously. Nor were they able to cope with the non-linearity of spatial phenomena like agglomeration or proximity economies. Finally, introduction of the variable 'space' required the discarding of the simplifying hypotheses (always dear to economists) of constant returns and perfect competition. According to the logic of a spatial market divided among producers, firms do not compete with all other firms, but only with those closest to them. Spatial distance is thus a barrier to entry which imposes a system of monopolistic competition – which too has only recently been formalized in analytical growth models.[2]

Regional economics therefore seeks to answer the following fundamental questions. What economic logic explains the location choices of firms and households in space? What economic logic explains the configuration of large territorial systems (e.g. city systems)? Why are certain areas – regions, cities, individual territories – more developed than others?

Answers to these questions have been put forward by the two large groups of theories that make up regional economics:

1 *location theory*, the oldest branch of regional economics, first developed in the early 1900s, which deals with the economic mechanisms that distribute activities in space;
2 *regional growth (and development) theory*, which focuses on spatial aspects of economic growth and the territorial distribution of income.

Location theory gives regional economics its scientific-disciplinary identity and constitutes its theoretical-methodological core. It has typically microeconomic foundations and it adopts a traditionally static approach. It deals with the location choices of firms and households. Linked with it are a variety of metaphors, cross-fertilizations and theoretical inputs (from macroeconomics, interregional trade theory, development theory, mathematical ecology, systems theory) which have refined the tools of regional economics and extended its range of inquiry. In microeconomic terms, location theory involves investigation into the location choices of firms and households; but it also involves analysis of disparities in the spatial distribution of activities – inquiry that enables interpretation of territorial disequilibria and hierarchies. Location theory uses the concepts of externalities and agglomeration economies to shed light on such macro-territorial phenomena as disparities in the spatial distribution of activities, thereby laying the territorial bases for dynamic approaches.

Regional growth theory is instead intrinsically macroeconomic. However, it differs from the purely macroeconomic approaches of political economy in its concern with territorial features. Just as we speak of the micro-foundations of macroeconomics, so we may speak of the locational foundations of regional growth theory.

Numerous cross-fertilizations have taken place between these two branches of regional economics, and they have brought the traditional conceptions of space on each side – *physical-metric* for location theory, *uniform-abstract* for regional growth theory – closer together. I call the more recent conception of space *diversified-relational*: this is the bridge and the point of maximum cross-fertilization between the two traditional branches of regional economics. It yields an authentic theory of regional development based on the intrinsic relationalities present in local areas. These three conceptions of space are still today separate, however, and their integration has only been partly accomplished by the more modern notion of *diversified-stylized* space used by recent theories of local growth.

2 Location and physical-metric space

The first and earliest group of theories in regional economics falls under the heading of 'location theory'. This group adopts a purely geographical conception of continuous, *physical-metric* space definable in terms of physical distance and transportation costs. Thus interpreted are the regularities of price and cost variations in space, and their consequences in terms of location choices and the dividing of the market among firms. This was the conception of space used by the great geographers of the first half of the twentieth century.

Location theory seeks to explain the distribution of activities in space, the aim being to identify the factors that influence the location of individual activities, the allocation of different portions of territory among different types of production, the dividing of a spatial market among producers and the functional distribution of activities in space. These various phenomena are analysed by removing any geographical (physical) feature that might explain the territorial concentration of activities,[3] so that location choices are interpreted by considering only the great economic forces that drive location processes: transportation costs, which diffuse activities in space, and agglomeration economies, which instead cause activities to concentrate. By balancing these two opposing forces, these models are able to account for the existence of agglomerations of economic activities even on the hypothesis of perfectly uniform space.

Location models differ according to hypotheses on the spatial structure of demand and supply which reflect the aims that the models pursue.

There are models whose aim is to interpret the *location choices of firms*, on the assumption of punctiform final and raw materials markets with given locations. Choice of location is determined in this case by an endeavour to minimize transportation costs between alternative locations and under the influence of agglomeration economies (theories of minimum-cost location). Here the obligatory reference is to the models developed by Alfred Weber and Melvin Greenhut. There are then models that seek to identify the *market areas* of firms, that is the division of a spatial market among producers. In this case, the models hypothesize a demand evenly distributed across the territory which determines the location choices of firms, these being assumed to be punctiform. Locational equilibrium is determined by a logic of profit maximization whereby each producer controls its own market area (theories of profit-maximizing location); the reference here being to the market area models developed by, for example, August Lösch and Harold Hotelling.

There are then models which seek to identify *production areas*. That is, they seek to identify the economic logic whereby a physical territory (land) is allocated among alternative types of production. In this case, the models are based on assumptions about the structure of demand and supply which are the reverse of those made by theories of market areas. The final market is punctiform in space (the town or city centre), while supply extends across the territory. Activities are organized spatially according to access to the final market, and locational equilibrium arises from a balancing between transportation costs on the one hand, and the costs of acquiring land for a central location on the other. The models developed by Johann Heinrich von Thünen, William Alonso and the 'new urban economics' school express this logic.

Finally, location theory analyses the economic and spatial mechanisms that regulate the size of territorial agglomerations, their functional specialization and their territorial distribution. These models put forward a more complex and general theory of location and the structure of the underlying economic relations able to account for the existence of diverse territorial agglomerations within a framework of general spatial equilibrium. The principal contributions to development of this theory have been made by Walter Christaller and August Lösch.

3 Regional growth and uniform-abstract space

The second large group of theories pertaining to regional economics seek to explain why growth and economic development come about at local level. Why are there rich regions and poor ones; regions which grow more than others, and regions that grow less? What factors determine economic growth at local level? In other words, in this case regional economics analyses the capacity of a subnational system – a region, a province, a city, an area with specific economic features – to develop economic activities, to attract them and to generate the conditions for long-lasting development. Here by 'regional economic development' is meant the ability of a local economic system to find, and constantly to re-create, a specific and appropriate role in the international division of labour through the efficient and creative use of the resources that it possesses. By emphasizing the more economic elements of this definition, regional development can be defined as the ability of a region to produce, with a (comparative or absolute) advantage, the goods and services demanded by the national and international economic system to which it belongs.[4]

The first theories of regional growth were developed midway through the last century. They used a conception of space – as *uniform-abstract*, no longer physical and continuous but abstract and discrete – entirely different from the physical-metric space of location theory. Geographic space was divided into 'regions', areas of limited physical-geographical size (largely matching administrative units) considered to be internally uniform and therefore synthesizable into a vector of aggregate characteristics of a social-economic-demographic nature: 'small countries' in the terminology of international trade but, unlike nations, characterized by marked external openness to the movement of production factors.[5]

The advantage of this conception of space is that it enables the use of macroeconomic models to interpret local growth phenomena. But although these models fit the above-mentioned features, they nevertheless, and it seems inexorably, require the analyst to exclude any mechanism of interregional agglomeration, to discard location theory, to ignore the advantages of local proximity, and instead to assume unequal endowments of resources and production factors, unequal demand conditions and interregional disparities in productive structures as the determinants of local development. Space is thus no more than the physical container of growth and performs a purely passive role in economic growth paths, while some macroeconomic theories reduce regional growth to the simple regional allocation of aggregate national growth.

Theories that take this view of space are *growth theories* developed to explain the trend of a synthetic development indicator – income for instance. Although this approach inevitably entails the loss of qualitative information, its undeniable advantage is that it makes modelling of the growth path possible. These theories differ sharply in their conceptions of growth: there are those that conceive growth as a short-term increase in output and employment, and others that instead identify the growth path in a long-period increase in output associated with higher levels of individual well-being (high wages and per capita incomes, more favourable prices on the interregional market).

This conception of space has been adopted by the neoclassical regional growth theory, the export-base theory, and the interregional trade theory that developed from various branches of mainstream economics in the 1950s and 1960s: macroeconomics, neoclassical economics, development economics and economics of international trade.

4 Local development and diversified-relational space

Interpretation of space as *diversified-relational* has restored to theories of regional development one of the key concepts of location theory, namely agglomeration economies – and made them the core of local development processes. According to this conception, which received its fullest development in the 1970s and 1980s, space generates economic advantages through large-scale mechanisms of synergy and cumulative feedback operating at local level.

A number of seminal theories of the early 1960s for the first time conceived space as diversified-relational. Development was defined, in the words of Perroux, as 'a selective, cumulative process which does not appear everywhere at the same time but becomes manifest at certain points in space with variable intensity'.[6] Perroux's definition affirmed the existence of 'poles' at which development concentrates because of synergic and cumulative forces generated by stable and enduring local input/output relations facilitated by physical proximity. Space is thus conceived as diversified and 'relational'.

But it was during the 1970s that studies on 'bottom-up' processes of development, on districts and local *milieux*, gave the notion of diversified-relational space its most thorough formulation. The conceptual leap consisted in interpreting space as 'territory' or, in economic terms, as a system of localized technological externalities: a set of tangible and intangible factors that, because of proximity and reduced transaction costs, act upon the productivity and innovativeness of firms. Moreover, the territory is conceived as a system of local governance that unites a community, a set of private actors and a set of local institutions. Finally, the territory is a system of economic and social relations constituting the relational or social capital of a particular geographical space.[7]

Any connection with abstract or administrative space is thus obviously discounted. Adopted instead is a more intangible account of space which emphasizes – by focusing on the economic and social relations among actors in a territorial area – more complex phenomena that arise in local economic systems.

Precisely because the diversified-relational space theories of the 1970s and 1980s viewed development as depending decisively on territorial externalities in the form of location and spatial proximity economies, they stressed (for the first time in the history of economic thought) the role of endogenous conditions and factors in local development. These theories adopted a micro-territorial and micro-behavioural approach; they can be called *theories of development* because their purpose was not to explain the aggregate growth rate of income and employment – as in the case of the above-mentioned uniform-abstract space theories – but instead to identify all the tangible and intangible elements of the growth process.

In the theories that conceived space as diversified-relational, location theory was inextricably and interestingly wedded with local development theory. By pointing out that concentration generates locational advantages, which in their turn create development and attract new firms whose presence further boosts the advantages of agglomeration, these theories elegantly revealed the genuinely 'spatial' nature of the development mechanism.

In this sense, diversified-relational space theories form the core of regional economics, the heart of a discipline where maximum cross-fertilization between location theory and development theory permits analysis of regional development as *generative* development: the national growth rate is the sum of the growth rates achieved by individual regions – as opposed to the *competitive* development envisaged by certain uniform-abstract space theories, where regional development is nothing but the simple regional allocation of aggregate national development.

The intriguing objective of these theories is to explain the competitiveness of territorial systems, the local determinants of development, and the capacity of an area to achieve and maintain a role in the international division of labour. They thus seek to identify the local conditions that enable an economic system to achieve and maintain high rates of development.

Figure I.1 summarizes the principles underpinning location theory and regional development theory. The two large theoretical blocks in regional economics – location theory and local growth/development theory – rest on different initial hypotheses: location theory assumes a given factor endowment; local growth/development theory assumes the localization of firms and households. The theories within each group are differentiated by their economic assumptions (transportation costs, agglomeration economies, and the spatial distribution of resources and the productive system) and their conceptions of space (differing spatial structures of demand and supply

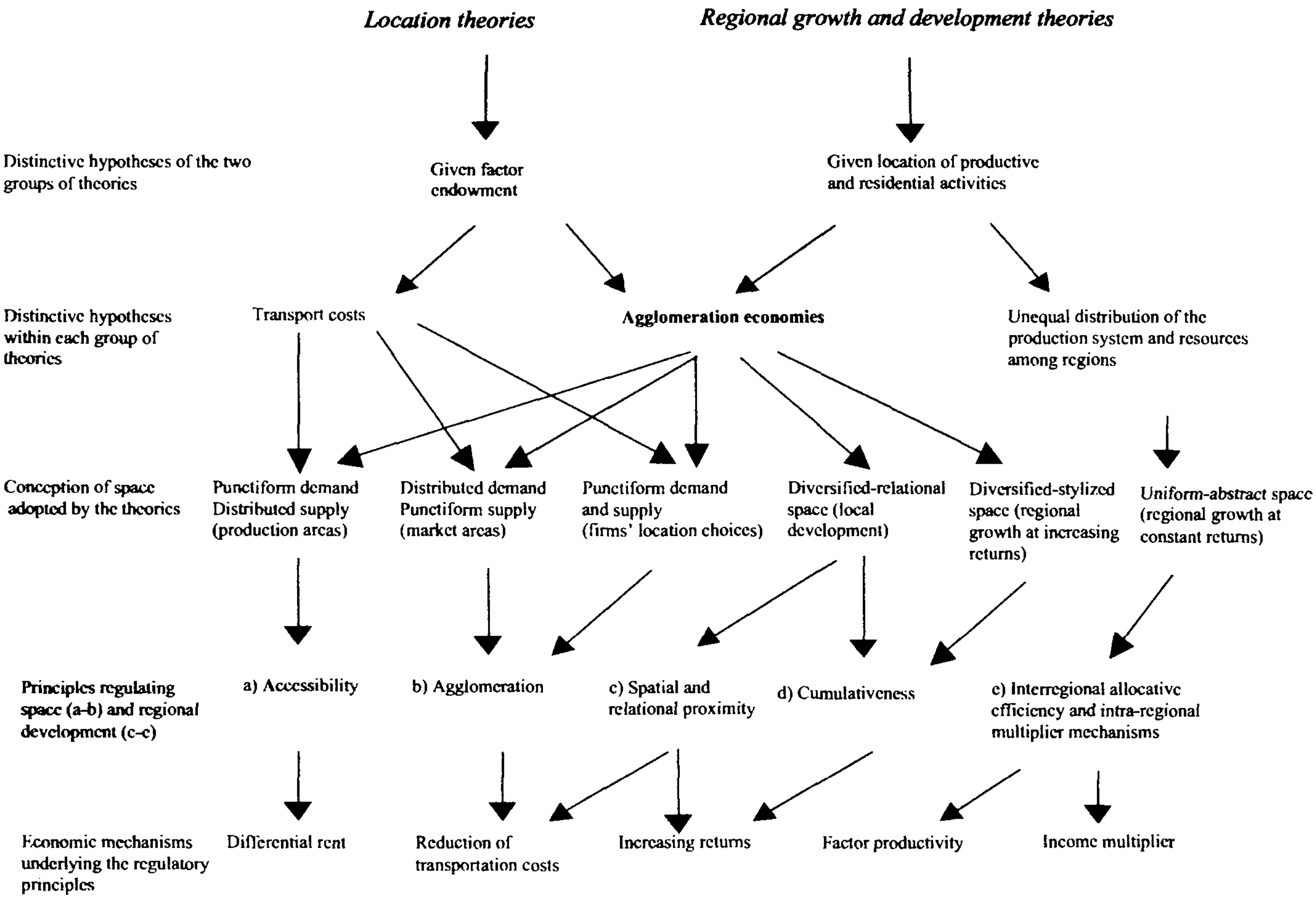


Figure I.1 The principles and hypotheses underlying theories of location and of regional growth and local development for location theory; uniform-abstract, diversified-relational and, as we shall see, diversified-stylized space for local growth/development theory). Thus evidenced by Figure I.1 are the governing principles of space and regional growth/development that buttressed the approaches: agglomeration and accessibility for location theory; interregional allocative efficiency and intraregional multiplier mechanisms, relational proximity, and cumulative growth processes for theories of local development and regional growth. Figure I.1 also highlights the role of agglomeration economies as the hinge between these two broad components of regional economics.

5 Regional growth and diversified-stylized space

Until the end of the 1980s these different conceptions of space developed within regional economics without the slightest convergence between them. In the words of Edwin von Böventer, 'within regional economics one can distinguish between "pure and exact" regional theory without agglomeration economies, on the one hand, and "applied regional theory" which is inexact but takes agglomeration factors into account, on the other hand'.[8] Von Böventer was referring, in the former case, to a rigorously economic and formalized theory of growth, one closer to mainstream economics and envisaging a uniform-abstract space. In the latter case, he had in mind a theory of development without the formal rigour of macroeconomics and predicated on a conception of space where agglomeration economies drive local development.

The 1990s saw the development of more advanced mathematical tools for analysis of the qualitative behaviour of dynamic non-linear systems (bifurcation, catastrophe and chaos theory) together with the advent of formalized economic models that abandoned the hypotheses of constant returns and perfect competition. These advances made it possible to incorporate agglomeration economies – stylized in the form of increasing returns – into elegant models of a strictly macroeconomic nature.

The reference is in particular to the models of 'new economic geography' and endogenous growth in which space becomes diversified-stylized. These theories anchored their logic on the assumption that productive activities concentrate around particular 'poles' of development, so that the level and growth rate of income is diversified even within the same region. Moreover, these models stylized areas as points or abstract dichotomies in which neither physical-geographical features (e.g. morphology, physical size) nor territorial ones (e.g. the local-level system of economic and social relations) play a role.

These theories achieved considerable success and acclaim in the academic community because they showed that territorial phenomena can be analysed using the traditional tools of economic theory (optimizing choices by individual firms and people), and that the various conceptions of space can – apparently – be synthesized. These models in fact conceived growth as an endogenous growth generated by the advantages of the spatial concentration of activities, and by the agglomeration economies typical of diversified space theories. They counterposed dynamic growth mechanisms with increasing returns and transportation costs, thus reprising the economic-locational processes analysed by location theory.

Though diversified (inasmuch as there exist territorial poles of concentrated development), space in these models is stylized into points devoid of any territorial dimension. Thus inevitably abandoned is the concept of space as territory so favoured by regional economists. This stylized space does not comprise localized technological externalities, nor the set of tangible and intangible factors which, thanks to proximity and reduced transaction costs, act upon the productivity and innovative capacity of firms; nor the system of economic and social relations constituting the relational or social capital of a particular geographical area. Yet these are all elements that differentiate among territorial entities on the basis of specifically localized features. As a consequence, these approaches are deprived of the most interesting, and in a certain sense intriguing, interpretation of space as an additional resource for development and as a free-standing production factor. Predominant instead is a straightforward, somewhat banal, view of space as simply the physical/geographical container of development.

6 Towards a theoretical convergence: territorial foundations of macroeconomic growth models

In the final chapter of the first version of this textbook it was concluded that a certain convergence has come about between the large groups of theories discussed. Diversified-relational space theories, in particular those of (endogenous) local development, merge together ideas put forward by the theories of development and of location. Diversified-stylized space theories (in particular new economic geography) amalgamate growth and location theories (Figure I.2). At that time, the impression was that still required was the further step forward that would produce an approach combining the economic laws and mechanisms that explain growth on the one hand, with the territorial features that spring from the intrinsic relationality present at local level on the other. Such an approach would represent the maximum of cross-fertilization among location theory, development theory and macroeconomic growth theories; a synthesis that would bring out the territorial micro-foundations of macroeconomic growth models.

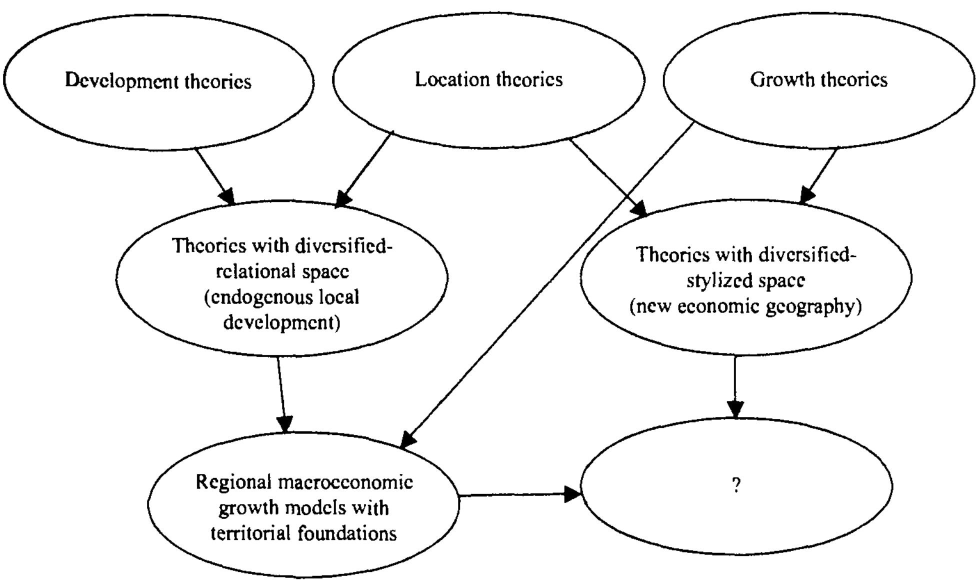


Figure I.2 Convergence among theoretical approaches

This version of the book makes a first attempt in this direction (Figure I.2). The economic crisis of the past five to seven years requires interpretation of growth (and development) through models in which the macroeconomic elements play a prominent role; it is in fact today clear that macroeconomic conditions and trends decisively condition national growth trajectories, and with them inevitably those of the regions belonging to the nation. As the regional growth and development models of the past 60 years have shown, these assets are only in part material. Regional economists are increasingly convinced that the elements – like knowledge and creativity – which drive local competitiveness are of an intangible nature. They are linked to the ways in which actors perceive the reality, to relational elements, and to co-operation attitudes that arise and grow thanks to local socio-economic specificities present in the local context.

Such an approach would represent the maximum of cross-fertilization among location theory, development theory and growth macroeconomics; a synthesis that would bring out the territorial foundations of macroeconomic growth models (Figure I.2). An undertaking of this kind, however, would require analysis of variables besides the cost of transport, which annuls the territory's role in the development process. Also necessary would be variables that give the territory prime place – even in purely economic models – among local growth mechanisms.

I am aware that, with an endeavour of this kind, a step forward has been made in merging macroeconomic growth theories and theories of local development; I am also aware that what is still lacking is an attempt to merge local development theories with theories in which micro-founded location choices are at the basis of regional macroeconomic growth models. This is the challenge that awaits regional economists in the years to come.

7 Theories of convergence and divergence: a distinction by now superseded

Handbooks on 'regional economics' have often drawn a distinction, indeed a dichotomy, between theories of convergence and divergence; that is, between theories that examine the reasons for diminishing disparities between rich and backward regions, and theories that, on the contrary, explain the persistence of those disparities.[9]

Ranged on the convergence side are theories originating within the neoclassical paradigm and that interpret (in their initial formulation) development as a process tending to equilibrium because of market forces. In equilibrium, not only is there an optimum allocation of resources but also an equal distribution of the production factors in space which guarantees, at least tendentially, the same level of development among regions.

On the divergence side stand theories of Keynesian origin which, by introducing positive and negative feedback mechanisms and the cumulative attraction and repulsion of productive resources respectively in a country's rich and poor areas, envisage not only the persistence but also the worsening of disparities among regions.[10]

In recent years, more refined mathematical and modelling tools have demonstrated that the same theories are able to explain both divergence and convergence. By introducing, for example, scale economies and agglomeration economies into a production function – obviously more complex than that of the 1960s model – the neoclassical model successfully simulates a series of behaviours and tendencies, both continuous and 'catastrophic', very distant from the mechanicism and univocity of the convergence predictions of the original neoclassical model. In the same way, the divergence yielded by Keynesian models (à la Myrdal and Kaldor in particular) is called into question if the model's dynamic properties are analysed; according to the parameter values of the dynamic equations describing the model's economic logic, the local system either converges on a constant growth rate or explosively or implosively diverges from it.

It is therefore possible to conclude that there are no longer grounds for any dichotomy to be drawn between theories of convergence and divergence, between optimistic theories and pessimistic ones. However, the problem in and of itself is still very much present, and it is much more complex than was believed in the past. The neoclassical model, elegant in its formulation and consistent in its economic logic, has been frequently criticised as unsuited (in its original formulation) to interpretation of constant and persistent regional disparities. The Keynesian model, in its turn, has been faulted for being unable to foresee territorial limits to the evolution of the cumulative process, although these limits have substantial effects on territorial development paths. But if the 'theories of divergence/convergence' dichotomy is abandoned, the explanatory capacity of each theory can be recovered, to produce a broad array of conceptual tools with which to interpret the complex processes of territorial development. Moreover, I submit, it is much more interesting, as we shall see in the next section, to divide theories according to other and more meaningful features – the definition of space and the goals implicitly pursued by each theory.

8 The elements distinctive of theories: the structure of the book

This book abandons the distinction between theories of convergence and divergence. It instead chooses new elements around which to organize theories of growth and development. These elements throw the interpretative capacities and objectives of theories into sharp relief.

As said, the first element is the conception of space, which enables theories to be grouped according to their approach (micro or macro); the roles performed by space in the development process (passive or active); their interpretative focus (growth or development); and the principles determining development and growth (allocative efficiency, cumulativeness, spatial proximity).

A second element distinguishing among theories is their interpretation of growth. There are theories that associate growth with employment creation, and which have as policy objective the reduction of unemployment in a context of given but largely under-utilized resources. It is thus easy to disregard the problem of endowment, the allocation of resources and factor productivity, and instead take a short-term perspective that envisages *current* competitiveness of production and structure – a condition that can be extrapolated only for a brief period. There are then models and theories that associate growth with increased individual well-being (unitary wages, per capita income) achievable either through higher levels of productivity (and therefore higher levels of wages and per capita income) or through the productive specialization that permits interregional trade and the purchase of goods on the interregional market at prices lower than they would be if the goods were produced internally. Associated with this view of growth are policy problems concerning poverty, underdevelopment and inequalities in the spatial distribution of income. The long-period objective of these approaches is to achieve growth of per capita incomes through higher productivity. Finally, there are models and theories whose policy objective is to identify the determinants of an economic system's real competitiveness and its constancy in time.

Emphasizing the different interpretations given to the concept of growth yields more thorough understanding of each theory's objectives, strengths and weaknesses, and real interpretative capacity. A theory developed with a view to short-period employment is unable to demonstrate the determinants of regional competitiveness (which if anything it presupposes), or the elements that give a region a role in the international division of labour, or the mechanisms that enable the region to maintain that role in the long period. Vice versa, a theory that seeks to define the key factors in long-period regional competitiveness is unlikely to be interested in quantitative changes in income and their effects on individual well-being (if anything, it implicitly associates better well-being with greater development).

The book is structured according to these elements distinctive of the theories examined. Differing conceptions of space account for the division of the book into parts; and differing definitions of growth and development account for its division into chapters.

The first part of the book examines location theory. It is organized into chapters that reflect the various hypotheses put forward on space and the objectives pursued by analysing it (Table I.1). The first chapter sets out theories that envisage punctiform supply

Table I.1 Elements distinctive of the theories examined and structure of the book

Definition of space	*Elements distinctive of the theories examined*	*Structure of the book*
Physical-metric space	**Location theories**	**PART I**
	Definition of firms' location choices and of market areas	Chapter 1
	Definition of production areas	Chapter 2
	Definition of the structure of urban systems	Chapter 3
Uniform-abstract space	**Theories of regional growth at constant returns**	**PART II**
	Preconditions for development	Chapter 4
	Short-term growth of employment and income (with given but largely under-utilized resources)	Chapter 5
	Growth of well-being and per capita income	Chapter 6
Diversified-relational space	**Local development theories**	**PART III**
	Determinants of competitiveness (exogenous microeconomic factors)	Chapter 7
	Determinants of competitiveness (agglomeration economies)	Chapter 8
	Determinants of competitiveness (proximity and innovation)	Chapter 9
Diversified-stylized space	**Theories of regional growth at increasing returns**	**PART IV**
	Determinants of competitiveness (endogenous macroeconomic factors of demand/supply interaction)	Chapter 10
	Determinants of competitiveness (endogenous macroeconomic supply-side factors)	Chapter 11
	Towards a theoretical convergence	Chapter 12

and demand and seek to identify the forces that determine the locational choices of firms. It then deals with theories that assume punctiform supply and spatially distributed demand in order to explain the formation of market areas. The second chapter describes theories that assume spatially distributed supply and punctiform demand to explain the formation of production areas. Finally, the third chapter discusses theories of general spatial equilibrium developed to explain the economic processes that configure large territorial systems, urban systems in particular.

The subsequent parts of the book examine regional growth and development models (Table I.1). Part 2 deals with *uniform-abstract space theories of growth* at constant returns. Belonging to this first group are theories of regional growth that derive directly from various branches of mainstream economics: macroeconomics, neoclassical economics, development economics and international trade economics. For these theories, the engine of development is the unequal distribution among regions of factor endowments and the sectoral structure of supply. This part of the book first examines theories that investigate the preconditions for local development (Chapter 4). It continues with theories concerned with short-period development with given resources (Chapter 5), and it concludes with theories that shift the focus to supply, from the point of view of well-being and full employment (Chapter 6).

Part 3 of the book examines *diversified-relational space theories of development* related to location theory. Microeconomic and micro-territorial in their approach, these theories seek to identify the determinants of local competitiveness. They are sharply distinguished between theories that conceive competitiveness as generated by factors exogenous to the local area (Chapter 7) and ones that, from a more modern perspective, consider endogenous development factors (Chapters 8 and 9). The latter are the most 'spatial' theories of economic development; and they are the first to have furnished an economic interpretation of the 'territory' as comprising co-operation and synergy relations among local actors that influence the productive efficiency and innovative capacity of firms, and therefore the local-level rate of development. Space becomes a source of increasing returns in the form of agglomeration economies; the highest growth rate is registered in local production systems in which increasing returns act on the local productive efficiency, reducing production and transaction costs and increasing the efficiency of production factors. In this logic, both localization theories (stemming from the local district theory) and urbanization economies (in the theory of the optimal city size and all its modern variants) will be analysed (Chapter 8).

Space, or better territory, becomes also sources of uncertainty reduction associated with all innovative processes, and therefore it generates dynamic advantages. In the theories belonging to this stream of thought, dynamic efficiency of firms, defined as firms' efficiency associated with innovative activities, finds its roots in the exchange of knowledge and information facilitated in local areas from the proximity among actors, that over time has lost its pure geographical meaning; relational, cognitive and institutional proximities come to the fore in the explanation of knowledge exchange and in the interpretation of the innovative dynamics of an area (Chapter 9).

Finally, Part 4 of the book discusses *diversified-stylized space theories of growth*. This group comprises the most recent theories, the distinctive feature of which is that they include increasing returns in macroeconomic growth models. Put otherwise: they represent the first attempt to explain local development by combining purely economic and dynamic equilibrium processes with spatial and locational features. The great merit of these theories, in fact, is that they construct elegant economic models

5 Ohlin defines a 'region' as a territory characterized by perfect mobility of production factors. See Ohlin, 1933.
6 See Perroux, 1955, p. 308.
7 See Camagni, 2002.
8 Von Böventer, 1975, p. 3.
9 The notion of 'backwardness' employed by regional economics should not be confused with the underdevelopment analysed by development economics. Although there are points of contact between the two disciplines – indeed, some of the early models of regional economics were decisively influenced by those of economic development theory – there are also important differences. The underdevelopment treated by regional economics is contextualized within a broader economic system (the country as a whole) with an already advanced level of industrialization on which backwardness can count: the 'Objective 1' regions of the European Union, termed such because they have levels of per capita income below the average of European regions, are parts of economically advanced countries with infrastructures, technologies, labour forces and industrial systems typical of the industrialized world. The concern of development economics is instead with the underdevelopment of entire countries, and therefore also with the 'preconditions' for development: industrialization, population support, the creation of basic infrastructures and services for people and firms. Moreover, because regional economics deals with subnational territorial areas, it must disregard certain macroeconomic policy instruments, like the exchange rate or the interest rate, which belong among the public policy instruments available for country-level development.
10 See Isard, 1956; Meyer, 1963.

Part I

Location theory

Physical-metric space

comprising agglomeration economies, in the form of increasing returns, which drive virtuous demand/supply processes (Chapter 10), or counteract the decreasing marginal productivity of the individual factors in a virtuous supply-side circle (Chapter 11). Finally, Chapter 12 conducts a concluding discussion which summarizes the book's analyses, gives a modern interpretation of local development and presents a recent macroeconomic regional growth model that contains the territorial, behavioural and intangible elements of growth. As indicated by the most recent theories, the latter play a role of levers of local development; at the same time, they also have the role of catalysers of exogenous shocks able to diversify at territorial level aggregate territorial processes.

Review questions

1 How do you define regional economics and what are the main topics addressed by this discipline?
2 Would you define regional economics as that part of economics addressing economic problems of administrative regions?
3 What are the theoretical aspects addressed in regional economics?
4 What are the main topics addressed by location theory? How is space conceived within location theory?
5 What are the main topics addressed by regional growth theory? How is space conceived within regional growth theory?
6 What is the difference between regional growth and local development theories?
7 What is the difference in the way space is conceived in the theories of regional growth and local development?

Further reading

Alonso W. (1964), 'Location Theory', in J. Friedmann and W. Alonso (eds), *Regional Development and Planning: a Reader*, MIT Press, Cambridge, Mass., pp. 78–106.
Böventer E. von (1975), 'Regional Growth Theory', *Urban Studies*, vol. 12, no. 1, pp. 1–29.
Hoover E.M. (1948) *The Location of Economic Activity*, McGraw-Hill, New York.
Isard W. (1949), 'The General Theory of Location and Space', *Quarterly Journal of Economics*, vol. 63, no. 4, pp. 476–506.
Isard W. (1956), *Location and Space-Economy*, MIT Press, Cambridge, Mass.
Meyer J.R. (1963), 'Regional Economics: a Survey', *American Economic Review*, vol. 53, no. 1, pp. 19–54.
Richardson H. (1973), *Regional Growth Theory*, Macmillan, London.

Notes

1 See Isard, 1956.
2 See the well-known model of Dixit and Stiglitz, 1977.
3 Geographical (physical) features are removed from models and theories by assuming the existence of a homogeneous plain with equal fertility of land (von Thünen, 1826) or uniform infrastructural endowment (Christaller, 1933; Palander, 1935; Hoover, 1948; Lösch, 1954; Alonso, 1964a).
4 See Camagni, 1999a. This definition has the merit of combining the two concepts of growth and development in a programme of research typical of the current phase of development of regional economics.

1 Agglomeration and location

1.1 Agglomeration economies and transportation costs

Space is inextricably bound up with economic activity. This statement is prompted by the rather banal observation that all forms of production require space. But it also derives from the fact that not all geographical areas afford the same opportunities for production and development. The uneven distribution of raw materials, production factors (capital and labour) and demand (final goods markets) requires firms, and productive activities in general, to select their locations just as they select their production factors and technology. And just as the choice of the factors and technology decisively influences the productive capacity of firms and their position on the market, so location crucially determines the productive capacities of firms and, in aggregate terms, of the geographical areas in which they are located. To ignore this dimension – as traditional economic theory does – is to disregard a factor that sheds significant light on the mechanisms underlying firms' behaviour and economic activities in general, which drive economic development.[1]

The notion of space was first introduced into economic analysis by theories on industrial location. The aim of these theories was to explain location choices by considering the two great economic forces that organize activities in space: transportation costs and agglomeration economies. These forces push the location process in opposite directions since they simultaneously induce both the dispersion and the spatial concentration of production.[2]

It is because of agglomeration economies that spatial concentration comes about. Widely used in regional economics, the term 'agglomeration economies' denotes all economic advantages accruing to firms from concentrated location close to other firms: reduced production costs due to large plant size; the presence of advanced and specialized services; the availability of fixed social capital (e.g. infrastructures); the presence of skilled labour and of managerial expertise, and of a broad and specialized intermediate goods market. All of these are resources whose availability, or production, require a high level of demand.

The advantages that induce firms to opt for concentrated location can be grouped into three broad categories:[3]

1 *economies internal to the firm, also called economies of scale*. These arise from large-scale production processes yielding lower costs per unit of output.[4] In order to reap the advantages of large-scale production, the firm concentrates all its plants in a single location. The advantages in this category derive, not from proximity to other firms, but from the pure concentration of activity in space;

2 *economies external to the firm but internal to the sector, or localization economies*. These spring from location in an area densely populated by firms operating in the same sector. Whereas scale economies depend on the size of the firm (of its plants), localization economies are determined by the size of the sector in a particular area with a wide range of specialized suppliers and in which skilled labour and specific managerial and technical expertise are available;
3 *economies external to the firm and external to the sector, or urbanization economies*. These derive from the high density and variety of productive and residential activities in an area; features which typify urban environments. The advantages in this category accrue from the presence of large-scale fixed social capital (urban and long-distance transport infrastructures, advanced telecommunication systems) and a broad and diversified intermediate and final goods market. These advantages increase with the physical size of the city.

All the above advantages result from the concentration of economic activities in space. However, there are two forces that work in the reverse direction and give rise to dispersed location. The first is the formation in the agglomeration area of increasing costs or diseconomies, these being (i) the prices of less mobile and scarcer factors (land and labour), and (ii) the congestion costs (noise and air pollution, crime, social malaise) distinctive of large agglomerations. These diseconomies are generated above a certain critical threshold.[5] However, the second factor – transportation costs – is of greater interest, because these costs countervail the spatial concentration of activities whatever level of agglomeration has been reached. For in conditions of perfect competition, perfectly mobile production factors, fixed raw materials and demand perfectly distributed across the territory, the existence of transportation costs may erode the advantages of agglomeration until activities are geographically dispersed and the market becomes divided among firms, each of which caters to a local market.

The theory of localization defines 'transportation costs' as all the forms of spatial friction that give greater attractiveness to a location that reduce the distance between two points in space (e.g. production site and the final market; place of residence and the workplace; the raw materials market and the production site). Transportation costs are accordingly the economic cost of shipping goods (the pure cost of transporting and distributing them); the opportunity cost represented by the time taken to cover the distance which could instead be put to other uses; the psychological cost of the journey; the cost and difficulty of communication over distances; the risk of failing to acquire vital information.

Transportation costs are therefore essential to location theory in its entirety, for they differentiate space and enable its treatment in economic terms. They are, moreover, comprised in the concept of agglomeration economies as the costs of interaction and distance: if transportation costs were nil, there would be no reason to concentrate activities, because doing so would not produce 'economies'. In this sense, agglomeration economies are 'proximity economies'; they are, that is to say, advantages that arise from the interaction (often involuntary) among economic agents made possible by the lower amount of spatial friction in concentrated locations.

As a later chapter will show (Chapter 9), agglomeration and proximity form the linkage between location theory and the theory of regional development. Indeed, development theory in the 1970s and 1980s took agglomeration, in the sense of proximity, to be the decisive endogenous factor in cumulative and territorialized processes of economic development.[6]

Two distinct groups of theories on the location of industrial activities can be identified on the basis of objectives that they set themselves, and according to the hypotheses that they assume about the spatial structure of the market:

1. *cost minimization theories*. These hypothesize a punctiform outlet market and a punctiform source of raw materials supply located at different points of space, in order to investigate the location choices of firms at minimum transportation costs. In that they analyse the location choices of individual firms, these theories are based on a partial equilibrium framework;[7]
2. *profit maximization theories*. On the hypothesis that demand is geographically dispersed and supply is concentrated in some points of the market, these theories account for the division of the market among several firms in terms of profit maximization. They assume that the extent of each firm's market and its location depend on consumer behaviour and on the location choices of other firms. These theories are conceived largely within a partial equilibrium framework; an exception is Lösch's model, which envisages a general spatial equilibrium (several firms simultaneously in economic-location equilibrium).

Cost-minimization theories offer answers to questions such as the following: Given the price and location of raw materials and the outlet market, where does the firm locate? How do location choices change when one hypothesizes a place in which agglomeration economies (e.g. the greater availability and higher quality of labour, broader outlet markets) exist? Profit-maximization theories seek to answer questions such as these: Given a certain spatial distribution of demand, how do firms divide up the market? Once the firm's location has been defined, how does it change with variations in the initial production conditions (e.g. variations in production or transportation costs) or in the location choices of other firms?

This chapter sets out the main theories that endeavour to answer these questions. It will begin by showing that when demand and supply are punctiform in space, (i) agglomeration economies (in the form of localization economies) influence firms' location choices even when these are intended to minimize costs (of production and transport) (Section 1.2), and (ii) (in their nature as urbanization economies) they may give rise to location choices which appear illogical if considered solely in terms of costs minimization (Section 1.3). The chapter then shows that, when supply is punctiform and demand is distributed uniformly in space, transportation costs influence the division of the market among firms; and moreover that, in the presence of scale economies or variations in their magnitude, market areas change in their extent (Sections 1.4 and 1.5). Finally, the chapter will explain how location choices depend closely on the choices of other firms, and also on consumer behaviour (Section 1.6).

1.2 Localization economies and transportation costs

1.2.1 Weber's model

One of the first and best-known studies on the spatial concentration of industry dates back to 1909. In that year, the economist Alfred Weber constructed an elegant location model where the costs of transportation among production site, raw materials markets, and the final goods market (which together define a minimum transportation cost) are directly compared against localization economies. The prevalence of one element over another determines the geography of industry location.[8]

Weber's model is based on the following simplifying assumptions:

a) there is a punctiform market for the good (C in Figure 1.1a);
b) two raw materials markets, these too punctiform, are located at a certain distance from each other (M_1 and M_2 in Figure 1.1a);
c) there is perfect competition in the market, i.e. firms are unable to gain monopolistic advantages from their choice of location;

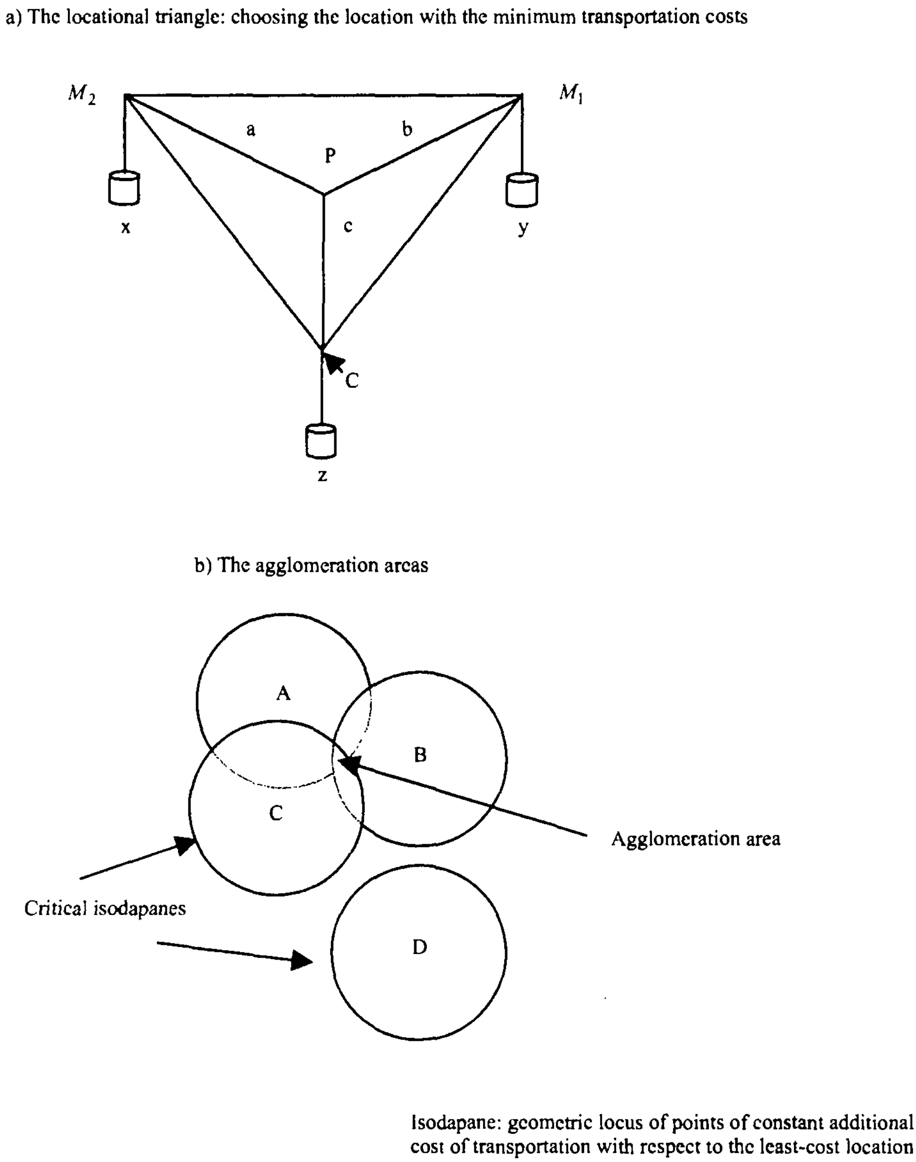


Figure 1.1 Weber's location equilibrium

d) demand for the final good is price-inelastic;[9]
e) the same production technique is used in every possible location. Production costs are therefore given and constant.

The location choice results from a complex calculation performed in two stages. In the first, the firm looks for the location that assures the minimum transportation cost between the production site, the raw materials market and the final market for the good produced. In the second stage, the firm compares the advantages of agglomeration (localization economies) against the higher transportation costs that it would incur by choosing the new location instead of the one with minimum transportation costs.

The first stage of calculation identifies the location that assures minimum transportation costs. Let x and y be the tonnes of raw materials present respectively in markets M_1 and M_2 and required to produce one unit of output, and let z be the tonnes of the finished good to be transported to the final market C. Total transportation costs (CT) are expressed as a function of the weight of the good to be transported and the distance to cover:[10]

$$CT = xa + yb + zc \tag{1.1}$$

where a, b and c are respectively the distances in kilometres between the raw materials markets and the production site, and between the latter and the final market; xa, yb and zc represent the 'forces of attraction' that push the firm respectively towards points M_1, M_2 and C (Figure 1.1a).

The minimum cost location solution can be identified:[11]

- at a point inside the triangle formed by joining M_1, M_2 and C if none of the 'forces of attraction' exceeds the sum of the other two. In economic terms, this situation occurs when the cost of transporting the z tonnes of the good one kilometre further away from the outlet market is less than the costs of transporting the x and y tonnes of raw materials one kilometre further away from their source market;
- at corner C of the triangle, i.e. the final market, if the sum of the costs of transporting the x and y tonnes of raw materials one kilometre further away from their market is less than the cost of transporting the z tonnes of final good produced one extra kilometre. This situation comes about because of the greater relative weight, in the composition of the finished product, of ubiquitous raw materials with respect to those that must be transported. Weber calls this condition 'market-oriented';
- at a point closer to the raw materials markets if the sum of the costs of transporting the x and y tonnes of raw materials one kilometre more is greater than the extra cost of transporting the z tones of the finished good. This situation can be explained by the lesser relative weight, in the composition of the final good, of ubiquitous raw materials with respect to localized raw materials, and/or the product's loss of weight during the manufacturing process. Weber calls this location 'raw-material oriented'.

Weber provides a practical solution to the problem of identifying the minimum point. He hypothesizes a triangular board (the location triangle) in which three holes are drilled at the vertexes M_1, M_2 and C. Threads are passed through these holes (Figure 1.1a) and their ends are knotted together on the upper surface of the board. Weights respectively proportional to x, y and z are attached to the other ends of the threads below the board. The point at which the knot of the three threads lies on the upper surface of the board corresponds to the point of minimum transportation costs.

In the second stage of the location choice process, the firm compares the least-cost location with an alternative one where it can enjoy localization economies – for instance the availability of labour at lower cost and/or better quality.

Assuming that P in Figure 1.1a is the location point with the lowest transportation costs, Weber describes the 'isodapanes': curves along which the additional transportation cost that the firm must pay in order to cover a certain distance from the least-cost location remains constant.[12] On the assumption that other firms operate in the same sector, and that these firms obtain advantages from concentrated location such that they have a pecuniary advantage equal to v, the decision to relocate will be taken if and only if each firm's isodapane measuring an extra transportation cost equal to the agglomerative advantage (v) intersects with the isodapanes of the other firms. In this case, in fact, within the area of intersection the additional transportation costs are less than the advantages generated by concentrated location. In Figure 1.1b, firms A, B and C find themselves in this situation and they relocate. But not so firm D, for which the agglomerative advantage is no greater than the additional transportation cost.[13]

1.2.2 Criticisms of the model

Weber's model has made a permanent and major contribution to industrial location theory. Its principal merit is that it uses entirely rational modes of reasoning; for instance, comparison between the advantages of an alternative location and the additional transportation costs that it would generate. Nevertheless, the model has a number of shortcomings:

- its static nature. The model identifies the least-cost location on the basis of productive efficiency, but it ignores dynamic aspects such as innovation at the microeconomic level, while, at the macroeconomic one, it neglects changes in income distribution and in the relationships among agglomeration advantages, rents and wages;[14]
- its transport-oriented nature. The cost of transportation defines first and foremost the most efficient location; only subsequently does it identify alternative locations. Some critics have claimed that this approach is less efficient than one based on the direct search for a point of minimum total production cost;[15]
- its abstractness, which makes the least-cost location difficult to calculate in real settings. It is rather unlikely, in fact, that the weight of raw materials in the final weight of the good can be calculated, distinguishing *inter alia* the weight of the raw materials to be transported from those present at the production site;[16]
- its nature as a partial equilibrium model which entirely neglects possible interactions among firms;
- its supply-side bias. The criticism most frequently made of the model is that it is excessively oriented to the supply side; it makes no mention of demand factors, assuming that demand is unlimited and inelastic to price variations.

1.3 Market size and transport costs

Weber's model assumes that demand is punctiform and therefore has no physical or economic dimensions. But this is to gainsay the existence of population agglomerations where outlet markets for goods and production factors are larger and of better quality than elsewhere. In other words, Weber's model ignores large urban agglomerations whose existence is due to the advantages ('urbanization economies') that residential and productive activities obtain from habitative density. For people, these advantages are the availability of a wide range of services, from recreational amenities (cinemas, theatres) to transport facilities (international communication hubs, airports and stations). For firms, they are broad and diversified markets for production factors and final goods, a concentration of social fixed capital and the efficient production of public services.

If the existence of final markets of different sizes (and densities) is hypothesized, it can be easily shown that the location choices of industries change with respect to those made both when final markets are punctiform and when their distribution is geographically uniform.

Melvin Greenhut has conducted interesting analyses of how the market's physical size determines the location of industrial firms. On the hypothesis that the distribution of demand is geographically homogeneous, Greenhut finds that firms do not always decide to locate in the region with the least distance between the final market and the raw materials market – as they would do if minimum cost were their only consideration.[17]

In proof of this statement, Greenhut assumes that:

a) there exist two areas, regions *A* and *B*, each of which has a final market and a source of raw materials;
b) region *A* has a larger final market than region *B*;
c) there exist punctiform raw materials markets;
d) firms can operate in one or other of the areas, but they cannot purchase raw materials in one region to sell on the other region's final market; the markets of the two regions are entirely separate;
e) the unit costs of production are the same in the two regions;
f) unit costs of transportation are constant, so that total transportation costs are proportional to distance;
g) the unit costs of transportation are the same in the two regions.

If the distance between the outlet market and the raw materials market is less in region *A* than it is in region *B*, the location choice of a firm in region *A* – which has greater density of productive and residential activity – is conditioned by two factors: (i) the lower transportation costs in region A; (ii) the greater earnings available in that region from higher local demand for the good. Both these factors generate higher profits in region *A*.

If, conversely, the distance between the outlet market and the raw materials market is greater in region *A* than it is in region *B*, the existence of a larger-sized market in *A* explains why firms choose to locate in that region even though the distance between the final market and the raw materials market is greater than it is in region B. The presence of a large market, in fact, may amply offset the higher transportation cost that the firm must pay in *A*, thus yielding profits greater than those obtainable in the alternative location.

1.4 Economies of scale and transportation costs

1.4.1 Market areas

The models discussed thus far account for the existence of industrial agglomerations by weighing localization or urbanization economies against transportation costs. Now examined is a second group of industrial location models which are instead intended to show that the co-existence of economies of scale (these being the first form taken by agglomeration economies and arising solely from the concentration of industry in a point of space) with transportation costs gives rise to a spatial division of the market among firms.[18] It is now necessary to abandon the hypothesis of a punctiform market structure and to assume that demand is uniformly distributed geographically.

How market areas are formed for each firm is demonstrated on the basis of the following assumptions:

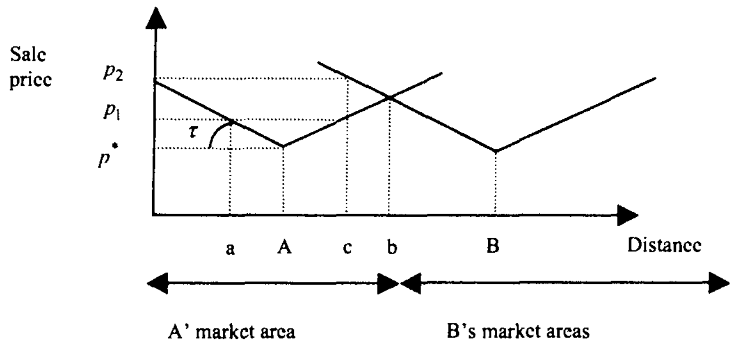
a) demand is distributed uniformly along a linear market and is entirely price-inelastic;[19]
b) two firms offer the same product with identical cost functions (an assumption that, as we shall see, the model makes only initially);
c) the locations of the two firms are given;
d) the cost of transportation per unit of distance (e.g. the cost of transportation per km) is constant, so that the total cost of transportation is proportional only to the distance covered;
e) the cost of transportation is paid by the consumer.

Defining the market areas of the two firms is straightforward. Let *A* and *B* be the two firms located at two points on a linear market (for example a beach or a straight road) (Figure 1.2a). The price at which the firm sells the good on the market is the sum of the good's production price (p^*), and the cost of transportation:

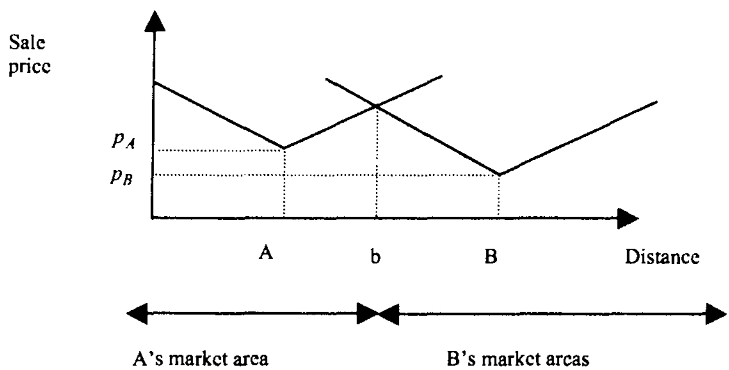
$$p = p^* + \tau d \tag{1.2}$$

where τ denotes the unit cost of transportation per unit of distance, and d the distance covered by the consumer to purchase the good.

The greater the distance from the production site, the more the purchase price of the product increases because of the transportation cost incurred by the consumer in travelling to purchase the product. The distance from the production site obliges consumers located in a to purchase the good at a higher price, equal to p_1 in Figure 1.2a. Attracted by a lower price, the consumers choose to buy the good from the firm located closer to them. In Figure 1.2a, for example, the difference in price between p_1 and p_2 induces consumers located in c to purchase the product from firm *A* rather than from firm *B*. It is obvious that this condition applies to all consumers located between points a and b: throughout this area, firm *A* offers the good at a cheaper price than does firm *B*. The same reasoning holds for consumers located from b onwards: they find it more economical to patronize firm *B*, which offers the same good as *A* at a *a) Supply and transportation costs equal for the two producers*



b) Transportation costs equal for the two producers and economies of scale for producer B



c) Reduced transport costs and economies of scale for producer B

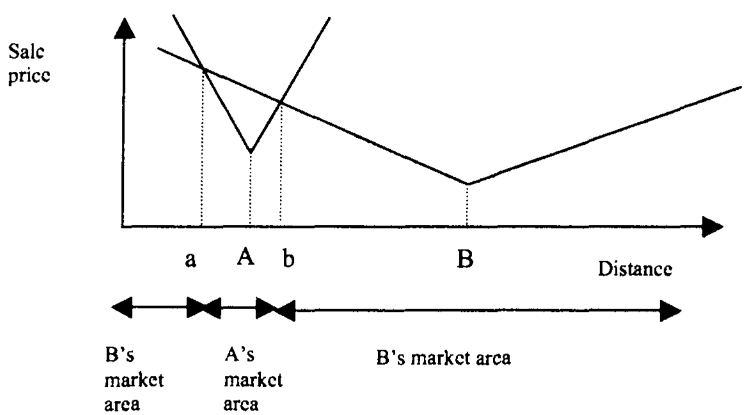


Figure 1.2 The division of the market between producers

d) Economies of scale for producer B such to force producer A out of the market

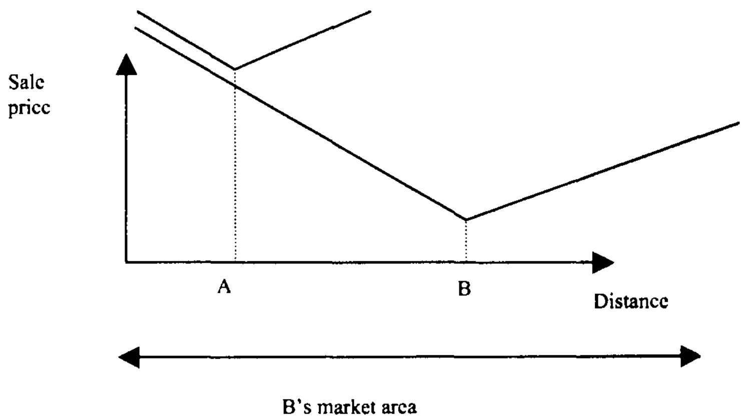


Figure 1.2 (Continued)

lower price. Point *b* represents the threshold between the two market areas; the point, that is, where the consumer is indifferent between purchasing from firm *A* or firm *B* because both charge the same price for the good.

The model just described rests on the hypothesis that the cost functions of the two firms are the same, and that transportation costs are paid by the consumers. But what happens if we assume that the firms benefit from economies of scale and that they pay the transportation costs?

If one of the two firms (in this case firm *B*) enjoys economies of scale (i.e. its production cost is lower than that of firm *A*), the threshold separating the two firms' markets, represented by point *b*, shifts in position so that it marks out a larger market area for firm *B* (Figure 1.2b). Firm *A* is able to stay in the market because of the distance that separates it from *B*.

In Figure 1.2c, firm *B* enjoys both economies of scale ($p_B{}^* < p_A{}^*$) and lower transportation costs ($\tau_B < \tau_A$) (more efficient transport and packaging technologies), and it takes over large part of *A*'s market. *A*'s control is further reduced and now covers a small area (*a–b*) adjoining its production site. Interestingly, owing to reduced spatial friction (expressed by low transportation costs), *B* even deprives *A* of market areas which were previously its undisputed monopoly. The extreme case is that in which *B*'s advantage in terms of economies of scale is so overwhelming that firm *A* is forced out of the market (Figure 1.2d).

Although simple, this model yields results of considerable interest:

a) consumers located closer to the production site obtain an economic advantage in terms of lower transportation costs (when these are assumed to be borne by consumers) and therefore pay a lower overall price for the good, provided the firm does not engage in price discrimination;
b) the firm may discriminate on price within its market area, where it operates a monopoly, without losing market shares. Numerous methods of price discrimination can be used if the cost of transportation is paid by the firm and not directly by the consumer. The firm may impose the same sale price – equal to

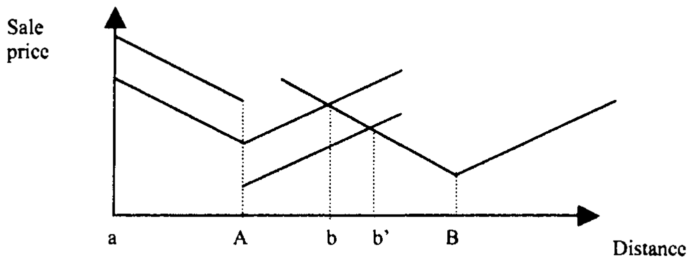


Figure 1.3 Effects of price discrimination on the market areas

the price at which the good is sold to the remotest consumer in its area – on all consumers, and thus appropriate all the surplus earned from those located closest to it; or it may discriminate among groups of consumers by charging higher prices to consumers located in one area (*a–A* in Figure 1.3) than to those in another area (*A–b*), thus extending its market area from *b* to *b*';[20]

c) it follows from points (a) and (b) that physical distance is a barrier to entry into local markets: a firm does not compete with all the others, but with those located closest to it. Consequently, the model of competition among firms in a spatial market is the model of monopolistic competition à la Chamberlin and Lancaster[21] where price discrimination is based, not on product differentiation as in traditional microeconomic models, but on the distance that separates purchaser from producer.

1.4.2 *A subsequent refinement: Hoover's theory*

The model just described was subsequently refined by Edgar Hoover, who found a simple way to endogenize economies of scale in the model and have them depend indirectly on distance. He did so as follows.

Hoover's assumptions are similar to those of the previous model. Demand for a good is uniformly distributed along a linear market. Two firms, *A* and *B*, are located at two extremes of the market and they produce a homogeneous good. Unlike in the previous model, however, transportation costs are paid by the firm, and production by both firms is characterized by economies of scale until a certain level of output has been reached. Beyond this level, the economies of scale turn into diseconomies which – as is usual in neoclassical microeconomic models[22] – push up average production costs if the quantity of output increases.

In order to sell their goods at a distance from the production site, firms must pay a transportation cost proportional to the distance over which the goods must be shipped. A decision by firm *A* to increase its market changes its production costs with respect to the initial level *a*: on the one hand, in fact, by extending its market, the firm obtains economies of scale and produces at a lower cost per unit of output, equal to *b*; on the other, the distance between the new market areas and the production site requires the firm to pay the transportation costs represented by the line *bb*' (Figure 1.4a). At distance *D*, the cost of distribution (or sale price) – which comprises both the new production cost and the transportation cost – is equal to *E*.

If this reasoning is applied to a variety of distances from the production site, lower levels of production cost are obtained as combinations of the different sizes of the markets being served at different distances. The distance/cost relation is negative as long as economies of scale operate; but once the point of most efficient production (the lowest least average cost point) has been reached, the economies of scale change – according to the model's hypothesis – into diseconomies, and the distance/cost relation becomes positive. In fact, an increase in production comes about at production costs higher than previously because of diseconomies of scale and transportation costs. For distance *F* in Figure 1.4a, the sale price is now *E*". By combining the various

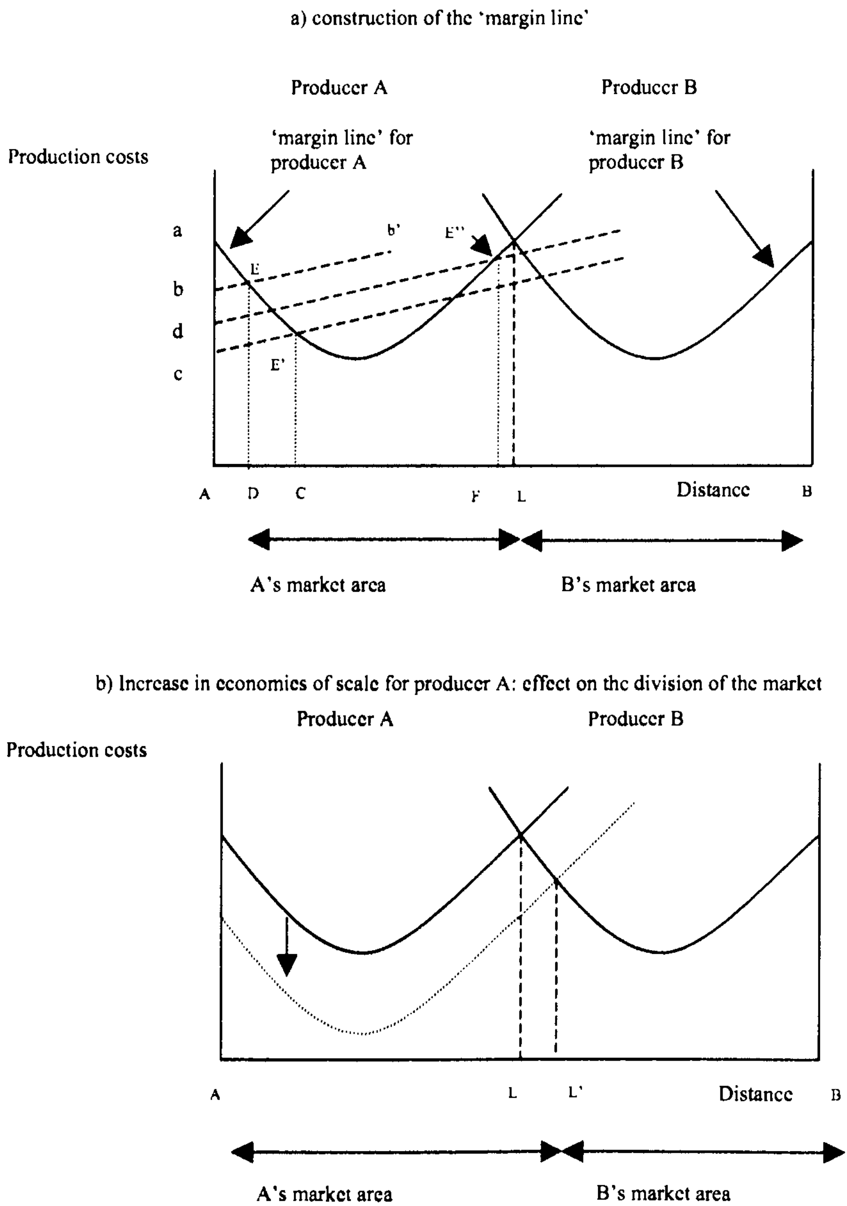


Figure 1.4 Division of the market in the presence of economies of scale

distribution costs thus obtained, a curve, U-shaped with respect to distance, can be constructed. This curve Hoover calls the 'margin line', and it represents an average production cost given by the sum of production and transportation costs.

The same procedure can be used to construct a 'margin line' for firm *B*. The intersection between the two 'margin lines' obviously represents the demarcation line between the two firms' markets: *A* serves the market extending from its production site to *L*; *B* controls the market extending from *L* to its production site.

It is possible to determine in the case of Hoover's model, too, what happens to the division of the market between the two firms if one of them (firm *A*) manages to achieve economies of scale – for instance by introducing a technological innovation: distance remaining equal, *A*'s production costs will be lower, and the entire margin line will shift downwards. The final outcome will be that the boundary between the two firms' market areas moves to *L*', and therefore in favour of firm *A* (Figure 1.4b).[23]

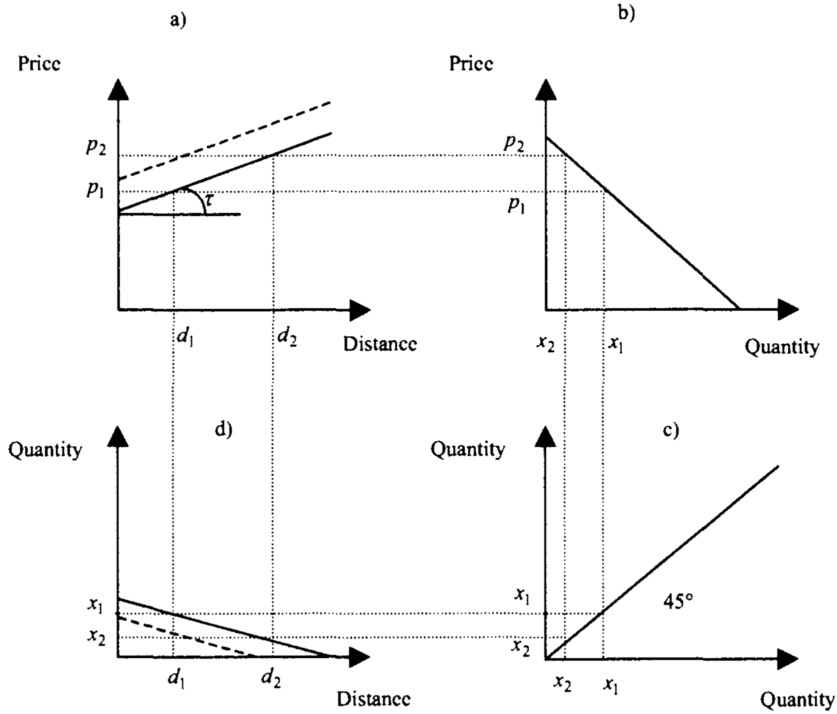
1.5 Spatial demand, market equilibrium and firm location

The next step is to define the quantities of the good offered by the firm that consumers are willing to purchase when they have to cover different distances to do so – given a certain cost of producing the good and a certain transportation cost (again assumed proportional to the distance). Necessary for this purpose is construction of an *individual spatial demand curve* which shows the different quantities of the good x that individuals are willing to purchase from firm i according to their distance from the firm and according to the production price (or factory price) set by the firm on the basis of production conditions (costs profile, economies of scale). Once the individual spatial demand curve has been plotted, the *market's* spatial demand curve can be calculated as the simple sum of the individual quantities demanded at the various distances from the firm by the n consumers existing in the market. When analysed together with the usual firm's supply curve of microeconomic theory, the market's spatial demand curve defines the market equilibrium.[24]

The spatial demand curve constructed by August Lösch is shown in Figure 1.5, which consists of four graphs. Graph a) represents the price/distance relation: a straight line, the slope of which depends on the value of the unit cost of transportation (τ), as in Figure 1.2. Graph b) shows the individual demand curve of traditional microeconomics, which evinces a negative relation between price and quantity: as the price of the good increases, the quantity that the consumer is willing to buy diminishes. Graph c) has the simple function of mapping the variable on the axes. Finally, graph d) plots the individual spatial demand curve.

At a distance of d_1, the firm offers good x to the consumer at price p_1. At this price, the consumer is interested in purchasing quantity x_1 of the good, as shown by the individual demand curve. Thanks to the transposition of the axes in the third graph, the quantity can be easily transferred to the fourth one. The same reasoning applies to other distances: moving clockwise, identified at d_2 is the quantity x_2 demanded by the consumer. Uniting the various quantity/distance combinations in the fourth graph produces the individual spatial demand curve.[25]

On the assumption that all consumers have identical individual spatial demand curves, total demand for the firm's good at every distance will be the sum of the individual quantities demanded at the various distances by the n consumers existing in the market. Assuming a uniform density of consumers for each unit of distance, q, the



a) price/distance relation
d) individual spatial demand curve

b) individual demand curve
c) transposition

Figure 1.5 Construction of the individual spatial demand curve

a) in the case of the linear market, spatial market demand is represented by the area below the individual demand curve multiplied by population density q.

b) in the case of the circular market, spatial market demand is represented by the volume of the Lösch demand cone multiplied by population density q.

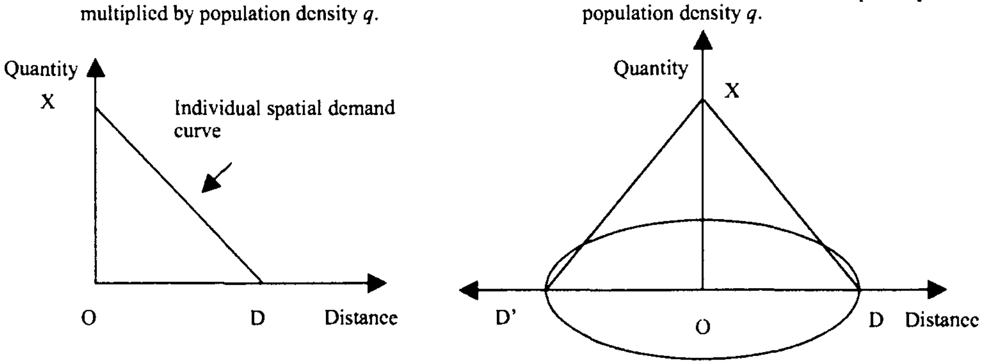


Figure 1.6 Spatial market demand

total demanded quantity of the good will be equal to the area below the individual demand curve multiplied by the density q (area ODX in Figure 1.6a).

Assuming a homogeneous plane, and no longer a simple linear market, Lösch uses the same procedure to identify the firm's market areas. Rotating the triangle formed by the individual demand curve 360 degrees around the vertical axis plots a circular market. Multiplying the volume of the cone thus obtained by the density q yields the total quantity of the good demanded in a circular market (Figure 1.6b).[26]

Interestingly, the size of the market area delineated by Lösch's 'demand cone' depends – given a certain structure of demand – on transportation costs and on the conditions under which the good is offered. An increase in the transportation costs inclines the individual spatial demand curve and restricts the firm's market area (Figure 1.7a). A higher sale price reduces the quantity of the good demanded by the consumer, distance remaining equal, and the individual spatial curve consequently undergoes a parallel shift downwards that reduces the firm's market area (Figure 1.7b).[27]

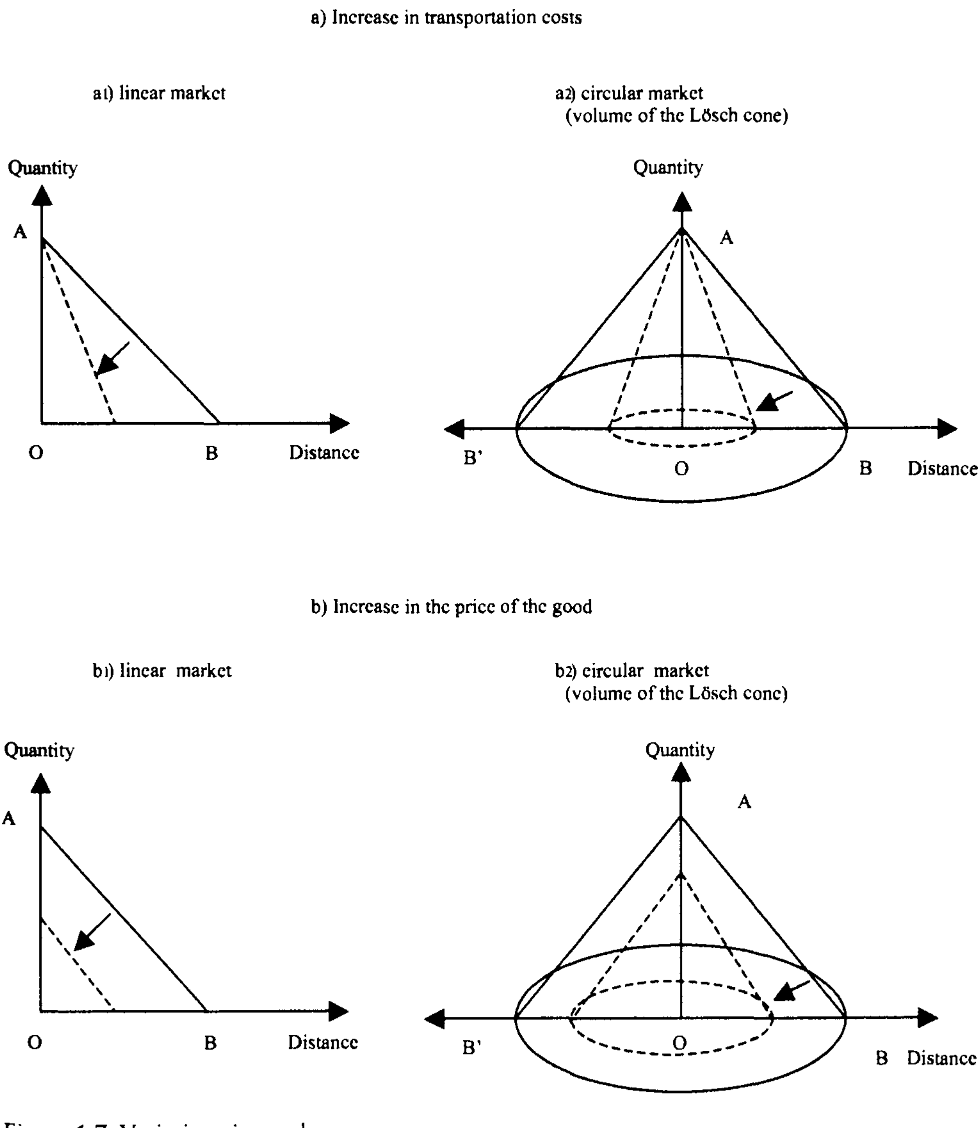


Figure 1.7 Variations in market areas

Having defined the demand curve, Lösch now describes the firm's economic-spatial equilibrium[28] and the firm's location. The firm acts as a monopolist in its market area, which is protected by distance (externally to it demand for the good falls to zero owing to the excessive transportation cost) and produces in conditions of profit maximization and surplus profits.[29] The market, at the spatial level, is made up of numerous non-overlapping market areas with broad spaces in which there is unsatisfied demand (Figure 1.8a).

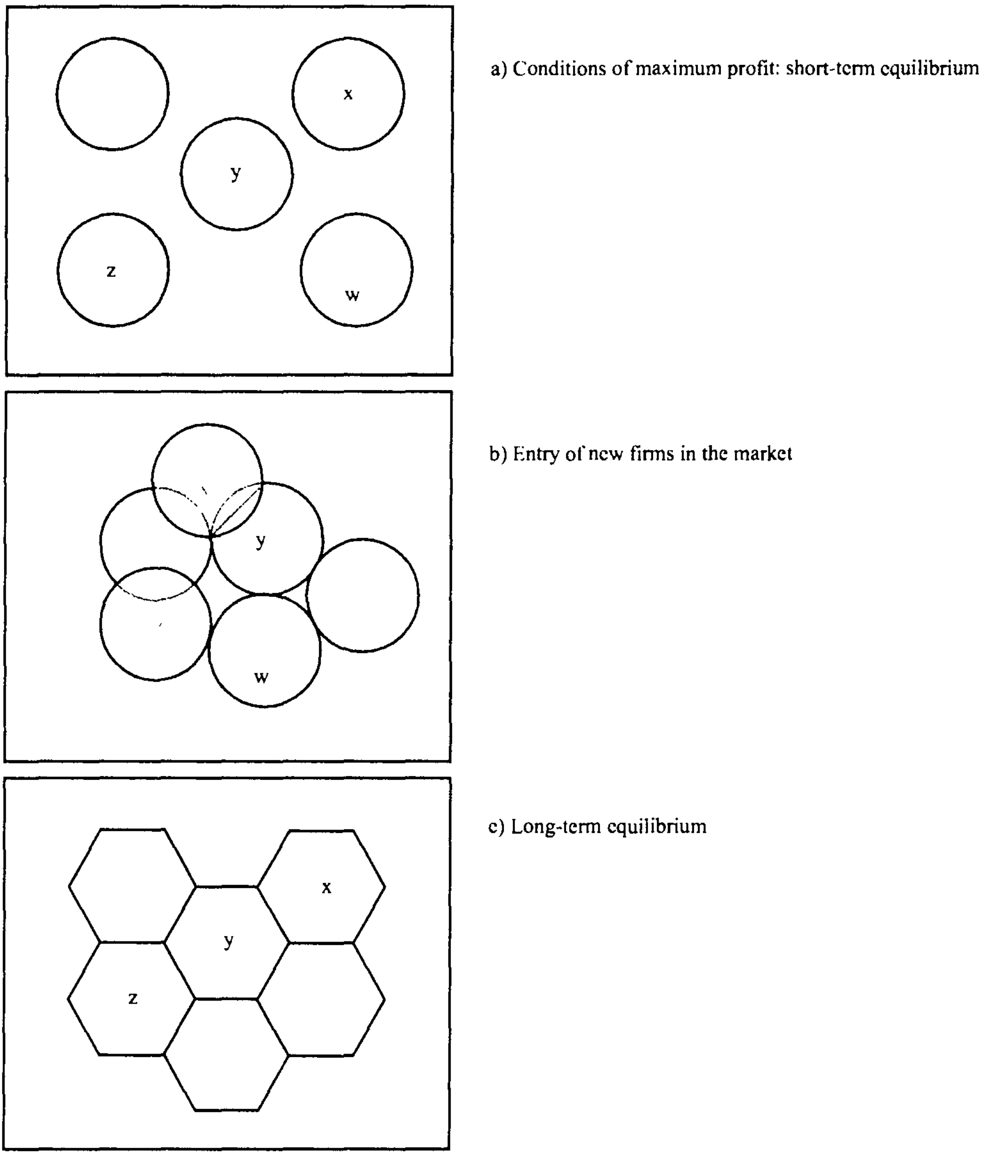


Figure 1.8 Evolution towards a long-term spatial market equilibrium

However, this is a short-term equilibrium. The existence of surplus profits from production of the good and of still unexploited market areas induces new firms to enter the market and to locate in the areas not yet covered by supply. The market entry by new firms has two joint effects: (i) the spatial market is occupied until areas come to overlap; (ii) the profit margins of individual firms are eroded by a decrease in demand, which is now divided among several firms, and by an increase in costs due to greater demand for production factors and intermediate goods with which to produce additional quantities of the good. The situation soon arises in which firms are no longer interested in entering the market because the surplus profits have been absorbed by the increasing production costs. The long-term equilibrium of the market has thus been reached.

Faced with overlapping market areas, and if the good is homogeneous, consumers choose to purchase the good offered at the lower price, which is the price determined by the shortest distance between buyer and producer (Figure 1.8b). The result of the process is a long-term market equilibrium in which a spatial market assumes a pattern of regular hexagons with no overlapping areas (Figure 1.8c).

1.6 Interdependency in location choices: the Hotelling model

The models described thus far assume that – given the locations of producers, and given demand uniformly distributed geographically (in linear or circular form) – the market is divided into areas within each of which there operates a single firm. None of these models hypothesizes that, once the market has been divided up, the firms will consider the possibility of relocating. Nor do they consider the existence of interdependency mechanisms operating in the location choices of firms, and which, as we shall see, give greater density to the spatial distribution of activities.

The earliest theory of interdependence among location choices was set out in the well-known duopoly model developed by Harold Hotelling – although various other authors have contributed to the theory.[30]

The model's assumptions are very similar to those of the market areas model:

a) the existence of only two firms (duopoly);
b) a linear market (a beach, for example) homogeneously distributed along which is demand for the good produced, which is also homogeneous (the same brand of ice cream, for example);
c) nil costs of relocation;
d) demand entirely inelastic to price: that is, a quantity of the good demanded by the consumer which does not change with variations in price (tourists on the beach purchase the same quantity of ice cream regardless of the price at which it is sold).[31]

Assuming that the firms are initially located at *A* and *B* (Figure 1.9a), if one of them – for example *A* – relocates to *A*', the division of the market will obviously change in favour of *A*, which acquires a new market share by appropriating it from firm *B*. It is likewise in the interest of firm *B* to relocate, for example to *B*', because by so doing it can take over a portion of *A*'s market. The process continues until the firms are located at the centre of the spatial market, each sharing one half of it (Figure 1.9b). Only this arrangement gives stability to a situation which otherwise, according to Hotelling, would be indeterminate under these hypotheses.[32]

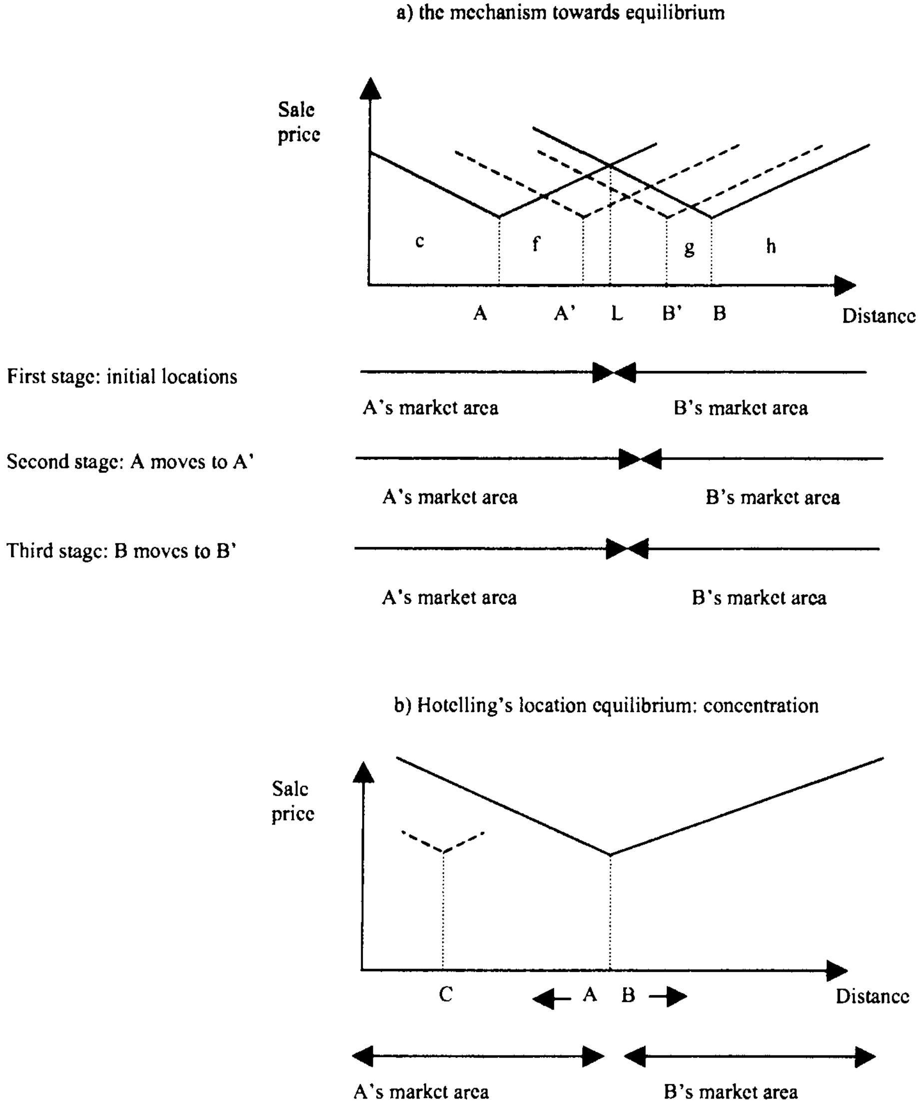


Figure 1.9 The Hotelling duopoly

Hotelling's model highlights the following two important considerations:

a) even in the presence of transportation costs, there is a natural tendency for firms to cluster in space; a tendency which may help explain the existence of large agglomerations, most notably cities;
b) the competitive solution obtained by market forces does not coincide with the public interest: the average distance that consumers must cover to purchase the good once the firms have achieved location equilibrium (Figure 1.9b) is – assuming an initial location pattern like that in Figure 1.9a – greater than the

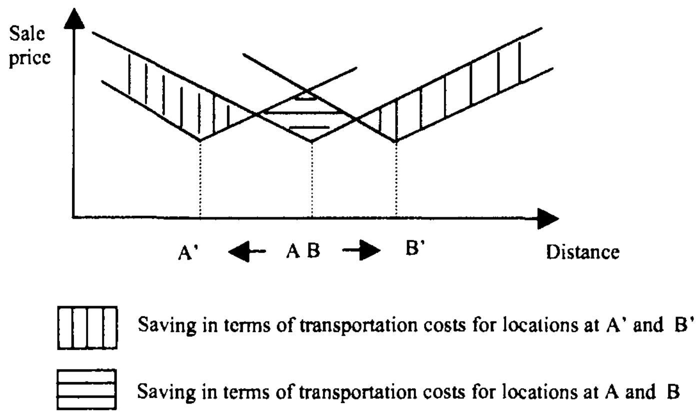


Figure 1.10 Advantages in terms of transportation costs for alternative locations

distance that they previously had to travel. This has long provided theoretical justification for planning actions intended to influence location choices. However, this is not to imply that private initiatives never coincide with the public interest; rather that the coincidence between private and public interest should be proved and not taken for granted.

Criticisms of the Hotelling model have centred on the close dependence of its results on its initial hypotheses.[33] First, if a co-operative solution is admitted, the two firms may agree not to change their initial locations (which generate market shares exactly equal to the final ones) and thus avoid the costs of relocating. Second, if the possibility of new market entrants is admitted, the spatial concentration disappears; for it is in the new firms' interest to avoid a central location and instead exploit more peripheral ones (for example at C in Figure 1.9b). From these locations they are able to take over portions of A's market, upsetting the location equilibrium as a consequence. Third, if the hypothesis of demand curve rigidity is discarded, the result of a central location is once again doubtful. In fact, if the price influenced sales (i.e. on the assumption that demand is elastic to price), the firms would select locations closer to the final consumer in order to minimize the transportation cost borne by the latter (which is reflected in the good's final price) and thus maximize their revenues. A shift by both firms from a central location to a more peripheral one would yield greater revenues. In fact, at the new locations A' and B' in Figure 1.10, consumers would obtain a greater saving in terms of transportation costs (depicted by the area cross-hatched with vertical lines in Figure 1.10) than they would obtain from a central location (depicted by the area cross-hatched with horizontal lines). Their demand for the good would consequently increase.

1.7 Critical remarks

As is frequently the case, models developed to define the market areas are characterized both by substantial interpretative capacity and by weaknesses due to abstraction and to the hypotheses necessary to reduce the complexity of the real world.

The great merit of these models is that they interpret the location choices of firms solely in the light of the large-scale forces that drive location processes: transportation costs on the one hand, agglomeration economies on the other. By balancing these two opposing forces, these models are able to explain the existence of agglomerations of firms on the hypothesis of a perfectly homogeneous space. In other words, they do not resort to geographic factors, which furnish an excessively banal explanation of the concentration of economic activities in space.

A second salient feature of these models is their ability to incorporate spatially extended demand into the location choices of firms. The existence of this spatial market obliges firms to take location choices which extend well beyond the logic of minimizing transportation costs between production sites and distant points of sale, and which as a consequence are more realistically oriented to controlling the market.

These models have the further merit of conceiving – within a framework of location choices – interdependency among the behaviours of different firms. A location choice does indeed depend on the size of the market to which it gives access; but it is not taken regardless of the choices of other firms. This feature turns analysis of location equilibrium into an iterative game whose solution depends on the game's hypotheses themselves.[34]

Another strength of these models is their demonstration of the real role of distance in economic analysis: distance is a barrier to market entry which enables each individual firm to exercise a monopoly within its market area.

However, although these positive features give significant interest to these models, a number of weaknesses intrinsic to them should not be overlooked. Mention has already been made, in regard to Hotelling's duopoly, of the strong influence exerted by the initial hypotheses on the final result. The most influential of these hypotheses is the price-rigidity of demand. Once it has been discarded, a series of alternative location equilibria arise. More peripheral with respect to the centre of the market, from the point of view of sales maximization these equilibria yield competitiveness gains in market areas more distant from the centre and where prices are higher.

As said in Section 1.3, if the hypothesis of a homogeneous spatial market is discarded, it may be possible to explain location choices that are counter-intuitive if viewed solely in terms of transportation costs. Because of these weaknesses, it has been pointed out that although these models assume the existence of demand distributed in space – and although they thus have sales depend on transportation costs – they still inadequately account for the effects of demand on the final equilibrium.[35]

A final consideration concerns the low incidence of transportation costs in the final value of the goods typically produced by present-day industry. This incidence has been estimated at only 3–8 per cent, which suggests that these models have a rather limited capacity to interpret reality: it is difficult to argue, in fact, that geographical agglomerations exist owing to the centre of gravity of certain industries. These models seem more realistically applicable to tertiary activities. Because of the low unit value of many services, for example commercial ones – which increases the incidence of transportation costs (we always choose the nearest baker's shop to buy our bread) – and the importance of face-to-face relations in many advanced services (law, accountancy, health care), in the case of services the costs of transportation and closeness to the firm of the service significantly influence consumer choices.[36]

1.8 Conclusions

This chapter has surveyed the earliest theories of location developed to explain the determinants of location choices by industrial firms.

The oldest of the models discussed – Weber's model – assumes demand and supply structures that are punctiform in space. On this hypothesis, the model elegantly and convincingly explains the existence of territorial agglomerations on the basis of two great economic forces that induce either the concentration or the dispersion of activities in space: agglomeration economies on the one hand, transportation costs on the other. Still today, these forces are components of more modern, and in certain respects more complex, models that seek to conjugate location choices with local growth dynamics (see Chapter 9), and it is on the balancing of them that the geographical organization itself of activities depends.

The chapter has also described models that envisage a punctiform source of demand and a supply uniformly distributed in space. Starting from these assumptions on the spatial structure of demand and supply, these models simply but convincingly demonstrate the importance of distance in determining the behaviour of consumers and firms in the market. Distance is a spatial friction – a cost associated with movement in space – which in economic terms raises a barrier to entry in the local market, for it protects the firm against competition and permits monopolistic behaviour, like price discrimination, within the local area.

Still to be examined are location models based on the hypothesis of a punctiform source of demand and a supply distributed in space. This is the topic of the next chapter.

Review questions

1. What are the economic forces explaining the location of activities in space?
2. What are the different categories of agglomeration economies?
3. What is the definition of transport costs in regional economics?
4. Why do transport costs play an important role in location theory?
5. In Weber's model, what are the elements on which location choice is based? How is the location point achieved by the firm in Weber's model?
6. If final markets with different size are conceived, how and why does the location choice of a firm change?
7. How are the market areas of different producers identified? What is the role played by distance in the identification of the market areas?
8. What is the definition of a spatial demand curve? How is it obtained?
9. How does the location choice of firms change when interdependence in location choices is assumed?
10. What spatial structure of demand and supply is behind the market areas theory? Is it the same structure as Weber's model?

Selected reading on empirical findings

About industrial location choices

Li S. and Ho Park S. (2006), 'Determinants of Locations of Foreign Direct Investment in China', *Management and Organization Review*, vol. 2, no. 1, pp. 95–119.

Urata S. and Kawai H. (2000), 'The Determinants of the Location of Foreign Direct Investment by Japanese Small and Medium-sized Enterprises', *Small Business Economics*, vol. 15, no. 2, pp. 79–103.

Woodward D.P. (1992), 'Locational Determinants of Japanese Manufacturing Start-ups in the United States', *Southern Economic Journal*, vol. 58, no. 3, pp. 690–708.

Further reading

Beckmann M. (1968), *Location Theory*, Random House, New York.

Chamberlin E.H. (1936), *The Theory of Monopolistic Competition*, Harvard University Press, Cambridge, Mass.

Greenhut M. (1959), 'Size of Markets versus Transport Costs in Industrial Location Surveys and Theory', *Journal of Industrial Economics*, vol. 8, pp. 172–184.

Hoover E.M. (1948), *The Location of Economic Activity*, McGraw-Hill, New York.

Hotelling H. (1929), 'Stability in Competition', *The Economic Journal*, vol. 39, no. 153, pp. 41–57.

Isard W. (1956), *Location and Space-Economy*, MIT Press, Cambridge, Mass.

Smith D.M. (1971), *Industrial Location: an Economic Geographical Analysis*, Wiley & Sons, London.

Weber A. (1929), *Alfred Weber's Theory of the Location of Industries*, University of Chicago Press, Chicago.

Notes

1 Location theory – the branch of regional economics that explains the economic mechanisms responsible for location choice – concerns itself not only with the location of economic activities, in particular industrial production, but also with the location of residential activities, and with the economic processes that configure large territorial systems like cities. This chapter, however, will examine only the mechanisms that condition the location choice of productive activities, leaving treatment of the other topics to subsequent chapters.

2 Isard, 1956.

3 Hoover, 1933, 1936 and 1948; Isard, 1949 and 1956.

4 Economies of scale arise when, on increasing the inputs to the production process, output grows more than proportionally.

5 Optimal city size is achieved when an urban agglomeration is able to maximize the advantages of agglomeration while also minimizing its costs, thereby obtaining a maximum net advantage (Alonso, 1971; Richardson, 1972). However, the constant physical growth of cities (whether small, medium or large) suggests that the benefits and costs of agglomeration stem from factors other than size; for instance, the city's functional specialization and its spatial organization of production (Camagni et al., 1986). These theories are presented in Chapter 8. The urban environmental diseconomies often exacerbated by the large size of cities now figure on the agendas of policy-makers. Local and national governments, and indeed the European Union, are currently considering the economic and territorial policies best able to render urban economic growth compatible with the natural and social environment. The intention is to achieve urban sustainability, for this is one of the factors on which the competitiveness of cities – and of the regions where cities lie – depends; see Chapter 8.4. On the concept of urban sustainability see, e.g., Nijkamp and Perrels, 1994; Banister, 1998; Camagni, 1998; Nijkamp, 1999.

6 The theory of industrial districts has made much use of the concept of agglomeration economies to explain local development. It has thus prepared the ground for analysis of the role of proximity in local development. Over time, the concept of proximity has assumed increasingly less a physical connotation and increasingly more an economic one. See Chapter 9.

7 Only if demand is uniformly distributed in space does the minimization of costs equal the maximization of profits. If, instead, demand is concentrated at some points in space, a

location with higher costs than another may coincide with better revenue conditions (sales) and thus increase the firm's profits.

8 Weber's original work was published in 1909. However, it became widely known in 1929, when it was translated for the first time into English. Weber drew on a previous work by Launhardt (1882 and 1885) in developing his theory.

9 Demand (or supply) is said to be inelastic when the price of a good changes but the quantity of the good demanded (or supplied) varies less than proportionally or remains the same.

10 The per unit costs of transportation are assumed to be the same in all directions. In a subsequent version of his model, Weber substituted the concept of real weight with that of ideal weight, which is the actual weight multiplied by the unit cost of transportation in a certain direction. The reasoning does not change, however, and the location solution is found using the same procedure; the only difference is that a further element of realism has been introduced by giving greater importance to the cost per unit of distance than to the weight per distance to cover.

11 Hypothesized here is a situation in which P is the production site, M_1 and M_2 are the raw material sources, and C is the market for the final good, these being located at distances from P which are respectively a, b and c (Figure 1.1a). It is also hypothesized that, in these conditions, the location of P guarantees a minimum total transportation cost. Hence, the forces of attraction ax, by and cz must stand in the relation $ax + by = cz$, which is the equilibrium condition of the forces of attraction. If $cz > ax + by$, for point P to be the point of minimum transportation costs, it must shift towards C, while if $cz < ax + by$, it must shift closer to the raw materials market, as described in the main text.

12 Also Palander and Hoover have worked on the concept of isodapane. Palander shows that isodapanes change shape when the hypothesis of non-uniform transportation costs is introduced. See Palander, 1935. Hoover uses the concept of isotime (curve along which the sale price of the good is constant) to define the division of market areas among producers. He shows how different firms, whose production is characterized by different transport and production costs, and therefore by different sale prices and different distances from the production site, divide up the market. See Hoover, 1937a.

13 Weber uses a very similar procedure to find the best location for a firm on the hypothesis that there exists a specific place, coinciding neither with the raw materials market nor with the final goods market, where there is an abundance of low-cost labour. Once the least-cost location has been found, the decision to relocate to the point of lowest labour cost is taken on the basis of a comparison between the saving that the firm obtains in terms of labour cost from the new location and the increased transportation cost that the latter entails. If the advantage is greater than the increased cost, the firm relocates.

14 Weber was nevertheless aware of the importance of distributive aspects and the manner in which they countervail agglomerative forces to disperse activities geographically. He argues, in fact, that deglomerative factors 'all depend on the growth of the value of land, which is caused by the increased demand for land which accompanies all agglomerations', and that they operate through a redistribution of the advantages of agglomeration in favour of rents and wages at the expense of profits. See Weber, 1929.

15 See Smith, 1971.

16 Hoover argues that Weber's model is more easily applicable to the location of firms operating in some sectors rather than others, for example the steel and coal industries. In the case of these sectors, in fact, it is easier to identify raw materials and their incidence in the production of the final good. See Hoover, 1933, 1937b and 1948.

17 Greenhut, 1959a, 1964 and 1966.

18 The first studies on the subject date as far back as Launhart, 1882; Fetter, 1924; Hotelling, 1929. Subsequently, Palander analysed market size and spatial competition in order to produce a solution that incorporated the partial findings of previous scholars in a more general explanation. For this reason, Palander is widely considered to be the first theoretician able to conceptualize market areas. See Palander, 1935. The economist Lösch was the first to formulate a general spatial equilibrium on the basis of definition of market areas. See Lösch, 1954.

19 As said in a previous note, by rigidity of demand is meant a situation in which, if the price of the good varies, consumers alter the quantity of the good that they want to purchase less than proportionally (or they do not change it at all).

20 For analysis of all the location equilibria in the various cases of price discrimination, see Beckmann, 1968. For critical discussion of the effects of price discrimination on the location equilibrium, see Smith 1971, chap. 8. On the consequences of different price elasticities of demand on price discrimination and on the location equilibrium, see Hotelling, 1929.
21 Chamberlin, 1936; Lancaster, 1975.
22 This hypothesis amounts to saying that above a certain threshold of production, the law of decreasing marginal returns to the production factors applies. As more and more units of production factors are introduced, their contribution in terms of additional output grows less than proportionally. When this is the case, above a certain threshold there arise decreasing marginal returns and increasing marginal and (average) production costs.
23 For detailed analysis of the effects of changes in factor prices, production techniques and the combination of production factors, on the position and slope of the overall average cost curve, see Smith, 1966.
24 The supply curve is defined by the firm's operating conditions (costs profile and economies of scale), which the model assumes as given.
25 See Cappellin, 1980; Camagni, 1992a, chap. 1. In analytical terms, with x_i denoting individual demand, we have:

$$p = p^* + \tau d$$ (1.1n) Price-distance: Fig. 1.5a

$$p = a - bx_i$$ (1.2n) Individual demand curve: Fig. 1.5b. Therefore:

$$P^* + \tau d = a - bx_i$$

$$x_i = \frac{a - p^*}{b} - \frac{\tau}{b} d$$ (1.3n) Individual spatial demand curve: Fig. 1.5d

26 For the analytical solution of the quantity demanded in a linear and circular market, see Segal, 1977, chap. 2; Camagni, 1992a, chap. 1.
27 An increase in the production price p^* produces a parallel upward shift of the curve in panel *a*, and a parallel downward shift of the firm's spatial demand curve in panel *d*. An increase in τ instead causes the individual spatial demand curve to slope more steeply (Figure 1.4d) and reduces the market area. Interestingly, these results correspond to those obtained by the market areas model.
28 Microeconomics defines 'market equilibrium' as a situation in which, given certain demand conditions, the firm can produce a quantity of goods that enables it to maximize profits. The equilibrium achieved may be short- or long-period. In the former case, the equilibrium is altered by conditions in the market (e.g. the existence of surplus profits) that attract new firms, with which the firm finds itself having to share the market. In the latter case, the equilibrium persists over time until conditions external to the market (technological innovations, variations in the prices of raw materials and the production factors) change the initial rules of the game (costs structure, relative prices of the factors).
29 Because profit is defined as total revenues minus total costs ($\pi = R - C$), profit maximization requires equality between marginal cost (C') and marginal revenue (R'). In fact, $\pi' = 0$ entails that $R' - C' = 0$, i.e. $R' = C'$. In this situation, the firm maximizes the so-called 'normal' profit, i.e. the entrepreneur's normal remuneration. However, there exist cases in which the producer obtains a surplus profit, in addition to its normal profit deriving from market conditions, whereby the good can be sold at prices higher than the average cost. This situation is one of short-term equilibrium, however, because the availability of surplus profits induces new firms to enter the market. In the long period, as the surplus profits are absorbed, the incentive for firms to enter the market disappears, and a stable equilibrium is reached.
30 Before Hotelling, only Fetter had envisaged the possibility that two firms might compete to control the broadest market area possible. Thereafter, other economists conducted

interesting analyses on the workings of a market with imperfect competition or a monopolistic market. See Fetter, 1924; Hotelling, 1929; Robinson, 1934; Chamberlin, 1936.

31 We shall see later that discarding this hypothesis changes the final result of the model.

32 Hotelling addressed the problem of location choice also in analytical terms (see Hotelling, 1929). Let τ be the unit cost of transportation, p_a and p_b the sale prices of firm *A* and *B* respectively, and x_a and x_b the quantities of the good manufactured by the two firms. Moreover, let *c*, *f*, *g* and *h* be, respectively, the distances between the origin and *A*, between *A* and *L*, between *L* and *B*, and between *B* and the end of the linear market, as in Figure 1.9. The lengths of the segments *f* and *g* depend on the difference between the two producers' prices: *f* increases if $p_a - p_b$ increases; vice versa, *g* decreases when $p_a - p_b$ increases.

The location point *L* in Figure 1.9 represents the division of the market between the two producers in so far as the following relation holds:

$$p_a + \tau f = p_b + \tau g \tag{1.4n}$$

In *L*, consumers are indifferent between purchasing from *A* or *B* because their sale prices are the same. The market, whose total size is *l*, is divided into four segments (Figure 1.9):

$$l = c + f + g + h \tag{1.5n}$$

Rewriting (1.5n) for *f* and (1.4n) for *g*, and substituting the latter in (1.5n), yields:

$$f = \frac{1}{2}\left(l - c - h + \frac{p_b - p_a}{\tau}\right) \tag{1.6n}$$

and likewise:

$$g = \frac{1}{2}\left(l - c - h + \frac{p_a - p_b}{\tau}\right) \tag{1.7n}$$

With these lengths of the segments *f* and *g* (or, in economic terms, with these market sizes served by the two firms) their profits become:

$$\pi_a = p_a x_a = p_a(c + f) = \frac{1}{2}(l + c - h)p_a - \frac{p_a^2}{2\tau} + \frac{p_a p_b}{2\tau} \tag{1.8n}$$

$$\pi_b = p_b x_b = p_b(g + h) = \frac{1}{2}(l - c + h)p_b - \frac{p_b^2}{2\tau} + \frac{p_a p_b}{2\tau} \tag{1.9n}$$

Each firm decides to sell the good at the price that enables it to maximize profit. In analytical terms, the price is decided by differentiating the profit functions from the price and setting them equal to zero:

$$\frac{\delta \pi_a}{\delta p_a} = \frac{1}{2}(l + c - h) - \frac{p_a}{\tau} + \frac{p_b}{2\tau} = 0 \tag{1.10n}$$

$$\frac{\delta \pi_b}{\delta p_b} = \frac{1}{2}(l - c + h) - \frac{p_b}{\tau} + \frac{p_a}{2\tau} = 0 \tag{1.11n}$$

from which one obtains:

$$p_a = \tau\left(l + \frac{c - h}{3}\right) \tag{1.12n}$$

$$p_b = \tau\left(l - \frac{c-h}{3}\right) \tag{1.13n}$$

and

$$x_a = c + f = \frac{1}{2}\left(l + \frac{c-h}{3}\right) \tag{1.14n}$$

$$x_b = g + h = \frac{1}{2}\left(l - \frac{c-h}{3}\right) \tag{1.15n}$$

The maximum profits become:

$$\pi_a = p_a x_a = \frac{\tau}{2}\left(l + \frac{c-h}{3}\right)^2 \tag{1.16n}$$

$$\pi_b = p_b x_b = \frac{\tau}{2}\left(l - \frac{c-h}{3}\right)^2 \tag{1.17n}$$

It is evident from equations (1.16n) and (1.17n) that firm *A*, given the location of *B*, increases its profit by relocating to *A*' (Figure 1.19): in fact, *c* increases and *h* decreases, so that *B*'s profit diminishes and *A*'s profit increases. Firm A will therefore seek to maximize segment *c*. In its turn, *B* wants to relocate closer to *A*, for example to *B*' (Figure 1.9), so that it can increase segment *h* to the detriment of segment *c*. By so doing, it increases its profit by appropriating from *A*. Once point *L* has been reached, neither of the two firms has any further interest in relocating.

33 See Chamberlin, 1936; Lösch, 1954.

34 See Isard, 1970.

35 See Greenhut, 1959a and 1964.

36 Cf. Camagni, 1992a, chap. 1. Recent studies have highlighted the role of logistic costs, in addition to transportation costs, in the location choices of industrial activities. On this see McCann, 1998. For a recent critical review on industrial location theory, see McCann and Sheppard, 2003.

2 Accessibility and location

2.1 Accessibility and transportation costs: land value and use

The previous chapter showed how some location models interpret the location choices of firms solely in terms of the two strategic economic forces that characterize location processes: transportation costs, which induce the dispersion of activities, and agglomeration economies, which instead give rise to concentration. By balancing these opposing forces, the models examined were able to account for the existence of agglomerations of economic activities. They did so, moreover, by hypothesizing a perfectly uniform space without the geographic features that can straightforwardly explain the spatial concentration of economic activities.

The previous chapter also showed that the underlying logic of models that define market areas entails specific assumptions concerning the spatial structure of supply and demand: production develops at specific points in space, and it supplies geographically dispersed markets.

The theories examined in this chapter reverse these hypotheses on the spatial structure of demand and supply. For the models now discussed, in fact, the production site assumes a spatial dimension and extends across a territory, while the consumption site (the market) is punctiform. This reversal of assumptions about the territorial structure of production and the market is not a purely academic exercise. Rather, it is entailed by the problem that these models set out to solve, for they abandon the endeavour to identify the market areas of each producer and address an issue that has not yet been mentioned: how to define a 'production area', meaning by this the physical space (the land) occupied by an individual economic activity.

In these theories, location choices are dictated by a specific principle of spatial organization of activity: namely 'accessibility', and in particular accessibility to a market or a 'centre'. For firms, high accessibility means that they have easy access to broad and diversified markets for final goods and production factors, to information and to the hubs of international infrastructures. For people, accessibility to a 'central business district' and therefore to jobs means that their commuting costs are minimal, while at the same time they enjoy easy access to a wide range of recreational services restricted to specific locations (e.g. theatres, museums, libraries) and proximity to specific services (e.g. universities), without having to pay the cost of long-distance travel.

High demand for accessibility to central areas triggers competition between industrial and residential activities for locations closer to the market, or, more generally, closer to the hypothetical central business district (the city centre).

All the location choice models described in this chapter have an important feature in common: the *cost of land*, or *land rent*. Assuming the existence of a single central business district, owing to high demand for central locations with their minimum transportation costs, land closer to the centre costs more; a condition accentuated by the total rigidity, at least in the short-to-medium period, of the urban land supply. The models described in this chapter resolve the competition among activities on the basis of a strict economic principle: *firms able to locate in more central areas are those able to pay higher rents for those areas.*

Unlike the industry location theories (in particular Weber's model) that identify a different location equilibrium according to the spatial principle that patterns activities in space (agglomeration economies rather than minimum transport costs), these models envisage just one factor organizing activity in space: land rent, this being the sole principle that explains location choices by all activities, whether agricultural, productive or residential.

The strength of these models is the elegant and irrefutable logic with which they account for the distribution of productive, agricultural and residential activities in a geographic space from which they eliminate every differentiating effect except for physical distance from the centre. Given their assumptions on the structure of demand and supply in space, these models are particularly well suited to analysis of the location of industrial and residential activities in urban space. In an urban environment, in fact, it is easy to hypothesize the existence of a single business district (a city centre) that, for firms, performs the function of collecting, distributing and exporting the city's products, and for households is the place where jobs are available. These models are able to establish where an individual firm or household will locate.

The first model analysing the spatial distribution of alternative production activities was developed in the early nineteenth century by Johann von Thünen. Only in the 1960s did pioneering studies by Walter Isard, Martin Beckmann and Lowdon Wingo prepare the ground for Alonso's formulation of von Thünen's historical model applied to an urban context.[1] The model of the monocentric city soon became a free-standing school of thought within location theory, where it was labelled 'new urban economics'. This corpus of theories endeavoured to develop general equilibrium location models in which the main interest is no longer decisions by individual firms or households. Instead, the main areas of inquiry become definition of the size and density of cities, and identification of the particular pattern of land costs at differing distances from the city that guarantees achievement of a location equilibrium for all individuals and firms in the city.[2]

As we shall see, these theories are set forth as elegant models demonstrating the economic nature of spatial phenomena and showing that they can be analysed with the conventional tools of economic theory. They are in fact an application of microeconomic theory to the study of the intra-urban structure.

The chapter first describes von Thünen's basic model, which in simple terms and with strict economic logic explains the spatial distribution of agricultural production around a medieval town (Section 2.2). It then presents the models subsequently developed on the same theoretical bases to examine the location of firms and households in urban areas[3] (Sections 2.3 and 2.4). Moreover, brief discussion will be made of the recent developments of the general equilibrium model developed in this regard, since they have constituted one of the most important fields of urban economics since the 1980s (Sections 2.5 and 2.6). The chapter concludes by outlining a method that can be used to measure the attractiveness of an urban centre for firms and households located in its surroundings (Section 2.7).

2.2 The location of agricultural activities: the von Thünen model

Johann Heinrich von Thünen developed the first location model based on the hypothesis of a continuous production space and a single punctiform final market.[4] His model has generated the entire corpus of theories on the urban location of economic activities.

Von Thünen's model is based on a set of assumptions which all subsequent theories would adopt:

a) there exists a uniform space where all land is equally fertile and transport infrastructures are identical in all directions (isotropic space);
b) there is a single centre, the medieval town, where all goods are traded (i.e. there is a specific market place);
c) demand is unlimited, an assumption which reflects the supply-oriented nature of the model: the location equilibrium depends solely on the conditions of supply;
d) the production factors are perfectly distributed in space: the allocation of land among alternative production activities does not derive from an uneven spatial distribution of the production factors;
e) there is a specific production function, with fixed coefficients and constant returns to scale, for each agricultural good; this assumption entails that the quantity of output obtainable from each unit of land and the unit cost of production are fixed in space;
f) perfect competition exists in the agricultural goods market: farmers therefore take the prices of the goods they produce to be given;
g) unit transport costs are constant in space: the total cost of transportation depends on the distance between the production site and the town, and on the volume of production. Transport costs may vary according to the crop.

Assuming the existence of a certain number of farmers, von Thünen addresses the problem of how to determine the allocation of land among farmers working in the area surrounding the market place.[5]

He bases his model on a concept of rent obtained as a residual that would also characterize subsequent models: the price that farmers are willing to pay for land is the remainder left when transport and production costs, including a certain remuneration (profit) for the farmer, have been subtracted from revenues.

In formal terms, if x is the quantity of a good produced by a farmer, c the unit cost of production, p the price of the agricultural good, τ the unit cost of transportation, and d the distance to the market, rent r is defined as:

$$r(d) = (p - c - \tau d)x \qquad (2.1)$$

This equation states the levels of rent that farmers are willing to pay for land at different distances from the market place where goods are traded. It is represented graphically by a straight line, with slope $-\tau x$ and intercepts equal to $(p-c)x$ and $(p-c)/\tau$, respectively denoting the maximum value of rent in the town and the maximum distance from the town, where land value is nil.

From equation (2.1) one can straightforwardly obtain the impact on rent due to a shift in space (e.g. of one kilometre) by calculating the first derivative of rent with respect to distance:

$$\frac{dr(d)}{dd} = -\tau x \tag{2.2}$$

As (2.2) shows, the variation in rent is exactly equal to $-\tau x$: a shorter distance from the centre generates a saving in total transport costs equal to the increase in the rent required to occupy more central locations.[6]

On the assumption that there are three categories of farmers (A, B and C), each of them producing a specific agricultural product with a differing degree of perishability, a rent supply curve can be constructed for each category. Partially because goods are perishable to differing extents, the rent supply curves assume different positions and slopes (Figure 2.1). The category of farmers who produce the most perishable good will have a productive process that uses the land in the most intensive and

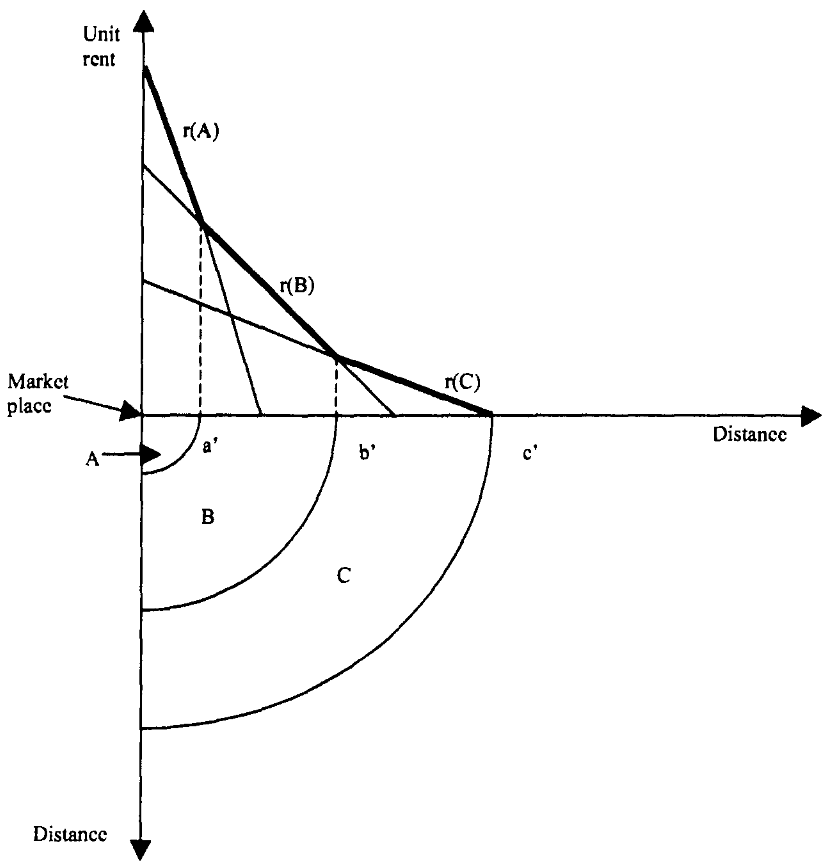


Figure 2.1 Land allocation among three farmers: the von Thünen model

economically efficient way (geometrically, the highest intercept on the Y-axis, equal to $(p-c)x$); this category will be more willing to pay the rent charged for land one kilometre closer to the town (geometrically, the steeper slope of the straight line, equal to $(-\tau x)$). As the categories of farmers compete for the more accessible land, each unit of surface area will be allocated to the category willing to pay the highest rent for that land. As far as a', the land will be allocated to category A, who offers the highest rent for the most central locations, from a' to b' to category B, and from b' to c' to category C: the actual rent realized by the landowner from cultivation of his land is the envelope of the three rent supply curves.

It is interesting to ask where, within each area attributed to a category of farmers, a single farmer will locate. The reply is that farmers will be indifferent to whatever location; that is, they will not compete with one another, since each location guarantees the same level of (normal) profit; by moving towards the village, in fact, transportation costs decrease for the exact amount that is necessary to pay a higher rent.

One of the main strengths of this model is its ability to demonstrate that it is simple distance from, or accessibility to, the town (expressed by transport costs) that accounts for differences in land rent. It thus departs from the classical Ricardian view that differences in land profitability are due to different degrees of fertility.[7] By so doing, it is able indirectly to explain the location of economic activities in space – a result that is a significant achievement.[8]

In this perspective, the Von Thünen model can be defined as a general spatial equilibrium model. Starting from the assumption of a homogeneous space (an un-cultivated plain), and giving for granted a normal profit level, common to all categories of farmer, the model is able to identify at the same time the location of the different categories and the rent level.

2.3 The urban location of firms: the Alonso model

In the early 1960s, first William Alonso and then Richard Muth reconsidered von Thünen's model and adapted it to an urban context,[9] thus paving the way for numerous subsequent studies. Alonso and Muth extended the bases of von Thünen's pioneering model, making it more specific to the urban case; but they also made it more general by abandoning the hypothesis that only transport costs express spatial friction and the preference for more central locations.

Based on a similar logic to von Thünen's model, Alonso's simplest version of the model is in reality a partial spatial equilibrium model. In fact, differently from von Thünen who assumed an un-cultivated plain and a normal profit equal among all categories of farmer, and derived the location choices and the level of unit rent, Alonso assumes the existence of a city that cannot be built instantaneously, and therefore of an effective rent curve from the city centre to the periphery; as we will see, from these assumptions Alonso determines the location for a new firm willing to locate in the city, and the profit the firm can obtain, which might also be different from the normal or average profit.

The assumptions of Alonso's model are the same as those of von Thünen's model of agricultural activity described above. It envisages a city (no longer a plain) characterized by uniform space (homogeneous spatial distribution of the production factors) and endowed with infrastructures which cover the entire city in all directions (isotropic space). The city has a single centre – the city centre or business district – that is generically defined as the most attractive location for all firms and households.

Given these assumptions, the city is analysed along only one dimension: a radius comprising different distances from the city centre to the periphery.

Also, Alonso's model defines rent as the remainder left when the entrepreneur has subtracted production costs (including transport costs) and a desired level of profit from the revenue obtained by selling the good. Formally, rent is expressed as:

$$r(d) = (p_x - \pi - c(d))x(d) \tag{2.3}$$

where r denotes the rent, p_x the unit price of the good produced by the entrepreneur, c unit production costs (including transport costs), π the profit, d the distance from the centre, and x the quantity of the good produced.

Because production costs include transport costs, in the Alonso model they depend on distance, as they do in the von Thünen model. However, unlike in the latter, revenues too depend on distance: a less suburban location gives greater proximity to broader markets, and consequently access to higher earnings (consider the sales of a shop located in the city centre compared to one in the periphery, especially if they sell luxury items).

Equation (2.3) expresses the 'bid rent', or the rent (by square metre) that the entrepreneur is willing to pay at differing distances from the centre, once costs and the entrepreneur's intended profit have been subtracted from revenues. Profits remaining equal, a more central location implies a willingness to pay higher rent because the entrepreneur incurs lower transport costs and obtains higher revenues. Likewise, a suburban location can yield the same profit if and only if less rent is paid for the land: the saving on land cost must offset the higher transport costs and the lower revenues that less central locations entail (Figure 2.2a).

The slope of the bid-rent curve, which expresses the variation in the cost of land due to a one unit of variation in the distance from the centre, is given by:

$$\frac{\partial r(d)}{\partial d} = (p_x - \pi - c(d))\frac{\partial x(d)}{\partial d} - \frac{\partial c(d)}{\partial d}x(d) \tag{2.4}$$

This shows that, at one unit of distance further away from the centre, the rent offered to maintain the same profit level π diminishes because of increased transport costs and decreased revenues.

At every distance from the centre (for example d_0 in Figure 2.2b), if the firm wants to increase its profits, it must offer a lower rent. Vice versa, at the same distance, it can offer a higher rent if it is willing to accept lower profits. It is therefore possible to plot different bid-rent curves for an individual firm, all of them with the same slope, and each of them defined on the basis of a different profit level which increases towards the origin (Figure 2.2b).

In a partial equilibrium framework, which assumes as known the 'market land rent curve' (i.e. the real market cost of land: curve *re* in Figure 2.2b), it is possible to define the optimal location for the firm. Along the market land rent curve (*re*), the firm will choose the location yielding the highest profit, which is expressed by the tangency of the market rent curve with the lowest bid-rent curve. In Figure 2.2b the location

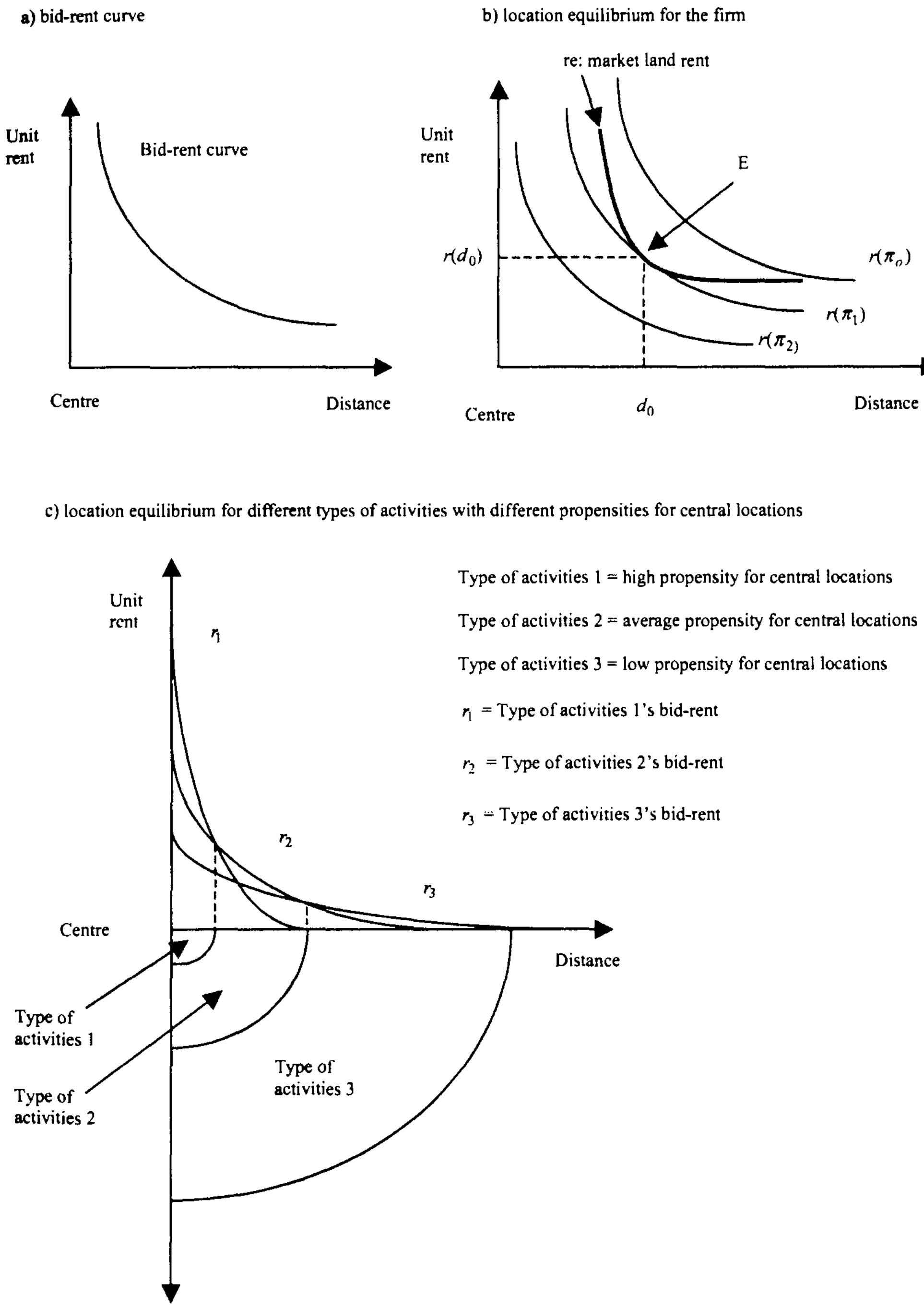


Figure 2.2 The bid-rent curve and the location equilibrium for firms

equilibrium is reached at point E, and thus at a distance d_0 from the centre and with a rent equal to r_0.

The differing 'fertility' of various types of land has traditionally been viewed as determining differences among agricultural rents.[10] In this model differing fertility can be interpreted in the modern sense as the differing 'productivity' of more central urban land due to its greater accessibility to information. If revenues and costs vary with distance, rent is the value that reduces all net revenues to those obtained from marginal lands.[11] If we discard the assumption that a city already exists (a market land exogenously determined), and therefore if we move away from the identification of an urban location for a new firm entering a city, we abandon a partial spatial equilibrium model, and go back to von Thünen's general spatial equilibrium framework, which entails an interesting interpretation of the allocation of the urban space between alternative production activities or between production and residential activities. In this case, Alonso's model addresses a problem similar to the one that preoccupied von Thünen. As firms compete for central locations, the model shows how the urban space is allocated among alternative production activities once the market cost of land at different distances from the centre is known. Suppose the existence of a point in space (a centre) attracting activities belonging to industries with different propensity for central locations (headquarters of banks, specialized shops and manufacturing firms). The different types of activities compete for locating closest to the central location, in a homogeneous space around the centre. The slopes of the bid-rent curves will differ according to the different levels of propensity for central location; as the propensity increases, firms will be willing to pay more for a unit of land in order to locate (one unit of distance) closer to the centre (Figure 2.2c).

The three categories of activities are distributed across the urban area as in von Thünen's model; each area will be occupied by the category that makes the highest rent bid. Market land rent will be the envelope of the bid-rent curves at each distance from the centre, so that the city can be depicted as a set of concentric rings each containing the type of activities willing to pay the highest rent for that distance (Figure 2.2c).

But what determines the propensity for a central location? To reply to this question a reasoning on the slope of the bid-rent curve, which expresses the variation in the cost of land due to one unit of variation in the distance from the centre, is helpful. The slope is given by eq. 2.4. This shows that, at one unit of distance further away from the centre, the rent offered to maintain the same profit level π diminishes because of increased transportation costs and decreased revenues. Eq. (2.4) contains the four elements that, on their own or in combination, theoretically explain higher propensity of activities to central location; an activity will be in fact interested to locate in the centre if:[12]

a) the influence of distance on the demand of goods $\left(\frac{\partial x(d)}{\partial d}\right)$ is high;
b) extra profits $(p_x - \pi - c(d))$ are high;
c) the quantity of goods/services produced, i.e. the value of the goods/services, for unit of land (x) is high;
d) the influence of distance on the production costs of one unit of goods/services $\left(\frac{\partial c(d)}{\partial d}\right)$ is high.

Table 2.1 Taxonomy of activities with high propensity for central location

Influence of distance on the demand for goods/services $\frac{\partial x(d)}{\partial d}$	*Extra profits* $p_x - \pi - c(d)$	*Value of goods/services for unit of land* x	*Influence of distance on the production costs of goods/services* $\frac{\partial c(d)}{\partial d}$	*Examples of activities*
High	Normal	Normal	Low	Commercial activities, shopping centres, supermarkets
Normal	High	Normal	Low	Advanced service functions (e.g. lawyers, specialised doctors), or activities that require a prestigious location
Normal	Normal	High	Low	Travel agencies, insurance brokers
Normal	Normal	Normal	High	Activities dependent on a central market characterized by high transportation costs of the final good

Source: adapted from Camagni (1992a)

Table 2.1 presents some examples of activities characterized by the major values of the slope of the bid-rent curve:

a) activities oriented towards a high demand density, like commercial activities, shopping centres, supermarkets, all characterized by a strong influence of distance on the demand for goods;
b) advanced service functions (e.g. lawyers, specialized doctors), or activities that require a prestigious location that can obtain thanks to their oligopolistic position (headquarters of banks and of insurance companies, public and private managerial functions), whose costs of moving one unit of service/good towards the periphery and the influence of distance on the demand of goods sold by unit of land are low but the extra profits of a central location are very high; through a central location these activities abandon a perfect competition market and differentiate the product quality through the use of a traditional urban input factor, like information;
c) activities like travel agencies, insurance brokers, all characterized by a very high value of their activity per unit of land;
d) activities that depend on a central market, with a high transportation cost of the final output: all industrial and service activities that depend on population and central activities.

These reflections provide already the first evidence of how these models, strongly abstract in their nature, are able to describe conditions which closely match actual reality.

2.4 The urban location of households

The model for households, this too formulated by William Alonso, is entirely similar in its theoretical basis and analytical structure to the model just discussed for firms. In this model, location choices are no longer made by firms, but by households. The main difference from Alonso's model for firms is the influence exerted on location choices by a new variable: the size of the house. A household may in fact decide to forfeit housing space in order to move closer to the city centre. The saving obtained by purchasing a smaller house, and from the lower transport costs associated with more central locations, enables the household to pay the higher unit cost of land typical of less peripheral urban areas.

The households model therefore comprises three variables: the unit cost of land (or of the house), the size of the house and transport costs.

Assume the following utility function[13] for the household:

$$u = u(d, z, q) \tag{2.5}$$

where d is the distance from the centre, q is the size of the house, and z is the set of all the other goods that the household needs.

Given a certain distance from the centre and any combination of house size q and other goods z, a reduction (even slight) in the size of the house causes a loss of satisfaction for the household that must be offset – in order to maintain his/her utility constant – by an increase in the quantity of other goods in the household's possession. This condition is depicted by Figure 2.3a, which shows the so-called indifference curves for the various 'house size/quantity of other goods' combinations that leave the household's utility unchanged.[14]

Each indifference curve represents a certain level of utility, which rises as the distance from the origin increases (Figure 2.3a).[15] The household will seek to position itself on the highest indifference curve, subject to its budget constraint; that is, provided its income equals the expenditure that it must undertake. The budget constraint is written as:

$$y = p_z z + r(d_0)q + \tau d_0 \tag{2.6}$$

where y is the household's income, $r(d_0)q$ and τd_0 are respectively expenditure on the house and transport costs at distance d_0 , and $p_z z$ represents the cost of purchasing other goods. On solving (2.6) for z, we obtain:

$$z = \frac{y - r(d_0)q - \tau d_0}{p_z} \tag{2.7}$$

Graphically, the budget constraint is represented by a straight line with intercept $(y - \tau d_0) / r(d_0)$ if the household decides to spend all its income on the house, not on purchasing units of other goods, and $(y - \tau d_0) / p_z$ when, conversely, the household decides to use its entire income to purchase other goods.

The condition that maximizes the household's utility, under the budget constraint, is represented by the point of tangency between the budget line and the indifference curve. The household cannot go beyond that level of utility because it lacks the income

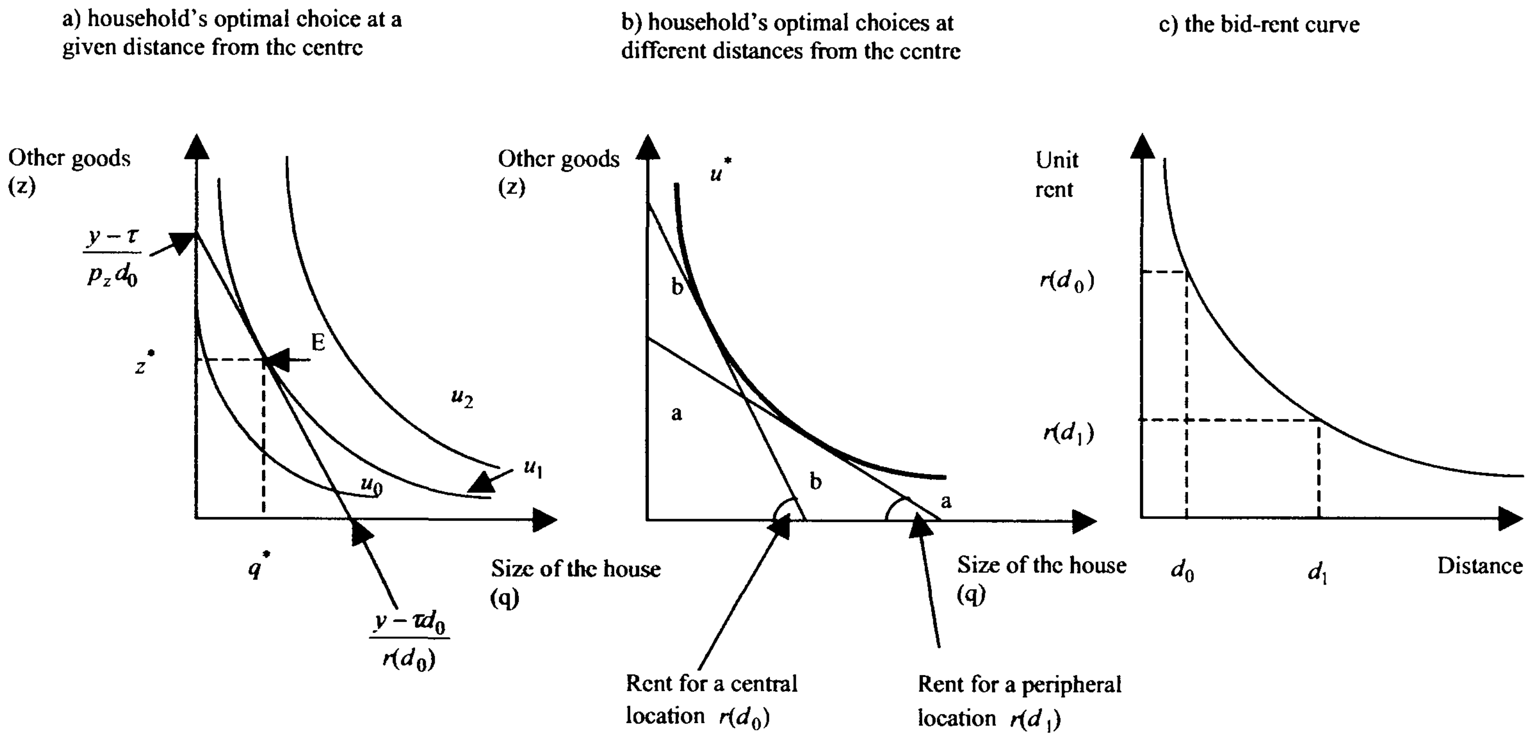


Figure 2.3 The household's optimal choice and the bid-rent curve

to do so; at the same time, it is irrational for the household to position itself below that level of utility, given the income available to it (Figure 2.3a).

In mathematical terms, the same condition holds if the two curves are equal:[16]

$$\frac{u'_q}{u'_z} = -\frac{r(d_0)}{p_z} \tag{2.8}$$

Assuming that good z is expressed in numeraire – i.e. its price is equal to one – the slope of the budget line is exactly equivalent to the rent $r(d_0)$. Equation (2.8) shows that, in equilibrium, the household is indifferent to substituting the other goods with size of the house when the relative utility deriving therefrom $\left(u'_q / u'_z\right)$ is equal to the relative cost of an additional square metre of the house, that is the unit rent.[17] Each household thus expresses that maximum amount that it can pay *for each distance from the centre* compatibly with a certain utility level that it wishes to attain (u^* in Figure 2.3b).[18] In order for the household to express less demand for the good 'house' (i.e. it settles for a smaller house) and spend income on other goods, maintaining utility constant, the price of the house must increase. The other goods thus become relatively less costly and therefore relatively more attractive to the consumer. In Figure 2.3b, this means that for the household to remain on the same indifference curve, the budget constraint must slope like line *bb*.

We now move an important step forward. Given the value of income spent in buying a house and not in purchasing other goods (the intercept $(y - \tau d_0) / r(d_0)$) and the tangent condition, we are able to obtain the value of the bid-rent: given a certain level of utility, the bid-rent is expressed by the different slopes of the budget line for different distances from the centre. Along the curve, as the distance from the centre decreases, the budget line becomes more rigid, as the result of an increase in the maximum rent offered for a location at that distance, as shown by Figure 2.3c. Along the bid-rent curve in Figure 2.3c, the size of the house (q) decreases by moving towards the city centre since the square meter of physical space becomes more expensive.

The analysis thus far yields the following important result: *bid-rent curves are nothing but a transformation that maps indifference curves in consumption space (the trade-off between goods z and house size q in Figure 2.3b) into corresponding indifference curves in urban space (the trade-off between rent and distance in Figure 2.3c).*[19]

The slope of the rent curve with respect to distance, or the bid-rent gradient, expresses the increase in the cost of land (of the house) that the household is willing to sustain in order to move one unit of distance closer to the centre, maintaining its utility constant. In formal terms it corresponds to:[20]

$$\frac{q\partial r(d)}{\partial d} = -\tau \quad or \quad \frac{\partial r(d)}{\partial d} = -\frac{\tau}{q} \tag{2.9}$$

Equation (2.9), known as the 'Muth condition', defines indifference to alternative locations by expressing a condition whereby alternative locations maintain the household's utility constant. The increase in utility due to higher accessibility to the city centre equals the loss in utility deriving from a smaller size of the house and higher unit land costs, typical of more central locations.

Eq (2.9) underlines that a location indifference condition is reached when the increase in the rent for a higher house, stemming from higher unit rent and only partially compensated for a lower size of the house, is equal to the reduction of the unit transportation costs that characterizes more central locations.

Eq. (2.9) reminds us that a condition of indifference location is reached when the increase in the cost of the house in more central locations, stemming from an increase in the unit land cost $\left(\frac{\partial r(d)}{\partial d}\right)$ and only partially compensated by a decrease in the size of the house (q), equals a decrease in unit transportation costs that characterize more central locations. Eq. (2.9) is nothing other than (2.2) in von Thünen's model, from which it differs only by including an additional variable: the size of the house (q), which decreases on moving closer to the centre because physical space becomes more expensive. In this case, the household is indifferent to less peripheral locations when the saving made possible by lower transport costs *and the purchase of a smaller house* equals the higher unit costs of land typical of central locations. The shape of the bid-rent curve is therefore not linear (as in von Thünen) but exponential, while still exhibiting a negative slope. The utility that an individual obtains from the two alternative locations remains constant: a higher accessibility to the centre equals the loss of utility stemming from *a smaller house* and higher unit land costs, associated to more central location.

As in the case of firms (Figure 2.2b), location equilibrium is obtained by superimposing the market land rent curve (which expresses the real market prices of land, defined exogenously) on the bid-rent curves. The point of tangency between the actual rent curve and the lowest bid-rent curve (relative to the highest utility) represents the household's optimal location choice (Figure 2.4).

An interesting final element for analysis is the effects of an increase in household income on the location equilibrium. Let us take a certain house size as given. The household will choose to locate at a distance from the centre where a further shift towards the periphery (the cost of transportation c in Figure 2.5a) would equal the marginal advantage represented by the saving on the cost of land (v in Figure 2.5a). An increase in income may give rise to reverse relocation choices: if the household is more interested in a larger-sized house, the advantage in terms of saving on the price of the house increases (curve v shifts to v' in Figure 2.5b) and the household will locate in a more peripheral area. When the household is instead more concerned about the

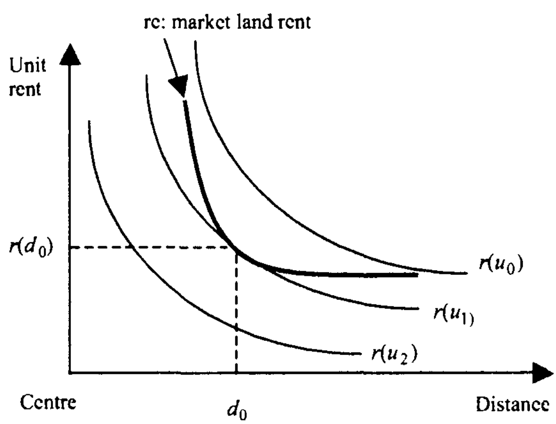


Figure 2.4 The location equilibrium for households

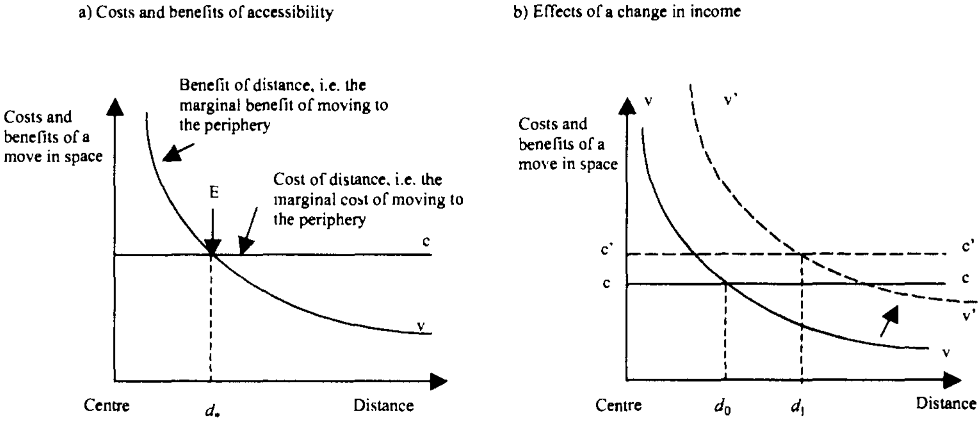


Figure 2.5 Costs and benefits of accessibility and the effects of a change in income on the location choice

greater opportunity cost of transport (curve *c* rises to *c'* in Figure 2.5b), he/she will choose a more central location.[21] In the presence of both effects, it is generally believed in the American literature that the former (the size of the house) prevails, with the consequence that the location equilibrium shifts towards the periphery.

2.5 Recent developments on the general equilibrium model

2.5.1 *Characteristics of the general equilibrium model*

The models just described have guided us through the location choices of an individual firm and an individual household, but they have done so in most cases within a partial equilibrium framework. Moreover, given their inability to derive actual land prices, in order to define the location equilibrium, they have assumed as known the market rent curve expressing the real market cost of land.

The general equilibrium approach developed since the mid-1980s in 'new urban economics' – or less emphatically 'economics of the monocentric city' – enables us to remedy both shortcomings of the previous partial equilibrium models. The modern versions of the general equilibrium model, in fact, seek to endogenize the market land rent value; that is, to obtain it from the location equilibrium conditions themselves. They are conceived with the aim to identify the urban density, city size and land prices that produce a location equilibrium where all households are equally satisfied (or all firms earn equal profits), keeping the logical-conceptual framework identical to that of the partial equilibrium models.[22] Numerous studies have been produced on these matters, their intention being no longer to identify the location of an individual household or firm, as in the partial equilibrium models, but rather to identify the land price conditions that produce a location equilibrium characterized by equal utility for all households or equal profit for all firms, and to determine the residential density of each city under these conditions.

In the case of firms, on the hypothesis that the value of land at the edge of the city is equal to the value of agricultural rent and known, the equilibrium quantity of the good (i.e. the quantity that simultaneously satisfies the demand and supply conditions for the good), the prices of the other production factors (besides land) and the level of profit desired by firms, the models identify the maximum size of the city, the density of firms and the trend of land prices in space (the actual rent curve). The same characteristics (maximum size, density and land price in space) are identified by the models in the case of households, once the population wanting to locate in the city, the price of other goods and household's desired utility have been given exogenously.[23] This endeavour has been carried forward on various hypotheses: the existence of a 'closed city', that is a city whose demographic size is given in the case of households (or market equilibrium in the case of firms), and an 'open city', that is the case in which also the size of a city is endogenously defined.

Although the models now discussed are striking in their elegance and economic logic, they are often extremely complex. Consequently, the descriptions that follow have been simplified as far as possible, and they are accompanied by illustrations.

2.5.2 *The general equilibrium model for firms*

The purpose of the general equilibrium model is to identify – given a percentage (s) of urban land used for productive activities – the equilibrium density of the n firms located on the urban land, all of them specialized in production of the same good (and therefore characterized by the same production function).[24]

The model hypothesizes a Cobb–Douglas production function[25] comprising only two production factors (land and capital), which can be substituted for each other, and with constant returns to scale:[26]

$$Y_d = aT_d^{\alpha} K_d^{1-\alpha} \tag{2.10}$$

in which Y is the quantity of the good produced by the firm, a is a constant representing technical progress, T and K are respectively the quantities of land and capital used in the production process, while α and $(\alpha - 1)$ respectively denote the efficiency of the production factors 'land' and 'capital' in the production process.

The various combinations of production factors required to achieve a certain level of production (Y) are represented in Figure 2.6a by the isoquant curves. These represent higher levels of production at increasing distances from the origin. When they are compared against the isocost line, which represents the factor combinations that keep total production costs constant, they identify the land/capital combination that maximizes the firm's profit (or, revenues remaining equal, minimize its costs) given a certain distance (d_0) from the centre (point E in Figure 2.6a).[27]

If the firm wants to increase the amount of urban land that it uses, but producing the same quantity at the same costs, it will have to reduce the amount of capital invested in the production process. The firm will have an incentive to choose this option when the land is more attractive; that is, when its cost is lower (Figure 2.6b). Thus, as previously obtained with the households model, the different slopes of the isocost line for different distances from the centre, assuming p_K as numeraire, identify the bid-rent curve (Figure 2.6c) – given a certain profit level (i.e. along a given isoquant curve). As the distance from the centre decreases, the budget line grows more rigid because the maximum bid rent for that distance increases, as shown in Figure 2.6.

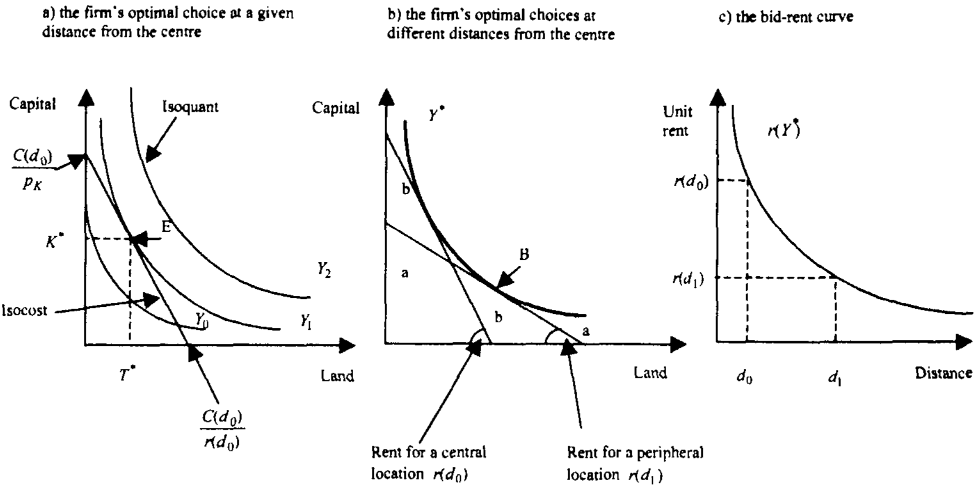


Figure 2.6 The firm's optimal choice and the bid-rent curve

The rent curve is defined as:[28]

$$\frac{T_d}{Y_d}\frac{\partial r(d)}{\partial d} = -\tau(\lambda) \tag{2.11}$$

(2.11) is the Muth condition in the case of firms. It states that a firm is indifferent to location when the new location's advantage in terms of saving on transport costs is equal to the increase in the cost of urban land. In order to cover the increase in land cost, the firm will be tempted to use a lesser amount of land, which has become relatively more costly, and to replace it with additional quantities of the other good, capital, for example by constructing taller buildings. It will thus use less land per unit of output (T_d/Y_d will be lower).[29]

We may thus once again state that each firm is indifferent to alternative locations along the bid-rent curve. In other words, the cost of moving in space is nil, so that any alternative location along the bid-rent curve leaves the firm's profit unchanged.

Moving to the general equilibrium case, on the assumption that the *n* firms all have the same production function, and that the equilibrium quantity of the good in the city (Y^*) is known, it is possible to use the 'boundary rent curve' to identify the intensity of land use, the maximum profit level achievable by each firm, and the total size of the city.[30] The boundary rent curve defines, for each maximum profit level of firms (and therefore for each bid-rent curve), at which size of the city the quantity produced at equilibrium in the urban market exactly equals the quantity given exogenously.[31]

Assuming land value at the edge of the city as known and equal to the value of agricultural land, the equilibrium bid-rent curve can be identified. It will be the bid-rent curve that intersects the boundary rent curve for the value of land at the city's edge (*brc* in Figure 2.7a).

The reasons for this statement are the following. Higher bid-rent curves are logically excluded from the definition of an equilibrium condition, since they define levels of

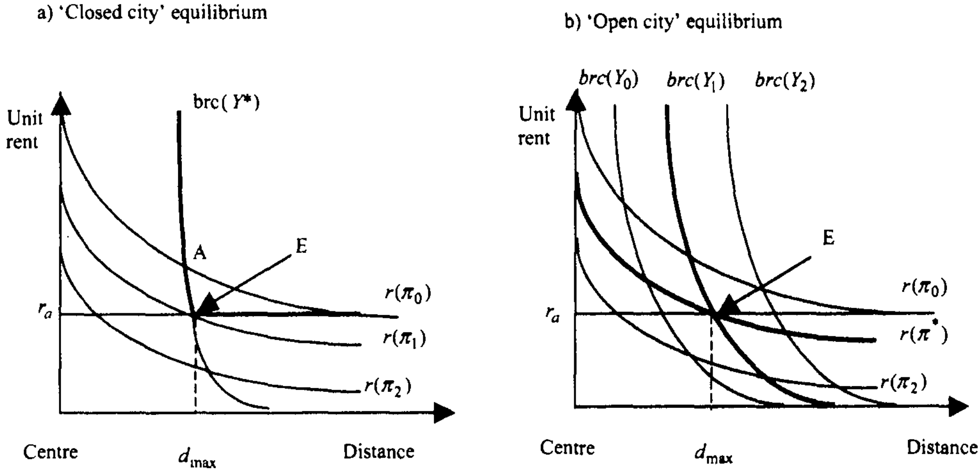


Figure 2.7 General equilibrium for firms

individual profits lower than the one that firms can actually achieve (in Figure 2.7a, in fact, point E is determined for higher profit levels than point A). Lower bid-rent curves are also logically unacceptable for defining an equilibrium condition since they determine a land value at the urban edge which is lower than the agricultural rent. The only acceptable bid-rent guaranteeing an equilibrium is therefore the one which crosses the boundary rent curve at the level of agricultural rent.

Point *E* in Figure 2.7a therefore defines:

- the market land rent curve, previously defined exogenously and coincident with the bid-rent curve, for city sizes smaller than d_{max}, and with r_a for larger sizes;
- the individual profit level (π_1) achieved by firms in the city;
- the maximum size reached by the city (d_{max}).

This model also shows that firms locate in the urban area according to a specific equilibrium density distribution that is obtained once the maximum size has been identified. Indifference to alternative locations obtains for every distance from the centre, because the lower rent charged for suburban locations exactly offsets higher transport costs and allows savings on capital costs (reducing the amount of land used per unit of output).

An interesting case is the one in which an open city – where firms can relocate to other urban areas – is hypothesized. There are several boundary rent curves in this case, each of them expressing an equilibrium production level in the goods market (Figure 2.7b). The location equilibrium is obtained by exogenously defining a profit level externally and internally to the city equal to π^*. The point at which the bid-rent curve guaranteeing the profit level π^* intersects with the agricultural rent defines the size of the city (and its density). The intersection of the bid-rent curve with one of the boundary rent curves shows the equilibrium quantity actually produced ($brc(Y_1)$ in Figure 2.7b).

2.5.3 The general equilibrium model for households

Reasoning which is very similar (in its logical structure if not in all its hypotheses) to that used by the general equilibrium model for firms defines the location equilibrium of *n* households in a city. The general equilibrium for households first requires that the indifference to alternative locations, as expressed by Muth's condition, must hold for all the *n* households. On the hypothesis that households have the same incomes and the same preference structures, they exhibit location indifference along the same bid-rent curve.

Figure 2.8a shows various bid-rent curves, with higher levels of utility for curves closer to the origin.

However, the market land rent curve is no longer known, as it was in the partial equilibrium model. It must be determined by comparing the bid-rent curves of the various households, just as the market land rent curve was defined in von Thünen's model by the envelope of the bid-rent curves of the three farmers.

As in the above case of firms, it is possible to use the boundary rent curve to define the market land rent curve (and therefore the utility of households). For every level of household utility (and therefore for every bid-rent curve and for every maximum urban size at every distance from the centre), this curve defines what size of the city ensures that the total population is equal to that given exogenously (Figure 2.8).

This curve delimits the urban area for different levels of utility and different maximum sizes of the city, subject to the condition that the total population is equal to the population given exogenously; if the utility increases, so does the amount of space required by each household, with the consequence that residential density diminishes.[32]

As in the case of firms, assuming the value of land on the urban edge as known and equal to the value of agricultural land, it is possible to identify the equilibrium bid-rent curve. This will be the bid-rent curve that cuts the boundary rent curve for land value at the city's edge (*brc* in Figure 2.8a). If the boundary rent curve intersects with the highest bid-rent curve (point A in Figure 2.8a), households will be forced to shift down to lower rent curves, which express greater utilities. Instead, if it intersects with lower

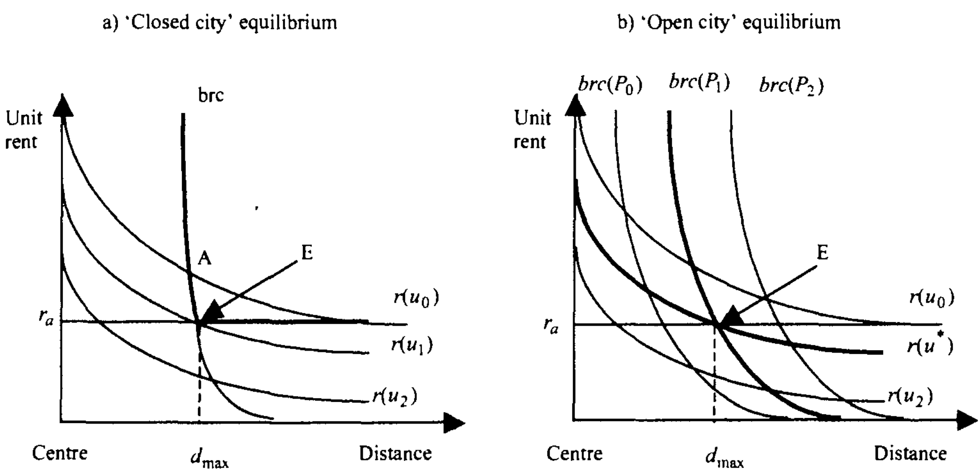


Figure 2.8 General equilibrium for households

bid-rent curves, there will be no equilibrium because the land value offered by households at the outermost edge of the city will be less than the agricultural value.

As in the case of firms, point *E* in Figure 2.8a defines:

- the market land rent curve, previously defined exogenously and coincident with the bid-rent curve, for city of sizes smaller than d_{max}, and with r_a for larger sizes;
- the utility level (u_1) achieved by households in the city;
- the maximum size, and the density, reached by the city (d_{max}).

In the case of an open city – where households can relocate to other urban areas – there exist (as previously in the case of the model of firms) several boundary rent curves, each of them defining a different level of population (Figure 2.8b). Equilibrium is obtained by exogenously defining a level of utility externally and internally to the city which equals u^*. The point at which the bid-rent curve guaranteeing utility level u^* intersects with agricultural rent defines the size of the city (and its density). The intersection of the bid-rent curve with one of the boundary rent curves shows the equilibrium ($brc(P_1)$ in Figure 2.8b).

Finally, we may discard the hypothesis that all households have the same income and assume that there exist three classes with different incomes and different preference structures. The slopes of the bid-rent curves will differ according to level of income. As income increases, the different classes of households will be willing to pay more for houses in order to locate (one unit of distance) closer to the centre (Figure 2.9).[33] The three classes of households are distributed across the urban area as in

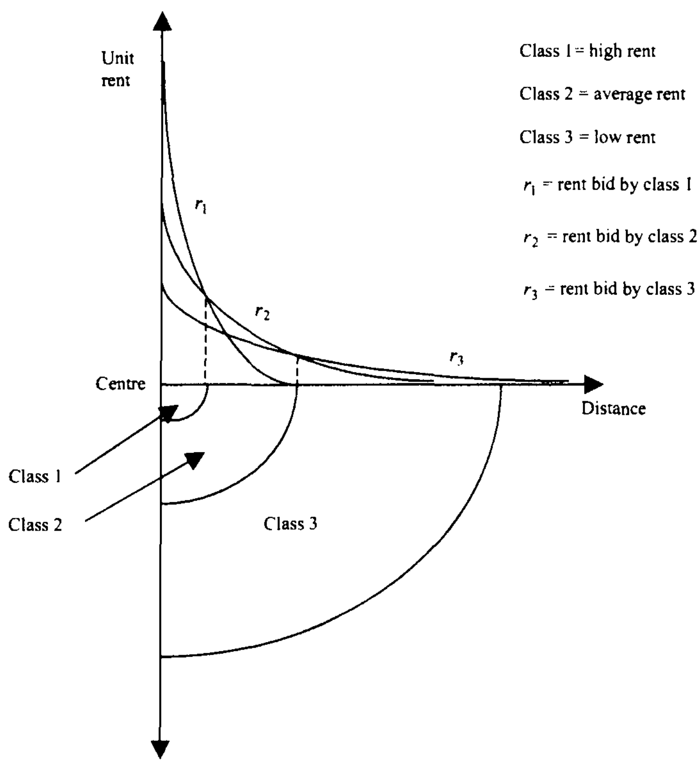


Figure 2.9 Location equilibrium for different classes of households

von Thünen's model: each area will be occupied by the class of households that makes the highest rent bid. Market land rent will be the envelope of the bid-rent curves at each distance from the centre, so that the city can be depicted as a set of concentric rings each containing the class of households willing to pay the highest rent for that distance (Figure 2.9).

2.5.4 *Firms and households*

Finally provided is a brief description of some models of monocentric cities which, with minor adjustments, enable simultaneous analysis of the location of firms and households, with a general spatial equilibrium logic. On the hypothesis that the rent gradient of firms is higher than that of households (i.e. firms are willing to pay higher unit rents in order to move one unit of distance closer to the centre), the bid-rent curves for firms and households will be those shown in Figure 2.10.

These models lead to two important results. The first is that they identify the bid-rent curves of firms and households endogenously. Let us assume that at time t_0 households choose a level of rent $r_0(d_0)$ characterized by a certain level of utility. For equilibrium to come about, the level of utility must be such that it determines an amount of population and a labour supply equal to the labour demand of firms. If households have chosen too a high level of utility, and therefore make rent bids which are too low, the population located in the city (in the range $d_1 - d'_{max}$) may be insufficient to satisfy the labour demand by firms. The availability of work will attract new households into the city, with a consequent increase in demand for urban land which pushes the bid-rent curve up to $r_1(d_1)$ in Figure 2.10. The city will expand (d''_{max}) until labour-market equilibrium has been re-established at a lower level of utility.

The second important result of these models is that they divide the urban area between productive and residential activities. Urban land will be allocated to the

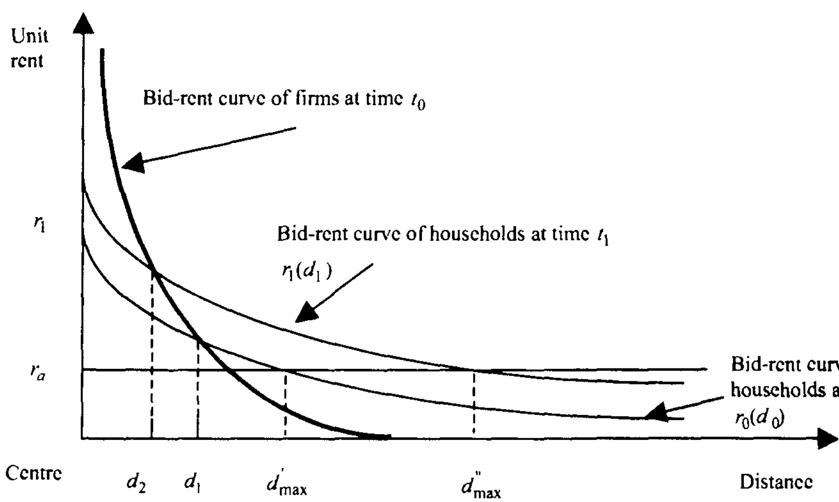


Figure 2.10 Location equilibrium for households and firms

activities able to pay a higher rent for each distance from the centre – as in von Thünen's model. In this case, the central areas will be occupied by firms, while households will be pushed towards suburban areas; a theoretical result that closely reflects what actually happens in reality.

2.6 Critical remarks

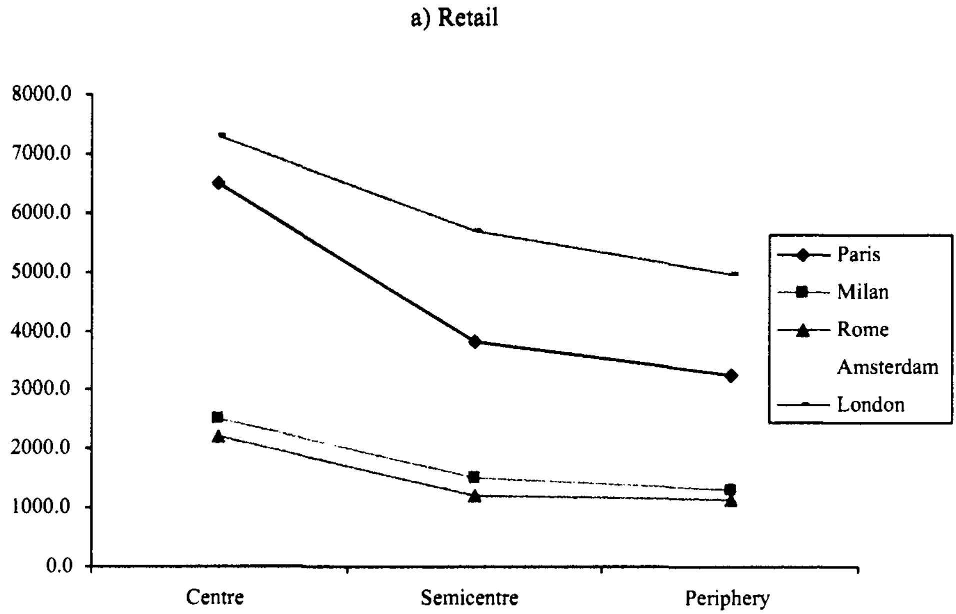
Starting from hypotheses similar to those of von Thünen's classic model, with considerable formal elegance and economic rigour, the models described in this chapter adopt a neoclassical framework of profit maximization for firms, and utility maximization for households, to identify the location equilibrium conditions for productive and residential activities, the spatial pattern of urban land prices, and the density and size of the city.

The models draw on microeconomic concepts to explain purely spatial phenomena, such as the distribution of activities in space and the location choices of households and firms. They are thus able to break down the disciplinary barrier between mainstream and urban economics that long hampered development of a general theory of space. Traditional economics has consequently been enriched with theories that interpret space in purely microeconomic terms; while urban economics has acquired a traditional economic logic with which to interpret location choices.

Moreover, although these models are highly abstract, owing to their unrealistic hypotheses (isotropic space, a city with a single centre), they are able to describe conditions that closely match actual reality: an urban land rent gradient negative with respect to distance from the centre (Figure 2.11), the central location of activities with high value added (business and management), broad suburban spaces for residential activities.

However, despite their logic, elegance and economic rigour, these models have a number of theoretical elements that weaken their overall logical structure. One of them is the decisive role played by commuting in determining location equilibrium. If real behaviour does not comply with the perfect rationality envisaged by the models, so that commuting is of less importance for a person's utility, the entire theoretical-conceptual edifice collapses. This shortcoming can be partly remedied, however, if we acknowledge that the costs of transport to the centre and the desire to reduce them may reflect other important aspects of an individual's utility function when s/he makes location choices, like accessibility to information, recreational services and opportunities for social interaction.

A second shortcoming is more serious. These models concern themselves neither with how a city centre is organized nor with what happens outside the city itself. They restrict themselves to interpreting locational behaviour within the area extending between an hypothetical aspatial centre and the physical boundary of the city. Moreover, when these models are used to interpret location equilibrium, not internally to a city but among cities, and therefore on the hypothesis that the city is part of an urban system and that firms and households may decide to relocate to other cities with attractively higher levels of utility or profit, they display a clear interpretative weakness. On the hypothesis that households have equal preference structures and firms have equal production functions, there can only be indifference to alternative locations in other cities if all these exhibit – in the logic applied here to describe them – the same bid-rent curve and the same boundary rent curve, and are therefore all of the same size.



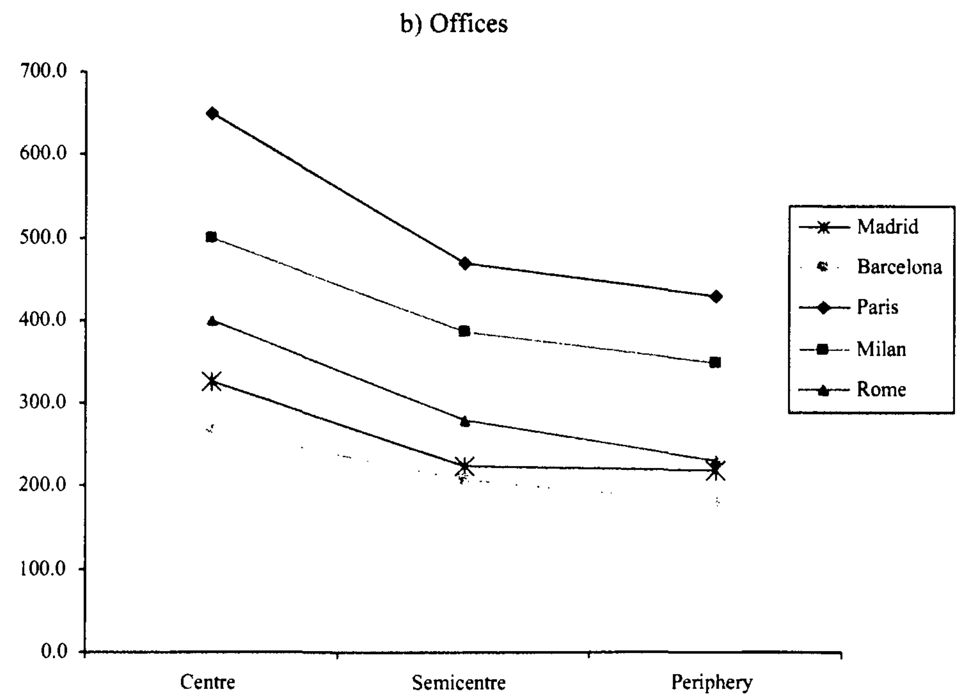


Figure 2.11 Urban rent gradients in some major European cities (rental values per sqm. per annum in euros, 2003)

Source: our elaborations on ONCOR European Property Bulletin data, 2004

If this is the case, there will be an urban system made up of cities that are all of equal size;[34] but this circumstance is amply contradicted in the real world. In order to deal with this defect, the conceptual framework should be able to accommodate the hypothesis that locational advantages differ according to the size of the city, and that rents – the monetary counterpart of the advantages that households and firms obtain from central urban locations – vary (distance from the centre remaining equal) from city to city.

Only thus is it possible to conceive a location equilibrium with cities of different sizes. Yet this also requires acceptance of the idea that large, medium-sized and small cities are structurally different and perform different functions in the overall economy, and consequently have specific production specializations; a hypothesis at odds with the basic features of these models, and that instead opens the way for the general equilibrium models discussed in the next chapter.

2.7 Generalized accessibility and the gravity model

Throughout this chapter, the location of activities in urban space has been explained in terms of accessibility to a generic central business district – the locus of trade, information, and social and economic interaction. The centre attracts firms and households which, in their turn, influence the centre in diverse ways: commuter movements, diffusion of knowledge and information, co-operation networks and personal relationships. These forces of attractiveness (and repulsion) invest not only the centre but every couple of points in space (two different zones of the same city, but also two different cities) and they are generated by the intensity of the flows of people, information and goods between those points. In fact, these flows measure the relations that may arise between the activities located at the two points in space and, at aggregate level, the greater attractiveness of one zone with respect to the other (one city with respect to the other).

Surprisingly, flows of people, goods and information across a territory seem to organize themselves on the basis of gravitational fields sensitive to the amount of activities located on the territory, and to their relative distance.[35]

Since the work of George Kingsley Zipf in the years after the Second World War,[36] territorial flows of information, goods and people have been estimated using the gravity model, so called because it is based on Newton's law of universal gravitation. The model states that every point in space undergoes (or exerts) an influence proportional to its mass and inversely proportional to the distance that separates it from the other point in space.[37]

In general, the *model of flow* (so called because of its ability to estimate flows between two territorial entities) is expressed as:

$$T_{ab} = K(P_a^{\alpha} P_b^{\beta}) \,/\, d_{ab}^{\gamma} \tag{2.12}$$

where T is the intensity of interaction between a generic pair a and b of points in space, K is a constant of proportionality, P is the mass of points a and b respectively, d is the distance between a and b. α and β are assumed equal to 1. The distance exponent, γ, represents the impedance or friction exerted by physical space on movement (of goods or people). Its value therefore differs according to the phenomenon studied (consider, for example, the greater weight of the distance covered between home and shop to purchase staple goods like bread or milk, compared with luxury goods like jewellery). The mass P of the territorial entities is often expressed in terms of population.

When the gravity model is generalized so that it can be applied to estimation of flows between a generic point and n points in space (between one zone of a city and all the other zones of the city), the analogy still holds between gravitational physics and interpretation of territorial phenomena, as well depicted by the following gravity model (labelled *model of potential* for its capacity to measure the attractiveness potential of a place):

$$E_a = K \sum_j P_j \,/\, d_{aj}^{\gamma} \tag{2.13}$$

where E represents the potential energy produced on a by a set of masses P_j each at a distance d from a.[38] Once again, parameter γ denotes spatial friction, which may differ from the value of 1 assumed in the original gravitational physics model, thereby enabling application of the model to territorial phenomena with an extra degree of freedom.

In economic terms, (2.13) represents 'generalized accessibility or interaction'. It measures the accessibility (or attractiveness) of every point with respect to the space that surrounds it.[39] A great deal of information is comprised in the generalized accessibility of a hypothetical zone i. It expresses 'demographic potential' when it estimates the commuter flows from all the other zones of the city to zone i; 'market potential' when it interprets the flow of people (potential customers) from all the other zones to zone i if it is an area of commercial activity; 'income potential' when a per-capita income is associated with individuals (potential customers); and 'location potential' when it is used to explain location choices (made according to the place with the greatest location potential) and the set of flows away from that location (demand for mobility and transport). Finally, location potential also explains the value attributable to a particular location (the rent of the previous models) in view of its attractiveness and generalized accessibility.

(2.12) and (2.13) can be easily estimated with simple econometric models, following their conversion into logarithmic form and the consequent linearization of the equations. Knowing the physical distance between two cities (or between two places in a city), the populations of the two places and interaction flows between two places, T – or of one place with j other places, E – it is statistically possible to obtain the values of K, α, β and γ.

In operational terms, these models have obvious predictive capacity if they are used to estimate the potential impact of the location of a new productive activity in a particular area. In the case of a project to build a shopping centre in a generic zone i, for example, once the values of K, α, β and γ have been estimated, and once the distance between two points in space and the increase in population expected in the area (e.g. new jobs) are known, equation (2.12) is able to predict the amount of people who will move from a zone of the city to zone i. Furthermore, the model of potential (equation 2.13) is able to predict the demand for transport, the market potential (potential number of shoppers at the shopping centre) and the income potential associated with construction of the shopping centre.

2.8 Conclusions

The chapter has described models, of a strictly neoclassical nature, that seek to account for the allocation of land between alternative activities within a spatial structure of uniform supply in space and a punctiform source of demand. High demand for access to central areas triggers competition between firms and households to obtain locations closer to the market, or more generally to a hypothetical central business district.

Land rent is the main factor that organizes activities in urban space. According to strict economic logic, competition for land closer to the centre is resolved by its allocation to activities able to pay higher rents.

The virtues of these models are their rigour and their stringent economic logic. Their main weakness emerges when they set out to explain the location choices made by households and firms between cities with different levels of utility or profit.

Indifference to alternative locations, which is the long-period equilibrium condition, is guaranteed if and only if cities offer the same utility and the same profit; and therefore, according to the model's logic, if and only if cities are of the same size. Yet this implies the existence of an urban system consisting of cities that are all of the same size – a circumstance widely belied by reality. In order to understand the economic reasons for the existence of urban systems with cities of different sizes, consideration must be made of the functional characteristics of cities. This is an aspect that the models described thus far are unable to handle, and that is instead addressed by the models discussed in the next chapter.

Review questions

1. What principle is behind the organization of activities in space in a monocentric city?
2. How is competition for central locations among alternative activities solved?
3. How is rent defined in the von Thünen model?
4. How is location equilibrium of productive activities in Alonso's model achieved?
5. How does Alonso's model change when residential activities are taken into consideration?
6. How would you define the 'Muth condition'?
7. Why is the bid rent also defined 'indifference curve in urban space'?
8. What are the main purposes of general location equilibrium models?
9. What is the 'boundary rent curve' and how is the location choice achieved in a general equilibrium model?
10. Who are the main critics who moved to the 'new urban economics'?

Selected reading on empirical findings

About land use and land prices

Brueckner J.K., Thisse J.-F. and Zenou Y. (1999), 'Why Is Central Paris Rich and Downtown Detroit Poor: An Amenity-Based Theory', *European Economic Review*, vol. 43, no.1, pp. 91–107.

Fanning Madden J. (1981), 'Why Women Work Closer to Home', *Urban Studies*, vol. 18, no. 2, pp. 181–194.

Gin A. and Sonstelie J. (1992), 'The Streetcar and Residential Location in Nineteenth Century Philadelphia', *Journal of Urban Economics*, vol. 32, no. 1, pp. 92–107.
Wheaton W. (1977), 'Income and Urban Residence: An Analysis of Consumer Demand for Location', *American Economic Review*, vol. 67, no. 4. pp. 620–631.

Further reading

Alonso W. (1964), *Location and Land Use: Towards a General Theory of Land Rent*, Harvard University Press, Cambridge, Mass.
Beckmann M.J. (1969), 'On the Distribution of Urban Rent and Residential Density', *Journal of Economic Theory*, vol. 1, no. 1, pp. 60–68.
Fujita M. (1989), *Urban Economic Theory: Land Use and City Size*, Cambridge University Press, Cambridge, Mass.
Huriot J.-M. (1988), *Von Thünen: économie et éspace*, Economica, Paris.
Richardson H. (1977), *The New Urban Economcs: and Alternatives*, Pion, London.

Notes

1 See Isard, 1956, Wingo, 1961 and Beckmann, 1969. An unpublished version of Beckmann's study dates back to 1957, which was prior to Alonso's work. See Alonso, 1964b.
2 See Beckmann, 1969; Montesano, 1972; Mills, 1972; Mirrlees, 1972; Solow, 1972; Anas and Dendrinos, 1976; Richardson, 1977, Fujita, 1989.
3 For firms see Alonso 1960 and 1964b; for households see Muth, 1961 and 1969; Alonso, 1964b.
4 See von Thünen, 1826.
5 Interestingly, the problem that induced von Thünen to develop his theoretical model was a highly practical one: how to organize agricultural production on his own estates.
6 This result is important, because it has been obtained by all the models developed since von Thünen's. It states that rent is nothing other than a saving in transport costs made possible by more central locations. From this follows the 'indifference to alternative locations' condition, which is reached by an individual or a firm when a move in space costs nothing; that is, when the saving in transport costs obtained by moving one kilometre closer to the centre equals the cost of the land that must be purchased to do so. See Samuelson, 1983.
7 See Ricardo, 1971; orig. edn, 1817.
8 As repeatedly pointed out, by eliminating everything except the distance between land and the town from concrete geographic space, von Thünen defined a new type of space: namely economic space. See Huriot, 1988.
9 See Alonso, 1960 and 1964b; Muth 1961, 1968 and 1969.
10 See Ricardo, 1971.
11 See Camagni, 1992a, chap. 9.
12 See Camagni, 1992a.
13 Microeconomics defines the utility function as the relation between a person's well-being (expressed in terms of the level of satisfaction – i.e. utility – that possession of a good generates for him/her) and the quantity of goods that s/he possesses.
14 The curve slopes downwards and is convex, thus indicating that the law of the diminishing marginal utility of goods applies. This law states that if a person possesses a large quantity of a particular good, the utility accruing to him/her from acquiring additional units of that good is so small that s/he is willing to exchange units of goods for even very small quantities of another good without this altering his/her satisfaction (utility). Vice versa, if the person possesses limited quantities of a good, his/her utility will be increased by possession of additional units to such an extent that a very large quantity of the other good must be offered before s/he is willing to exchange units of the first good. Only thus will his/her utility remain constant.
15 House size remaining equal, if the individual obtains greater quantities of other goods, his/her satisfaction (expressed in terms of utility) will increase.

16 Equation (2.8) is obtained by using the Lagrangian (£):

$$£ = u(q,z) - \lambda(r_d q + p_z z + \tau d - Y) \qquad (2.1\text{n})$$

Setting the partial derivatives with respect to q and z equal to zero, we obtain:

$$\frac{\partial £}{dz} = u_z' - \lambda p_z \qquad u_z' = \lambda p_z \qquad (2.2\text{n})$$

$$\frac{\partial £}{dq} = u_q' - \lambda r_d \qquad u_q' = \lambda r_d \qquad (2.3\text{n})$$

Dividing (2.2n) by (2.3n) and bearing in mind that by definition of indifference curve $u_z' dz = -u_q' dq$, we obtain (2.8).

17 This situation reflects the traditional consumer optimal choice equilibrium of microeconomic consumption theory. Here the situation is complicated by the fact that the price of one of the two goods (the size of the house) is influenced by the quantity of the other good, distance, chosen by the household.

18 In analytical terms, this problem can also be solved by taking (2.5) as the constraint and maximizing, for that level of utility, the rent offered by households:

$$\max r(d) = \frac{Y - p_z z - \tau d}{q(d)} \qquad (2.4\text{n})$$

$$\text{s.t. } u^* = u(q,z,d) \qquad (2.5\text{n})$$

The system can be solved with the Lagrangian in function of distance and u^*. This system is the dual of the one outlined at note 15. The results do not change, and in particular (2.9) still holds. See Alonso, 1964b, chap. 2.

19 See Fujita, 1989, chap. 2, p. 23.

20 This condition is obtained by setting the third partial derivative of the Lagrangian (2.1.n) equal to zero with respect to distance:

$$\frac{\partial £}{\partial d} = -\lambda\left(\frac{\partial r(d)}{\partial d} q + \tau\right) = 0 \qquad (2.6\text{n})$$

Solving (2.6n) for $\partial r(d) / \partial d$, we obtain:

$$\frac{\partial r(d)}{\partial d} = -\frac{\tau}{q} \qquad (2.7\text{n})$$

which is equation (2.9).

21 When an individual obtains a higher income, the time taken to commute rather than work (called the opportunity cost of transport) is of greater value.

22 Because of the enormous number of studies published in this field, it is impossible to provide an exhaustive list of references. To be mentioned in particular, however, are Miyao, 1981, 1984 and 1987a; Fujita, 1985 and 1989; Kanemoto, 1987; Miyao and Kanemoto, 1987. For surveys see Wheaton, 1979; Huriot, 1994; Derycke, 1996.

23 Hypotheses on the model's exogenous variables vary according to the approach. Solow and Mills take the population wanting to settle as given and thereby obtain, *inter alia*, individual utility. See Mills, 1972; Solow, 1972. Fujita instead hypothesizes a situation in which utility is known, instead of the population. See Fujita, 1989, chap. 2.

24 The first part of this model, which identifies the indifference to alternative locations conditions for an individual firm (to be then extended to n firms) has an entirely similar logical structure to that of the partial equilibrium location model for households described in Section 2.4. Here the size of the house is substituted by the firm's intensity of land use.

25 A production function is the relation between the quantity of a firm's output and the quantity of production factors used by the firm in its productive process. The functional form expressed by (2.10) was first proposed by Cobb and Douglas. See Cobb and Douglas, 1928.

26 The expression 'constant returns to scale' refers to the relation between variations in output and equiproportional variations in all the inputs. If the inputs are increased and the output increases proportionally, the firm produces at constant returns to scale; if the output increases more than proportionally, the firm achieves increasing returns to scale (economies of scale); if the output increases less than proportionally, the firm has decreasing returns to scale (diseconomies of scale).

27 This problems translates into a system that minimizes costs (C) at each distance (d) from the centre, under the constraint of achieving a certain total revenue ($\overline{V}$):

$$\min C_d = r(d)T_d + p_K K_d \tag{2.8n}$$

$$s.t.: \overline{V} = (p_y - \tau d)Y_d = (p_y - \tau d)aT_d^{\alpha}K_d^{\alpha-1} \tag{2.9n}$$

where p_y represents the price of the good y, τ the unit transport costs, and d the distance from the centre. The solution of the system is obtained by means of the Lagrangian:

$$£ = r(d)T_d + p_K K_d - \lambda(p_y - \tau d)\overline{Y}_d - \overline{V} \tag{2.10n}$$

Setting the partial derivatives equal to zero, we obtain:

$$\frac{\partial £}{\partial T_d} = r(d) - \lambda(p_y - \tau d)\frac{\partial Y_d}{\partial T_d} = 0 \tag{2.11n}$$

$$\frac{\partial £}{\partial K_d} = p_K - \lambda(p_y - \tau d)\frac{\partial Y_d}{\partial K_d} = 0 \tag{2.12n}$$

from which we get:

$$r(d) = \lambda(p_y - \tau d)\frac{\partial Y_d}{\partial T_d} = \lambda(p_y - \tau d)MaP_{T_d} \tag{2.13n}$$

$$p_K = \lambda(p_y - \tau d)\frac{\partial Y_d}{\partial K_d} = \lambda(p_y - \tau d)MaP_{K_d} \tag{2.14n}$$

where MaP represents the marginal productivity of land and capital respectively; i.e. the variation in production when one additional unit of a factor is included in the input. Dividing (2.13n) by (2.14n), we obtain:

$$\frac{r(d)}{p_k} = \frac{\partial K_d}{\partial T_d} = \frac{MaP_T}{MaP_K} \tag{2.15n}$$

The firm will obtain maximum profit (or minimum cost, given a certain revenue) when the ratio between the marginal productivity of the two goods (MaP_{T_d} / MaP_{K_d}) equals the ratio between the prices. In other words, the firm gains by substituting one factor with the other until the productivity of the latter compared with that of the former is higher than its relative price.

28 (2.11) is obtained by deriving the Lagrangian expressed in (2.10n) with respect to d and setting the first derivative equal to zero. This yields:

$$\frac{\partial £}{\partial d} = \frac{\partial r(d)}{\partial d}T_d + \lambda Y_d \tau = 0 \tag{2.16n}$$

whence:

$$\frac{\partial r(d)}{\partial d} = -\tau \frac{Y_d}{T_d}(\lambda) \text{ or } \frac{\partial r(d)}{\partial d}\frac{T_d}{Y_d} = -\tau(\lambda) \qquad (2.17n)$$

which is (2.11).

29 A distinctive feature of 'new urban economics' models is their introduction of a production function that allows for substitutability between production factors. The bid-rent curve is thus convex and slopes downward as the distance from the centre increases. This pattern shows that closer to the centre of the city (von Thünen's market place), the location equilibrium is also defined by the elasticity of substitution between the two factors (goods in the residential activity equilibrium). The term 'elasticity of substitution' between factors (goods) denotes the percentage by which the quantity of one of the two factors (goods) must be increased in order to offset the reduction by one unit of the other factor (good) so that the producer's (consumer's) total production (utility) remains the same.

30 The boundary rent curve was first proposed by Fujita in his outstanding book of 1989.

31 In mathematical terms, this means that in order to obtain a general equilibrium, two further conditions besides those already stated must be imposed. The first concerns land use: the demand for urban land use must exhaust the supply of land:

$$T_d = 2\pi ds \qquad (2.18n)$$

where s is the percentage of urban land used for production, and $2\pi d$ represents the area of the circle.

The second condition is that the market for the good must be in equilibrium, thereby ensuring stability in the number of firms operating in the market:

$$\int_{d=0}^{d_{max}} Y_d \partial d = D(p) \text{ with } D'(p) < 0 \qquad (2.19n)$$

where $D(p)$ is demand for the good, which is a function of its price p. Conditions (2.18n) and (2.19n) are both fulfilled along the boundary rent curve: hence, the firms in the market occupy a quantity of land equal to the available supply of urban land; at the same time the number of firms exactly covers demand for the good in the market.

Assuming as known the rent on the edge of the city, which is equal to agricultural rent:

$$r(d_{\max}) = r_a \qquad (2.20n)$$

it is possible to solve the system of equations (2.10 – 2.11; 2.13n – 2.14n; and 2.18n – 2.20n) and obtain the equilibrium values: from equation (2.10), (2.18n) and (2.13n) the values for land and capital (T and K) and for production (Y) at every distance from the centre in function of the price of the good (p), of the price of capital (P_K) and of the cost of land (r); from equations (2.14n) and (2.11) the profit rate and rent for every distance (except for one constant, the urban edge rent, which is given as exogenous); finally, from (2.19n) the price of the good for which demand equals supply. (2.20n) defines the extreme boundary of the city, its size, and closes the model with definition of the absolute value of rent.

32 The size of the city whereby total population equals the population given exogenously, equal to $\bar{n}$, is given by:

$$\int_{d=0}^{d_{max}} \frac{2\pi d}{q(d,u)} \partial d = \bar{n} \qquad (2.21n)$$

If the amount of space required by each household (q) increases – for instance because the utility of location in the city increases – residential density ($1/q$) decreases. See Fujita, 1989, chap. 3, p. 57.

33 Once again surprising is how these models turn out to be only more modern versions of von Thünen's model.

34 See Camagni, 1992a, chap. 6.

35 Various theories have been propounded regarding the relation between territorial phenomena and the law of gravitation of celestial bodies. It was not until the early 1970s, however, that the first convincing account appeared in the form of Wilson's entropy principle. See Wilson, 1970 and 1971. For previous theories see Stouffer, 1940 and 1960, and his interposed opportunities approach. See also Niedercorn and Bechdolt, 1969 and 1972, who developed the individual utility approach to movements in space.

36 Zipf, 1949. Several theories had been formulated before Zipf, most notably Reilly's 'law of retail gravitation'. See Reilly, 1931. For recent empirical examinations of the Zipf's law, see Ioannides and Overman, 2003; Soo, 2005.

37 According to the law of universal gravitation, two celestial bodies attract each other with a force proportional to the product of their masses and inversely proportional to the square of the distance between them:

$$T_{ab} = K(M_a M_b) / d_{ab}^2 \qquad (2.22n)$$

On inserting a generic parameter γ as the exponent of distance, and substituting population for the mass of the two bodies, we obtain (2.12).

38 Also (2.13) derives from an analogy with gravitational physics that states that every unit mass *a* in the gravitational field of a mass *b* has a potential energy *E* equal to the work that *a* would yield by falling to *b*:

$$E_{ab} = KM_b / d_{ab} \qquad (2.23n)$$

Assuming different force fields, the total potential produced on *a* by a set of masses is defined as:

$$E_a = K\sum_j M_j / d_{aj} \qquad (2.24n)$$

which is equation (2.13) when the exponent of distance assumes value 1.

39 See Camagni, 1992a, chap. 3.

3 Hierarchy and location

3.1 Hierarchy and urban systems

The location theories discussed in the previous chapters analysed the location choices of individual firms or people. They disregarded, however, the existence of other activities or individuals and of dichotomous location alternatives: urban or non-urban areas, central or peripheral ones, areas with high or low concentrations of activities. When they considered the existence of several activities, they ruled out the possibility that these might locate in alternative urban centres. And when they dealt with several cities, they reached the somewhat paradoxical conclusion that the existence of urban systems apparently in equilibrium entailed that those cities must all be of the same size. Only thus could indifference to alternative locations be guaranteed because the levels of profit and utility were the same in all the cities (see Section 2.5).

Thus far, therefore, we have not met theories able to explain the location choices of several firms and households among alternative urban centres. Nor, consequently, have we found a theory of location able to explain why in reality there exist numerous cities, of different sizes and performing different functions, that depend partly or wholly on larger cities for higher-quality services and activities. In other words, we have so far been unable to explain why an urban hierarchy exists.

The theories presented in this chapter seek to account for the existence of urban systems made up of cities of different sizes. The aim of these theories is to formulate rules able to interpret the *urban hierarchy* by explaining:

- the size and frequency of urban centres at every level in the hierarchy, and therefore the market area of each of them;
- the distance between a particular city and those at the levels immediately below or above it, and therefore the geographical distribution of all the urban centres.

The founders of this school of thought, known as 'central place theory', were the geographer Walter Christaller and the economist August Lösch.[1] These were the first to formulate models able to explain the urban hierarchy (Sections 3.2. to 3.4) and they prepared the ground for subsequent analyses (Sections 3.5 to 3.6).

3.2 The geographical approach: Christaller's model

3.2.1 The original model

Christaller's model is based on the assumption that an urban centre exists where there are goods and services to be traded. This central 'place' (hence the name 'central place theory' given to the literature that Christaller's model inspired[2]) must produce or supply goods or services to a population spatially dispersed across a uniform and isotropic surrounding territory.[3] The aim of the model is to show how products and services (especially tertiary functions) come to be territorially organized into an urban hierarchy.

For this purpose Christaller introduces the concepts of *threshold* and *range*. These express in geographical terms the economic forces that organize activities in space: transport costs and agglomeration economies, or economies of scale. The range of a service is the maximum distance that consumers are willing to travel to purchase it (which includes the maximum transport costs that they are willing to pay in doing so). The threshold of a service is the distance that, when rotated around the supply centre, marks out a circular area with sufficient population to generate a level of demand such that the service can be produced profitably. A service is produced only if the range exceeds the threshold. In other words, a service is produced only if there is sufficient demand for it to be supplied at a profit.[4]

The central place is located at the centre of a circular market area, which is the optimal location because consumers located in the area are able to minimize their total transport costs.[5]

In equilibrium, the circular market areas defined by the range of the service assume the shape of a hexagon. This geometric shape enables Christaller to maintain three fundamental assumptions: (i) minimization of transport costs for consumers (the hexagon, in fact, is the geometric shape closest to a circle); (ii) even distribution of the service supply, so that the territory is covered without areas being left unserved; and (iii) competition among producers, which requires that market areas must not overlap.[6] In equilibrium, a 'honeycomb' lattice consisting of n centres producing for n hexagonal market areas, all of the same size, arises in space (Figure 3.1).

According to Christaller, moreover, each service has a range that determines the size of its market area: high-quality services, produced and supplied in large urban centres,

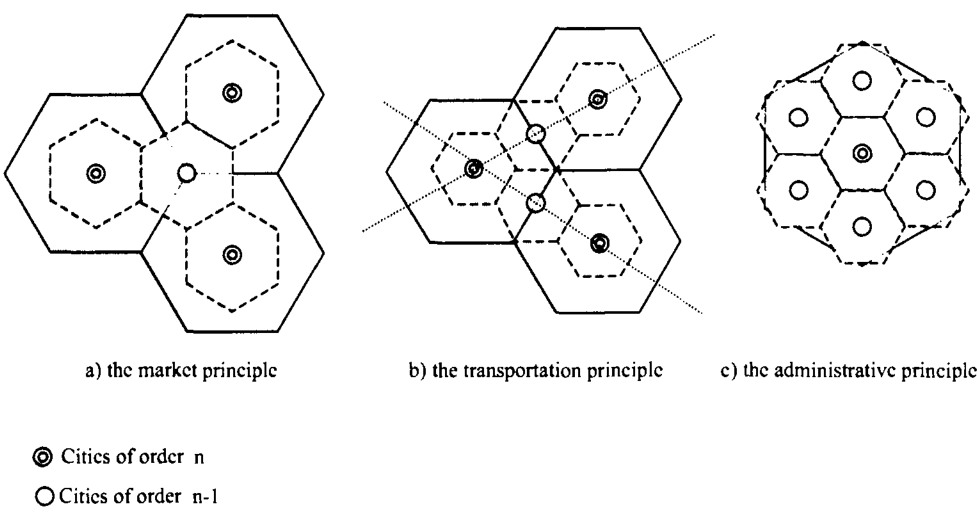


Figure 3.1 Organization of market areas according to Christaller's three principles

have more extensive ranges that delineate market areas broader than those in which lower-quality services are supplied.

Having defined the regular hexagon-shaped market areas – where an *n*-order service is supplied – Christaller identifies the market areas of the immediately lower-order service. He hypothesizes for this purpose that the first relative production units of the lower-order service choose to locate in the central place where higher-order services are already being produced – that is, the centres of the hexagons – so that they can benefit from agglomeration economies.

Because the range of the lower-order service is by definition less than that of the higher-order service, the market area served by production units located at the centre of the hexagon is smaller than the hexagon itself. Consequently, a part of the territory is left uncovered. This unsatisfied demand attracts new service production units into the area. These choose their locations according to three different principles that Christaller envisages as shaping market areas in space:

- the *market principle*, which postulates location equidistant from a triad of higher-order centres represented by the vertex of the larger-sized hexagon (Figure 3.1a). Optimization of this location fulfils the criterion of minimizing the number of centres able to cover all the territory of the higher-order market. According to this location pattern, there are 1 + 6/3 = 3 lower-order centres in a higher-order market area;
- the *transportation principle*, which applies to a location equidistant from a pair of higher-order centres (Figure 3.1b). This choice optimizes the location of lower-order centres on the basis of minimization of transport costs to the higher-order centres. In each higher-order market area there are 1 + 6/2 = lower-order centres;
- the *administrative principle*, identified by location in the centre of the triangles making up the hexagon (Figure 3.1c), so that the purpose of optimization is to prevent higher-order centres from competing to administer lower-order ones. This aim is achieved if the lower-order centres pertain to a single higher-order centre. In this pattern, there are 1 + 6 = 7 lower-order centres for each market area of a certain order.

The model thus generates a hierarchy of urban centres: for each centre (or market area) of order *n* there are *k* centres (market areas) of order *n-1*; *k* is the factor of proportionality between the centre of a certain order and the one immediately below it, and it assumes value 3, 4 or 7 according to the predominant location principle (market, transportation or administrative).[7] In Christaller's model, this proportionality factor is constant throughout the urban hierarchy. For each *k*, simple rules can be applied to obtain the number of centres of each order, the distance between the centres of each order and the size of the market area.[8]

The model reaches an important conclusion: each large centre produces the goods/services relative to its hierarchical level and all lower-order goods/services. The large centre's advantages therefore derive from the functional level typical of its hierarchical order. Hence, the size of the city becomes a proxy for the urban function, and each higher-order centre has a descending array of lower-order centres until the lowest-level agglomeration is reached.[9]

Christaller's model therefore generates a system of hierarchical spatial relations that gravitate on the surrounding market area. Although the model is mainly geographical in nature, it gains robust internal consistency from the economic postulates on which it is based:[10]

- *optimal behaviour by consumers*, who minimize transport costs so that they can purchase the service offered. The market areas are separate from each other and do not overlap;
- *homogeneous geographical space* in which the agglomeration of activities comes about for economic, not physical-geographic, reasons;
- *a transportation cost proportional to the distance covered*;
- *existence of economies of scale*, these being implicit in the concept of threshold;
- *equity in supply of the service*, which is implicit in the statement that the territory must be covered so that all consumers have access to all services/goods.

When Christaller applied his model to reality, he obtained surprising results. He first analysed the urban structure of Southern Germany, exogenously defining six levels of centres with a centrality indicator consisting of the number of telephones connected to the interurban network. When he then applied the market principle, he found a striking correspondence between the number of centres identified by his model and the number that actually existed in reality:

Hierarchical level	1	2	3	4	5	6	7
Theoretical number of centres	1	2	6	18	54	162	486
Observed number of centres	1	2	10	23	60	105	462

It should be stressed that Christaller's model is able to answer the questions put at the beginning of the chapter: it demonstrates the existence of an urban hierarchy in which each city of a certain size performs a specific function. Moreover, the model is able to furnish rules with which to identify the number of centres of a certain order, the size of each market area of each centre, the distances among centres of the same order, and therefore their geographical distribution.

3.2.2 *Mathematical formalization*

The Christaller model was purely qualitative in its original formulation. However, a very simple quantitative version of it has recently been proposed.[11] Let p_1 denote the population of the lower-order settlement, and r the population of the rural area depending on p_1. The population of the area served by p_1, called P_1, is easily identified:

$$P_1 = p_1 + r \tag{3.1}$$

On the hypothesis that each city has a population in its area which is a constant fraction c of the area, i.e.

$$p_j = cP_j \text{ con } 0<c<1 \tag{3.2}$$

(3.1) can be rewritten as:

$$p_1 = c(p_1 + r) \tag{3.3}$$

so that:

$$p_1 = \frac{cr}{1-c} \tag{3.4}$$

(3.4) states that the population of the city of order 1 is equal to *c/(1 – c)* times the population of the rural area. In the central place literature *c/(1 – c)* is termed the 'urban multiplier'.

Let us assume that there are n levels of urban centres and that each centre serves itself and s 'satellite' centres around it.[12] The population of a region served by a higher-order city, called P_n, is obtained from the population of the lower-order area which it controls *(1+s)*, considering that centre n maintains a population of order n, not n – *1*, within its area:

$$P_n = P_{n-1}(1+s) - p_{n-1} + p_n \tag{3.5}$$

Bearing (3.2) in mind, (3.5) can be rewritten as:

$$P_n = P_{n-1}(1+s) - cP_{n-1} + cP_n \tag{3.6}$$

so that:

$$P_n = \left(\frac{1+s-c}{1-c}\right) P_{n-1} \tag{3.7}$$

Because in Christaller's model s and c are constant throughout the urban hierarchy, (3.7) identifies a constant relation between the size of the region's population and that of the lower-order region.

(3.7) can be written in generic form:

$$P_n = \left(\frac{1+s-c}{1-c}\right)^{n-1} P_1 \tag{3.8}$$

Substituting P_1 with (3.4), when we multiply the right-hand member by *(1+s – c)/(1+s – c)*, we obtain:

$$p_n = \left(\frac{1+s-c}{1-c}\right)^{n} \frac{rc}{1+s-c} \tag{3.9}$$

(3.9) states that, knowing the population of rural settlements r, we are able to find the size of the market area and the population of centres of any order whatever.

3.3 The economic approach: Lösch's model

3.3.1 The original model

In 1940, Lösch developed a general equilibrium model in order to remedy a major shortcoming of Christaller's model: its assumption of a proportionality factor constant throughout the urban hierarchy.

Lösch's model also generates a hexagonal structure of market areas, but it does so on the basis of purely economic principles:

- competition among firms: this does not permit the existence of uncovered market areas, since the potential profits available in non-controlled spatial markets attract new firms into those areas;
- consumer rationality: when consumers have to choose between two possible suppliers, their rationality induces them to select the one which offers the good at the lowest price; and therefore, according to the logic of the model, they select the producer located closest to them (see Section 1.5).

Lösch's model defines market areas by explicit (though exogenous) cost and demand curves of goods, and thus achieves spatial equilibrium of an individual sector with explicit reference to Chamberlin's monopolistic competition market (see Section 1.5). The model identifies a stable spatial economic equilibrium in the hexagonal market areas that arise when firms no longer have incentives to enter the market.

Unlike Christaller, Lösch identifies several factors of proportionality – also called nesting coefficients – that operate up the urban hierarchy: Christaller's k = 3, 4 and 7 is still valid, but other values of the coefficient of proportionality are considered, in particular 9, 12, 13, 16, 19, 21 (Figure 3.2). Lösch assumes, in fact, that there is a specific value of the nesting coefficient, and therefore a specific size of the hexagonal market areas, corresponding to each type of good or service. Lösch's coefficients are simple geographic multiples of Christaller's coefficients (3, 4 and 7), and they therefore comply with Christaller's three principles:[13]

- 9 and 21 with the market principle: 9 = 1 + 6 + 6/3, and 21 = 1 + 6 + 6 + 6 + 6/3;
- 16 with the transportation principle: 16 = 1 + 6 + 6 + 6/2;
- 13 and 19 with the administrative principle: 13 = 1 + 6 + 6 and 19 = 1 + 6 + 6 + 6;
- 12 with the market principle and then the transportation principle: 12 = 1 + 6 + 6/3 + 6/2.

Relaxing the assumption of a proportionality factor constant throughout the urban hierarchy has significant consequences. It eliminates the two-way relation between size of the centre and specialization, and it enables consideration to be made of such important empirical situations as:

- the different functional specializations of centres of the same size;
- the possibility that centres have functional specializations: hence a centre may perform only the function of its order and not of all the others, as the Christaller model imposes.

The way in which Lösch arrives at the overall structuring of the territory on the basis of his hypotheses is interesting but analytically unsatisfactory. For Lösch, the organization of economic space results from the superimposition of several hexagons, of different sizes and structures (corresponding to different types of goods and services). All hexagons have a common centre, which produces all the goods (Figure 3.3a).

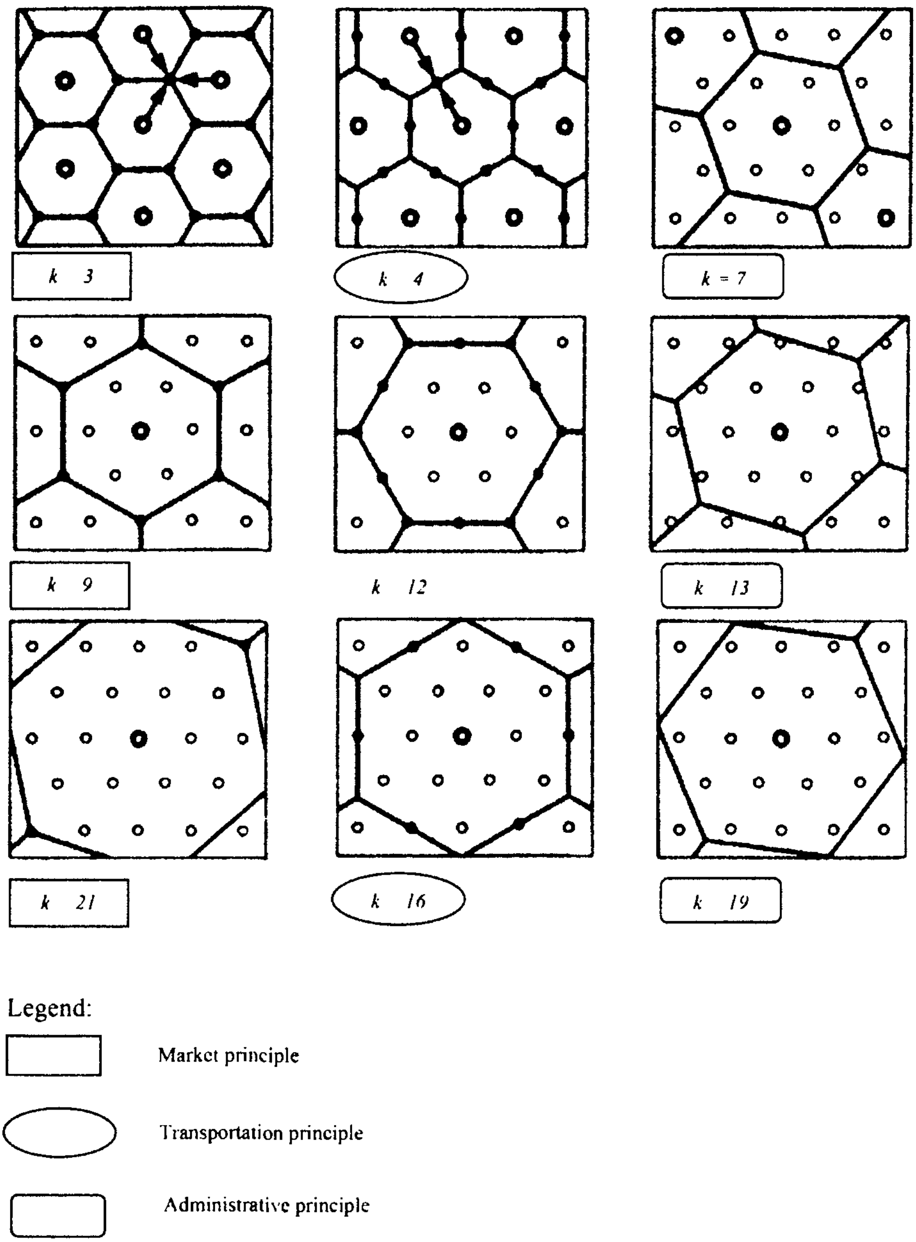
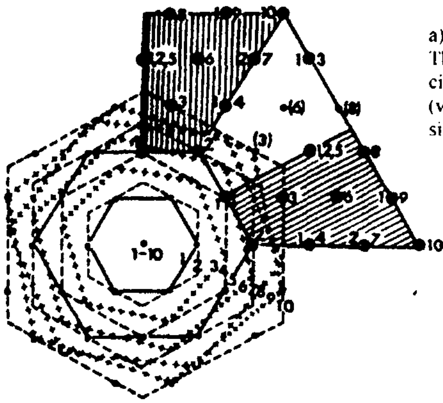
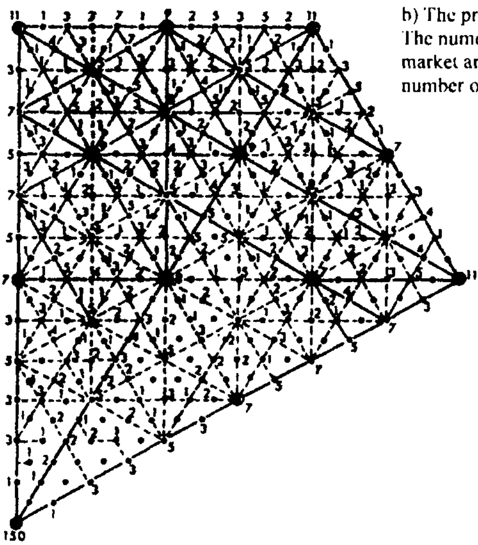


Figure 3.2 The nine most compact patterns of the organization of centres

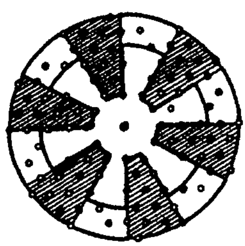
The final structure of the territory is obtained by rotating the superimposed hexagons to obtain the maximum density of centres in some areas, and the maximum coincidence of different production locations (Figure 3.3b). The result is a series of circular alternate sectors of high and low settlement densities. These sectors radiate out from a large city to form a structure that complies with an efficiency principle for the transport system (Figure 3.3c).[14]



a) The first 10 smallest hexagonal structures.
The hatching indicates the areas where the largest number of cities are located. The dots represent original settlements (villages); the circled dots are the market area centres of the size indicated (from 1 to 10).



b) The production activities present in the various centres.
The numerals refer to the number of centres of specific market areas that coincide at that point, and therefore to the number of production sectors present.



c) The overall landscape.
The hatching indicates the areas where the highest number of cities is located.

Figure 3.3 Lösch's model

Lösch's model yields more realistic and less paradoxical patterns than those produced by Christaller's model. However, they are obtained at the expense of blurring the concept of urban hierarchy: there are no elements in the high density spaces of urban centres that can be used to identify a hierarchical structure of centres – due to a lack, amongst other things, of a clear division of labour among them.[15]

Also Lösch demonstrates the validity of his model empirically, in his case by applying it to the territory of Iowa in the United States. Hypothesizing an urban structure comprising six hierarchical levels and a nesting factor of 4, he obtains the following results:[16]

Hierarchical level	0	1	2	3	4	5
Theoretical no. of centres	0–1	2–3	9–10	39	154	615*
Actual no. of centres	0	3	9	39	153	615
Theoretical dist. among centres (miles)	179	90	45	22	11	5.6*
Observed dist. among centres (miles)		94	50	24	10	5.6
(*) = observed value						

As said, for empirical verification of his model Lösch preferred to use a nesting coefficient of 4, which corresponds to Christaller's proportionality coefficient for his transportation principle. This principle is compatible with the territory selected by Lösch for his empirical analysis, which was a plain traversed by rectilinear transport infrastructures.[17]

3.3.2 *Mathematical formalization*

Like Christaller's model, Lösch's has also attracted the interest of numerous economists, some of whom have proposed a mathematical formalization of it.[18]

The formal approach uses five equations to state the economic conditions that the model regards as crucial for achievement of general spatial economic equilibrium. The latter arises from two specific factors: each producer seeks to maximize his/her profit; and each consumer seeks to maximize his/her utility by accessing the least costly market. Moreover, the existence of several firms in the same sector gives rise to competition among them which nullifies extra-profits.

The conditions are the following:

1. The locations of producers must be the most advantageous possible, so that no movement in space improves profitability;
2. the number of possible locations must be such that coverage of the entire territory is guaranteed;
3. extra-profit must be nullified, so that stability is ensured in the market and the entry of new firms is prevented;
4. the firm's volume of production and the size of its market area must be as small as is compatible with its average production costs. If change in the size of the market is associated with a price increase greater than the increase in average production costs, this will generate extra-profit and therefore create room for new firms to enter the market;
5. the good's sale prices on the border between market areas must be equal in order that consumer indifference between alternative locations is guaranteed at the border.

These economic conditions can be straightforwardly translated into equations.
Defining:

m	=	generic good, in which $m = 1, 2, \ldots$ M;
n	=	generic centre of goods production, in which $n = 1, 2, \ldots$ N;
x_{mn} e y_{mn}	=	the spatial co-ordinates indicating each production centre's position in space. There are two of them for each place n;
X_{mn}	=	quantity produced of good m in the production centre n;
Π_{mn}	=	total profit obtained from sale of the good m in the production centre n;
c_m	=	average production cost of good m;
p_m	=	sale price of good m;
A	=	total size of the territory, while $A_{m1}, A_{m2}, \ldots, A_{mn}$ represent the sizes of the market areas in which the generic good m is sold;
C_{mn}	=	boundary of the market area of the production centre n for good m;
d_{mn}	=	the distance between the production centre n and the boundary of its market area;
τ	=	the unit cost of transportation for the good m;

we can rewrite the five conditions as the following equations:

1 First-order conditions ensuring profit maximization in each location:

$$\frac{\delta\Pi_{mn}}{\delta x_{mn}} = 0 \text{ and } \frac{\delta\Pi_{mn}}{\delta y_{mn}} = 0 \qquad \text{2N in number (2 for each market area } n\text{)}$$

2 Condition for the territory to be entirely covered by the supply of each individual good:

$$\sum_{m=1}^{M} A_m = A \qquad \text{M in number}$$

3 Condition for the absence of extra-profit in each production centre:

$$p_{mn} = c_{mn} \qquad \text{N in number}$$

4 Condition of minimum market area size so that marginal changes in average costs equal marginal changes in the good's price in each production place:

$$\frac{\delta p_{mn}}{\delta X_{mn}} = \frac{\delta c_{mn}}{\delta X_{mn}} \qquad \text{N in number}$$

5 Condition of equality in prices at each market-area boundary:

$$p_{mn} + \tau d_{mn} = p_{mn'} + \tau d_{mn'} \qquad \text{C in number}$$

The number of equations to be estimated in the model is therefore 4N + M + C, which is equal to the number of unknowns, these being:

- the size of the market areas A_{mn}, which are equal in number to the centres N;
- the prices P_{mn} for each centre, N in number;
- the number of production centres N_m, for each good *m*, M in number;
- the co-ordinates of each centre x_{mn} and y_{mn}, 2N in number;
- the boundaries of the market areas, C.

Equality between the number of unknowns and the number of equations is the necessary condition for a solution to exist. However, because the system is non-linear, there may be multiple solutions or no solution at all. There are various ways to simplify the system of equations. The number of unknowns can be reduced by imposing a regular shape on the market areas, as in Lösch's original model. In this case, each area's number of boundaries becomes known. Moreover, with $N_m - 1$ market areas of the good *m* being known, the *n*-th market area is the difference from the total size of market A.

The formalized version of Lösch's model has a number of shortcomings, however:

- because the number of centres in which production takes place is not known, solution of the model is highly complex, in that the number of equations to estimate in order to satisfy the first condition is *a priori* unknown;
- even more than in the original non-formalized model, no account is taken of the possible existence of agglomeration economies in the form of both the urbanization economies typical of a higher-order centre and the localization economies of more specialized centres;
- the non-linearity of the system is such that it guarantees neither the existence nor the uniqueness, nor the stability of the equilibrium solution.

3.4 Critical appraisal of the two models

Christaller's and Lösch's models are generally recognized to be the first models of general spatial equilibrium. Assuming uniform space, they were able to explain (i) the existence of sets of different-sized cities, (ii) the role of each of them and (iii) the distance among them. In short, these models were the first to make interpretation of city systems possible; and only very recently have new models of general location equilibrium been proposed. However, their logical structure appears unable to account for the existence, in equilibrium conditions, of urban centres of different sizes.[19]

Still today, therefore, Christaller's and Lösch's models are unique tools with which to interpret the structure of city systems. On the basis of the simple existence of the two well-known economic forces determining location choices (agglomeration economies and transport costs), location equilibrium arises from a logic of profit optimization for firms and utility optimization for customers.

However, the two models have a number of weaknesses widely emphasized in the literature.[20] Some of them stem from the abstract nature of certain assumptions – for example the spatial homogeneity of resources and constant unit transport costs – but they do not undermine the interpretative capacity of the models, and they can be easily justified by the need to simplify the complexity of the real world. But other shortcomings are more critical because they affect the models' inner consistency, namely:

- the lack of analysis of demand, this being assumed to be evenly distributed in space and immobile.[21] Likewise, these models exogenously define the quantity of each good demanded by each individual. In other words, they are approaches that largely pertain to production theory;
- the lack of interdependence between the production and location choices of firms. Given the models' hypothesis of geographically homogeneous production resources and demand, on the supply side proximity to other firms does not influence cost functions; on the demand side there is no interdependence – of complementarity or substitutability – between goods.[22] Demand for a good is independent of the price of other goods and independent of the locations of the producers of other goods;
- the static nature of the models. This restricts their use in analysis of the evolution and dynamics of the urban hierarchy.

The first two of these shortcomings are serious, for they reflect contradictions between the results and the initial hypotheses which undermine the interpretative logic of the models. Indeed:[23]

- the assumption that demand is uniformly distributed across the territory conflicts with the models' pivotal result that population concentrates in urban centres. The concentration of producers at certain points in space, in fact, rules out any possibility of homogeneous demand, because it induces consumers to resort to a central place offering optimal conditions (goods which are cheap because of low transport costs);
- the models ignore any location interdependence on both the supply and the demand sides. On the supply side, they fail to consider any input/output relation that might favour location choices dictated by proximity to other suppliers. As a consequence, they have no production function that ties the production of one good to another. On the demand side, they have no utility function which associates one good with another. The models admit no substitutability or complementarity and consequently take the form of superimposed partial equilibria. Paradoxically, Christaller, and more specifically Lösch, invoke a concept of agglomeration economies in order to explain firms' choices; yet they subsequently ignore the advantages of such economies, thereby undermining their models' inner consistency.

Despite these defects, however, given their uniqueness, and in many respects their unsurpassed economic and geographic conceptualization of the urban hierarchy, Christaller's and Lösch's models still today occupy a central place in spatial economics.

3.5 Some recent developments

Christaller and Lösch were the pioneers of what has been called the 'central place theory'. Since their ground-breaking studies, considerable efforts have been made to improve the original models and to remedy their shortcomings and internal contradictions. This section outlines some of these theories, providing references in the endnotes to texts that deal with formal aspects.

The first important achievement of these more recent theories has been to develop models of a more strictly economic nature, and of increasing analytical complexity, which enable demand aspects to be included in the general equilibrium. Among them, the model of Martin Beckmann and John McPherson has been widely successful.[24] This model is able to overcome the restriction imposed by a constant nesting coefficient (the k in Christaller's model, which Lösch also sought to eliminate), and the constant ratio between the population of the centre and the population of the surrounding area up the urban hierarchy (or within the same hierarchical level of services, when these are supplied by centres of different sizes). Beckmann and McPherson's model hypothesizes that the proportionality factor differs according to the hierarchical level of centres and the service supplied at level n. The significance of the proportionality factor is no longer geographic but economic (although it is measured in terms of population), and it states the total number of individuals resident in the centre of order n necessary to supply the corresponding service to each inhabitant of its market area (including the centre itself). In other words, it expresses the number of people necessary to produce the good for the entire market area.

More recently, Hubert Beguin has extended the Beckmann and McPherson model to include two economic aspects that the original model identified as determinants of the ratio between the centre's population and the population of the surrounding area:[25] labour productivity (i.e. the number of people necessary to produce one unit of the good m), and the structure of individual consumption (i.e. the quantity of good m demanded by each inhabitant of the market area of centre n). Beguin is able to show that, in reality, the structure of the urban hierarchy – in terms of the ratio between the central area's population and that of the rural area surrounding it – depends upon:

- variation of labour productivity across the various levels of the centre's hierarchy (increasing, decreasing and constant returns to labour);
- the distribution of the various types of goods/services in overall demand;
- the income elasticity of demand for various goods.

'Central place' theory has accomplished a second step forward by introducing the advantages of localization economies into Christaller's and Lösch's models. A noteworthy contribution in this regard has been made by W. Long,[26] who incorporates interdependence among goods into a model à la Christaller. He hypothesizes that the quantity purchased of a good m does not necessarily diminish with distance from the centre. Although the price of good m increases as one moves further away from the centre, it may subsequently decrease as one approaches other centres, owing to the localization economies deriving from proximity to these new centres. Long also states that the range of a good may change according to whether it is produced in a smaller centre or a larger one. To the latter he attributes a shorter range due to the presence of a larger number of substitute goods on the broader urban market. It is evident that the introduction of these hypotheses disrupts the regularity of Christaller's and Lösch's patterns, and that interdependence mechanisms, on both the demand and supply sides, may distort the results obtained by the original models. However, Long's model is no more than a preliminary exercise, whose mathematical complexity does not yet allow analytical solutions to be obtained.

Finally, central place theory has taken a third step forward in response to the need to understand not only the urban structure but also its evolution and its dynamics. Here the main contribution has been made by John Parr with his comparative statics model,[27] which analyses the evolution of the spatial organization of the urban hierarchy on the following hypotheses:

- the formation of successive levels of the hierarchy from the lowest to the highest;
- change in the allocation of economic functions at the hierarchy's various levels;
- alterations in the hierarchical structure, i.e. in the number of levels associated with the various sizes of centres: formation of a new level in the hierarchy, change in the extension in the market area of a hierarchical level, disappearance of a hierarchical level.

Parr modifies the structure of the hexagonal market areas envisaged by Christaller's original model; in the cases described above, they are transformed into rectangular or triangular or again hexagonal areas of varying sizes up the urban hierarchy.

3.6 Towards a new theory of urban systems: city networks

In recent years, the urban systems of the advanced countries have evolved in a manner evidently at odds with Christaller's hierarchy model. Medium-sized cities (40,000 to 200,000 inhabitants) have undergone marked development and are now characterized by close interdependencies among centres of the same order, pronounced productive specializations and the absence of hierarchical relationships within individual urban systems. Developments over the past 20 years therefore show that urban systems have little in common with Christaller's hierarchical structure. They instead display the following features:[28]

- urban specialization, especially in industry but also in services, which contradicts the Christaller model's prediction of the hierarchical despecialization of each centre;
- incomplete presence of the entire mix of functions in each city;
- high-rank functions in lower-order centres;
- horizontal linkages among cities performing similar functions: for example, the network of cities specializing in international financial services;
- synergies among similar centres performing advanced production functions and services, as exemplified by sub-regional industrial districts.

These empirical findings demonstrate the inadequacy of Christaller's traditional model, which fails to explain phenomena widely apparent in the evolution of the urban systems of the advanced countries. They have led to the development of a new conceptual paradigm – that of *city networks* – which furnishes a more convincing and coherent interpretation of emergent territorial patterns.

This new paradigm envisages the possibility that close relationships may arise among urban centres that co-operate and interact on the basis of specific economic relations. These relations may be vertical among cities of different orders, or (and this is the much more innovative aspect) they may be horizontal among cities of the same order that interact on the basis of complementarity or synergy. In light of these two processes, it is possible to identify two types of city network:[29]

- *complementarity networks* consisting of specialized and complementary centres linked by a set of input–output relations. Sectoral specialization guarantees economies of scale and agglomeration economies even in centres of small size. Examples of this type or urban network are the specialized cities of Randstad in Holland, or the polycentric structure of the cities of the Veneto region in Italy;
- *synergy networks* consisting of similar and mutually co-operating centres; economies of scale are guaranteed by the co-operation network itself, which links the markets of the individual centres together. Examples of this type of network are financial centres operating worldwide, whose markets are virtually linked by advanced telecommunications networks, or networks of cities connected by the religious tourism itineraries created during the Vatican Jubilee celebrations.

A third category, which can also be conceived as a sub-category of the second one, can be identified in:

- *innovation networks* consisting of centres that co-operate on specific infrastructural or productive projects in order to achieve a critical mass in terms of both demand and supply. Examples of this type of network are agreements among French cities for the construction of infrastructures.

The new paradigm for interpretation of city systems has numerous novel features. First, the concept of 'city network' abandons the territorial logic of hierarchical relations among cities controlling non-overlapping market areas and closely embedded in each other. It instead focuses on long-distance relationships among cities of the same size performing very similar functions, which by definition cannot exist in Christaller's model.

Furthermore, the model based on 'network' relationships among cities discards the principles of economic efficiency (minimization of transport costs and maximization of the market area controlled from the centre) that underpin the organization of urban centres in hierarchy models. Now of prime importance are new principles of economic efficiency that govern the organization of urban systems and originate from the positive effects of co-operative or complementary activities. In the case of synergy networks, the advantages are termed 'network externalities',[30] which accrue to all and only the members of the network. An example is provided by the advantages obtained by international financial centres from the creation of the ICTs networks and the 'virtual' market that enable them to operate across great distances and enjoy the relative economies of scale. In the case of complementarity networks, the advantages are those of the territorial division of labour and specialization that can be achieved through economies of horizontal integration among production units, and vertical integration in specific specialization chains.

According to the city network paradigm, therefore, relations among urban centres are no longer governed by a clear hierarchy among centres, or by competition among them in which localization economies and input–output relations strengthen the growth of one centre necessarily to the detriment of another. Economic relations among centres are now based on co-operative links that enable urban economies of scale to be achieved, without cities necessarily having to grow in terms of physical size.

We are now able to give a definition of the network paradigm. City networks consist of sets of horizontal, not hierarchical, relationships among complementary or similar centres. These relationships generate economies or externalities of, respectively, specialization/division of labour and of synergy/co-operation/innovation.

Empirical studies have amply demonstrated the tendency of urban systems to organize themselves into networks, but they have not investigated the magnitude of the advantages created by this type of organization.[31] At the end of the 1990s the first attempts have been made to verify the positive effects arising from co-operative or complementary activities. An econometric study applied to the 'Healthy Cities Network' of the World Health Organization – an institutional network of urban centres set up in order to promote and co-ordinate urban policies for the protection of the quality of life in cities – has demonstrated that a network organization has positive effects. The cities most closely involved in the network were most successful in terms of their urban policies implemented.[32]

It is still too early for the city network paradigm to be called a theory, given that it still lacks adequate theorization. However, it seems certain to flank (and in certain respects supersede) Christaller's traditional territorial approach to the study of urban systems and their evolution. It is beyond doubt, in fact, that the paradigm represents a major theoretical advance on the urban hierarchy model. A good example of its interpretative force is its ability to sever the mechanistic relation between specialization and city size imposed by Christaller's model. It explains, for example, why Zurich, a city of only 360,000 inhabitants, fulfils a function of prime importance in international finance together with megalopolises like New York and Tokyo; a circumstance impossible to conceive or explain on the logic of Christaller's hierarchical model.

3.7 Conclusions

The chapter has analysed theories able to interpret the economic reasons for the existence of urban systems made up of cities of different sizes, and in doing so to remedy the evident limitations of other approaches, in particular those described in Chapter 2. Central place theories and the seminal works of Christaller and Lösch explain how urban systems organize themselves on the territory according to strict economic principles. Their merits and shortcomings have been analysed, and the most recent advances in central place theory rectifying the weaknesses of the original models have been outlined.

The chapter has concluded with discussion of 'city networks theory', which is the most recent conceptualization of the organization of urban systems. This approach no longer interprets urban systems in terms of purely hierarchical relations among cities (which allow for control over non-overlapping market areas embedded in each other). It instead envisages horizontal relations among urban areas of the same size and performing similar functions. These relations exist in the real world but do not find theoretical explanation in Christaller's model. Although this approach yields a conceptual interpretation of evident real phenomena, it is still a paradigm that lacks an adequate theory that would give it wider scientific recognition.

Review questions

1. What is the main aim of the 'central place theory'?
2. What is the meaning of threshold and range in Christaller's model?
3. How does Christaller identify hexagonal market areas?
4. What are in Christaller's model the principles governing activities in space and how do they differentiate one from the other?
5. What are the economic postulates of Christaller's model?
6. What are the main differences between Lösch's and Christaller's models?
7. How does Lösch identify hexagonal market areas?
8. What are the main limits envisaged in Lösch's and Christaller's models? Have they been overcome? If yes, how?
9. What are the reasons behind the development of the city network theory?
10. What are the main conceptual elements contained in the city network theory? What empirical results have been achieved to support this conceptual framework?

Selected reading on empirical findings

About city networks and urban systems

Boix R. (2004), 'Redes de Ciudades y Externalidades', *Investigaciones Regionales*, no. 4, pp. 5–27.

Capello R. (2000), 'The City-Network Paradigm: Measuring Urban Network Externalities', *Urban Studies*, vol. 37, no. 11, October, pp. 1925–1945.

Dematteis G. (1994), 'Global Networks, Local Cities', *Flux*, no. 15, pp. 17–13.

Gottman J. (1991), 'The Dynamics of City Networks in an Expanding World', *Ekisticks*, no. 350–351, pp. 227–281.

Pumain D. and Saint-Julien T. (eds) (1996), *Urban Networks in Europe*, John Libbey Eurotext, Paris.

Taylor P.S. (2001), 'Specification of the World City Network', *Geographical Analysis*, vol. 33, no. 2, pp. 181–194.

Further reading

Beckmann M.J. and McPherson J. (1970), 'City Size Distribution in a Central Place Hierarchy: an Alternative Approach', *Journal of Regional Science*, vol. 10, pp. 25–33.

Camagni R. and Capello R. (2004), 'The City Network Paradigm: Theory and Empirical Evidence', in R. Capello and P. Nijkamp (eds), *Urban Dynamics and Growth: Advances in Urban Economics*, Elsevier, Amsterdam, pp. 495–529.

Capello R. (2000), 'The City-Network Paradigm: Measuring Urban Network Externalities', *Urban Studies*, vol. 37, no. 11, October, pp. 1925–1945.

Christaller W. (1933), *Die Zentralen Orte in Süddeutschland*, Wissenschaftlische Buchgesellschaft, Darmstadt; English edition (1966), *The Central Places in Southern Germany*, Prentice-Hall, Englewood Cliffs, NJ.

Lösch A. (1954), *The Economics of Location*, Yale University Press, New Haven, CT.

Notes

1. See Christaller, 1933; Lösch, 1954, although the original work dates to 1940.
2. For surveys of this literature see Mulligan, 1984; Beguin, 1988. The latter also conducts interesting critical analysis of the base models.

the role played by transport costs in defining the market area of a baker compared to that of a specialist doctor.

16 See Lösch, 1954, p. 435. Lösch numbers the ranges of cities in an order that is the reverse of the one used here, in that he gives higher values to higher-order centres. However, in order to maintain the notation of Christaller's model, here Lösch's numbering is inverted.

17 See Segal, 1977.

18 Presented here is the mathematical formulation by Paelink and Nijkamp, 1976.

19 The reference is to the models of 'new urban economics' (see Section 2.5), which were developed much later chronologically than those of Christaller and Lösch. As said, there are doubts concerning the ability of the 'new urban economics' models to explain the existence of city systems: in order for firms and households to be indifferent among locations in different cities, these models require equal levels of profit and utility, but these are conditions fulfilled only on the hypothesis that all cities are of the same size.

20 For criticisms see Valavanis, 1955; Beckmann, 1958; Berry and Garrison, 1958; Mills and Lav, 1964; Mulligan, 1979; Eaton and Lipsey, 1982; Evans, 1985; Fujita et al., 1999a.

21 The lack of analysis of demand in these models has always been regarded as one of their major weaknesses. Even recently it has prompted attempts to develop – at the cost of considerable analytical complexity – general equilibrium models comprising both demand and supply components. See Fujita, Krugman and Venables, 1999b.

22 In economics, two goods are said to be 'complementary' when a percentage increase in the price of one of them causes a percentage decrease in the quantity sold of the other. Goods are instead termed 'substitute goods' when a percentage increase in the price of one of them causes a percentage increase in the quantity sold of the other.

23 See Beguin, 1988.

24 Beckmann and McPherson, 1970.

25 Beguin, 1984.

26 Long, 1971.

27 Parr, 1978, 1981 and 1985.

28 Camagni, 1994.

29 Camagni, 1994.

30 In microeconomics, 'externalities' are the advantages (or disadvantages) that arise from activities of exchange or production among two actors and involuntarily affect a third actor extraneous to the transaction, which gains an advantage (or suffers a disadvantage) without monetary compensation. For detailed treatment of the concept of externality see Meade 1952, Scitovski, 1954; Mishan, 1971. In the present case, co-operation among cities generates advantages for which the cities do not pay the exact monetary equivalent. For example, they achieve a critical mass of demand for jointly produced services, and they enjoy the relative economies of scale in supplying and managing the service. For a critical survey of the concept of network externalities see Capello, 1994, chap. 2.

31 For empirical analyses of the network organization of urban systems see Cappellin and Grillenzoni, 1983; Emanuel, 1988; Camagni, 1994; Dematteis, 1994; Emanuel and Dematteis, 1990; Gottman, 1991; Pumain and Saint-Julien, 1996; Taylor, 2001; Subirats, 2002; Boix, 2004. For a synthesis of theoretical and empirical aspects on city networks, see Camagni and Capello, 2004.

32 See Capello, 2000.

Part II

Theories of regional growth

Uniform-abstract space

3 Unlike the models described in Chapter 2, which hypothesized punctiform demand and spatially distributed supply, the models discussed here envisage punctiform supply and demand uniformly distributed in space. They thus recall those described in Chapter 1 which sought to explain how market areas are divided between firms.
4 Christaller defines the threshold as the minimum range within which a service can be supplied from the central place. See Christaller, 1933.
5 See Mills and Lav, 1964. It has been erroneously claimed that Christaller's market areas hark back to von Thünen's circular areas (see Ullman, 1941): in Christaller's theory the circular areas are market areas (of demand), whereas in von Thünen's they are areas of production (of supply).
6 An equilateral triangle or a square also guarantees the complete coverage of a space without overlaps, but it does not also guarantee the minimization of transport costs, which a hexagonal market area instead does.
7 The constant k may also be defined as the equivalent number of market areas of a given level that nest with a market area of the next higher level.
8 If a is the distance between two original settlements and R is the number of ranks in the hierarchy (excluding the original agricultural settlements), the various progressions are:
No. of market areas: $k^0, k^1, k^2, k^3 \ldots$
No. of central places of a certain order R: $1, k^0(k-1), k^1(k-1), k^2(k-1) \ldots$
Distances among central places of the same order: $a\sqrt{k^{R-0}}, a\sqrt{k^{R-1}}, a\sqrt{k^{R-2}} \ldots$
On applying Christaller's three principles, we obtain in numerical terms:

Order	*Market principle*		*Transportation principle*		*Administrative principle*	
	Central places	*Market areas*	*Central places*	*Market areas*	*Central places*	*Market areas*
n	1	1	1	1	1	1
n - 1	2	3	3	4	6	7
n - 2	6	9	12	16	42	49
n - 3	18	27	48	64	294	343
...						

9 The relationship between specialization and urban size had already been analysed by previous studies. See Clark, 1945.
10 See Beguin, 1988.
11 The first attempt at mathematical formalization was made by Beckmann, 1958. The definitive version was produced by Beckmann and McPherson, 1970, who were able to solve the problem of the demographic size of centres of different orders – a problem that Christaller did not address. Described here is the version set out in Segal, 1977, which is derived from Beckmann and McPherson.
12 $s+1$ is k, the proportionality factor in Christaller's model.
13 See Dacey, 1964.
14 See Segal, 1977.
15 For this reason it has been suggested that Lösch's model is better suited to describing urban systems with a predominant industrial sector in which internal, scale or localization economies prevail to produce more 'specialized' agglomerations. Christaller's system is better suited to analysis of tertiary urban systems: in this case, transport costs, which are mainly paid by consumers, still decisively determine the size of a service's market. An example is

4 Productive structure and development

4.1 The different interpretations of regional growth and development

This chapter and those that follow examine the second broad area of regional economics: *regional development theory*. Although there are numerous and markedly different approaches to regional development, all of them endeavour to identify the factors responsible for the development path assumed by a local system. They analyse local development in terms of (i) absolute growth (from the viewpoint of the *efficient allocation of local resources*) and (ii) relative growth (among regions) in order to interpret regional disparities and possible paths of convergence and divergence in levels and rates of income growth, doing so from the viewpoint of *even income distribution*.

Regional economics shifts the focus of analysis from location choices – so far examined in location theories – to the processes involved in the economic development of subnational areas. It seeks to explain, given a certain quantitative and qualitative distribution in space of resources and activities, the capacity of a local system – whether a region, a city, a province or a geographical area with specific economic features – to develop economic activities or to attract new ones from outside, and to generate local well-being, wealth and enduring growth.

Consequently, the theories and models discussed in this and the following chapters deal with regional development, by which expression is meant the capacity of a region to find (and constantly to re-create) a specific and appropriate role within the international division of labour through the efficient and creative use of the resources possessed by the local economic system. Regional underdevelopment and regional imbalances arise from differing capacities to exploit and to organize local resources (environmental, economic, physical and human) and to attract new resources and activities into an area. Regional development theory seeks to identify the factors that generate this capacity, and the external processes and the relations that either strengthen or weaken it. The level and evolution of these tangible and intangible factors determine the *development* path of a region and its well-being. However, for the sake of brevity, the theories and models often sum up the various elements determining the development patterns of an economic system in a single indicator – the *growth* of a region's per capita output or income. Although this approach to development has the obvious drawback that qualitative information is lost, it has the indubitable advantage that analytical modelling of the development path becomes possible. In this case, we shall be dealing in what follows with *regional growth theories*. When the analysis instead concerns the tangible and intangible elements (often difficult to formalize) which define and conserve the well-being of a society, we shall be dealing with *local development theories.*

As we shall see, no single definition has been given to the concept of regional growth. Rather, the various theories on the subject pertain to three 'philosophies' that have interpreted economic dynamics. The first, that of the classical (and neoclassical) economists of the eighteenth and nineteenth centuries, interprets the growth process in terms of productive efficiency, of the division of labour in a Smithian sense, and of production factor productivity, and hence examines the dynamics of wages, incomes and individual well-being. The second philosophy adopts a short-term view of growth and concentrates on the exploitation of given and unused capital resources and of large labour reserves. The third philosophy – the most modern of them – interprets the growth path as a problem concerning competitiveness and long-term dynamics and therefore takes the constant innovation of an economic system to be essential for development patterns.

We can use these three philosophies and their three views of the economic dynamic to classify the theories analysed later into three groups and highlight their normative aims:

1 the theories belonging to the first group aim to identify the factors that generate employment and income in a local system over the short period. They hypothesize the existence of unused production capacity (capital stock) and large labour reserves. In these conditions, local economic growth does not depend on the structure and dynamic of supply (which by definition is able to expand and respond rapidly to market requirements); rather, it is driven by growing demand for locally produced goods which exerts an income multiplier effect through increases in consumption and employment.[1] This was the definition given to growth by the first theories of the 1950s (see Chapter 5), which presupposed a problem of unemployment;
2 a second group of theories seeks to identify the economic mechanisms that enable a region to move out of poverty, start along a growth path, and ensure a certain level of *well-being and per capita income for its inhabitants*. Growth is a problem of individual well-being to be addressed in two ways: by acting upon factor productivity, thereby obtaining increases in real per capita wages and incomes, and by fostering processes of production specialization that yield advantages deriving from the purchase of goods on interregional markets at prices lower than they would be if the goods were produced internally to the region. These theories also comprise the notion of relative growth – of divergence/convergence in levels and rates of growth among regions – in that they measure the magnitude and trend of disparities among per capita incomes.[2] Growth was viewed in this way by most of the theories developed in the 1970s (see Chapter 6). Problems of poverty, underdevelopment and inequalities in the spatial distribution of income are the normative aspects of concern to these models;
3 the theories in the third group embrace a more modern conception of growth. They investigate the local conditions that enable the economic system to achieve high levels *of competitiveness and innovativeness* and, more crucially, to maintain those levels over time. Growth is defined as an increase in a region's real production capacity and its ability to maintain that increase. This conception is adopted by present-day theories and models of regional growth (see Chapters 7 to 11).

This classification is useful for two reasons. First, it prevents the attribution to theories and models of aims that they do not in fact set for themselves. For example, it is wrong and misleading to think that theories that seek to identify processes of employment growth on the assumption of given but unused resources are able to suggest policies for long-term development. Indeed, it is hazardous to base normative action intended to foster a long-term dynamic on theories that concern themselves with the short period.

Second, the distinction drawn by the above classification of conceptions of growth dispels some apparent contradictions in theories and models of regional development. According to the conception of short-period income growth, an increase in exports is a development mechanism because it creates income. Yet from the viewpoint of individual well-being, it removes goods from final consumption and consequently hampers growth. Likewise, when development is viewed in terms of a short-period increase in income, emigration from a region is a cost because it deprives the area of effective demand (although it does so only at the level of subsistence consumption). But if the concern is with individual well-being, emigration is viewed as a positive factor in a region's development because it redresses imbalances (and consequently inefficiencies and income differentials) in the local labour market. On this view, surplus labour has nil marginal productivity and tends to spend any increase in income on consumption, rather than on savings and production investments.[3] Far from being a resource for production development, it is an obstacle to growth, and its reduction statistically increases per capita income.[4] Finally, if the focus is on an area's potential for long-period development, the population is once again viewed as a resource that should not be wasted on emigration.

The element that triggers the growth process can be deduced from these various interpretations of development. A short-period increase in income can be straightforwardly achieved through growth in demand for locally produced goods and services. The latter takes the form of effective sectoral demand, also external to the local economy and possibly dynamic, which sets off a virtuous 'demand/supply' mechanism through Keynesian multiplier effects on income. In this case, the engine of development is *demand*. From this point of view, therefore, no consideration is made of the ability of supply to keep up with growing demand, given the assumption that there are no limits on local production capacity. But although this assumption may well be realistic in the short period, it is unsustainable in the long one. By contrast, if the focus is on individual well-being and long-period competitiveness, the engine of development must necessarily lie on the *supply* side, and specifically in the availability of production factors (labour, capital, entrepreneurship), and in the absolute and comparative advantages of the local firms that determine an area's production capacity and its position on the world market.

Finally, it is evident that those who set out to analyse development and growth must necessarily assume a dynamic perspective, whatever conception of development they may have – short or long period, posited on employment, on per capita income and individual well-being, or on competitiveness.

In order to simplify their formal treatment, less recent models often assume that the effects of development only last for the period in which they arise. The use of a single-period framework makes it possible to employ *static or comparative statics models* in which changes in the levels of the variables, in a single period, come about independently of the time variable. More recently, however, much use has been made of *dynamic, initially linear, models* that enable analysis of how equilibrium conditions change over time. In this case, reference is no longer made to static equilibrium conditions defined by the *level* of a certain variable. Rather, the intention is to identify stationary equilibrium conditions in which it is the *rate of growth of the variable* that remains constant over time. *Dynamic stable or unstable dynamic equilibria* can be identified according to whether the system is able to return to equilibrium when it has deviated from the equilibrium growth rate, or whether it has departed from it permanently. Most recently, regional development has been studied using non-linear dynamic models which yield, as we shall see in Chapter 10, multiple, stable or unstable, oscillating or even 'chaotic' equilibria. It would be of great interest if proof were forthcoming that the results on equilibrium stability obtained with a static or dynamic linear model change when non-linearities are introduced into the structural relations that characterize the model. This outcome would demonstrate that the traditional dichotomy between 'divergence theories' and 'convergence theories', so often stressed in the literature and economics handbooks, has been superseded.

4.2 The different conceptions of space

A further important element in understanding the theories and models set out in the literature is the rather different *conception of space* that they use, for it plays a crucial role in identification of the determinants.

The earliest theories of regional development were growth theories that sought to explain trends in income and employment over the short and medium-to-long period. To do so they abandoned the concept of physical-metric space employed by location theory and replaced it with a notion of *uniform-abstract* space – a space in which supply conditions (factor endowment, sectoral and productive structure) and demand conditions (consumer tastes and preferences) are identical everywhere in the region. This is the case of the neoclassical theories of regional growth, the export-base theory and the theory of factor endowments, all of which, with this definition of space, deliberately disregard any economic diversity within a region. They instead hypothesize a uniform territory in which production processes have no cumulative and synergic effects, and in which there are none of the agglomeration economies that instead play such a major role in the location theories examined in previous chapters.

A space of this kind allows local growth phenomena to be interpreted using macroeconomic models adapted to the specificities of the local area. In fact, on the assumption of a uniform-abstract space in which the economic variables assume the same values throughout the region (conceived as a point in space), it is possible to stylize the region's economic behaviour in aggregate macroeconomic models and theories. The analyst is thus able to predict the economy's development on the basis of interactions among certain variables (for example, the propensity to import or to consume, or the capital/output ratio). These theories are *theories of regional growth* that seek to interpret the trend of a synthetic development indicator like income, with an inevitable loss of qualitative information but the undoubted advantage that they make it possible to model the development path analytically.

This conception of space was adopted by the first theories of regional growth. These theories were facilitated in their task of interpreting local development paths by already well-established economic approaches, although these had to be adapted to specific interpretative needs. The neoclassical theory of regional development, the

Harrod–Domar model, and the theory of factor endowments discussed in this part of the book derive in fact from macroeconomics, neoclassical economics, development economics and the economics of international trade. These theories view economic growth as driven by economic differences among regions and by the interregional relations generated by those differences: weak regions with poor factor endowments, low resource productivity and limited production capacity are matched by regions with high endowments of capital, technologies and know-how.

A second interpretation of space is comprised in the notion of *diversified-relational* space. Unlike the previous interpretation, this one hypothesizes the existence of marked polarities in geographical space, and of specificities in the relationships among people, society and the territory on which development is based. This conception of space both allows and requires analysis to shift from a macroeconomic and macro-territorial approach to a micro-territorial and micro-behavioural one. These theories can therefore be defined as theories of development that seek not so much to explain a rate of aggregate growth of income or output as to identify all the elements – tangible and intangible, exogenous or endogenous – that characterize the development process. This conception of space is adopted by the theories examined in the next part of the book (Part III). The growth pole theory, analysis of the role of multinational companies in local development and studies on the diffusion of innovation in space endeavour to identify the (*exogenous*) causes of territorial polarities on which development depends. The heavy emphasis placed upon the role of local relations in development explains why these theories conceive space as '*relational*' as well as diversified. Such relations are local input–output relationships between a leader firm and other local firms, between a large multinational company and the local industrial system, between innovators (external to the region) and local imitators.

This interpretation of space is expressed most forcefully by theories on industrial districts, *milieux*, or 'learning regions' which look for the endogenous determinants of development. These theories maintain that cumulative development processes stem from the concentration itself of activities in space. This is the source of economic and social relations that – facilitated and strengthened by proximity – act upon the productivity and innovativeness of local firms. Once again, the emphasis on local economic and social relations leads to definition of space as a relational space. For these theories, it is territorial concentration itself that generates development and the increasing returns that (in the form of agglomeration economies) make the growth process self-fuelling and give rise to a virtuous circle of development. However, development is selective: it only comes about in areas where the spatial concentration of production exerts its positive effects on the efficiency parameters of production processes. Space thus becomes an independent economic resource and production factor. It generates static and dynamic advantages for the firms situated within it; and it crucially determines the competitiveness of a local production system.

Because theories of endogenous development are mainly concerned with externalities, and localization and 'district' economies, we may say that they represent the 'core' of regional economics – the hub of the discipline where location theories and development most closely interweave and merge. These theories permit definitive abandonment of the notion of *competitive development* (embraced by some neoclassical theories of the 1960s), which derived from the simple regional distribution of an aggregate growth rate, and the adoption instead of a notion of *generative* development where the national growth rate is the sum of the growth rates of individual regions.

Finally, the most recent theories (described in the last part of the book) conceive space as *diversified-stylized* in that it comprises development-generating polarities. These polarities have no territorial dimension, however, because they are stylized into simple points in space. This conception has been adopted by the theories of new geographical economics and endogenous growth theories, and it enables them to construct elegant economic models that include the synergies and cumulative feedback processes that arise in space. Because polarities are punctiform, they can be handled by traditional macroeconomic models (in fact, they once again become regional growth models), while economic growth is selective and cumulative because of the presence of increasing returns which stylize the advantages of concentrated location.

This new conception of space has partly resolved the problem from which regional development theories have always suffered: their inability to construct formal models which combine specifically territorial features, like externalities and agglomeration economies, with macroeconomic laws and processes of growth. However, it should be pointed out that the assumption of a stylized rather than relational space deprives the polarities envisaged by such models of a territorial dimension able to give space – through synergy, co-operation, relationality and collective learning – an active role in the growth process. The introduction of agglomeration advantages in stylized form, through increasing returns, cancels out the territorial dimension. And in so doing it divests these theories of the aspect of greatest importance to regional economists: namely space as territory defined as a system of localized technological externalities, or as a set of material and non-material factors which by virtue of proximity and reduced transaction costs act upon firms' productivity and innovativeness.[5] Finding a way to incorporate the territorial dimension into theories already able to merge physical-metric, uniform-abstract and diversified space is the challenge that now faces regional economists.

Before early theories of regional growth are described, this chapter reviews the theories that have identified the *industrial structure* and *geographical location* as the *preconditions for local development*. These are theories that seek to determine the tangible and intangible elements necessary for the growth process to begin.

They conceive space as uniform but not abstract (here the reference is to the theory of the stages of development, and to the theory that associates phases of development with levels of regional disparity). By identifying the economic and social characteristics of the development path, in fact, they deal with a real rather than abstract space, although they still treat the territory as being internally uniform.

These theories are of interest for their simplicity, and also for their long-sightedness. They contain in embryonic form the development-generating factors – local production specialization, transport infrastructures, capital, advanced services, location close to large outlet markets – that later theories would amplify and elaborate.

As we shall see, these theories link closely with those on underdevelopment (for example the theory of the stages of development), and in so far as they take their theoretical framework from those theories, they share their strengths and weaknesses.

4.3 The theory of the stages of development

One of the earliest theories of development applicable to territories of all sizes – from nations to regions to local economies – is the theory of stages of development. This was the earliest attempt by location theorists to combine analysis of the location patterns of firms with interpretation of the effects of location choices on development.[6]

The simplicity of the theory is both its strength and its weakness. It depicts regional development as a natural sequence of phases, each characterized by growing factor productivity and an increasing capital/labour ratio that yield ever higher levels of well-being and per capita wealth.[7]

The theory identifies the following sequence of development stages:

a) *autarky*, when the local economic system is self-sufficient within a subsistence economy: everything produced locally is used (and produced in sufficient quantities) for local consumption;
b) *specialization*, when the creation of transport infrastructures makes trade in agricultural goods possible and the local economy begins to specialize in certain primary goods;
c) *transformation* of the local economy from agriculture to industry as a result of the take-off of industrial activities closely connected with the processing of primary goods (agricultural and mining products), and with the needs of a growing population (building construction). These industrial activities often develop on the basis of knowledge and expertise external to the area;
d) *diversification* of manufacturing activity due to increasing demand for intermediate goods, the growth of income, and the consequent appearance of new sectors catering to the consumption needs of a growing and increasingly diversified population;
e) *tertiarization*, the expansion of tertiary activities in response to what has by now grown into an advanced industrial system.

This simple theory captures a number of important features of a development process. First, it highlights the productive specialization that – according to the standard Smithian conception of the division of labour – is the source of greater labour productivity. Productivity increases derive from what have been termed 'roundabout methods': increasingly indirect production processes divided into vertically specialized phases, cycles and processes that enable the simplification and mechanization of every production phase.[8] As a result of debate within development theory on the underdeveloped countries, this theory has also emphasized the importance for development of the simultaneous growth of diverse sectors and infrastructures in a process of 'balanced development'. The latter has a number of advantages and externalities[9] deemed to be the main sources of increasing returns at territorial level, and the engine of local growth. The most significant of them are the following:

- externalities deriving from *interdependencies among sectors* that, via input–output linkages, ensure development of the local economy as a whole if there is an initial growth impulse in a single sector;[10]
- externalities arising from *interdependencies between demand and supply* that generate cumulative development processes based on a growth of supply in line with the preferences structure of local consumers;[11]
- externalities arising from *investments in different infrastructures* for integrated projects selected on the basis of the infrastructural needs of local demand, whether this is planned or even only potential. Investments in transport infrastructures are of particular importance because expansion of the market controlled by local firms depends on them.[12]

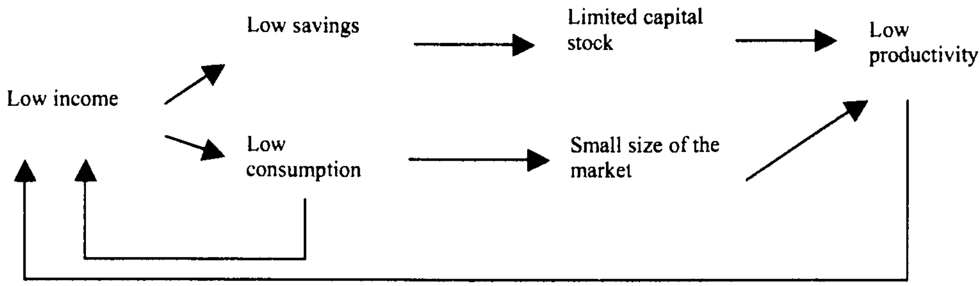


Figure 4.1 The vicious circle of underdevelopment

According to this theory, underdevelopment is an area's forced persistence in a particular phase.[13] The causes of this situation are conditions internal and external to the area. The internal ones relate to a lack of the sources of territorial-level increasing returns just mentioned. If a local economy does not have sufficient savings to invest in capital or infrastructures, or if its market is too small, its productivity level will remain extremely low and will fuel a vicious circle of underdevelopment – limited market expansion, low savings and low consumption, reduced stock of capital in the economy and low income (Figure 4.1). Besides an insufficient critical mass of demand, savings and infrastructures, there are various external constraints that operate as well. If the region belongs to a system of more developed regions, it is likely that demand/supply interactions or sectoral interdependencies, stimulated by internal demand for relatively advanced products, will be set in motion externally to the region. The leakages to more advanced regions may be so strong that they restrict the local effects of expansion in demand – a risk, as we shall see, that other theories of regional development have considered carefully.[14]

The proponents of 'balanced development' accordingly suggest that, in early phases, development policies should channel public investments into a few large and diversified sectors with significant weights at local level (strong sectors). The purpose is twofold: (i) to minimize leakages to advanced areas and (ii) to remedy the insufficient saving formation typical of backward economies. Later, when the take-off of the strong sectors has expanded private resources, a share of public investments may be directed to other sectors.

This simple theory highlights a number of important features of the development process: the role of infrastructures and their development, with particular regard to demand; the role of the production specialization as the basis of increasing returns in factor productivity; and the importance of transport in increasing the size of the market and production. However, it is rather difficult to accept the idea that development moves through necessarily identical stages in all regions, when these are characterized by different economic, social and historical conditions; and, moreover, when they are embedded in very different economic contexts. In light of these considerations, we shall see how the economist Douglass North rejected the idea of a natural development process and developed his export-base theory, furnishing a more convincing explanation of the development of the western American regions in the 1950s.[15]

4.4 Stages of development and disparities

In the mid-1960s, J. G. Williamson reprised the idea that development proceeds through stages and analysed how regional disparities evolve within a country.[16]

Williamson's thesis, with which other regional economists agree,[17] is that development in its early stages is concentrated and polarized in a country's central area. Only subsequently does it spread to more peripheral areas and to weaker sectors. The consequence of this 'two-speed' development is that the regional gap widens in the early phases of a country's economic development, and then narrows when the national income reaches a certain level. It therefore follows an inverted U-shaped trajectory (Figure 4.2).

The reasons for the widening gap between strong and weak regions in the early phases of development relate to the following well-known 'crowding-out' effects that favour the strong economy over the weak one:

- emigration of skilled labour from weak areas to strong ones;
- capital flows to the wealthier regions, to which they are attracted by higher demand, by the availability of infrastructures, services and a potential market, and by better environmental conditions for firms;
- allocation of a larger share of public investments to strong areas, in response to explicit actual or potential demand;
- limited interregional trade in resources, so that, in early stages, the rich area does not exert pull effects on the poor one.

Over time, these processes exacerbate regional disparities within a country until mechanisms working in the opposite direction begin to operate, for example:

- the creation of new jobs in less developed areas, as well, with the consequence that emigration diminishes or even ceases;
- reduced attractiveness of the more advanced areas due to the saturation of markets and physical congestion, with the consequent prohibitive costs of land and an inevitable fall in the average profit rate;
- growth of public investments in the weak areas, which has a twofold effect: the birth of a local production system that requires major investments in social capital, and the growth of private investments in the strong areas;
- the onset of pull effects exerted by the strong area on the weak one.

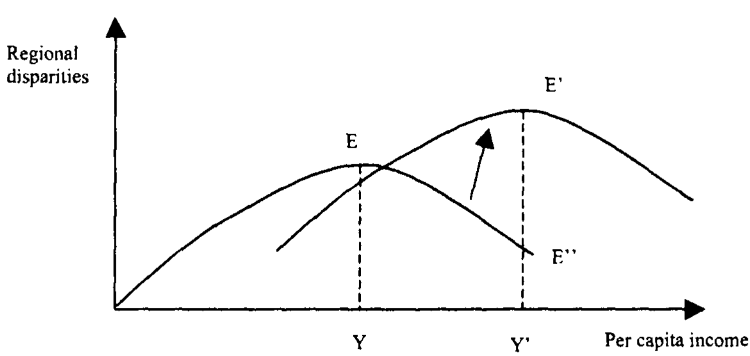


Figure 4.2 Williamson's curve of regional disparities

Although the empirical evidence confirms that regional disparities increase during the early phases of a country's development, it does not bear out the hypothesis that the growth differentials subsequently diminish. The theory therefore seems excessively optimistic in its interpretation of disparities as following a natural, deterministic and universal law. Technological progress, social changes and the evolution of knowledge are all factors that may give advanced regions a greater capacity to attract capital and labour from the weaker regions, and to obtain public investments in modern social capital and advanced infrastructures (e.g. hub airports, high-speed trains). Consequently, in the advanced regions, the frontier of decreasing returns on investments is reached at higher levels of income. In graphic terms, this means that the U-curve of regional disparities moves rightwards and upwards, as in Figure 4.2, so that, given a level of income Y', the country may find itself with a higher level of regional disparities: E' rather than E".

It is also likely that development in the weak regions will come about on the basis of 'traditional' industries that require non-innovative production processes and standard technologies – as implied by Vernon's life-cycle theory.[18] Accordingly, the gap between the leader and follower regions may persist in qualitative rather than quantitative terms.

4.5 Industrial structure and regional growth: shift-share analysis

Its simplicity notwithstanding, the theory of development stages yielded an important finding, which analysts reprised and amplified at the end of the 1950s: the sectoral composition of a region explains its rate of growth. Given their low levels of factor productivity and small capital/labour ratios, mainly agricultural regions experience low growth rates, whilst industrialized regions, by contrast, record high factor productivities and therefore high rates of development.

However, when the sectoral structure of a region is taken to be the main determinant of economic growth, some of the simplifications in the above line of reasoning must be removed. Interpretation of a region's growth rate solely in terms of sectoral composition aggregated into only three sectoral macro-categories (agriculture, industry and services) – as in the theory of the stages of development – entails the hypotheses that each sector within a macro category has the same productivity, and that the latter does not vary according to the region in which the sector produces.

It is obvious that both these hypotheses are somewhat unrealistic. Sectors in a particular macro-category have very different productivities. Consider, within the category 'industry', the differing capital intensities (capital/labour ratios) that distinguish heavy industries (chemicals, pharmaceuticals) from light industries such as textiles, clothing and food. It is likewise evident that a sector located in two regions that differ in their infrastructure endowments, quality of production factors and technological knowledge will achieve different levels of productivity in the two regions.

At the end of the 1950s these considerations prompted a group of economists to develop more composite analysis of the relation between production structure and regional growth. This gave rise to the well-known statistical method for determining a region's relative growth rate known as 'shift-share analysis'.[19]

The theory's basic idea is that the regional growth rate is influenced by three factors: (i) the industrial structure, (ii) sector productivity, (iii) the dynamics of demand and consumer preferences.

On the assumptions that the same sectors have the same productivities regardless of their location, and that the region has the same sectoral composition as the country as a whole, the region's rate of growth should be equal to that of the country.

However, the regional growth rate often differs from the value that it should assume were it to grow at the same rate as that of the country. In formal terms, it is equal to:

$$y_r = y^* + s \tag{4.1}$$

where y denotes the income growth rate, r the region, and s the difference between the national and regional growth rates, while the asterisk on the variable y indicates the growth rate that the region should achieve if it is to be the same as that of the country as a whole. The difference between the national and regional growth rates – called 'shift' (s) – may depend on two effects:

- the *composition effect* (proportion effect) exerted by the region's sectoral structure – also termed the 'MIX effect' – and deriving from the presence in the region of sectors with more marked dynamics at national level due to increasing demand in those sectors. The composition effect can be measured as:

$$MIX = \sum_{i=1}^{n} \frac{E_{ir}^0}{E_r^0} \left(\frac{E_{in}^1}{E_{in}^0} - \frac{E_n^1}{E_n^0} \right) \tag{4.2}$$

  where E represents the sectoral variable analysed (employment or value added), i denotes the sector, while n and r respectively stand for the country and the region. The term in brackets measures the difference, in the period of time from 0 to 1, between national-level employment in sector i and the average national increase in employment. This is multiplied by the relative weight of the sector in the local economy;
- the *competition effect* (differential shift) of the region's sectoral structure – or the 'DIF effect' – that derives from the regional economy's capacity to develop each of its sectors at greater average rates than those achieved by the corresponding national sectors. The DIF effect can be calculated as follows:

$$DIF = \sum_{i=1}^{n} \frac{E_{ir}^0}{E_r^0} \left(\frac{E_{ir}^1}{E_{ir}^0} - \frac{E_{in}^1}{E_{in}^0} \right) \tag{4.3}$$

  In this case the term in brackets measures the increase in sector i at regional level compared to the increase in the same sector at national level. As in the case of the MIX effect, the increase is multiplied by the sector's relative weight in the local economy.

When applied at a disaggregated sectoral level to a region, the shift-share analysis highlights the aspects that the theory of the stages of development simplistically ignored: on the one hand, the differing productivity of the same sector in different areas (measured by the DIF effect); on the other, each sector's contribution to regional growth differentials.

The strength of this approach is its ability to distinguish between structural factors (MIX effect) and short-term ones (DIF effect) in regional growth differentials, and to isolate those that drive regional development: demand-side elements measured by the MIX effect on the one hand, and supply-side elements of local competitiveness measured by the DIF effect on the other.

It is possible to illustrate the foregoing analysis by means of a graph where the national rate of employment growth (although other sectoral variables, like value added, can be used) is plotted on the X-axis, and the regional one on the Y-axis. Each sector is represented by a point indicating its growth at the national and regional level respectively. Moreover, by showing the average national and regional growth rates on the graph, and by drawing a 45-degree line from the origin (along which the sectors record a rate of regional growth equal to the national rate), it is possible to mark out different areas representing different development conditions (Figure 4.3):

- development conditions favourable to the region are represented by a large number of sectors lying above the 45-degree line (areas A, D, E and F): these represent a capacity for local growth superior to the national capacity, and therefore a favourable DIF effect. In this case, development is driven by the competitiveness of the local sectors;
- positive growth conditions are represented by a large number of sectors to the right of the line representing the national sectoral average (areas A, B, C and H): these are sectors that achieve growth rates above the national average. Specialization by the region in these sectors denotes local growth driven by increasing demand at national level: that is, by a favourable MIX effect;
- conditions favourable to a region's growth are represented by a large number of sections occupying area A of the graph, which indicates that both the MIX and DIF effects are favourable; or area B, where the regional dynamic is weaker but nevertheless sufficient to maintain a generally high level of development: these, in fact, are sectors above the regional average;
- conditions equally positive for a region are represented by a large number of sectors located in areas D and E, where the competitiveness of local sectors is sufficiently high to offset the 'crisis' in which the sectors find themselves at national level. Limited national demand for these goods is more than offset by

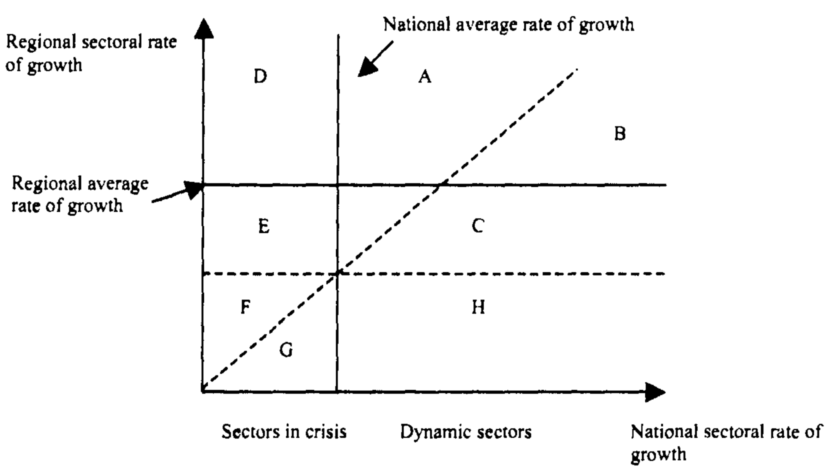


Figure 4.3 Relative sectoral development: composition and competition effects (shift-share analysis)

Source: Camagni (1992a: 165)

the competitiveness of local firms, which are able to conserve and increase their market shares. This was the case of the Italian industrial districts in the 1970s, when they maintained positive growth rates despite the general crisis that hit the economies of the industrialized countries after the oil shocks;[20]

- crisis conditions for a local economy are instead represented by a large number of sectors in areas F and G of Figure 4.3: these are sectors in crisis at the national level and which have even lower growth rates at the local one;
- also indicative of a crisis is the presence of a large number of sectors in areas C and H, where the growth of local demand is not enough to offset the limited competitiveness of local sectors.

The usefulness of graphics in depicting the results of shift-share analysis is demonstrated by Figure 4.4, which shows such analysis applied to three different geographical areas.[21] One can see at a glance the differences in competitive capacities recorded between 1995 and 2001 for the European areas of Vienna (Austria), Provence (France) and South Yorkshire (UK). Vienna and South Yorkshire show a clearly unfavourable DIF in numerous sectors; Provence instead has a large number of sectors with local growth rates superior to national ones (Figure 4.4 a, b and c).

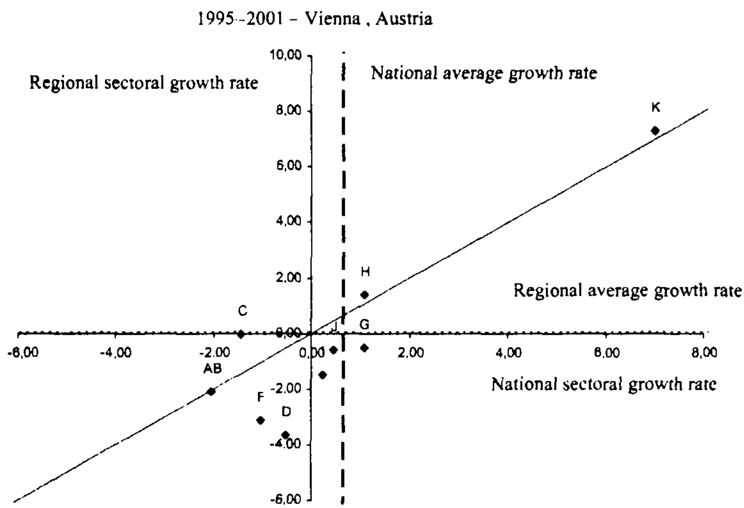


Figure 4.4 Relative sectoral development: composition and competition effects (shift-share analysis) in three different geographical areas

NACE classification of sectors:

A, B Agriculture, hunting, forestry and fishing
C Mining and quarrying
D Manufacturing
F Construction
G Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods
H Hotels and restaurants
I Transport, storage and communication
J Financial intermediation
K Real estate, renting and business activities

Source: our elaborations on Eurostat REGIO data

(*Continued*)

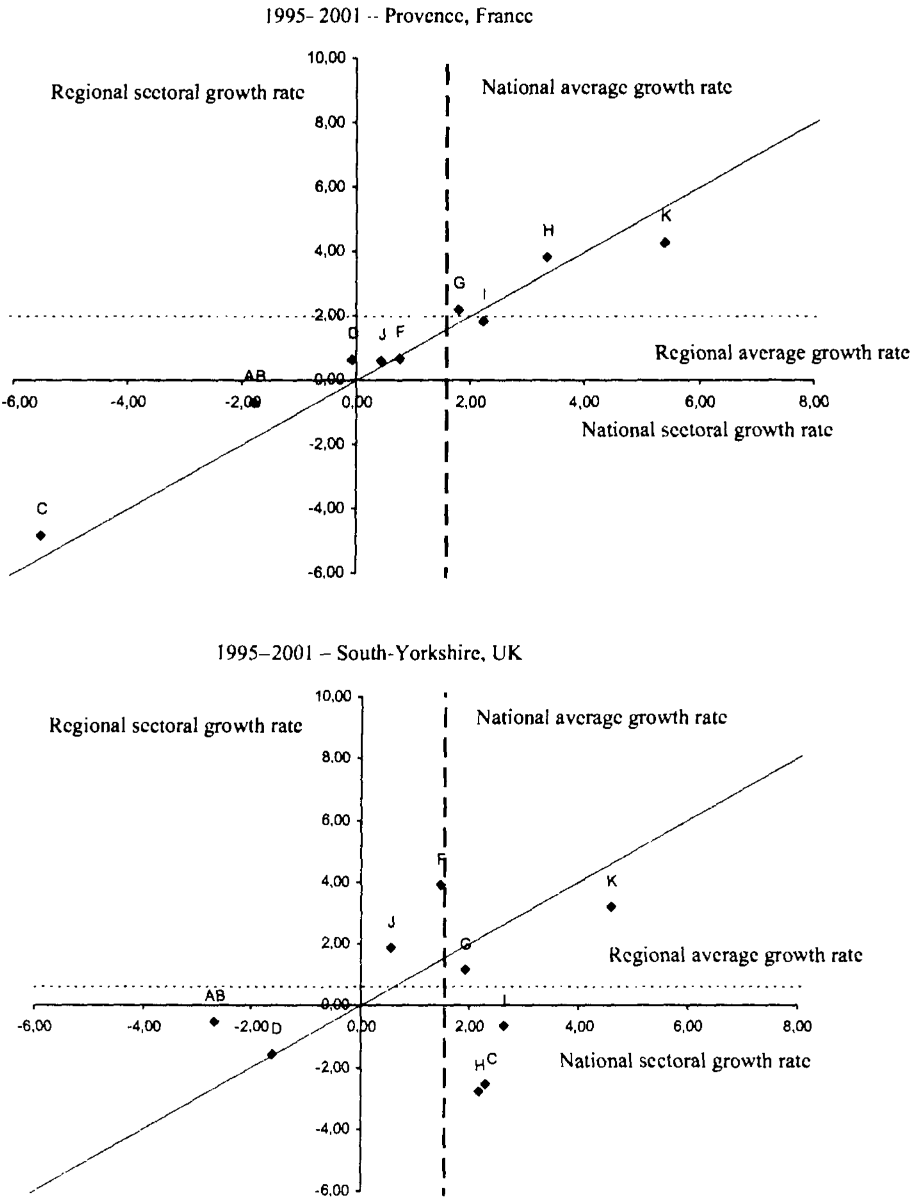


Figure 4.4 (Continued)

The limitations of this method are well known. The results are highly sensitive to the degree of sectoral disaggregation used in the analysis, and also to the method employed to calculate the region's weight relative to the country.[22] Moreover, the method is purely descriptive; it is entirely unable to explain the relative performances of regions. Its application, in fact, only allows analysis of the growth conditions of regions; it cannot be used to show the causes of their development paths. Identification of the determinants of growth is still the primary goal of all theories of regional development, and it is to these that the following chapters are devoted.

4.6 The centrality/peripherality approach

Strictly geographical in nature,[23] the centrality/peripherality approach regards distance from the centre of economic activities as the cause of delayed development. The approach originated in Walter Isard's 1950s theory of development potential, and in H. Giersch's analysis of the barycentric location of core European regions, which was subsequently reprised by J. Friedmann in his theories of underdevelopment.[24]

The simplicity of this approach is also its strength. It simply points out that geographic centrality is in itself a factor fostering development, while peripherality hampers it. Access to information, technological knowledge, outlet markets of goods and markets for production factors is a necessary condition for the growth of a local market, while peripherality – defined as the distance from a hypothetical economic centre (or 'core') – entails higher transport costs for finished goods, raw materials, semi-finished goods, greater costs of information acquisition, delays in the adoption of innovations: all of which features hamper income growth and competitiveness.

This model works very well for Europe, where a strong, developed and highly industrialized centre has formed over time, contrasting with more peripheral, less dynamic and more backward peripheral areas. Mediterranean regions, but also Nordic ones, or regions in Western Europe, have always recorded lower levels of development than have central regions. The European Union's 'Objective 1' covers the least advanced regions in the Union, and all of them are geographically peripheral. However, it is not always the case that the geographical centre of a country is also its economic centre. Consider the United States: a country with a less developed 'centre', and where geographically peripheral regions located on the coasts comprise the country's main economic activities, areas of development and wealth.

4.7 Conclusions

This chapter has reviewed the second broad area of analysis conducted by regional economics: regional growth and development. It has shown the difference between growth theories and development theories, the former being concerned to explain trends in a single indicator, a region's per capita income; the latter to analyse trends in the tangible and intangible aspects of development. It has also examined the different definitions given to the concept of growth or development; this can be conceived as short-term growth of employment and income, growth of well-being and per capita income, or long-period growth of competitiveness. Identifying how theories conceive growth aids the comprehension of theories, and the removal of apparent contradictions among them. Also important for understanding the theories and models of regional economics is the concept of space. This differs considerably, ranging from uniform-abstract space (which represents a geographical-administrative conception of space) to diversified-relational space (where territorial factors account for local development) to diversified-stylized space. This last conception hypothesizes development-generating polarities, and therefore the existence of agglomeration economies and increasing returns in growth processes, but it removes the spatial dimension from those polarities. By so doing, it enables space to be incorporated into macroeconomic models of growth; yet at the same time it deprives space of an active role in explanation of development. The chapter has concluded by outlining the early theories that sought to explain the conditions necessary for growth to come about.

Review questions

1 What is the difference between regional growth and local development?
2 What are the different interpretations of growth given in the different theories of local development and regional growth?
3 Why is it sensible to divide theories on the basis of their conception of growth?
4 What are the different conceptions of space in the different theories?
5 What are the strengths and weaknesses of the stages of development?
6 What do the MIX and DIF effects measure? How are they calculated?
7 Would you define the shift-share analysis as an interpretative or a descriptive methodology?

Selected reading on empirical findings

About accessibility and regional development

Bruinsma F.R. and Rietvield P. (1993), 'Urban Agglomerations in European Infrastructure Networks', *Urban Studies*, vol. 30, pp. 919–934.

Bruinsma F.R. and Rietveld P. (1998), 'The Accessibility of European Cities: Theoretical Framework and Comparison of Approaches', *Environment and Planning A*, vol. 30, no. 3, pp. 499–521.

Capello R. (2006), 'Knowledge and Accessibility for Regional Cohesion in the Enlarged Europe', special issue of *Scienze Regionali – Italian Journal of Regional Science*, vol. 5, no. 2.

Handy S.L. (1997), 'Measuring Accessibility: An Exploration of Issues and Alternatives', *Environment and Planning A*, vol. 29, no. 7, pp. 1175–1194.

Hansen W.G. (1959), 'How Accessibility Shapes Land Use', *Journal of the American Institute of Planners*, vol. 25, no. 2, pp. 73–76.

Linneker B.J. and Spence N.A. (1992), 'Accessibility Measures Compared in an Analysis of the Impact of the M25 London Orbital Motorway on Britain', *Environment and Planning A*, vol. 24, no. 8, pp. 1137–1154.

Shen Q. (1998), 'Location Characteristics of Inner-city Neighbourhoods and Employment Accessibility of Low-wage Workers', *Environment and Planning B*, vol. 25, no. 3, pp. 345–365.

Spiekermann K. and Wegener M. (2006), 'Accessibility and Spatial Development in Europe', *Scienze regionali – Italian Journal of Regional Science*, vol. 5, no. 2, pp. 15–46.

Vickerman R., Spiekermann K. and Wegner M. (1999), 'Accessibility and Economic Development in Europe', *Regional Studies*, vol. 33, no. 1, pp. 1–15.

Vickerman R.W. (1996), 'Location, Accessibility and Regional Development: The Appraisal of Trans-European Networks', *Transport Policy*, vol. 2, no. 4, pp. 225–234.

About the shift-share analysis

Andrikopoulos A., Brox J. and Carvalho E. (1990), 'Shift-share Analysis and the Potential for Predicting Regional Growth Patterns: Some Evidence for the Region of Quebec, Canada', *Growth and Change*, vol. 21, no. 1, pp. 1–10.

Barf R.A. and Knight III P. (1988), 'Dynamic Shift-Share Analysis', *Growth and Change*, vol. 19, no. 2, pp. 1–9.

Fritz O. and Streicher G. (2005), 'Measuring Changes in Regional Competitiveness over Time: A Shift-share Regression Exercise', Working Papers 243, WIFO Publications.

Lasuen J.R. (1971), 'Venezuela: an Industrial Shift-share Analysis 1941–1961', *Regional and Urban Economics*, vol. 1, no. 2, pp. 153–220.

Ray M.A. and Harvey J.T. (1995), 'Employment Changes in the European Economic Community: A Shift-share Analysis', *Review of Regional Studies*, vol. 25, no. 1, pp. 97–110.

Further reading

Capello R. (ed.) (2006), 'Knowledge and Accessibility for Regional Cohesion in the Enlarged Europe', special issue of *Scienze Regionali – Italian Journal of Regional Science*, vol. 5, no. 2.

Friedmann J. (1966), *Regional Development Policy: A Case Study of Venezuela*, MIT Press, Cambridge, Mass.

Isard W. (1954), 'Location Theory and Trade Theory: Short Run Analysis', *Quarterly Journal of Economics*, vol. 68, no. 2, pp. 305–320.

Keeble D., Owens P.L. and Thompson C. (1982), 'Regional Accessibility and Economic Potential in the European Community', *Regional Studies*, vol. 16, no. 1, 419–432.

Perloff H., Dunn E., Lampard E. and Muth R. (1960), *Regions, Resources and Economic Growth*, University of Nebraska Press, Lincoln, Neb.

Spiekermann K. and Wegener M. (2006), 'Accessibility and Spatial Development in Europe', *Scienze regionali – Italian Journal of Regional Science*, vol. 5, no. 2, pp. 15–46.

Notes

1 In macroeconomics, the income multiplier effect is generated by the following process: an increase in one of the components of aggregate demand – for example, demand for goods produced in the area (local consumption) – gives rise to a general increase in income. However, an increase in income in its turn generates an increase in consumption, and therefore in aggregate demand. The latter once again produces an increase in income, which once again generates increased consumption. The 'Keynesian multiplier' yields a value, by definition greater than unity, which measures the variation in output resulting from a unit change in some component of aggregate demand (consumption, investments, public spending, exports).

2 Note that per capita income as an indicator of disparity has the major shortcoming from the statistical point of view of associating better conditions of relative well-being with emigration from an area. In fact, increased per capita income is obtained either through real growth in regional income (the numerator in the income/population ratio) or through real growth in regional income (the denominator in the ratio). While the two effects are statistically recorded in the same way by the indicator, from the economic point of view they represent two very different cases: the former that of real economic growth; the latter that of possible social hardship and crisis.

3 The marginal productivity of a production factor, labour for example, measures the extent to which output varies with a change in one unit of labour. If the neoclassical law of decreasing marginal productivity holds, marginal productivity diminishes as the workforce of a firm (or an area) increases. Inevitably, therefore, surplus labour has nil marginal productivity. If new workers were included in the production process, they would be unable to produce additional units of output; for this reason, they remain unemployed.

4 On this see also Graziani, 1983.

5 See Introduction, Section 4, for a more detailed interpretation of the notion of 'territory' by economists.

6 See Fisher, 1933; Hoover, 1948; Hoover and Fisher, 1949. More recently see Rostow, 1960; Poratt, 1977.

7 At the time, the hypothesis that factor productivity grew from one stage to the next was empirically confirmed by Fisher's and Clark's studies on the linkage between growth in per

capita income and employment in agriculture, industry and services. See Fisher, 1933; Clark, 1940.
8 See Young, 1928.
9 For the definition of 'externalities' see Chapter 3, note 30, p. 93.
10 See Rosestein-Rodan, 1943 and 1959.
11 See Nurkse, 1952.
12 Nurkse writes: 'a balanced increase in production generates external economies by enlarging the size of the market for each firm or industry': Nurkse, 1952, p. 574; the same idea is put forward by Rosestein-Rodan, 1943 and 1959.
13 See Hirschmann, 1957; Hirschmann and Sirkin, 1958.
14 See the export-base theory discussed in Chapter 5.
15 The reference is the export-base theory discussed in Chapter 5.
16 See Williamson, 1965.
17 See Richardson, 1969.
18 See Vernon, 1957.
19 See Perloff, 1957; Perloff et al., 1960.
20 On this see Chapter 8.
21 The three geographical areas have been identified as NUTS2 areas in the Eurostat classification.
22 See Richardson, 1978.
23 The model was developed by geographers at Cambridge in the 1970s. See Keeble et al., 1982; Keeble et al., 1988.
24 See Giersch, 1949; Isard, 1954; Isard and Peck, 1954; Friedmann, 1966.

5 Demand

5.1 Demand and regional growth

The previous chapter described the conditions necessary for a growth process to begin. The existence of infrastructures and production services, a shift of the productive structure to sectors with greater value added and higher factor productivity, and the access to central markets that breaks down the barriers due to peripheral location: all these factors determine whether a region will be able to move to a growth path.

This chapter begins with examination of the theories that in the 1950s and 1960s (when regional economics was still in its infancy) investigated the economic determinants of development and the mechanisms that enable a system to grow and achieve higher rates of output, greater levels of per capita income, lower unemployment rates and higher levels of wealth. We shall see that these models interpret development by using a synthetic indicator: the *growth* of a region's output or per capita income. Although this approach has the indubitable advantage of making analytical modelling of the growth path possible, it requires the assumption of a uniform-abstract space wherein supply conditions (factor endowment, sectoral and productive structure) and demand conditions (consumer tastes and preferences) are everywhere identical and can be expressed with a vector of aggregate socio-economic-demographic characteristics. We may accordingly call the theories examined in this and the next chapter *theories of regional growth.*

There are numerous factors that may trigger a growth process: among them increased demand for locally produced goods; greater local production capacity; a more abundant endowment (quantitative and qualitative) of local resources and production factors; and a larger amount of savings available for investments in infrastructures and technologies intended to increase the efficiency of production processes.

This chapter will examine theories and models that conceive growth as resulting from greater *demand* for locally produced goods and that adopt the typically Keynesian notion that development consists in the growth of output, income and employment. According to this approach, greater demand for a locally produced good does not confine its positive effects to employment and the incomes of those employed in the sector producing that good. Because of interdependencies in production and consumption, greater demand also generates increases in employment and income in activities upstream from the expanding sector, and in service activities supplied to the local population as a whole. In the end, therefore, increased demand for a local good gives rise to higher income and employment in the entire area.

These models therefore envisage demand as the engine of growth; a hypothesis quite acceptable to regional economies. Regions are in fact small geographical entities where it is rarely the case that all necessary goods are produced locally; and, conversely, where those goods that are produced frequently exceed local demand for them and are sold on domestic or even international markets (consider the number of cars manufactured in Turin or Detroit: an amount certainly excessive to the needs of the city's residents!).

Demand is often external in these models, in fact, and stems from interest in a local good expressed on the world market. Hence, the growth of a region depends on the extent to which its productive structure specializes in goods demanded by consumers world-wide. There are numerous local economic systems in the world whose products are sold internationally: the textiles of Prato (near Florence, in Italy), the glassware of Murano (near Venice, in Italy), the cars of Turin, Detroit or Munich, the olive oil of Greek and Italian regions, the wines of areas in France and Italy, to mention only some. Expansion in demand for the goods produced in these areas determines whether or not the entire territory will grow. As shown by the export-base model (the best known in this family of theories), increased exports of a good generate greater local production, with positive effects on income and local employment and – via interdependencies in production and consumption – on employment and income in activities upstream and downstream from the production of that good. Considering that consumption usually grows with income, any additional expenditure will be transformed into income, the growth of which will in its turn augment expenditure, in a circular process characterized by increasingly smaller income increments.[1]

Reasoning in terms of demand-driven development has a number of consequences. First, an approach of this kind can only interpret a *short-term* process of growth, because it implicitly assumes the competitiveness of *current* production and the economic system; an assumption that can only be sustained in the short period.

Second, development is associated with the pursuit of higher levels of employment and income: no consideration is made of either individual well-being or the competitiveness of the local production system. The latter aspect is perhaps the most problematic, in that analysis centred on the demand components assumes the existence of unused capacity (capital stock) and large reserves of labour on which the system can draw to meet increasing demand; in other words, the competitiveness of the local system is taken for granted. Yet this is an assumption that can only hold for the short period. To return to the example of the Detroit (or Turin) car industry, it is true that local income and employment depend on world demand for cars. In the short period, therefore, it is possible to hypothesize that Turin's or Detroit's productive capacity will be able to satisfy increasing demand. But in the long period, the area's development will depend on the car industry's ability to maintain its position on the world market, and to compete on the basis of the quality and innovativeness of its products. These, elements, however, are entirely absent from the Keynesian models of demand that will be examined in this chapter.

Given the assumption of surplus in production resources, Keynesian theories should be used with caution when they are employed in interpretation of a long-period growth path – and especially when they are used to devise measures to support a local long-period dynamic. By contrast, when these theories are applied to the specific problem of high unemployment in the presence of given productive capacity, they have two evident merits: the simplicity and rigour of their economic logic, and the ease with which they can be applied to concrete situations. We shall see below that when Keynesian theories shed their short-period perspective and assume a long-term, multi-period one – as exemplified by the Harrod–Domar model – they are able to abandon strictly demand-related aspects and give due importance to supply elements (the availability of savings and capital formation) in the interpretation of growth processes. For all these reasons, Keynesian theories deserve specific treatment in a handbook on regional economics.

5.2 Interregional relations: accounting aspects and macroeconomic elements

5.2.1 The regional balance of payments

The growth models described in this and the next chapter argue that, whilst countries can rely on their internal capabilities to develop, regions are economic systems of small size and therefore have only limited markets for both goods and production factors. Moreover, because their productive structures are often highly specialized, their economic systems produce a surplus of specialized goods but are unable to furnish the local market with a wide range of resources and physical capital, which must therefore be wholly or partly purchased on external markets.

The relations that a region *qua* economic system establishes with the rest of the world influence its development, and the economic mechanisms underlying these exchanges determine the macroeconomic conditions that accompany the region's growth path. All these intertwined aspects can be easily understood by looking at the social accounting systems used to record the relations of an economic system (at national or regional level) with the rest of the world, and the effects of these relations on the levels of production, income and capital formation.

The *balance of payments* is the accounting instrument that records, at aggregate level, all the economic and financial transactions undertaken by a regional system with the rest of the world in a particular period of time, for example one year. It is compiled for national systems, but it can also be drawn up in simplified form for regional systems. In the latter case, too, it is an important logical device, and it will prove useful for understanding the models which follow.

A balance of payments consists of three distinct parts, in each of which receipts and payments are recorded (Table 5.1). The first is the *current account*, which is divided into the trade balance, services balance, and unilateral transfers. The trade balance records the values of exports (credits) and imports (debits).[2] Entered among receipts in the services balance is expenditure within the region by non-residents (e.g. tourists from outside the region), and among payments, expenditure by residents externally to the region. Also recorded as receipts are all revenues generated by local production factors outside the region (residents working in neighbouring regions, the profits of resident-owned businesses operating externally), and as payments, the earnings of production factors which produce in the region. Finally, the current account also includes, under the heading 'unilateral transfers', the region's gratuitous receipts or payments: the former include transfers by the central government to the region in the form of pensions, unemployment benefits, development aid and remittances by emigrants (funds sent regularly by emigrants to their families in the region); the latter consist of remittances sent by immigrants to other regions – in the case of wealthy

Table 5.1 The regional balance of payments

a) Current account

1) Trade balance	
Credits	*Debits*
1. Value of goods exported	1. Value of goods imported
* Value of goods exports	* Value of goods imports
2) Services balance	
2. Value of services exported	2. Value of services imported
* Expenditures by non-residents for services (e.g. expenditures by tourists in the region)	* Expenditures by residents for external services (e.g. expenditures by residents on tourism outside the region)
* Remuneration of employees and property income owned by residents outside the region (e.g. incomes of workers commuting to the region; profits of resident-owned firms located in other regions)	* Remuneration of employees and property income owned by non-residents in the region (e.g. incomes of workers commuting from other regions; profits of non resident-owned local firms)
3) Unilateral transfers	
3. Positive unilateral transfers	3. Negative unilateral transfers
* Remittances by emigrants	* Remittances by immigrants
* Public transfers to the region • pensions • unemployment benefits	

Current account balance: *Receipts – Payments*

b) Capital account

Credits	*Debits*
1. Commercial credits received by local importers	1. Commercial credits granted by local exporters to external importers
2. Direct investments from other regions and abroad	2. Direct investments in other regions and abroad
* investments by state-controlled enterprises in the region	* investments in external property assets (purchases of buildings and land outside the region)
* investments by private firms located outside the region	* investments by local firms outside the region
* investments in property assets by non-local private firms and public bodies (purchases of buildings and land)	

Capital account balance: *Receipts – Payments*

c) Balancing Account

Outflows	*Inflows*
* Money outflows	* Money inflows

Account balance: *Receipts – Payments*

regions with a large number of immigrant workers, this item may be conspicuous. The sum of the trade balance, the services balance and unilateral transfers is the current account balance; that is, the balance of all the *real* transactions undertaken by the region with the rest of the world.

The second component of the balance of payments, the *capital account*, records *financial* transactions in regard to the opening of debits or credits for payment of the goods recorded in the trade balance, and direct investments by or to the region. These investments take a variety of forms: as receipts, investments by state-owned enterprises, investments by privately owned firms in the region, investments in property assets; and, vice versa, public and private investments outside the region as payments.

The third component of the balance of payments, termed the 'balancing account', records the monetary counterparts (inflows and outflows of money) of the transactions in goods or capital performed by the region.

Individual transactions are recorded by means of the double-entry system. This is a method that records a credit and a debit for every transaction undertaken. For example, a purchase of goods from another region to the amount of 100,000 euros is recorded as a credit under the heading 'imports' in the trade account, and at the same time as a debit of 100,000 euros in the balancing account. If the same goods were bought with a business loan, the purchase would still be recorded as imports in the trade balance, but the loan would enter among the receipts of the capital account. Because of the double-entry principle, therefore, the balance (the difference between the totals of the credit and debit columns of an account) of the entire balance of payments is always zero: the overall balance is always 'in balance'.

5.2.2 The balance of payments and the value of regional output

The individual balances – of trade, services, the current account and the capital account – are used to calculate regional macroeconomic values by means of a series of closely interconnected social accounting schedules.[3]

The first is the gross domestic product account (or simply 'production account') which summarizes supply and demand items (Table 5.2a).[4] Recorded on the credit side are the region's resources (the value of domestic output and imports), and on the debit side, the uses made of those resources, which may be consumed, invested or exported.[5] A positive balance of payments (due to positive trade and service balances) signifies that a proportion of domestic production has been undertaken for an external market, and it is entered as a component of gross domestic product (the value of local production in a certain period of time), as in Table 5.2a.

A second schedule sets out regional gross disposable income account. For accounting purposes this is defined as the sum of aggregate production, net compensation of employees from outside the region, and unilateral transfers. The gross disposable income is either consumed or saved (Table 5.2b).[6] Public transfers to backward regions increase the gross disposable income via the item 'net current transfers from outside' without influencing gross domestic product.

A third schedule is the capital formation and financial account. Here, receipts consist of internal and external savings and capital transfers, while payments are real (not financial) investments made internally to the region (by local and external firms) and capital account taxes or transfers paid (Table 5.2c). The balance records the internal

Table 5.2 The main social accounts at regional level

a) Gross domestic product account

Resources	*Uses*
Gross domestic product at market prices	Final internal consumption
Imports of goods and services	• public
	• private
	Final consumption by other regions
	• public
	• private
	Gross fixed capital formation
	• net fixed capital formation
	• depreciation
	Changes in inventories
	Exports of goods and services

b) Gross disposable income account

Sources	*Uses*
Gross domestic product at market prices	Final internal consumption
+ *Net compensation of employees and*	• private
property income from outside the region	• public
Total (Gross income at market prices)	Final consumption *from the rest of the world*
Gross income at market prices	• private
+ *Net current transfers from outside*	• public
+ Net indirect taxes	Gross saving
- Subsidies	• internal to the region
Total (Gross disposable income)	• *from the rest of the world*

c) Capital formation and financial account

Receipts	*Payments*
Gross domestic savings	Gross fixed capital formation
• internal to the region	• internal capital formation
• from the rest of the world	• capital formation from the rest of the world
Capital transfers (inward – credits)	*Capital transfer (outwards – debits)*

Note: Items from the balance of payments are in italics.

financial resources in surplus (if the balance is positive) or in deficit (if it is negative). At the level of the aggregate economic system, deficit internal resources must equal the resources obtained from outside the region. Vice versa, if internal resources are in surplus, they must equal the resources employed externally to the region. This means that, in accounting terms, the current account balance always equals the balance of the capital formation account.[7]

5.2.3 Macroeconomic conditions in interregional relations

The mechanisms by which interregional relations determine the levels of output (gross domestic product), income (gross disposable income) and capital formation subsume very different macroeconomic conditions. A positive balance of the current and capital accounts – a balance that favours local growth – may in fact result from the following very different circumstances:

a) a large volume of exports, due to a highly competitive productive system able to finance the imports which the region requires. This situation suggests that positive macroeconomic conditions characterize the region; a competitive productive system, in fact, signals the existence of high levels of real production, employment and income;
b) large public transfers (which are included among current account items as net current transfers) generating an increase in income but not in local production. In this case, growth is financed by other regions, income does not reflect any real local productive capacity, and the region is 'living beyond its means'. If the flows of external financing cease, as a result of a political decision or because of a national economic crisis, local growth may halt and have no chance of spontaneous recovery. The macroeconomic conditions that accompany this development path are likely to be a limited local degree of competitiveness associated with unemployment and stagnation;[8]
c) inward interregional capital movements for the purchase of property assets like land and buildings, which increase regional wealth held in liquid form and may engender greater spending on consumption. However, in this case too, the macroeconomic accounting equilibrium may conceal a situation of unemployment and stagnation;
d) inward interregional capital movements for direct investments in the region, with a positive impact on gross domestic product due to greater real investments stimulating employment and the region's real productive capacity. It is very unlikely that a positive balance of payments obtained by this means will conceal unemployment, especially in the long run;
e) inward capital movements of a short-term financial nature (business loans) that give rise to greater import volumes. The macroeconomic conditions concealed behind this accounting relation are less clear-cut.

In the theories that follow we shall see how these conditions alternate among models of local growth. The first theory examined – the export-base theory – argues that growth depends substantially on the competitiveness of the local production system (case (a) above): an (exogenous) increase in exports raises income and employment levels. The next model discussed, the one developed by Harrod and Domar, stresses the importance of inflows of savings and capital for the growth of income and employment: the logic of this model corresponds to cases (c), (d) and (e) above. Finally, Thirlwall's Law warns that development may be hampered by a negative trade balance if exports are conceived as the sole means to finance imports; that is, if only case (a) above holds.

5.3 The exporter region: the export-base model

5.3.1 Hoyt's model

The export-base model is the best known of those developed to determine the role of demand in growth and development. The main idea behind this model, in all its versions, is that whilst large economic systems, such as those of large countries, are able to rely on their own internal forces for their development, smaller economic systems – regions or cities, many of them specialized – cannot rely solely on endogenous capacities to achieve development: their economic growth is closely conditioned by factors external to the local system.[9]

The origin of the export-base model is interesting. In the 1930s, the Federal Housing Administration asked a planner, Homer Hoyt, to provide a simple instrument to forecast the physical growth of cities. For this purpose, Hoyt developed the first export-base model at the urban level. He distinguished employment in the base sector (E_b) (the sector in which the area specializes) from employment in the services (or non-base) sector (E_s) and formulated the following relations:[10]

$$\begin{aligned} E_T &= E_b + E_s \\ E_s &= aE_T \qquad \textit{with } 0<a<1 \\ E_b &= \overline{E}_b \end{aligned} \tag{5.1}$$

Total employment (E_T) is by definition the sum of employment in the two sectors. Employment in the base sector is exogenous to the economic system, while employment in the services sector is a share *a* of total employment. With appropriate substitutions and after some simple steps, we obtain:

$$E_T = \frac{1}{1-a} E_b \tag{5.2}$$

and in growth rate terms, within a single period:

$$\Delta E_T = \frac{1}{1-a} \Delta E_b \tag{5.3}$$

Equation (5.3) states that when employment increases in the base sector, total employment undergoes a more than proportional increase, whose amount is defined by the urban multiplier ($1/(1-a)$ – which by definition assumes values greater than one.

Assuming a simple proportion, equal to *b*, between total employment and the population resident in the area, we can write:

$$P = bE_T \qquad \textit{with } b>1 \tag{5.4}$$

By unifying equations (5.4) and (5.3), the growth of the resident population (and therefore the physical growth of the area) can be straightforwardly calculated as:

$$\Delta P = b\Delta E_T = \frac{b}{1-a} \Delta E_b \tag{5.5}$$

5.3.2 *The export-led Keynesian model*

During the 1950s – it seems entirely independently of each other – the economist Douglass North, and subsequently Charles Tiebout and Richard Andrews, developed an economic version of the Hoyt model. They replaced the physical variables of Hoyt's model with aggregate macroeconomic variables – income, demand internal and external to the region – in order to determine the economic growth of areas rather than their physical development.[11]

The economic version of the model was based on a traditional Keynesian aggregate demand model where aggregate income or production, Y, equals the components of aggregate demand, consumption C, exports X and imports M (assuming for simplicity, for the time being, that there is no public sector, G and $T = 0$):

$$\begin{aligned} &Y = C + X - M \\ &\textit{where}: \\ &X = \overline{X} \qquad\qquad \textit{with } 0<c<1 \textit{ and } 0<m<1 \\ &C = cY \\ &M = mY \end{aligned} \tag{5.6}$$

While exports are hypothesized as exogenous to the model, consumption and imports depend on the level of income and on the respective propensities to consume, c, and to import, m.

With simple substitutions and the consequent logical steps, equation (5.6) can be rewritten as:

$$Y = \frac{1}{1-(c-m)} X \tag{5.7}$$

In growth rate terms, equation (5.7) becomes:

$$\Delta Y = \frac{1}{1-(c-m)} \Delta X \tag{5.8}$$

This states quite simply that when an area's exports increase, production and income increase more than proportionally, as long as the marginal propensity to spend $(c - m)$ is less than unity – a condition guaranteed by the values that c and m assume by definition.[12]

Equation (5.8) is analogous to (5.5) in the previous model. Both state that external demand – measured in terms of exports (expressed in values or in units of employment in the sector producing for sales outside the region) – generates and determines the amount of local growth, doing so through its multiplier effects on local income (in the economic model), and on employment in the base sector (in the model with physical variables).

Equation (5.8) states that more rapidly developing regions are those able to maintain a surplus of exports over time – unless the expansion of initial exports is cancelled out by an even greater volume of induced imports. In fact, a greater propensity to import signals that most of the multiplier effects fall outside the region.

The export-led model can be expanded in two directions. In the first, consideration is made of all the components making up aggregate demand; with respect to the model just described, also private investments, public spending and tax rates are considered.[13] In this version of the model, the possible determinants of growth are not only an increase in exports but also a growth of investments or public spending. The second direction – the one taken by interregional income theory – consists of a model that is similar to the previous one but constructed on interregional bases. Exports of a region depend on the income produced in other regions – with the advantage that the link between the growth of local income and the growth of income in other regions can be taken into consideration.[14]

It is now possible to examine certain key aspects of this theory, aspects that also highlight its limitations. First, the theory does not imply, nor does it elaborate, an equilibrium growth rate. If a region has resources and productive capacity, an expansion of activity in the base sector (of exports in the economic version) generates a regional rate of growth without economic or physical constraints on development. Entirely lacking, in fact, is any treatment of the supply structure. Second, the theory does not concern itself with processes of convergence or divergence among regions, and therefore with relative growth. Convergence is only possible in so far as low-income regions are more likely to increase their exports. Yet there is nothing in the model able to interpret that likelihood. Finally, the theory is unable to define the determinants of growth because it takes the growth of exports (or increased employment in the base sector) to be a matter of fact, not as a result of the model.

5.3.3 An early dynamic version of the model

An early dynamic version of the model was formulated towards the end of the 1970s[15] in order to deal with one of the main criticisms brought against the original model: the constancy of the ratio between employment in services and total employment.[16] Employment in services, in fact, may easily increase independently of the trend in the base sector, for example as the result of autonomous investments in the region, or of a growth in per capita income. This possibility is included in a model very similar to Hoyt's, where the variable 'income' replaces the variable 'employment', the aim being to study the time trends of the growth rates of the variables. Equation (5.1) therefore becomes:

$$\begin{aligned} &Y_T = Y_b + Y_s \\ &\textit{where} \\ &Y_b = \overline{Y}_b \\ &Y_s = a_0 + a_1 Y_T \end{aligned} \tag{5.9}$$

where Y_b and Y_s denote the incomes generated respectively by the base sector and the services sector. The latter is made to depend on total income, as in Hoyt's model, and on a constant a_0 which measures exogenous variations in the income of the services sector. Simple substitutions produce the static equilibrium equation:

$$Y_T = \frac{a_0}{1-a_1} + \frac{Y_b}{1-a_1} \tag{5.10}$$

In development rate terms, equation (5.10) becomes:[17]

$$\frac{\Delta Y_T}{Y_T} = \frac{\Delta Y_b}{1-a_1}\frac{1}{\dfrac{a_0+Y_b}{1-a_1}} = \frac{\Delta Y_b}{a_0+Y_b} = \frac{\Delta Y_b}{Y_b}\frac{Y_b}{a_0+Y_b} \tag{5.11}$$

Equation (5.11) demonstrates the important role of constant a_0 in a dynamic process. If it assumes zero value, total income increases at the same rate as the base sector's income, which bears out the hypothesis of the constant ratio between employment (or income) in the services sector and total employment (or income). If, instead, constant a_0 assumes values greater or less than zero, the growth rate of income differs from that of income in the base sector, assuming respectively higher values (in the case of negative values of a_0) or lower ones (in the case of positive values of a_0). A study of the American regions has shown that a_0 often assumes negative values, and therefore that the growth rate of a regional income is higher than that of the base sector, because of a higher income growth rate in the services sector. In the United States, the higher income growth rate in the services sector is mainly determined by public investments (in the construction industry, for example), by an income elasticity to demand for local public services greater than unity, and by import substitution mechanisms that develop as local activity grows.

5.3.4 *A recent dynamic version of the model*

More recent years have seen formulation of a dynamic version of the export-base model,[18] the purpose of which is to verify the stability conditions of the equilibrium solution.

Starting from the well-known aggregate demand relation:

$$Y(t) = C(t) + X(t) - M(t) \tag{5.12}$$

and introducing time lags in the relations between consumption and income, and between imports and income, it is possible to state that consumption and imports at time t are defined by income at $t - 1$:

$$C(t) = cY(t-1) \tag{5.13}$$

$$M(t) = mY(t-1) \tag{5.14}$$

Equation (5.12) thus becomes:

$$Y(t) - (c - m)Y(t-1) = X(t) \tag{5.15}$$

Assuming that external demand for locally produced goods increases exponentially over time at a constant rate $g > 0$, the dynamic of regional income assumes the time trend shown by Figure 5.1.[19] With an initial income level equal to Y', the growth of exports induces income to grow towards Y^*. The same tendency is apparent if the region has an initial income level of Y''.

Interestingly, according to the same logic, if $g < 0$, income converges on a negative level Y^{**}, declining at a constant rate g provided that $0 < c - m < 1 + g$. The condition for convergence to come about is that the propensity to consume locally $(c - m)$ must be less than 1, a condition already contained in the static model and respected by definition. The dynamic characteristics of this model are quite simple, given the linear structure of its underlying relations. However, if non-linearity is introduced into these relations, it is possible to obtain explosive trends or sudden crises, depending on the structural changes that occur in the system (see Chapter 10).

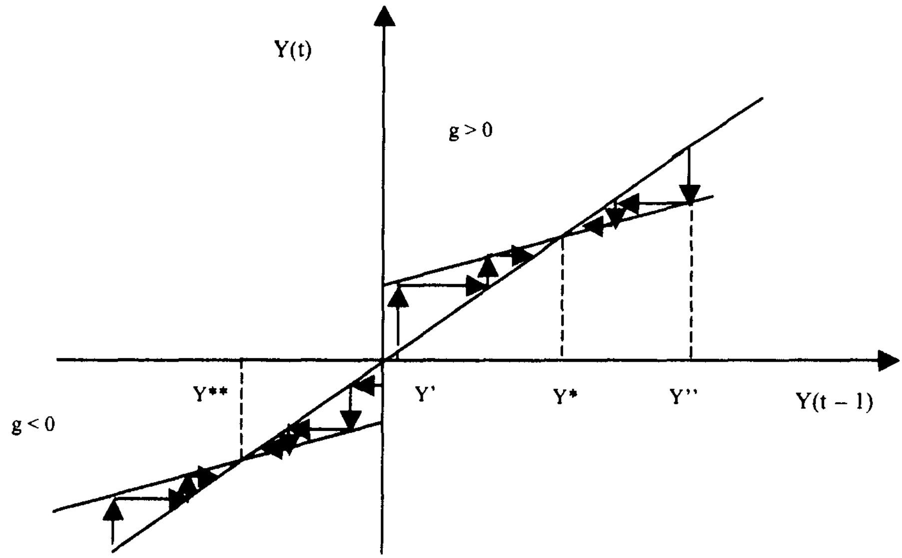


Figure 5.1 The dynamic equilibrium of the export-base model. Convergence towards development ($g > 0$) or decline ($g < 0$) at a constant rate g

5.4 A critical assessment of the model

5.4.1 General remarks

The great merit attributed to the export-base model, in all its versions, is that it has analysed regional development from the point of view of a small economic system. Using the purely economic logic of Keynesian macroeconomics, the model is able to demonstrate the decisive role performed by interregional trade relations in the growth of a small economic system. Given the difficulty of characterizing a local economic system as a self-sufficient economy, exports are treated in these models as a major component of aggregate demand, and autonomous investments only as a minor one.

The export-base model also reminds us that *productive specialization* is a key determinant of economic growth. The role assumed by a region in the international division of labour depends on its ability to identify the specific productive assets with which it can offer goods on a much broader market and acquire demand extending well beyond local barriers. This thesis is still largely valid today, and in the 1970s it was strongly supported by bottom-up development theories. Moreover, for the period in which this thesis was conceived, it represented a significant advance in regional development analysis. Given the way in which it is conceived, the export-base model also warns of the risks to specialized local economic systems when international demand for their products exhibits marked long-period fluctuations. On the logic of the model, in fact, just as an increase in demand for locally produced goods is a source of development, so a decrease in such demand may presage a recession.

This approach to local growth, with its conceptual simplicity, economic logic and simple application to real situations (given that it needed a relatively small dataset), enjoyed considerable success for a number of years in both regional and urban economics.

This success, however, induced analysts to ignore the intrinsic limitations of the model, which were mainly due to its a-spatial nature adapted from macroeconomics to a local context. A region is interpreted by this model as an internally uniform space that differs from an external space (like a nation does from the rest of the world) in terms of its different productive structure and specialization. But the model provides no explanation for this distinction. Its interregional version counterposes a region to others assumed to differ from it in their propensities to import and consume, and in their export capacities. All of these differences are hypothesized, but none of them are explained. The model is therefore able to identify, but not to interpret, the determinants of local development. For this reason, it is a model well suited to describing the development of areas historically specialized in certain industrial sectors, or in so-called 'Ricardian goods', or those goods connected with the availability of natural resources (e.g. minerals, specific agricultural products). Specialization in 'Ricardian goods' does not need to be explained, therefore, but is taken for granted. In all other cases, however, the way in which the export base is determined and converted into greater competitiveness has to be interpreted by analysing the structure and the dynamic of local supply, which are unfortunately entirely neglected by the model.

A second criticism concerns the high level of aggregation at which the analysis is conducted. No distinctions are drawn among different productive activities or different industrial specializations. The model implies that the multiplier effects of exports on income are of the same magnitude whatever sector produces the goods exported. But this implication is refuted when one considers that every sector of specialization generates a multiplier effect of greater or lesser proportions according to whether the demand for intermediate goods created by the production of export goods is addressed to internal producers, or whether, instead, it gives rise to greater imports from other regions.[20] Pioneering studies by Chenery in the early 1950s used a North/South Italy input–output table to show that the large-scale investments made in the Italian *Mezzogiorno* at the time benefited the industrialized North more than the developing South because of the marked leakage effects in the *Mezzogiorno*'s income multiplier.[21]

Moreover, the model assumes that there is no obstacle to an expansion of supply. If external demand increases, the system has the resources with which to augment productive capacity: unemployed production factors and productive capacity – either unused or expandable at nil cost – are assumed to be available. If this is not the case, an increase in demand engenders a short-term rise in prices, rather than a real physical expansion of production activity.

Use of the model for long-period predictions is restricted by its intrinsic assumption that multipliers are stable over time.[22] In the long period, in fact, it is easy to foresee – along a development path – that the productive diversification of the local economy will generate import substitution effects at odds with the multiplier's constancy. In the long period it is likely (and desirable) that an area's specialization will shift to more advanced sectors with higher value added, given the probability that, along its development path, the region will be able to transfer resources from declining sectors to emerging ones. If study of regional development confines itself to the logic of constancy in specialization, it may overlook one of the most interesting aspects of dynamic analysis: the structural change that accompanies long-period development.

For the same reason, a further shortcoming of the model is that it deliberately ignores the role of the residential (or services) sector in an area's growth, in that growth is solely dependent on the dynamic of the base sector. The level and quality of local, financial, managerial, marketing and technological services instead largely determine the long-period productivity and competitiveness of base sectors, especially in modern economies.

5.4.2 *Estimation of the 'economic base'*

Interest in application of the export-base model has led to the development of various methods with which to distinguish a region's base sector from its services sector, a distinction easy to draw in theory but difficult to apply in practice. The method most commonly suggested for this purpose is the location quotient technique, which distinguishes the two sectors according to the sector's employment share at regional level compared to the same sector's share at national level:[23]

$$QL_{ir} = \frac{E_{ir}}{E_r} \Big/ \frac{E_{in}}{E_n} \tag{5.16}$$

where i, r and n respectively denote the sector, the region and the nation, and E is total employment (although it can represent any sectorally disaggregated variable available at regional level – income for example). If the ratio between the shares is greater than unity, then the surplus is interpreted as a measure of production exceeding local demand, and therefore as consisting of net exports. With the exporting sectors thus established, summing employment (or income) yields an estimate of employment (or income) in the base sector.

However, the use of the location quotient to define the base sector has a number of shortcomings.[24] First, it assumes that consumers have the same tastes and preferences over space. If they did not, the surplus might not signify that production is more than enough to meet local demand, but rather that the structure of local demand is different from that of the country as a whole. Second, the method assumes that the economy concerned is a closed system; it presumes, in fact, that the nation does not export, so that the share of employment (or income) at the denominator represents only demand internal to the country. Because this is an unrealistic assumption, the location quotient underestimates the base sector. Finally, if the location quotient is instead calculated on the basis of units of production, the method assumes equal levels of productivity across space, which is a further unrealistic assumption.

Another method proposed for estimation of the base sector is the minimum requirements technique.[25] This starts from the assumption that the lowest share of employment in a sector in all regions is the minimum share necessary to satisfy the needs of a region, and that if the employment share in that sector is larger than in other regions, it signals that productive capacity is in excess of the region's requirements. The sum of employment in sectors with employment shares above the minimum determines employment in the base sector. This method too has its limitations. First of all, a region may have a very small proportion of employment in a sector, not because it produces only to satisfy its own needs, but because it is a net importer in that sector – an aspect that makes selection of the benchmark region highly arbitrary. Second, this method requires the same assumptions about the constancy of demand and productivity across space that have already been mentioned in regard to the location quotient technique.

5.4.3 *Estimation of the regional multiplier*

Assuming that the above-mentioned shortcomings are acceptable, application of the export-base model in prediction of a region's development requires estimation of the income multiplier. Early attempts in this direction began by identifying the base sector with the methods (with their shortcomings) just described.[26]

But there are at least two other methods commonly used to estimate the regional multiplier. The first of them consists in empirical estimation of the multiplier (i.e. the various marginal propensities). Originally proposed by Archibald, this method estimates the multiplier via direct estimation of the propensity to purchase goods locally.[27] It examines official figures on household consumption – detailed by spending category at national level – in order to identify goods and services with a high probability of being purchased locally (retail services, public services, cinemas, educational services, and so on). It then calculates the local shares compared to national ones and aggregates total spending for each. Repetition of the exercise for a certain number of years yields a time series of local spending which, when regressed on disposable income, produces an estimation of the marginal propensity to consume income at local level, $(c - m)$.[28] Once this value is known, the value of the regional multiplier can be easily obtained.

The second method was first proposed by Allen. It abandons the idea of directly estimating local spending, or the various marginal propensities (to consume and to import from other regions or abroad). Instead, it takes as a proxy for the regional multiplier the inverse of the 'leakages' on a region's gross domestic product. Allen identifies four channels of leakages from the multiplier effect on income: savings, interregional imports, imports from abroad, and direct and indirect taxes. Once the values of these are known, and their shares of income calculated, the inverse is nothing other than the value of the multiplier.[29]

5.5 Input–output analysis

While export-base methods are able to measure the extent to which local product changes with variations in external demand, there is a technique – called input–output analysis – that enables estimation to be made of the impact of growth in demand in a particular sector on output by each individual sector of the local economy, and on total output. Accordingly, export-base models can be classified as input–output models with only two sectors.

Using Wassily Leontief's model of sectoral interdependencies as its basis, input–output analysis can be used to predict the effects exerted by a growth of demand in a particular sector on the rest of the local economy. We shall see how.

Input–output analysis involves construction of a square $n \times n$ matrix. Recorded in this matrix are all the flows of sales (in the rows) and purchases (in the columns) that take place in a year among the n local production sectors; or in other words, the intermediate flows of goods (expressed in values) among the various sectors.[30] Completing the matrix are series of columns and rows. Recorded in the former are sales by each sector to final demand (public and private consumption, investments and exports);

Table 5.3 Simplified structure of an input–output table

	Intermediate demand Purchasing sectors 1 .... n					*Final demand*				*Total output*
						C	G	I	X	
Selling sectors										
1	A_{11}	.	A_{1j}	.	A_{1n}	C_1	G_1	I_1	X_1	R_1
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	
n	A_{n1}	.	A_{nj}	.	A_{nn}	C_n	G_n	I_n	X_n	R_n
Labour (wages)	W_1	.	W_t	.	W_n					W
Other components of value added (profits)	Π_1	.	Π_j	.	Π_n		Y			Π
Imports	M_1	.	M_j	.	M_n	M_c	M_g	M_i		M
Total output	R_1	.	R_j	.	R_n	C	G	I	X	

recorded in the latter are purchases of the original production factors, labour and capital (and therefore wages and profits) and purchases from abroad – and from outside, in the case of a sub-national input–output matrix (imports).[31]

By construction, the sum of each row represents each sector's revenue from sales of goods to other sectors and to final demand. The sum of each column represents the costs incurred by each sector in order to purchase intermediate goods, and goods produced externally, and to remunerate the production factors, wages and profits. The row values equal the column values. Moreover, the sum of final demand equals the sum of the income components or of value added, this being the regional output *Y* (Table 5.3).

With A_{ij} denoting the value of goods that sector *i* sells to sector *j*, and *C, G, I, X* and *R* respectively denoting private consumption, public spending, investments, exports and the value of output, the row sum (the revenue of a generic sector *i*) is given by:

$$\sum_j A_{ij} + (C_i + G_i + I_i + X_i) = R_i \qquad \forall i \tag{5.17}$$

and the column sum (the costs of a generic sector *j*) by:

$$\sum_i A_{ij} + W_j + \Pi_j + M_j = R_j \qquad \forall j \tag{5.18}$$

where *W* and *Π* are respectively wages and the other components of value added (profits). The regional gross domestic product *Y* is given by:

$$W + \Pi = Y = C + G + I + X - M = R - \sum_j \sum_i A_{ij} - M \tag{5.19}$$

The flows of goods between sector *i* and sector *j* can be expressed by the so-called 'technical coefficients' a_{ij}, which state the (technical/structural) relation between production by sector *i* and by sector *j*. In other words, the technical coefficients state how many euros of output by sector *i* are necessary for the production of one euro by sector *j*:[32]

$$A_{ij} = a_{ij}R_j \qquad \text{and} \qquad a_{ij} = A_{ij} / R_j \tag{5.20}$$

Substituting equation (5.20) in equation (5.17) and writing final demand as a single item *D*, we obtain, for every sector *i*:

$$\sum_j a_{ij}R_j + D_i = R_i \qquad \forall i \tag{5.21}$$

Equation (5.21) expresses the value of production (revenue) obtained by sector *i* selling output in part to final demand *D* and in part to other sectors.

After linear algebra operations on the matrix of the technical coefficients, equation (5.21) can be rewritten as:[33]

$$R_i = \sum_j b_{ij}D_j \qquad \forall i \tag{5.22}$$

where b_{ij} is the 'inverse Leontief matrix' or the 'multiplier matrix'. This matrix enables calculation of the value of output by each sector *i* generated directly or indirectly by one euro of final demand addressed to each sector *j*. In fact, it allows calculation not only of the direct effects of a certain amount of demand but also the indirect effects operating via final demand itself. For example, equation (5.22) is able to determine not only the production of wood generated by demand in the construction industry but also the production of wood generated by demand in the furniture industry, this demand being in its turn generated by increased demand in the construction industry. Whereas in export-base theory the Keynesian multiplier is synthesized into a single value, in input–output analysis it is disaggregated into a $n \times n$ set of multipliers relative to every sector or good demanded.

Input–output analysis is therefore a useful tool for forecasting the effects of a hypothetical increase in demand in a particular sector. If the technical coefficients matrix a_{ij} is known, it is possible to calculate the effect of an increase in external demand for a good (sector) ΔD_j on:

- the value of production by individual local sectors *i*: $\Delta R_i = b_{ij}D_j$
- local wages: $\Delta W = \sum_i \Delta R_i a_{wi} = \sum_i b_{ij} D_j a_{wi}$, where $a_{wi} = W_i / R_i$;
- local employment, assuming a constant sectoral average wage w_i^*: $\Delta L = \sum_i \Delta W_i / w_i^*$;
- local income: $\Delta Y = \Delta W + \Delta \Pi = \sum_i R_i(a_{wi} + a_{vi})$, where v denotes the other components of income besides wages.

The limitations of this methodology are well known. First, the constancy of the technical coefficients to increased production, as well as their constancy over time, requires all production to be at constant returns, and technical progress to be non-existent. The difficulties increase when input–output analysis is used at sub-national, regional or urban level. In this case, if the instrument is used to forecast the trend of the local economy, the technical coefficients matrix must be divided into an intraregional trade flow matrix and an interregional trade flow matrix. This enables measurement of the impact of increased demand in a particular sector on the local economy, and it excludes the 'leakages effects' on other areas of the country. But dividing the coefficients matrix in this way is a complex undertaking. It is usually performed with one of the following two methods: (a) the compilation of empirical survey-based tables, which is an accurate but extremely costly method; (b) desk research on the sector's specialization, on the assumption that the sector is able to meet increased demand only if it pertains to the specialization of the area.[34]

The use of input/output analysis to forecast and simulate the effects of economic policy measures at the local level entails the assumption that technical coefficients are constant over time. Consequently, the results of such analysis should be interpreted with caution.

5.6 The importer region: the Harrod–Domar model

5.6.1 The original model

In 1939 the economist Roy Harrod, and then entirely independently in 1957, Evsey Domar, investigated the rate of growth required for an economic system to maintain its initial macroeconomic equilibrium.[35] The model formulated by Harrod and Domar concludes that equilibrium of an economic system is more the exception than the rule, given that the growth path is highly unstable and very likely to diverge from the equilibrium growth rate.

The Harrod–Domar model was subsequently used to interpret the dynamics of regional economic systems.[36] The assumption behind this regional version is that imports – when analysed as the channel through which capital goods and savings are acquired from other regions – determine the local economy's growth rate; moreover, they allow for equilibrium growth conditions less restrictive – and therefore more easily sustainable over time – than those that applied to a national economy closed to foreign trade. Hence, while the export-base theory highlights the importance of external demand as the engine of development, the Harrod–Domar model emphasizes that the regional dynamic may also be driven by investments originating from other regions that stimulate local output and income: cases (c), (d) and (e) discussed in Section 5.2.3.

As we shall see, the Harrod–Domar model also differs from the export-base theory in that it is a multi-period model. It abandons the strictly short-period logic that we have seen thus far and assumes a longer-term perspective. Moreover, far from conceiving saving as a deduction from effective demand, this approach stresses the importance of saving as a means with which to acquire resources for productive investments.

The model is based on the following assumptions:

1 production of a single good, which can be used either as a final good (in which case it is consumed and exits the economic system) or as a capital good (in which case it remains in the system and engenders the production of other goods);
2 the non-perishable nature of the capital good, which means that there is no need to differentiate between investments in new capital and investments to amortize capital assets;
3 a constant propensity to save;
4 fixed production coefficients, i.e. constancy in the quantities of the production factors 'capital' and 'labour' necessary for one unit of output – which is equivalent to assuming the absence of technical progress;

5 growth of labour at a constant rate n equal to growth of the population;
6 investments proportional to the increase in demand as defined by the accelerator theory:

$$I_t = v_t(Y^*_{t+1} - Y_t) \textit{ with } 0<v_t<1 \tag{5.23}$$

where v_t – the 'acceleration coefficient', assumed constant by the model – measures the increase in capital (I_t) with a unit increase in demand ($Y^*_{t+1} - Y_t$). It therefore expresses the capital/output ratio;
7 income proportional to the increase in investments, as suggested by the Keynesian multiplier theory:

$$Y_{t+1} - Y_t = \frac{1}{s}(I_{t+1} - I_t) \textit{ with } 0<s<1 \tag{5.24}$$

where s is the propensity to save and $1/s$ is the factor of proportionality (or the Keynesian multiplier) between investments and income: unit increases in investment have a more than proportional effect (equal to $1/s$) on income.

An increment in aggregate demand entails an increase in the capital and labour production factors necessary to adjust production to the new level of consumption. Assuming full employment in order to maintain this equilibrium condition, on the hypothesis of constancy in the technical coefficient of production (L/Y), the labour growth rate must be equal to the population growth rate, n.

On the capital side, as suggested by equation (5.23), an increase in demand generates an increase in investments, the financing of which requires an amount of savings (S) equal to the necessary investment (I). If this is the case, the economy grows at a rate – called the 'warranted growth rate' (y_t) – equal to:

$$y_t = \frac{\Delta K}{K} = \frac{I}{K} = \frac{sY}{K} = \frac{sY}{Y}\frac{Y}{K} = \frac{s}{v} = n \tag{5.25}$$

Equation (5.25) states that the growth rate must be equal to the ratio between the propensity to save s and the acceleration coefficient v, which in turn must be equal to the rate of growth of the labour force. If this is the case, the initial equilibrium between aggregate demand and production will be maintained over time.[37]

However, if the system grows at the warranted growth rate y_t, it moves, in Harrod's words, along a 'knife edge' between the risk of explosion on the one hand, and of recession on the other. Any disequilibria generated by conditions external to the system, in the absence of exogenous interventions, tend to be aggravated by signals emitted by the market that induce firms to operate in the direction opposite to the one required for equilibrium to be re-established. In an economic system, in fact, planned investments in excess of actual savings ($I > S$) signal that effective consumption is greater than expected saving,[38] and therefore that effective demand is greater than expected demand. Firms react to the short supply of goods in the system by increasing their investments – a reaction that, paradoxically, worsens the disequilibrium. The increase in investments generates, via the multiplier effect on income, a more than proportional increase in income and aggregate demand. The latter increasingly diverges from expected demand, with severe inflationary effects in the long run. In the same way, if planned investments are less than effective saving ($I < S$), this signals that expected demand is greater than effective demand. Firms react to a surplus of supply in the system by reducing their investments, thereby slowing the growth of effective demand even further, and in the long period causing recession.

In short, the model shows that there are forces within an economic system in initial disequilibrium that push it further and further away from stationary equilibrium and, according to the initial conditions, towards either inflationary conditions or deep recession.

5.6.2 *The regional version of the model*

When the Harrod–Domar model is adapted to the regional context, it yields interesting results in addition to the ones furnished by the national version. The distinctive feature of the model's regional version is its macroeconomic equilibrium condition, which for a regional economy is:

$$S + M = I + X \tag{5.26}$$

where M and X respectively denote the imports and exports of capital from/to one region and the other. Regions are not closed economic systems (if they were, the regional model would not be different from the national one): they have close relations with other regions, with which they exchange goods and production factors.

For a generic region i, equation (5.26) can be rewritten as:

$$(s_i + m_i)Y_i = I_i + X_i \tag{5.27}$$

that is:

$$\frac{I_i}{Y_i} = s_i + m_i - \frac{X_i}{Y_i} \tag{5.28}$$

where m is the propensity to import capital proportional to income. Equation (5.25) thus becomes:

$$y_i = \frac{s_i + m_i - \dfrac{X_i}{Y_i}}{v_i} = n_i \tag{5.29}$$

Equation (5.29) states that – unlike in a closed economy – capital may grow at the same rate as output (thus guaranteeing the steady state) even if investments tend to outstrip savings, provided that the gap between savings and investments is covered by a surplus of net imports.[39] A regional economic system can finance investments not only with internal savings but by importing capital goods from other regions.[40] Net exports may likewise help maintain the steady-state equilibrium, when there is a surplus of internal saving, because they make up the shortfall between low internal consumption and the level of production corresponding to full use of productive capacity.

Similarly, full employment in a region with an internal shortage of labour may be maintained by an inflow of workers from other regions, while outflows of migrants to other regions may offset unemployment in the region. The labour-market equilibrium condition is therefore:

$$y_i = n_i - e_i \tag{5.30}$$

where e_i is the net migratory balance (emigrants minus immigrants) in each period of time as a percentage of the regional population P_i.

A first important result obtained by the regional version of the Harrod–Domar model is that the conditions for constant-rate growth are less restrictive – and therefore more easily sustainable over time – than those governing a national economy closed to foreign trade. However, once again the steady-state equilibrium can be interpreted as the exception rather than the rule, given that there are no conditions in the model ensuring interregional flows of labour and capital sufficient to guarantee growth at a constant rate. That is to say, there are no conditions within the model that guarantee that the interregional flows of production factors will equilibrate the system.

A second important result of the model in its 'regional' version is the following: regions characterized by net surpluses of imports, that is those for which:

$$m_i - \frac{\sum_i X_i}{Y_i} > 0 \tag{5.31}$$

are regions which grow more rapidly than others – propensity to save and capital/output ratio remaining equal. In fact, according to the logic of the model, a net surplus of imports gives rise to a higher growth rate because this surplus represents extra savings injected into the economic system from outside, as shown by equation (5.29).

Finally, the third important result of the Harrod–Domar model concerns the time trends of differences among regional growth rates. If, as is the case in the real world, there are initial differences among the growth rates of regions, the model shows that these differences not only persist but increase with the passage of time. In fact, when the initial growth rate of region i is higher than that of region j ($y_i > y_j$), it follows from equation (5.29) that

$$\frac{\sum_i m_j Y_i}{Y_i} \quad \textit{where by definition} \sum_i m_j Y_i = \sum_i X_i \tag{5.32}$$

diminishes, giving further impetus to y_i.

The two latter findings remind us, on the realistic assumption that a poor region is a net importer of capital from rich regions, that the model hypothesizes convergence by regional growth rates on steady-state equilibrium. In fact, equations (5.31) and (5.32) state that the growth rate of a poor region is constantly higher than the rates of the advanced ones: a situation brought about by convergent development and that demonstrates that, in contexts characterized by a scant propensity to save and a low capital/output ratio, the propensity to import performs a crucial role in intranational and international regional re-equilibrum processes.

Finally, more detailed analysis of the *s/v* ratio is required, given its importance in equation (5.29) for explanation of the regional growth rate. This equation states that greater growth is achieved by regions with high propensities to save and with low capital/output ratios (v); regions, therefore, that make efficient use of capital or have low capital-intensive sectoral structures (service sectors, for example). This latter feature has enabled the Italian region of Lombardy to attain high levels of development despite a long-standing low accumulation rate (1970s and 1980s). By contrast, the Italian *Mezzogiorno*, where investments were made in capital-intensive sectors for at least 30 years (1955–1985), has achieved limited growth despite large injections of (public) investments and external savings.

This observation enables us to state that, whereas in the single-period Keynesian model saving is detrimental to growth because it subtracts from effective demand and limits multiplier effects – as indicated by the term $1 - c$, equal to s, at the denominator – even in Keynesian models (if they adopt a genuinely dynamic, multi-period and long-term perspective) the availability of saving and capital formation, that is supply rather than demand elements, explain regional growth.

5.6.3 *Critical assessment of the model*

The Harrod–Domar model is well suited to describing and interpreting the growth of regional economic systems apparently characterized by critical macroeconomic conditions. Limited internal savings, a low capital/output ratio and a negative trade balance are expected to hamper macroeconomic growth. But this situation is contradicted in the real world by numerous regional systems, and it can only be explained if a decisive role is given to the external capital that makes up the internal saving shortfall.[41] When the same logic is used to identify the relative growth path, it is able to account for converging growth rates of regions which differ markedly in their levels of internal investment and saving – levels that would otherwise signal divergence.

Although the regional Harrod–Domar model furnishes useful insights, it can be criticized on various grounds, all of them concerning the fact that it was originally developed to interpret the macroeconomic conditions of a country's growth and was only subsequently adapted to a regional setting.

The first weakness of the model is its inability to predict whether interregional flows of production factors will restore equilibrium. This inability is due to the absence of mechanisms that regulate and interpret those flows. However, since the model was not originally formulated to explain flows of production factors, it is obvious that a theory of resource mobility is entirely lacking.

A second and evident limitation of the model is that it is unable to demonstrate clear tendencies towards divergence or convergence among regions. It is true that in the real world backward regions are generally net importers of capital, and therefore that, on the logic of the model, they are faster growing regions; as a consequence, they are able to converge on the growth rates of advanced regions. Yet it is equally true that in the real world backward regions are generally also net exporters of labour; a situation that, according to the logic of the model, is accompanied by lower rates of income growth, and divergence rather than convergence.

Finally, although we may accept the model's finding that backward regions are net importers of capital (as happens in the real world), the model provides no explanation as to the determinants of this greater capacity to attract capital. Whilst in the national version of the model, the reasons can be conceived as favourable macroeconomic conditions attracting capital (a higher interest rate), in its regional version they should be identified in location factors typical of any geographical area, which can only be identified using a microeconomic approach. Not surprisingly, therefore, they are entirely absent from a macroeconomic model with uniform-abstract space like the one developed by Harrod and Domar.

5.7 Balance of payments and local growth: Thirlwall's Law

At the beginning of the 1980s, the importance of exports for a region's growth was once again affirmed. Whilst the export-base model interprets exports as signalling the competitive advantage of a region, and therefore its potential development, the theory now discussed – known as Thirlwall's Law – gives exports a decisive role in the development process: they sustain the trade balance, and in the absence of other mechanisms, they enable the financing of the imports necessary to satisfy internal demand.[42] In fact, if exports were scant and failed to meet the area's import needs, a disequilibrium would arise in the regional trade balance; and this, over the long period, would restrict imports, the satisfaction of local demand, the needs of local industry and development. On this view, an area's failure to develop is due to disequilibrium in the regional trade balance.[43]

A region can therefore maintain a certain level of growth if there is equilibrium in its trade balance. Assuming that the growth rate of exports depends on the growth rate of world income (y_w), weighted by the elasticity of export demand to world income (α), and the growth rate of imports (m), in its turn dependent on the growth rate of regional income (y_r), weighted by the elasticity of import demand to local income (β),[44] equilibrium in the regional trade balance is ensured if the following equality holds:

$$\alpha y_w = \beta y_r \tag{5.33}$$

that is:

$$y_r = \frac{\alpha}{\beta} y_w \tag{5.34}$$

This last relation states that a regional growth rate depends on the growth rate of world income and on the ratio between the two elasticities of demand to income. In order to stimulate local development a region must foster an industrial structure with sectors whose exports have high elasticity of demand to world income (α) and, at the same time, whose imports have low elasticity of demand to local income (β). On this approach, therefore, local development is nothing other than a problem of industrial conversion to sectors whose exports and imports have respectively greater or lesser elasticities of demand to income.

The interest of this theory resides in its twofold contention that (a) imports are a *sine qua non* for internal production, which requires raw materials and non-locally produced intermediate goods; and (b) that exports are crucial for growth because they are the most immediate source of import financing.

However, there are other import-financing methods that enable the limits imposed on growth by a negative trade balance to be bypassed, as shown in Section 5.2. Numerous examples can be cited in the real world of rich exporting regions that co-exist with backward regions; in the latter, public transfers and private investments from advanced regions finance a negative trade balance and local income growth in the long run.

It therefore seems that, like any demand-driven theory, Thirlwall's Law has a certain validity if it is used to explain short-period development, given that its concern is not to interpret the specialization and competitiveness of the productive system.

5.8 Conclusions

This chapter has analysed Keynesian models of regional growth in which demand components are the engine of development. They are theories that interpret regional development as growth in employment and income driven by increased demand. The chapter has frequently pointed out that these theories restrict themselves to describing short-period growth: they entirely ignore the structure and dynamic of the production system, taking it for granted that a region can increase supply in response to expanding – also external – demand.

For these reasons, it is risky and misleading to use Keynesian theories in interpretation of a long-term growth path, or as the basis for policy measures to support a long-run dynamic. If they are instead used to overcome the specific normative problem that they address – reducing unemployment in the presence of a given productive capacity – they have obvious merits: they are simple and at the same time rigorous in their economic reasoning, and they are easy to apply to a real context.

Finally, the chapter has shown that when Keynesian theories discard the short-term perspective and adopt a long-term, multi-period one, as in the case of the Harrod–Domar model, they are able to abandon demand aspects and highlight the importance of supply components – such as the availability of saving and capital formation – for interpretation of the regional growth process.

Mathematical appendix

The differences equation (5.15) is solved, as usual, by separating the solutions of the homogeneous and particular equations.

The homogeneous equation, which we define as $Q(t)$, is:

$$Q(t) - (c - m)Q(t-1) = 0 \tag{5.1a}$$

Setting

$$Q(t) = K\phi^t \tag{5.2a}$$

we obtain

$$\begin{gathered} K\phi^t - (c-m)K\phi^{t-1} = 0 \\ \text{i.e.} \\ \phi = c - m \end{gathered} \tag{5.3a}$$

(5.2a) thus becomes:

$$Q(t) = K(c-m)^t \tag{5.4a}$$

with constant K, whose value is not yet known.

The particular equation is defined as:

$$R(t)-(c-m)R(t-1)=X(0)(1+g)^t \tag{5.5a}$$

Defining:

$$R(t)=h\psi^t \tag{5.6a}$$

(5.5a) becomes:

$$h\psi^t-(c-m)h\psi^{t-1}=X(0)(1+g)^t \tag{5.7a}$$

and therefore:

$$h\psi^t(1-\frac{(c-m)}{\psi})=X(0)(1+g)^t \tag{5.8a}$$

This holds if and only if:

$$\psi=(1+g) \tag{5.9a}$$

and

$$\begin{aligned} X(0)&=h-\frac{(c-m)h}{\psi}\\ h&=\frac{X(0)\psi}{1-\dfrac{c-m}{\psi}} \end{aligned} \tag{5.10a}$$

Recalling that $\psi=(1+g)$, (5.10a) becomes:

$$h=\frac{X(0)(1+g)}{(1+g)-(c-m)} \tag{5.11a}$$

Substituting the values of h and ψ in (5.6a), the particular solution of (5.5a) becomes:

$$R(t)=\frac{X(0)(1+g)^{t-1}}{(1+g)-(c-m)} \tag{5.12a}$$

The solution of (5.15) is given by the sum of the homogeneous solution $Q(t)$ and the particular solution $R(t)$, i.e.

$$Y(t)=K(c-m)^t+\frac{X(0)(1+g)^{t+1}}{(1+g)-(c-m)} \tag{5.13a}$$

which is nothing other than (5.17). Using the initial condition to define K, which is still unknown, we obtain:

$$Y(0)=K(c-m)^0+(1+g)^0\frac{X(0)(1+g)}{(1+g)-(c-m)} \tag{5.14a}$$

or:

$$K = Y(0) - \frac{X(0)(1+g)}{(1+g)-(c-m)} \tag{5.15a}$$

which is nothing other than (5.18).

If $c - m < 1$, the $\lim_{t \to \infty} K(c-m)^t = 0$. It follows that the regional income converges on a development path with constant equilibrium rate g, as shown graphically by Figure 5.1.

Review questions

1 What conception of space is used in the regional growth theories of the 1950s and 1960s and why?
2 How can one define a theory of regional growth driven by demand dynamics?
3 How is the balance of trade structured? How do the trade balance, the service balance and the capital transfer balance enter the regional social account balance?
4 What macroeconomic conditions can be hidden behind a regional income balance in the presence of an openness of the region to external trade?
5 What is argued by the export-base theory (in all its formulations) and what are the strengths and weaknesses of this theory?
6 What are the methodologies to measure the regional consumption multiplier?
7 How can an input–output table be built and for what purposes? What do the technical production coefficients and the coefficient matrix represent?
8 What additional element is contained in the Harrod–Domar model in its formulation at the regional level? What are the main weaknesses and strengths of this model?
9 Is it true that a regional growth can be hampered by a negative trade balance? Explain why.

Selected reading on empirical findings

About regional multipliers

Allen K.J. (1969), 'The Regional Multiplier: Some Problems in Estimation', in J.B. Cullingworth and S.C. Orr (eds), *Regional and Urban Studies: a Social Science Approach*, Allen & Unwin, London, pp. 80–96.

Archibald G. (1967), 'Regional Multiplier Effects in the United Kingdom', *Oxford Economic Papers*, vol. 19, no. 1, pp. 22–45.

Daley W.M. (1997), *Regional Multipliers: A User Handbook for the Regional Input–Output Modelling System (RIMS II)*, U.S. Department of Commerce, Economics and Statistics Administration, U.S. Government Printing Office, Washington, DC 20402, Third edn, available online, www.bea.gov/scb/pdf/regional/perinc/meth/rims2.pdf.

Faggian A. and Biagi B. (2003), 'Measuring Regional Multipliers: A Comparison Between Different Methodologies in the Case of the Italian Regions', *Scienze Regionali – Italian Journal of Regional Science*, vol. 2, no. 1, pp. 33–58.

Steele D.B. (1969), 'Regional Multipliers in Great Britain', *Oxford Economic Papers*, New Series, vol. 21, no. 2, pp. 268–292.

About input–output analysis

Camagni R. (1982), 'L'Impatto sull'Economia Sarda della Spesa e dell'Investimento Turistico in Costa Smeralda', *Quaderni Sardi di Economia*, no. 4, pp. 371–413.

Carter H.O. and Brody A. (eds) (1970), *Applications of Input–Output Analysis*, North Holland, Amsterdam.

Chenery H. (1962), 'Development Policies for Southern Italy', *Quarterly Journal of Economics*, vol. 76, no. 4, pp. 515–547.

Cuihong Y. (2002), 'The Impacts of Water Conservancy Investment by Using Input–Output Technique: A Case of China', paper presented at the 14th International Conference on Input–Output Technique, Montreal, 10–15 October.

Harringan J. (1982), 'The Relationship between Industry and Geographical Linkages: A Case Study of the United Kingdom', *Journal of Regional Science*, vol. 22, no. 1, pp. 19–31.

Hewings G.J.D., Israilevich P.R., Sonis M. and Schindler G.R. (1997), 'Structural Change in a Metropolitan Economy: The Chicago Region 1975–2011', in C.S. Bertuglia, S. Lombardo and P. Nijkamp (eds), *Innovative Behaviour in Space and Time*, Springer Verlag, Berlin, pp. 183–212.

Israilevich P.R., Hewings G.J.D., Schindler G. and Mahidhara R. (1996), 'The Choice of an Input–Output Table Embedded in Regional Input–Output Models', *Papers in Regional Science*, vol. 75, no. 2, pp. 103–119.

Ning A. and Polenske K.P. (2005), 'Application and Extension of Input–Output Analysis in Economic-impact Analysis of Dust Storms: A Case Study in Beijing, China', paper for presentation at the 15th International Input–Output Conference held in Beijing, 27 June–1 July.

Nyhus D. (1983), 'Observing Structural Change in the Japanese Economy: An Input–Output Approach', in A. Smyshlyaev (ed.), *Proceedings of the Fourth IIASA Task Force Meeting on Input–Output Modeling*, International Institute for Applied System Analysis, Laxenburg, Austria.

Polenske K. (1970), 'An Empirical Test of Interregional Input–Output Models: Estimation of 1963 Japanese Production', *American Economic Review*, vol. 60, no. 1, pp. 76–82.

Further reading

Chenery H. (1962), 'Development Policies for Southern Italy', *Quarterly Journal of Economics*, vol. 76, no. 4, pp. 515–547.

Domar E.D. (1957), *Essays in the Theory if Economic Growth*, Oxford University Press, London.

Harrod R.F. (1939), 'An Essay in Dynamic Theory', *The Economic Journal*, vol. 49, no. 193, pp. 14–33.

Hoyt H. (1954), 'Homer Hoyt on the Development of Economic Base Concept', *Land Economics*, vol. 30, no. 2, pp. 182–187.

North D. (1955), 'Location Theory and Regional Economic Growth', *Journal of Political Economy*, vol. 63, no. 3, pp. 243–258.

Richardson H. (1978), *Regional and Urban Economics*, Pitman, London.

Thirlwall A.P. (1980), 'Regional Problems Are Balance of Payments Problems', *Regional Studies*, vol. 14, pp. no. 5, 419–425.

Notes

1 This circular process is driven by the Keynesian multiplier as defined at Chapter 4, note 1.

2 In the case of a region, exports (imports) comprise sales to (purchases from) subjects in other regions, and even the country itself.

3 Henceforth, 'balance of payments' will be used to denote the current account and the capital account balances. The financial account balance, which, as explained, equalizes the balance of payments, will be excluded.

4 The national accounts are published by the national statistical offices. Most of them furnish statistics on value added, consumption, investments and incomes which can be used to calculate savings and therefore the gross disposable income and its appropriation account and the capital formation account. The availability of publications on regional accounts varies among countries. Single items of the accounts are made available by Eurostat at different geographical levels (NUTS0, NUTS1 and NUTS2).

5 In macroeconomics, the resources and uses account (Table 5.2a) is the balance between aggregate demand and supply expressed by the following relation:

$$Y + R = C + I + G + X - M + T \qquad (5.1n)$$

where Y denotes output, R public transfers, C consumption, I investments, G public spending, $X - M$ the trade balance, and T the tax yield.

6 This relation is written in macroeconomics as:

$$Y + R = C + S + T \qquad (5.2n)$$

7 Equalizing equations (5.1n) and (5.2n) we obtain:

$$I + G + X - M = S + T \qquad (5.3n)$$

and therefore:

$$I + (G - T) = S + (M - X) \qquad (5.4n)$$

If the public budget is in equilibrium ($G - T = 0$), investments can be financed with either internal savings or external savings (business loans and capital investments) ($S + (M - X) = I$). This becomes clear if we assume that an economic system is initially in macroeconomic equilibrium ($S + (M - X) = I$) and that a direct investment takes place from outside. This investment is recorded in the capital account as a receipt, while its monetary counterpart is recorded as a monetary inflow; thus the account is balanced. At national level, the increase in investment obtained is entered in the capital formation account as an inflow. Given the assumption of initial macroeconomic equilibrium, the saving is sufficient to cover only the internal investment; the investment from outside therefore engenders an increase in the imports (machinery, raw materials, etc.) necessary for the investment and equal to its value. The current account balance thus perfectly counterbalances the balance of the capital formation account.

8 This circumstance highlights that a fiscal policy intended to assist backward regions by means of greater public transfers only affects the income level. It gives no stimulus to the region's productive capacity, nor does it ameliorate unemployment and stagnation.

9 Stabler argues: 'The size of the area in question also has a major bearing on the importance of what phenomena are most important in generating growth.' See Stabler, 1970, p. 53 (1st edn 1968). See also Aydalot, 1985.

10 See Weimer and Hoyt, 1939; Hoyt, 1954.

11 See North, 1955; Tiebout 1956; and Andrews 1953 and 1954. Andrews is also the author of numerous articles on the subject published in issues of *Land Economics* between 1953 and 1956. North developed his export-base model in critical reaction to the theory of the stages of development (see Section 4.3). The latter was ill-suited to interpreting the growth of certain states (regions) of America, and especially those on the West Coast. In these regions, the earliest phases of development had not been characterized by a subsistence economy, but by the industrial production of large quantities of goods, the bulk of which were sold on external markets.

12 $1/(1-(c-m))$ is the analytical expression of the Keynesian multiplier; its economic meaning is defined at Chapter 4, note 1.
13 In this case, aggregate demand is defined as:

$$Y - T + R = C + I + G + X - M \qquad (5.5n)$$

and the multiplier becomes, if an income tax rate equal to t is hypothesised (i.e. setting $Y_d = Y - tY + R$):

$$\frac{1}{1-(1-c)(1-t)} \qquad (5.6n)$$

applicable to changes in any of the components of aggregate demand.
14 In this case, the aggregate demand of a generic region r is defined by the equation:

$$Y_r = C_r + I_r + G_r + X_r - M_r \qquad (5.7n)$$

Exports are the sum of imports by all the other regions, and income is disposable income, once taxes have been subtracted, as follows:

$$\begin{aligned} X_r &= \sum_j M_{rj} = \sum_j m_{rj} Y_j^d \\ Y_{d_r} &= Y_r - tY_r \\ T_r &= t_r Y_r \end{aligned} \qquad (5.8n)$$

where j denotes the generic other regions, m_{rj} the propensity to import from outside the region, t the tax rate, tY the tax revenue, and Y_d the disposable income. Defined mr_a as the propensity to import from abroad, the multiplier becomes:

$$\frac{1}{1-(c_r - m_{ra} - \sum_j m_{rj})(1-t_r)} \qquad (5.9n)$$

15 See Tiebout, 1960; Richardson, 1978, p. 87.
16 See Tiebout, 1956; Weiss and Gooding, 1968.
17 The income differential is given by:

$$\Delta Y = \frac{1}{1-a_1}\Delta Y_b \qquad (5.10n)$$

Bearing in mind that income is defined by equation (5.10), the percentage variation in income thus obtained is:

$$\frac{\Delta Y}{Y} = \frac{\Delta Y_b}{1-a_1}/Y = \frac{\Delta Y_b}{1-a_1}/\frac{a_0+Y_b}{1-a_1} = \frac{\Delta Y_b}{1-a_1}\frac{1}{\frac{a_0+Y_b}{1-a_1}} \qquad (5.11n)$$

18 See Miyao, 1984.
19 If external demand for locally produced goods is hypothesized as increasing exponentially over time at a constant rate $g > 0$:

$$X(t) = X(0)(1+g)^t \qquad (5.12n)$$

the solution of the differences equation (5.15) is given by:

$$Y(t)=\frac{X(0)(1+g)^{t+1}}{1+g-(c-m)}+K(c-m)^{t} \tag{5.13n}$$

where K is a constant defined by the initial income condition:

$$K=Y(0)-\frac{X(0)(1+g)}{1+g-(c-m)} \tag{5.14n}$$

If $c-m<1$, $\lim_{t\to\infty} K(c-m)^{t}=0$ the. It follows that the regional income converges on a development path at a constant equilibrium rate g, as illustrated by Figure 5.1. The appendix to this chapter contains the mathematical solution of the differences equation (5.15).

20 This weakness can be remedied by using an input–output table in which the sectoral and geographical (internal and external to the region) disaggregation of commercial interrelations shows the actual multiplier mechanism operating at local level.

21 See Chenery et al., 1953; Chenery, 1962. Sirkin (1959) stressed that multiplier effects change not only according to the sector of specialization but also according to the *level* of specialization. More specialized areas necessarily require greater openness to interregional trade, especially if there exists highly diversified internal demand that stimulates greater commercial exchanges. It is therefore highly unlikely that the multiplier effect will be the same in regions with different structural features.

22 Greenhut pointed out as early as 1959 that 'the region's export base is not a datum. That is to say, the base changes with time, as currently produced private and social goods help bring forth new goods that change the base.' See Greenhut, 1959b, p. 71; Greenhut, 1966.

23 Use of the location quotient to define the base sector was first proposed by Hildebrand and Mace, 1950.

24 See Pratt, 1968.

25 See Ullman and Dacey, 1960.

26 Knowing the size (in units of employment or in value) of the base sector and the total sector, from equation (5.7) it is possible to obtain the value of the multiplier: in fact, $Y/X = 1/(1 - c + m)$. Of course, this is an average value when instead a marginal value is required ($\Delta Y/\Delta X$). the two values are only equal, in fact, if there are no autonomous expenditure items apart from exports.

27 See Archibald, 1967. A similar approach has been applied by McGuire, 1983, who calculated the multiplier for two localities in Scotland. For an application to England, see Steele, 1969. For a critical review, see Wilson, 1968.

28 In our case, we assume that:

$$C_i=a+bY_d \tag{5.15n}$$

i.e. that a proportion (b) of local spending Ci depends on income, while a proportion (a) does not. The regression method allows estimation of the values of the parameter, and in particular b, which is the marginal propensity to spend disposable income locally.

29 See Allen, 1969. The logic of this method can be understood by recalling that if regions were closed systems, the only variable reducing the value of the multiplier would be the propensity to save the increase in income. However, the inverse of the propensity is nothing but the Keynesian multiplier. This method has been recently used to estimate the multiplier for Italian regions; see Faggian and Biagi, 2003.

30 In order to sum the flows of diverse kinds of goods, they must obviously be expressed in value terms, not in quantities.

31 The same method of analysis is used at national, regional or urban level. In the case of a regional or urban input–output matrix, meant by 'imports' and 'exports' are flows into or

out of the region (or city), not just flows to or from abroad. For a critical description of the theory see Tiebout, 1957.

32 For example, the technical coefficients between the car industry and the rubber industry express the value of the rubber necessary to produce the value of a car.

33 Expressing equation (5.21) in matrix form, where R and D are the two sectoral vectors of the value of production and the value of final demand, A is the matrix of the technical coefficients and I the identity matrix, produces:

$$R = AR + D \tag{5.16n}$$

$$(I - A)R = D \tag{5.17n}$$

$$R = (I - A)^{-1}D \tag{5.18n}$$

With $B = (I - A)^{-1}$ defined as the Leontief inverse matrix, we have:

$$R = BD \tag{5.19n}$$

which is nothing other than (5.22).

34 If the area happens to be an island, use can be made of harbour and airport statistics documenting the value of goods entering and leaving the island. These statistics are excellent means with which to separate local and interregional effects. They have been used in Italy to assess the impact of a building project for the Costa Smeralda on the growth of the Sardinian economy. See Camagni, 1982. For detailed discussion of input/output analysis see Hewings, 1977; Hewings et al., 2001; Martellato, 1982.

35 See Harrod, 1939 and Domar, 1957. For a critical examination of Harrod's theory see Hawtrey, 1939.

36 See Richardson, 1969.

37 Equation (5.25) is constructed by setting $I = sY$, i.e. $I = S$, which is the macroeconomic equilibrium condition. In fact, the $I = S$ equality is an accounting identity which always holds *ex post*. On the hypothesis of an economy closed to foreign trade and in the absence of a public sector, output is either wholly consumed or wholly invested ($Y = C + I$). Income, on the other hand, is allocated between consumption and investments ($Y = C + S$). If $C + S = C + I$ – a condition that holds only if $I = S$ – the output offered is equal to the output sold, and the value of output is equal to the income earned, which in its turn is either spent or saved.

38 The $Y = C + S$ equality states that income is either spent or saved. If actual saving is less than planned investment, this means that effective consumption is greater than planned investment, and therefore that effective demand (defined by the level of consumption) is greater than expected demand.

39 The constraint for the system as a whole is that interregional trade must be balanced, i.e. that:

$$\sum_i \sum_j M_{ij} = \sum_i \sum_j X_{ij} \qquad \text{for each } i \neq j \tag{5.20n}$$

40 This condition is analysed, and explained in accounting terms, in Section 5.2.

41 The importance of imports for the growth and competitiveness of countries has recently been re-examined by the eminent economist Paul Krugman. The real purpose of international trade, Krugman argues, is to obtain imports, not to export. Exporting is only a way to finance imports, which are less costly than the direct production of what one needs. See Krugman, 1996a, p. 19.

42 In Thirlwall's words: 'export demand is a vital element in regional demand, which is necessary to compensate for a region's appetite for imports, in the absence of other compensating

expenditure'. See Thirlwall, 1980, p. 422. See also Thirlwall, 1980 and McCombie, 1992; McGregor and Swales, 1985.

43 Thirlwall argues, in fact, that 'regional problems are balance of payments problems'; see Thirlwall, 1980.

44 In microeconomics, the elasticity of the quantity demanded (supplied) to income measures in percentage terms the extent to which the quantity demanded (supplied) varies with a 1 per cent change in income.

6 Factor endowment

6.1 Factor endowment and regional growth

This and the next parts of the book examine theories that focus exclusively on supply components to explain long-period regional dynamics. In the light of the theories described in the previous chapter (in particular the export-base model), they consequently not only view exports as the engine of development but take a step further by identifying the factors responsible for the greater export capacity, and therefore the competitiveness, of a local economic system. If an economic system is able to export – or in other words, if it is able to gain a role in the international division of labour – it must enjoy some form of advantage: it must be able to produce goods at lower prices, supply higher-quality products and place new goods on the market. An economic system can fulfil these various requirements if it has more efficient productive processes, a complex and advanced local industrial system, modern production services and infrastructures, good quality resources and advanced production technologies – and also if its area comprises broad, diversified and advanced knowledge developed by complex cultural, social and economic processes.

There are therefore numerous sources of territorial competitiveness; and not surprisingly very different approaches have been taken to their analysis. This chapter presents theories that have concentrated on *factor endowment* as the source of territorial competitiveness. Although they differ in certain of their basic assumptions, these theories comprise a broad corpus of strictly neoclassical models that adopt diverse hypotheses on the mobility of goods and production factors in their treatment of growth from a *resource-based* perspective. Imbalances in interregional factor endowments, and differences in levels of factor productivity, account for the advantage enjoyed by a local system in its relations with the rest of the world. These are the elements that underlie the growth path and that condition its timing and the form that it takes.

According to these theories, it is trade in goods or factors that explains the adjustment of the relative prices of goods and factors, increased productive capacity and the achievement of full employment. For theories that assume the perfect mobility of production factors among regions (neoclassical growth models), differing remunerations of the production factors reallocate resources in space, and thus generate a higher rate of growth – according to typically neoclassical reasoning.[1] For theories that instead conceive goods as mobile (theories of interregional trade), differing levels of factor productivity give the region a comparative advantage in the production of a particular good, which it is able to export owing to price differential. Moreover, it is in the region's interest to resort to the external market for the purchase of those goods that it produces at a lower level of productivity than other goods. These imported goods are sold on the external market at prices that are more competitive than they would be if the goods were produced internally to the region.

It should be noted that the concept of 'growth' is used here with a meaning other than that given to it by the theories discussed in the previous chapter. The reason for this difference in the meaning of growth is the fact that these models have different policy concerns: not high unemployment – to be reduced by increased demand for local goods – but problems of poverty, underdevelopment and inequalities in the distribution of income. Growth is consequently no longer interpreted as an increase in employment and short-term income; rather it is conceived as individual well-being (and its interregional convergence), which is achieved either through increases in factor productivity, and consequently in wage levels and per capita income (neoclassical macroeconomic models), or through specialization processes that generate interregional trade, and consequently advantages deriving from the purchases of goods offered on the external market at prices lower than they would be if the goods were produced internally.

These theories have a number of distinctive features that should be borne in mind. The first group of them – classical and neoclassical with factor mobility – are distinctive in that they make reference to a concept of 'relative growth', the purpose being to identify and explain paths of convergence or divergence in the levels and rates of output growth. In this respect, neoclassical models of factor mobility are still today erroneously viewed as only able to explain a tendency of local economies towards convergence. But the modern versions of these theories show that, if increasing returns are introduced into the neoclassical production function, behaviours and tendencies are produced that differ greatly from the original model's mechanistic and univocal result of re-equilibrium in income levels among regions (see Chapter 11). Moreover, after modification of the original model by its authors to comprise two sectors, it is able to explain divergent trends in income levels if an initial equilibrium condition is assumed (see Section 6.2.2).

The distinctive feature of the second group of theories – classical and neoclassical, on interregional trade – is that they employ the concept of *relative advantage*, or *comparative advantage*, first formulated by Ricardo in his classical model of international trade and on the basis of which it was possible to identify a region's specialization. Among all the goods that can be offered on the external market, the region exports those that it produces at relatively lower production costs. This difference in production costs is due to the differing relative productivities of the factors used to manufacture the goods. This statement essentially means the following: even if a region produces all goods at higher prices, so that it is generally more inefficient in its production processes than any other region in the country, it may nevertheless be *relatively less inefficient in producing one particular good*. The region will thus be able to obtain a role for itself in the international division of labour by specializing in production of the good in which it is relatively more efficient. As we shall see, this argument has major normative implications, for it asserts that there is always an automatic mechanism guaranteeing the existence of some specialization, regardless of productive efficiency, and therefore that economic policy measures to foster development are unnecessary. The significance of this assertion is so far-reaching that it requires total guarantee of its truthfulness, although, as we shall see, this truthfulness is undermined by the ease with which economic mechanisms operating at national level are automatically expected to apply at regional and local level as well (Section 6.4). The models described in this chapter draw their theoretical framework from the classical and neoclassical theories of growth and international trade. Once again, therefore, they are approaches to regional *growth* that envisage a uniform-abstract space in order to treat economic conditions, everywhere identical, in terms of aggregate economic indicators.

The next section discusses models constructed on the assumption of perfect mobility of the production factors, at nil transportation costs, and immobility of the goods produced.[2] Section 6.2 reverses the terms of this hypothesis and considers the idea that it is the production factors that are immobile, while the goods produced are perfectly mobile. In the case of production factors with nil transport costs, it will be shown that the neoclassical theory, besides being a theory of local growth, is also a theory of the mobility of the production factors; if goods are perfectly mobile, the theory of local growth is also a theory of interregional trade.[3] The latter derives from neoclassical theories of international trade that comprise models notable for their elegant economic logic but criticizable for the facility with which they are applied to a local setting.

6.2 Regional growth and factor mobility

6.2.1 The one-sector model

The pioneering neoclassical model of regional growth was formulated by the economists George Borts and Jerome Stein at the beginning of the 1960s. It makes the usual assumptions of a neoclassical growth model:

- perfect competition in the goods market;
- perfect competition in the production factors market, which means that production factors are remunerated at their marginal productivity, guaranteeing profit maximization for the entrepreneur;
- full employment achieved by means of flexibility in the remuneration of the production factors;
- perfect mobility of the production factors among regions, at nil cost;
- total immobility of the goods produced;
- adjustment of the capital/labour ratio according to the dynamics of the production factors; there is therefore perfect substitutability between the two factors in the production of two goods.

In neoclassical theory, economic development depends on technical progress on the one hand, and on growth of the production factors on the other. These components are synthesized into the regional aggregate production function, which is expressed by a Cobb–Douglas function with constant returns:[4]

$$Y = AK^{\alpha}L^{1-\alpha} \tag{6.1}$$

where $0 < \alpha < 1$. Y denotes income, A technical progress, K capital, L labour, and α and $1 - \alpha$ respectively the efficiency of capital and labour.

In logarithms, the change in income Y over time is:[5]

$$y = a + \alpha k + (1 - \alpha)l \quad (6.2)$$

where the lower-case symbols *y, a, k, l* respectively represent the growth rates over time of income, technical progress, capital and labour. Equation (6.2) states that whether income will grow over time depends on the growth of technical progress, and on the growth of capital and labour. Equation (6.2) can also be rewritten as:

$$y - l = a + \alpha(k - l) \quad (6.3)$$

which highlights a further important aspect: growth in the productivity of labour and/or per capita income (indicated by the left-hand member of the equation) is equal to growth in technical progress and the capital/labour ratio. In the absence of technical progress, per capita output can only increase if the growth of capital exceeds that of labour. For the same reason, the steady state – i.e. the dynamic equilibrium in which the capital/output ratio or per capita output remain unchanged as income increases – is guaranteed when the growth rate of capital equals that of labour.

According to the neoclassicals, growth is a matter of the optimal allocation of inter- and intra-regional resources. In an open economy with perfect factor mobility, a more efficient interregional allocation of resources requires the production factors to shift to where their productivity is highest, and where they receive the greatest remuneration. In a region, therefore, the growth rate of capital (k) depends on the amount of internal savings (sY) available to finance investment (ΔK), and on the differential between capital remuneration in the area (i_r) and capital remuneration in the rest of the world (i_w). In symbols, this means that:

$$k = \frac{sY}{K} + \mu(i_r - i_w) \quad (6.4)$$

In the same way, labour grows with the growth of the population (n) and the increase in the differential between wage remuneration in the region and the rest of the world ($w_r - w_w$):

$$l = n + \lambda(w_r - w_w) \quad (6.5)$$

μ and λ represent the extent to which capital and labour move according to remuneration differentials.

Assuming the existence of two regions – a poor South with more labour than capital, and a North with conversely more capital than labour – capital migrates from the rich area to the poor one; and, vice versa, labour migrates from the South to the North. As a consequence, owing to different levels of factor productivity, remunerations are higher in the region where the factor is less abundant (Figure 6.1). The outflow of labour from the South enables it to increase productivity and therefore to increase remuneration of the labour factor. The same positive effect ensues from the outflow of capital from the North. The reallocation process halts when the regions attain the same factor productivities, the same remunerations, the same factor endowments, and therefore the same levels of income, in full employment. In Figure 6.2, where the

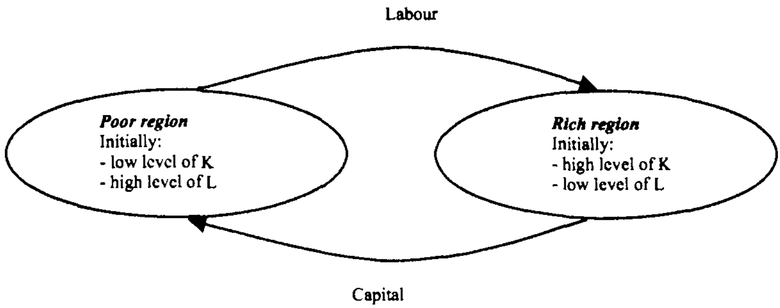


Figure 6.1 Interregional flows of production factors in the one-sector and two-regions model

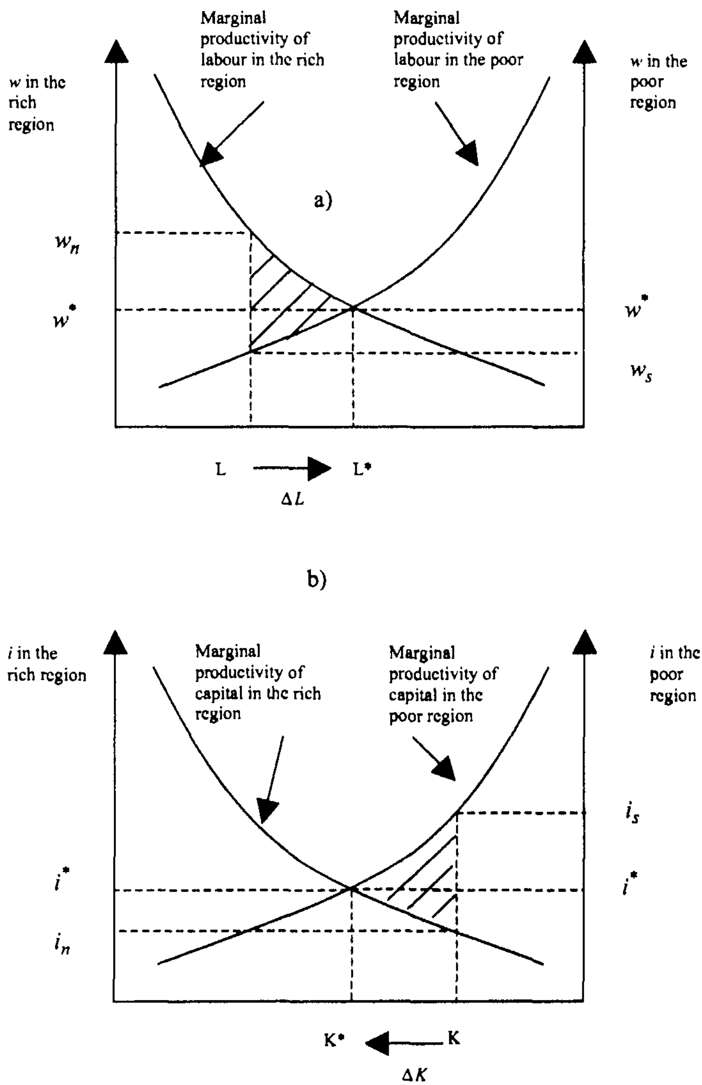


Figure 6.2 Production advantages from resource reallocation

Source: McCombie (1988)

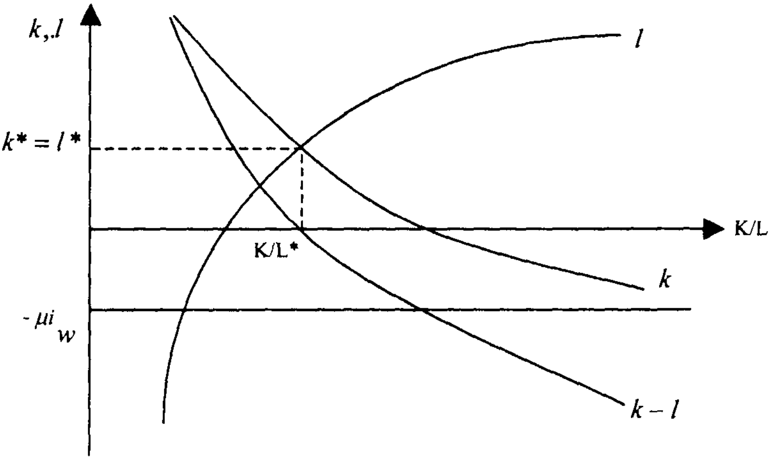


Figure 6.3 Steady-state equilibrium in a neoclassical model

complete availability of labour in the two regions is implicit in the x-axis, the area below the marginal productivity curve is the region's volume of output. The reallocation of resources generated by the differentials in factor remuneration therefore gives the North net advantages in terms of the increased production depicted by the dashed area in Figure 6.2a, while that of the South is equal to the dashed area in Figure 6.2b.

Put in dynamic terms, the model reaches steady-state equilibrium when capital and labour grow in exactly the same proportions. Figure 6.3 shows the shapes of the curves for which the growth rates of capital and labour are nil.[6] Straightforwardly obtained from these is the curve that represents constant growth of the capital/labour ratio used to analyse the dynamic properties of (6.3);[7] the steady-state equilibrium is reached at a certain positive value of the K/L ratio. Mathematical proof can be provided of the existence, uniqueness and stability of the equilibrium solution; the latter coincides with the point at which the growth rate of the capital/labour ratio is nil. If the capital growth rate curve meets the labour curve for negative values, the region does not grow but instead constantly declines. We shall see later how dynamic equilibrium may fail to come about if increasing or decreasing returns are incorporated into the model.

6.2.2 *The two-sector model*

When the neoclassical theorists were confronted by empirical evidence that apparently refuted their conclusion that capital flows to regions with low income levels, they were obliged to develop a different approach that would furnish a better interpretation of the real movement of the factors, and that would in particular confirm the tendency of capital to shift to areas with higher wage remunerations.[8]

The model that the authors of the original one developed for this purpose – known as the two-sector model – incorporates more realistic assumptions and emphasizes the role of the inefficient allocation of resources within the same region as the determinant of intra- and interregional flows of production factors.

The result is surprisingly different from that produced by the one-sector model: again according to neoclassical logic, the production factors migrate because they are attracted by higher remunerations. However, the subsequent reallocation of resources, due to an external shock that moves the regions far from the initial steady-state equilibrium, pushes local economies towards permanently different growth rates.[9]

The model is based on the following assumptions, some of which were already made by the one-sector model:

- the existence of two regions, each of which has two sectors producing two goods, one for export and one for domestic use, the former characterized by high labour productivity, the latter by low productivity. These sectors are often identified as industry (with high productivity) and agriculture (low productivity);
- disequilibria in the trade balance, which by hypothesis are offset by private capital movements;
- perfect competition in the goods market: the quantities sold by the individual regions do not influence the good's price on the world market, whilst the price of the domestic good is determined by local demand and supply;
- use of the capital factor only in the industrial sector: an assumption that does not affect the final result, as subsequently demonstrated by the authors;
- constant returns in the production of the goods;
- remuneration of the production factors at their marginal productivity;
- equality between the cost of the production factors and the value of the marginal product of the factors, which guarantees profit maximization for firms.[10]

Starting from a situation of initial equilibrium, in which the growth rate is stable and *uniform* between the regions, and in which capital and labour grow in each of them at a constant rate equal to that of income, the model shows how the growth rates of the two regions vary if an exogenous shock is introduced.[11] Suppose that demand for the good exported by one of the two regions increases: the price of the good rises as a direct consequence. This effect has a positive impact on the value of the marginal product of the factors in the region. The outcome is an intra- and interregional reallocation of production resources, as follows:

- capital stock in the sector producing for export increases as a result of the inflow of external capital attracted by greater remuneration;
- labour demand by local firms increases because of the increase in the value of the marginal product of labour (generated by the rise in the exported good's price);
- the greater demand for labour attracts workers both from the local agricultural sector and from other regions, given the higher remunerations available;
- finally, the expansion of production and employment in the sector producing for export has a backwash effect on the agricultural sector, which records an increase in demand for the good, and, consequently, in production and employment.

In this model, therefore, production growth results from a more efficient allocation of resources to the manufacturing sector, with its higher productivity. After an initial stimulus triggered by increased demand for the exported good, the endowment of productive resources in the manufacturing sector is augmented by investments from outside, and by migrations of workers from other regions and from the agricultural sector.[12]

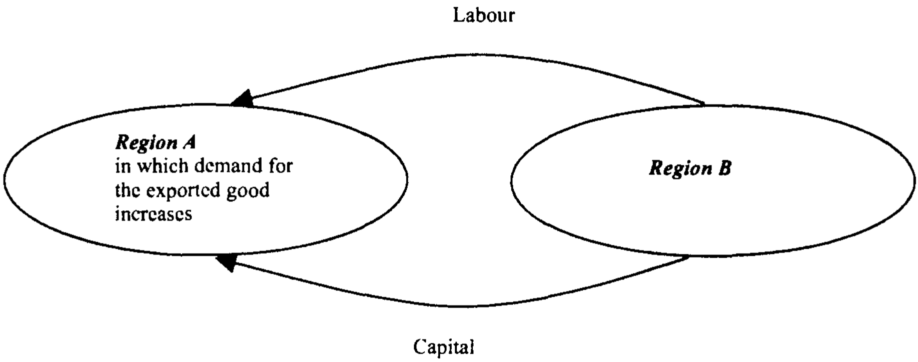


Figure 6.4 Interregional flows of production factors in the two-sectors and two-regions model

Two main conclusions ensue from this model. They differ from – indeed they conflict with – those of the single-sector model examined in the previous section. First, the production factors now move to the same high-wages region (Figure 6.4); the model is therefore supported by its authors' empirical results on the American regions.

The second conclusion is even more interesting: the model demonstrates that there is a tendency for regional growth rates to *diverge.* The reason for this is as follows: the income generated in the region exporting the manufacturing good differs from disposable income in an amount that equals the remuneration of the capital borrowed externally.[13] Internal saving, calculated as a share of disposable income, will therefore never be enough to finance local production. The shortage of capital guarantees high remuneration of this production factor, and this stimulates a constant inflows of capital from outside. As a result, the region's growth rate is persistently higher than that of other regions. The flow of workers into the exporter region alters the capital/labour ratio and thus attenuates the divergence in growth rates.[14] The agricultural sector acts upon growth rate disparities in two ways in the model: (i) it supplies labour to the exporter sector, so reducing the divergence between growth rates; (ii) in the growth process, its augmented goods demand stimulates production and attracts new workers from outside, once again mitigating growth rates disparities.

Interestingly, although the two-sector model starts from completely different assumptions and although it develops within a necessarily different conceptual framework, it reaches the same conclusion as the Harrod–Domar model. Both models argue, in fact, that if a region is a net importer of capital, it will have higher growth rates. Moreover, in line with Keynesian models of cumulative development à la Myrdal/Kaldor, the two-sector model demonstrates that this advantage persists over time, and thus aggravates regional disparities.

Contrary to the widely held belief, the neoclassical model also envisages divergence among growth rates, not just their convergence. It does so by assuming an initial steady-state, thus eschewing the one-sector model's hypothesis of initial disequilibrium between regions. More recently, when the model has been given dynamic formulation with increasing returns, it has been able to account for divergence in growth rates even on the basis of initial disequilibrium conditions (see Chapter 11).

6.2.3 *Critical assessment of the neoclassical approach*

The elegant and rigorous economic logic of the neoclassical models just described still today gives them wide currency among analysts of regional growth. Generally acknowledged as their main merit is their attribution of a prime role to production factor mobility in the regional growth process. This mobility has greater impact at the regional rather than national level because there are fewer spatial and social frictions impeding resource mobility between regions than between countries.

Wealthy regions are highly attractive to labour. But the decreasing returns consequent on the intensive use of labour may diminish their competitiveness. In the same way, backward regions offer locational advantages due to their relatively lower wages and unit labour costs, and therefore attract capital (or at least they do so in the one-sector model) which increases the competitiveness of local industry. Traditional labour-intensive manufactures may therefore be advantageous to backward or newly industrialized areas.[15]

However, the persistence of marked regional disequilibria suggests that these locational advantages are not enough to close the gap between advanced and backward regions. Strong areas are able to absorb the decreasing returns that accompany industrialization and high capital intensity, whilst the weak regions of the advanced countries have to compete with the low unit labour costs characteristic of the underdeveloped countries, and are therefore squeezed between the rich North and the poor South.[16]

The persistence of regional disequilibria also suggests that migratory flows, as the neoclassical theories interpret them, encounter a number of obstacles in reality, the first and perhaps most obvious of them being the economic and psychological costs of resources mobility. Assumed to be nil in the models examined above, these costs may instead explain why the factors do not move in the direction indicated by the model, or may not move at all.

Capital tends to remain in rich regions because of cumulative processes and synergies attendant on the process of development. Technical progress in the form of product and process innovations, new knowledge, processes of collective learning and agglomeration economies in general induces firms to invest only in rich regions already endowed with capital. Often supplementing these economic advantages are social and environmental conditions unfavourable to productive activities in regions with low per capita incomes.

Labour mobility, too, may encounter obstacles. First, the flow of labour to rich regions may well depend on the state of the strong region's economy; migratory flows may not take place to a rich but stagnant region with limited prospects of economic growth.[17] Moreover, migration from weak areas to strong ones is often a '*selective migration*' involving higher-skilled workers, who are able to find employment matching their expertise in the strong region. This type of migration inevitably deprives the weak area of more efficient and skilled resources, and thus works against possible convergence rather than for it. Finally, there may well exist 'imperfections' in the labour market that distort the perfect competition mechanism at the basis of the neoclassical logic, so that wages may increase even in the presence of unemployment in other regions; indeed, they may exacerbate that unemployment.

According to the theory of the 'Italian dual economy' propounded by Vera Lutz,[18] wage increases imposed by the trade unions in strong areas (though the argument applies to strong sectors or firms as well) create wage dualism and segmentation in the

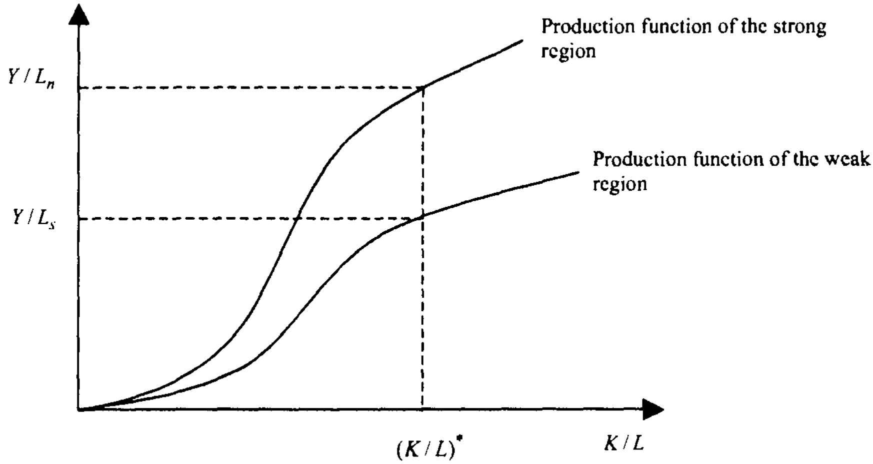


Figure 6.5 Different production functions in two regions

national labour market. According to Lutz, the strong area reacts to the increased labour cost by laying off workers and introducing more advanced technologies. The unemployed workers move to the weak area, where they are willing to accept lower-paid jobs with very low levels of productivity. Moreover, the homogenized wage levels among regions – brought about by national-level bargaining despite the presence of wide factor productivity differentials – are dangerous labour-market imperfections, in that they eliminate the economic incentives which induce interregional mobility, of capital in particular. On this neoclassical view, persistent underdevelopment may well be due to the presence of institutional and social factors that restrict the mobility of resources to more efficient allocations, besides the low specialization and scant competitiveness that, according to the two-sector model, determine the initial growth of regions.

Finally, the uniqueness of the production function for all regions is somewhat unrealistic. If it is removed, the results of the model change: in the presence of different technologies, an equal capital/output ratio among regions no longer guarantees an equal level of production (Figure 6.5). In the real world, regions are very likely to produce different goods; while the spatially uneven distribution of specific technological knowledge and factors necessary for firms also partly explains the limited mobility of capital among regions.

6.3 Factor immobility, specialization and well-being

6.3.1 *Specialization and comparative advantage: Ricardo's classical model*

It was stressed in the previous chapter that one limitation of the demand-driven theories of regional growth is their inability to explain regional specialization. One of the first attempts to find an endogenous explanation for a region's specialization consists in the application of international trade models to interregional trade. These models are applied at regional level on the belief that, since interregional trade flows are larger and freer than international ones (owing to shorter distances and the absence of economic barriers to trade), theories developed at international level can be applied at the regional one as well.[19]

However, this contention may be erroneous. Although a region tends to specialize as much as and perhaps more than a nation, and although trade with the rest of the world may therefore be extremely important for its growth and well-being, the economic factors that influence international trade cannot be transposed indiscriminately to the regional level. Either they do not work exactly as they do at national level (flexibility of prices and wages, equilibrium in the balance of payments) or they do not exist (exchange-rate fluctuations).

There is a further aspect that should be emphasized. These models were not developed to explain growth. They were instead conceived to interpret the specialization patterns of areas in the production of goods and the advantages deriving from trade – assuming the factoral endowment as given. They therefore determine the conditions that generate the greater sectoral specialization of regions, local production remaining equal. The advantage of specialization is the higher level of individual well-being that results from the lower relative prices of goods, which suggests a tendency to local development.

According to the models inspired by the theory of international trade, regions (and countries) exchange their goods on the basis of a *comparative advantage*, not an absolute one. This amounts to saying that even if a region produces all goods at higher costs and prices, and is thus generally more inefficient than the rest of country in the production of all goods, it may be *relatively less inefficient in the production of one particular good*. It thus acquires a role in the international division of labour by specializing in the good that it produces at relatively less inefficient conditions. This result, known as the 'Torrens–Ricardo paradox', becomes clear on examining the logic of the model that produced it: David Ricardo's theory of comparative costs.[20]

The assumptions of the model are the following:

- there are two regions, the North and the South, which produce two goods, A and B;
- there is only one production factor, labour, whose productivity differs between the two regions;
- goods are produced without increasing or decreasing returns: marginal costs are constant;
- there is perfect mobility of production factors within the region, and perfect immobility of them between regions;
- there is no money, so trade takes place in 'units of goods' according to a barter system.

Table 6.1a reports the production costs of the goods in terms of the hours of labour required to produce them. The North produces one unit of both A and B with one labour-hour, while the South requires two labour-hours to produce good A and fully four hours to produce good B. The South is therefore more inefficient than the North in the production of both goods. If we stop at this point, according to the logic of absolute advantage, the North has no reason whatsoever to trade with the South and to purchase goods 'more expensive' than those which it produces internally.

Table 6.1 Absolute and comparative advantages in the production of two goods in two regions

a) Absolute advantage
(labour-hour per unit of good)

Goods regions	*A*	*B*
North	1	1
South	2	4

b) Comparative advantage
(opportunity cost in terms of units of the good forfeited in order to obtain an additional unit of the other good)

Goods regions	*A*	*B*
North	1	1
South	1/2	2

According to the logic of the model, differences in absolute production costs are not enough to explain the advantages of trade between North and South: the analysis, according to Ricardo, should instead be made on the basis of *comparative costs*, or *opportunity costs*, defined as the quantity of another good that must be forfeited in order to be able to produce one unit more of a particular good. In our numerical example, in order to produce one extra unit of good A in the North, it is necessary to forfeit one unit of good B (and in the same way, to produce one extra unit of good B it is necessary to forfeit one unit of good A). In the South, to produce one extra unit A it is necessary to forfeit two units of good B (and to produce one extra unit of good B it is necessary to forfeit ½ a unit of good A).

Table 6.1b shows the comparative costs in production of the two goods in the two regions. The North is more efficient than the South in the production of good B; vice versa, the South is more efficient than the North in the production of good A. As a result, the North specializes in the production of only good B, and the South in good A, and the two regions exchange the amounts of those goods that are surplus to local demand.

How much do the regions gain from the trade? Let us suppose that the price of good B in the international market is fixed at 1.5 units of A, this being the intermediate price between the price of one unit of A in the North and two units of A in the South. Hence, if the North shifts one labour-hour from production of A to production of one unit of B, and exports the extra unit thus produced, it receives in exchange 1.5 units of A; whereas if it produces only A in that hour it will receive only one unit of A. The North thus saves half an hour of labour (0.5). Likewise, if the South specializes in the production of A, in one hour of extra labour (allocated to the production of A) it produces ½ a unit of A, forfeiting ¼ of a unit of B. By trading the half unit of A on the market at the ratio 1:1.5, the South obtains one-third (0.33) of a unit of B (0.5 * 1:1.5), instead of the one-quarter that it could produce internally: the South saves one-third (0.33) of a labour-hour.[21]

Both regions benefit from the exchange: their 'gains from trade' induce them to produce the good with which they enjoy a comparative advantage. Each region sees trade as an opportunity to obtain the imported good by resorting to a production technique superior to the one available within the region, and which allows the imported good to be produced at a lower 'labour value'. This increases the population's well-being.[22] Hence the reallocation of labour to more efficient uses yields greater individual well-being and a higher level of production.

The assumptions of constant costs and the unlimited availability of the production factor imply that regions achieve complete specialization; each region is induced to produce one single good, the good with which it enjoys a comparative advantage.

The main result of the model is that, according to the theorem of comparative advantage, there is an automatic mechanism – generated by the market – that ensures that a region will always have some specialization regardless of its real capacity to produce competitively. This obviously leads to the radical claim that regions are *always* able to attain a role in the national and international market whatever their real productive capacities may be. When this does not happen – as evinced by the wide regional disparities that still exist in the advanced countries – it is only because there are elements that distort the normal workings of the factors market. However, before dispensing with policies to support local competitiveness and the convergence of regional growth paths, as the theory of comparative advantage suggests, it is necessary to determine whether the validity of the theory of comparative advantage holds at regional level. It will be shown below that it is highly unlikely that the theory of comparative advantage does so, with the consequence that regions compete solely on the basis of absolute advantage.

The Ricardo model has a number of weaknesses. Principal among them is that the difference in labour productivity between the two regions, which generates the comparative advantage, is not explained. Nevertheless, the concept of comparative advantage has intrigued economists because of the rigorous, and highly counter-intuitive, logic on which it is constructed, with the result that it has been too hastily incorporated into regional economics.[23] Only recently has its applicability to the regional context been explicitly disputed.[24]

6.3.2 *The theory of factor endowments: the neoclassical Heckscher–Ohlin model*

Within a neoclassical framework, in 1933 a Swedish economist, Bertil Ohlin, reprised a study already begun by Eli Heckscher in 1919 to formulate a model of international trade that remedied some of the unrealistic assumptions of previous theories and is known as the 'Heckscher–Ohlin model'.[25]

The Heckscher–Ohlin model (also known as the model of factor-endowment) is structured on the assumption that production factors are immobile: an assumption typical of international trade models and the opposite of the assumption made by the neoclassical growth model. It accounts for the tendency of regions to assume sectoral specialization by evidencing the reasons for the differing factoral productivities of regions (or countries), which are assumed to be exogenous in Ricardo's model. Given immobile production factors and freely tradable goods, the factor-endowments model shows that it is more convenient for a region to specialize in manufactures that make the most intense use of the most abundant production factor in the area because it is relatively less costly. Of all the goods that the region can sell on the external market, it should specialize in the one that it can manufacture at relatively lower production costs and then export thanks to the price differential. In the same way, it is more convenient for the region to resort to the external market to purchase goods offered at prices lower than those that, because of the lower level of productivity, the region would be able to achieve by producing those goods internally.

To reach these important conclusions, the model starts from the following assumptions:

- there are two regions, North and South, each producing only two goods with only two factors, capital and labour;
- each good is produced with a different factor intensity: good A, steel, requires more capital than labour; good B, corn, requires more labour than capital;
- the production factors are qualitatively identical but differ in quantity between the two regions; the North has more capital than the South; the South has a greater quantity of labour;
- the production functions are identical in the two regions. This assumption precludes the possibility that the comparative advantage derives from interregional differences in production technology, contrary to the Ricardo model;
- there is perfect competition in the market for production factors. Consequently, the equilibrium price of each good is equal to the marginal cost required to produce it, and the price of each factor equals the value of marginal productivity;
- demand conditions are identical in the two regions: hence they do not alter the direct relation between the relative prices of the goods before and after trade. In other words, the different price of a good in the two regions does not reflect differences in consumer preferences, but is due only to differences in the relative prices of the factors;
- production factors are immobile; the regions have the same factor endowments before and after trade;
- the goods produced are traded on the national and international markets; trade is free of any obstacles, such as tariffs or transport costs.

One of the simplest versions of the model takes the following form. In the North, which is the high capital-intensity region, labour costs more than capital because it is available in limited quantities. Consequently, the North employs less labour and more capital to produce steel than does the South, where capital costs more than labour (in Figure 6.6 $(K / L)_A^N > (K / L)_A^S$).

The same applies to the production of corn: the North will tend to produce one unit of corn at a higher capital/labour ratio than that at which the unit is produced in the South. Note that in both regions, for any relative factor price, given that corn is a labour-intensive good, a unit of corn is produced at a capital/labour ratio lower than that of a unit of steel.

The argument thus far produces the following important result: in the South, where labour costs less than capital, it is obvious that the price of a unit of corn in relation to a unit of steel will be lower than in the North. Vice versa, in the North, where capital is more abundant, and therefore less costly, the price of the labour-intensive good, corn, will be higher in relation to steel than it is in the South.

These differences in relative prices generate comparative advantages for the two regions. It is economically convenient for the South to specialize in the production of corn and to trade the surplus for steel, which is less costly if purchased on the external

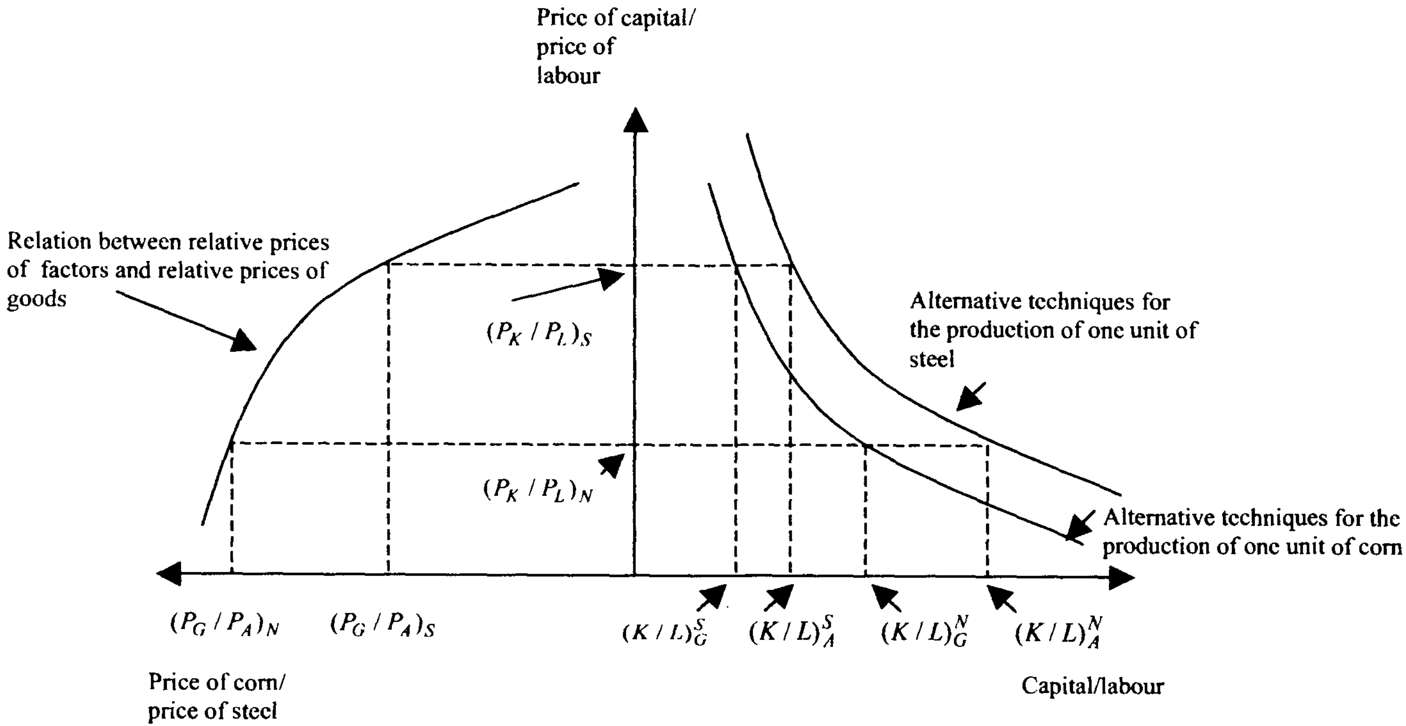


Figure 6.6 Relative prices of factors and goods for different capital/labour ratios

Source: Armstrong and Taylor (2000: 125) market. Conversely, it is economically convenient for the North to specialize in the production of steel, and to obtain corn on the international market at prices lower than domestic ones.

The adjustment process does not stop here, however. Specialization, even if only partial, in production of one of the two goods requires the regions to reallocate capital and labour between the two types of production, and this alters the relative prices of the factors. The North, which must shift resources from the production of corn to the production of steel – the high capital-intensive good – will now experience relatively greater demand for capital than for labour, with a consequent reduction in the availability of capital and an increase in its relative price. In the South, as a consequence of its specialization in corn, demand for labour will be relatively greater than demand for capital, producing a relatively greater increase in wages with respect to the cost of capital. The result, as illustrated by Figure 6.7, is the equality of the relative prices of the goods on the international and domestic trade markets ('law of one price').

Empirical verification of the Heckscher–Ohlin model has often produced results at odds with the theoretical conclusions. The best known of these contradictory findings is the Leontief paradox.[26] When testing the model in the case of the United States during the 1950s, Leontief found that the exporting sectors of the USA – a capital-abundant country – were in fact high labour-intensity sectors. Using a different methodology, Moroney and Walker obtained the same result as Leontief: the labour-abundant southern regions of the United States were in fact exporters of capital-intensive goods.[27] Furthermore, the industrialization of the northern Italian regions in the period 1960–1990 led to their predominant specialization in light industry – textiles, clothing, electronics – whilst the labour-abundant Italian South specialized in heavy industries like steel and petrochemicals.[28]

The first explanation for the empirical paradoxes of the Heckscher–Ohlin model is that production factors differ among regions not only quantitatively but also qualitatively. Leontief himself pointed out that it is impossible to treat labour as a homogeneous factor when testing the model. Different occupations and differing endowments of skilled labour may largely explain the specializations of regions – as demonstrated by the 'new factor-proportion theory' of international trade.[29] A second interpretation of the empirical paradoxes centres on the fact that the theoretical model does not allow for technical progress: product and process innovations may, in fact, generate substantial advantages even in traditional sectors, making them competitive in advanced regions endowed with modern and advanced capital – as pointed out by the proponents of the 'neo-technological' approach to international trade.[30] Finally, in regions where high public capital investments are made – like Southern Italy in the period 1960–1980 – or where large public incentives are offered in order to attract large firms – like the North of England and Ireland – industrial specialization is the result of these intervention policies rather than of market forces.

The results obtained by the model are interesting and they constitute its acknowledged merits. The model reminds us that interregional trade functions as a perfect substitute for factor mobility, because it equalizes the prices of the factors even in the absence of the geographical mobility of resources. Moreover, as said, the model is able to explain productive specialization (exogenous in the export-base model and in the classical Ricardian model), and to show increases in well-being.

It should be borne in mind, however, that under the model's intrinsic logic – with its assumptions of a given factor endowment, constant returns to scale and constant factor

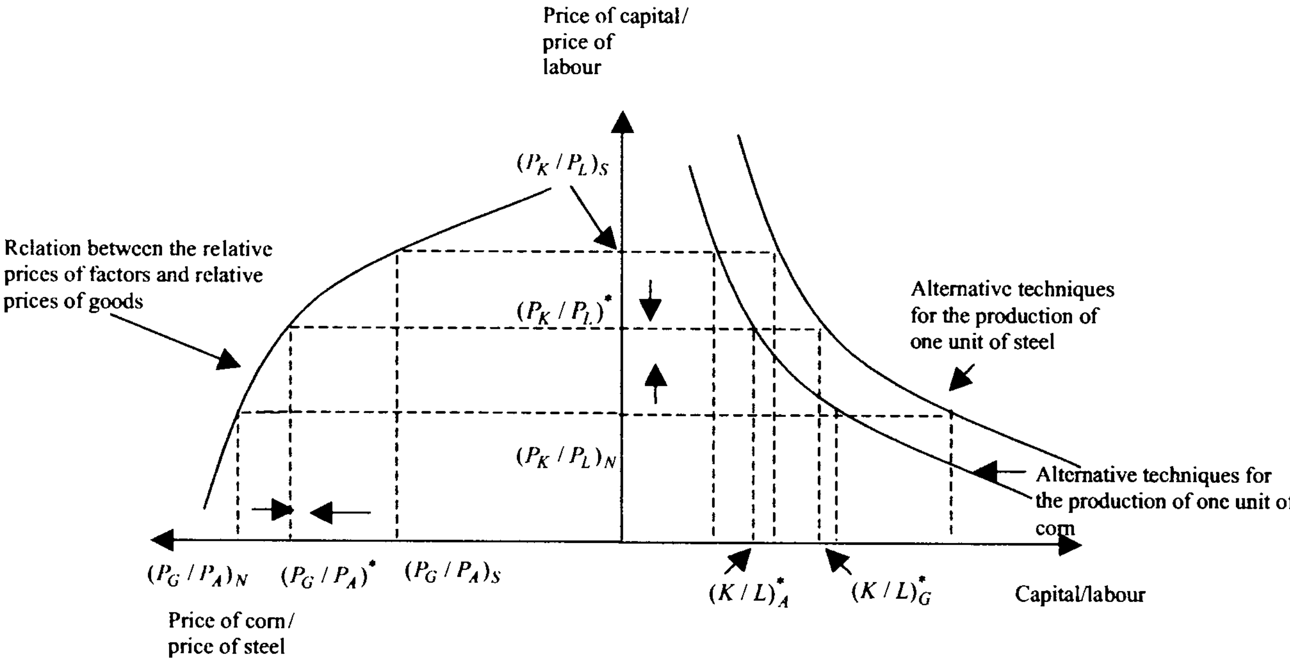


Figure 6.7 Equality in the relative prices of factors and goods as a result of interregional trade

Source: Armstrong and Taylor (2000, p. 130) intensity per unit of output – it is not possible to associate greater specialization with greater output. We may therefore conclude that the model, as it is formulated, is unable to define a process of regional growth. If anything, it suggests – implicitly and without proof – that there is a tendency towards regional development when this is understood in the sense of greater individual well-being (achieved in the model through 'gains from trade') and the obtaining and maintaining of a role in the division of labour.

6.4 Absolute vs. comparative advantage in regional growth

As we have seen, the model of interregional trade yields the important finding that regions, whatever their level of efficiency, always obtain a role in the international division of labour by specializing in production of the good which gives them a comparative advantage. There is therefore an automatic mechanism that guarantees that a region will have some sort of specialization regardless of its productive efficiency.

Given the importance of this assertion, one may legitimately enquire whether it applies to nations as it does to regions; or whether, instead, if a region is inefficient in the production of all goods, it may end up by producing nothing, leaving it to the region most efficient in absolute terms to produce all goods. In theoretical terms, this is to enquire whether regions compete on the basis of an absolute advantage, and not a comparative one.

It has been recently argued that regions differ from countries in that they compete on the basis of an absolute advantage.[31] To understand this assertion it is necessary to look at the adjustment processes that restore equilibrium in international trade, and at the operation of the principle of comparative advantages in the presence of exogenous shocks. The starting-point is the observation that, although the Ricardo model yields the result that trade is always in the interest of a country, it actually occurs only if there are absolute advantages in commerce[32] between economic actors which compare the (absolute) prices of a good in the two countries, given a certain exchange rate.[33] In the higher-productivity country, wages are necessarily higher than in the less efficient country, where factor remunerations are defined on the basis of lower levels of productivity and overall output. It is logically likely that productivity gaps will be on average perfectly offset by wage gaps (calculated in the same currency) – which demonstrates that comparative advantages are also absolute advantages.

However, let us assume the hypothesis that the monetary wage is kept artificially high in the less productive country so that it equals the wage in the efficient country, which is, say, 1 euro an hour (we are speaking here in terms of national, not regional economies). In this case, assuming that the exchange rate is 1:1, if labour productivity is equal to that shown in Table 6.1a, the monetary conditions of trade are those in Table 6.2; the North has absolute advantages in the production of both goods because of its labour force's greater productivity. The North produces everything; it achieves

Table 6.2 Monetary conditions of trade

Goods regions	*A*	*B*
North	1 euro	1 euro
South	2 euro	4 euro

more than full employment; and it has a positive trade balance because it exports the surplus goods produced to the South – which instead produces nothing; suffers high unemployment; and has a constant deficit in its balance of payments. At national level, three automatic re-equilibrating mechanisms reinstate a regime of comparative advantages:

a) in a fixed exchange rate regime, the North's persistently positive trade balance accumulates gold (or, in more modern terms, money) in the country, activating Hume's well-known 'price-specie-flow' mechanism:[34] an acceleration in the circulation of money generates inflation; this in its turn gives rise to an increase in prices and wages which erodes the country's absolute advantage in the production of all goods;
b) in a flexible exchange rate regime, the North's persistently positive trade balance induces revaluation of the exchange rate, and therefore induces an increase in the prices of exports and a decrease in the prices of imports. The result is again a loss of competitiveness by goods produced by the North which favours the South, which regains competitiveness in the goods for which it enjoys a comparative advantage;
c) the imbalance between demand and supply in the North's labour market pushes up wages, once again generating a loss of competitiveness by the North which favours the South.

These mechanisms restore a comparative advantage regime that enables the less competitive country to produce and thereby regain a role in the international division of labour.

However, although this holds for a country, it may not do so for a region. First, monetary wages may not reflect marginal productivities at regional level. On the one hand, wages are fixed at regional level on the basis of national-level agreements reflecting the country's average productivity; on the other, if low productivity is due to conditions external to firms (difficult accessibility, low quality of services), workers will not accept lower pay levels in a context where factor mobility is free. For this reason, the idea that wage gaps offset productivity gaps, protecting the comparative advantage, is unrealistic.

We may therefore argue that the automatic re-equilibrating mechanisms that operate between countries are less efficient if the territories analysed are regions, because:

a) at regional level, a positive trade balance may be maintained by outgoing interregional capital movements, and it does not necessarily generate increases in local prices and wages.[35] More evidently, this mechanism operates in backward regions with negative trade balances; a situation that may persist if it is financed by public transfers in the form of pensions and unemployment benefits, or by inflows of external capital. If this is the case, the trade balance is not a macroeconomic constraint; and as a consequence, the re-equilibrating mechanism is not activated;
b) there are no re-equilibrating mechanisms based on exchange rate flexibility at regional level because a single money regime operates;
c) at regional level, there are only re-equilibrating mechanisms generated by disequilibria in the labour market. But neither in this case does the labour market

> mechanism operate as efficiently as it does at national level. When there are labour-market disequilibria, in an intra-national context of high factor mobility it is more likely that workers will migrate to high-wage regions, and less likely that wages will fall in weak regions and rise in strong ones. The real-wage readjusting mechanisms do not have sufficient time to exert their positive effects.

The upshot, therefore, is that when some regions are more efficient in absolute terms than others, they tend to produce all goods, while the others are at risk of mass unemployment and 'desertification'. These conditions will persist because the pure macroeconomic re-equilibrating mechanisms that seemingly ensure the relative competitiveness of territories at national level do not exist, or do not work, at the regional level. Hence derives the need to 'safeguard' the real competitive capacities of regions, because it is on these that long-term development opportunities depend.

6.5 The theory of customs unions

There is an important area of analysis in the theory of international trade that concerns itself with the effects of the creation of customs unions, like the European Common Market of 1958 or the Single European Market of 1993.[36] Some studies have examined regional aspects, given curiosity in the effects of the creation of the Single Market on regional growth and disparities. In the last ten years, the decision to institute a 'Europe of 28' has sparked animated debate on how entry into the European Union of the former members of the Communist bloc has affected both regional disparities within each country of the East and the growth paths of the regions of the 15 original member-states of the Union.[37]

Creation of a customs union entails the abolition of economic and institutional barriers to international trade through the elimination of customs tariffs/duties, harmonization of technical standards in production and of rules on the quality certification of products, on their safety and transport, abolition of disparities in the indirect taxation of consumption goods and common regulation of the capital market. The main consequences are an expansion of outlet markets and the creation of a large integrated market in which geographic-institutional distance among local markets affords them increasingly less 'protection' – as testified by the large-scale globalization processes of today's economy.

According to the theory of customs unions, the expansion of markets produces a number of important macroeconomic effects:

a) a marked increase in competition on markets;
b) greater economies of scale in goods production because of the larger size of outlet markets;
c) the creation of trade in final and intermediate goods because local markets are no longer the only ones available. Each region purchases from the most efficient supplier in the European market;
d) increased investments prompted by forecasts of greater competition: an effect which comes about even before the creation of single market, in that it is the result of market expectations to which firms adjust;
e) demand for a greater variety of goods because of increased per capita income (income effect);

f) a shift of demand to goods produced with more efficient techniques guaranteeing lower prices for the same quality;
g) technology and knowledge transfers from strong to weak regions.

With the possible exception of the last of these effects, it is likely that all of them will favour the richer and more advanced regions – given that these possess the financial, productive and knowledge resources necessary to withstand increased competition on markets, to respond to a diversification of demand, to exploit increasing economies of scale, and to make decisive, targeted and timely investments as creation of the single market proceeds. An interesting example is provided by the widening of regional disparities in the former Communist countries consequent upon their entry into the European Union – between central and peripheral regions, and between ones lying more to the east or to the west (especially those bordering on the European Union). In these countries, in fact, the empirical evidence suggests that, although there is an evident diffusion of economic activity previously concentrated in metropolitan regions, the process works selectively in favour of the regions situated closest to the European Union.[38]

This is also apparent at the more micro-territorial level between the strong and weak areas of a particular region. For instance, the creation of the single European market in 1993 had positive effects on the large Italian cities of Milan, Rome and Naples, which in the years immediately prior to 1993 had undergone a marked process of economic recovery and development after years of recession.[39]

The last of the effects listed above – technology transfer from the 'centre' to the 'periphery' – can be interpreted as favourable to backward areas. Nevertheless, it requires a local capacity to exploit technologies in pursuit of specific local competitiveness targets which is often lacking in weak areas.

We may therefore conclude that customs unions theory warns that the creation of a large single market may have repercussions on regional growth; it offers major opportunities for development to local systems, but these opportunities may not be equal for all of them and they may work instead in favour of advanced and dynamic regions, thus widening regional disparities.

Customs unions theory states that, in broad integrated economic areas, the production factors, technical knowledge and consumption patterns circulate freely, generating substantial homogeneity in productive capacity and demand. But perfect homogeneity in incomes and factor endowments heightens the tendency towards productive despecialization. The distinction between international and domestic trade disappears, and all output by the integrated area is traded according to criteria pertaining to the internal market and interregional trade. Thus explained is the growth of intra-industry trade; that is, the exchange of similar goods 'in two directions' – or 'horizontal trade' – which for some time has developed greatly in the advanced countries. This process, too, has interesting effects on the development of local systems.

Explanations of the apparently paradoxical phenomenon of trade in the same goods among countries and regions have been based on two groups of components:

- *demand components*: following the pioneering studies of Linder, the well-known Lancaster model has explained horizontal trade as stemming from the existence of 'horizontally' differentiated products. Goods have specifications that depend on different mixes of the same characteristics – the overall quantity of these characteristics remaining equal – that consumers can obtain by changing brand, supplier or producer. Free choice by consumers among these differentiated products, which they purchase according to specific individual utility functions, is the basis of horizontal trade;[40]

- *supply components*: in this case, horizontal trade in identical goods is explained by analysing how a good is produced and, especially, distributed in modern economies. It is through its sales and distribution network that a firm establishes and defines its advantage on a particular market. On this view, product differentiation stems only from the ways in which firms sell their products and control their market shares.

The implications for regional development are obvious. If we accept the idea that supply components – different modes of production and distribution – explain horizontal trade, regions must compete on the basis of more efficient and less costly production achieved by exploiting economies of scale, technical progress, and process and product innovations; all of which are elements very distant from the macroeconomic advantage envisaged by the theory of comparative advantage, and which obviously bear out the idea that competitiveness is based on an absolute advantage for regions. This is the thesis argued by the theories treated in the next part of the book.

6.6 Conclusions

This chapter has examined neoclassical theories of regional growth, where 'growth' is understood as an increase in individual well-being (and its interregional convergence). For the first group of theories – the neoclassical macroeconomic models – greater individual well-being is achieved through increases in factor productivity, and consequently in wage levels and per capita income (neoclassical macroeconomic models). For the second group of theories – which comprises the classical and neoclassical models of trade – higher levels of individual well-being are achieved through processes of regional specialization. Greater specialization pushes towards interregional trade and consequently yields advantages from the purchase of goods offered on the external market at prices lower than if the goods were produced internally.

The chapter has examined a number of commonplace assumptions about these theories. As regards the neoclassical macroeconomic models, it has shown that their interpretation as 'theories of convergence' is too restrictive; the modern versions of these theories demonstrate that inclusion of increasing returns in the neoclassical production function produces a set of behaviours and tendencies at odds with the mechanistic and univocal result of re-equilibrium between regional income levels obtained by the initial model. Moreover, the original model, which was modified by its authors to envisage the existence of two sectors, is able to explain divergences in income levels if initial equilibrium is hypothesized.

As regards the second group of theories – classical and neoclassical – on interregional trade, the chapter has shown that it is not possible to make immediate use of Ricardo's concept of comparative advantage when explaining the competitiveness of regions. Indeed, the chapter has demonstrated that the economic mechanisms on which the concept of relative advantage rests at national level do not apply at the regional one. This means that regions compete on the basis of not a comparative advantage, but an absolute one; an advantage that must constantly be re-created over the long period. The theories examined in the next chapters are specifically concerned with identifying the factors responsible for this advantage.

Review questions

1 What conception of growth is at the basis of the neoclassical regional growth models?
2 What are the results achieved in the one-sector model and how do they differ from the results achieved by the two-sectors model?
3 What results of the two-sectors model are similar to results obtained in other models? Which models?
4 What are the weaknesses and strengths of neoclassical regional growth models?
5 Is the statement that regional growth models always interpret convergence processes valid?
6 What is meant by the statement that regions compete on the basis of a 'comparative advantage'?
7 What is the result achieved by Ricardo's model?
8 What is the main theory in the Heckscher–Ohlin model and what is the conception of growth that the model is able to interpret?
9 What are the reasons behind the idea that regions compete on the basis of an absolute advantage?
10 What does the theory of customs unions theorize?

Selected reading on empirical findings

About regional competitiveness

Department of Trade and Industry (DTI) (2005), *Regional Competitiveness and State of the Regions*, online edition, http://data.gov.uk/dataset/regional_competitiveness_and_state_of_the_regions.

Edmonds T. (2000), 'Regional Competitiveness and the Role of the Knowledge Economy', Research Paper 00/73, House of Commons.

Eskelinen H., Maskell P., Vatne E., Malmberg A. and Hannibalsson I. (1998), *Competitiveness, Localised Learning and Regional Development*, Routledge, London.

Huovari J., Kangasharju A. and Alanen A. (2003), 'Regional Competitiveness in Finland', paper presented at the ERSA conference, Finland. Available online, http://www-sre.wu-wien.ac.at/ersa/ersaconfs/ersa00/pdf-ersa/pdf/234.pdf.

Leontief W. (1953), 'Domestic Production and Foreigner Trade: The American Capital Position Re-examined', *Proceedings of the American Philosophical Society*, vol. 97, no. 4, pp. 332–349.

About advantages and disadvantages of customs unions

Emerson M. (1992), *One Market One Money: An Evaluation of the Potential Benefits and Costs of Forming an Economic and Monetary Union*, Oxford University Press, Oxford.

Resmini L. (2007), 'Regional Patterns of Industry Location in Transition Countries: Does the Economic Integration with the EU Matter?', *Regional Studies*, vol. 4, no. 6, pp. 747–764.

Further reading

Borts G.H. and Stein J.L. (1964), *Economic Growth in a Free Market*, New York, Columbia University Press.

Camagni R. (2002), 'On the Concept of Territorial Competitiveness: Sound or Misleading?', *Urban Studies*, vol. 39, no. 13, pp. 2395–2411.

Cecchini P. (1988), *The European Challenge: 1992*, Wildwood House, Aldershot.

Leontief W. (1953), 'Domestic Production and Foreigner Trade: The American Capital Position Re-examined', *Proceedings of the American Philosophical Society*, vol. 97, no. 4, pp. 332–349.

Lutz V. (1962), *Italy – A Study in Economic Development*, Oxford University Press, London.

McCombie J.S.L. (1988), 'A Synoptic View of Regional Growth and Unemployment: I – The Neoclassical Theory', *Regional Studies*, vol. 25, no. 4, pp. 267–281.

Ohlin B. (1933), *Interregional and International Trade*, Harvard University Press, Cambridge, Mass.

Ricardo D. (1971), *Principles of Political Taxonomy and Taxation*, Penguin Books, Harmondsworth.

Notes

1 Two important theoretical notions should be borne in mind if this reasoning is to be properly understood. First, in a neoclassical world, factor productivity is governed by the law of decreasing marginal returns: a larger quantity of factors entails lower factor productivity. Second, according to the neoclassical theory, the production factors can only be remunerated at their marginal productivity: the firm pays the additional factor exactly for the value of the good that the additional factor is able to produce, thus maximizing its profit. On this view, a region with a large endowment of a particular factor can only expect low productivity by, and therefore low remuneration of, that factor.

2 The reference is to Borts, 1960, Borts and Stein, 1964 and 1968 (original edition 1962).

3 See Krauss and Johnson, 1974; Camagni, 1999a.

4 See Cobb and Douglas, 1928.

5 Rewriting equation (6.1) in logarithms gives:

$$\ln Y = \ln A + \alpha \ln K + (1-\alpha)\ln L \tag{6.1n}$$

The derivative over time of a variable is calculated as:

$$\frac{d\ln Y}{dt} = \frac{1}{Y}\frac{\delta Y(t)}{\delta t} \tag{6.2n}$$

Setting $\dfrac{\delta Y(t)}{\delta t} = \dot{Y}$, equation (6.2n) can be rewritten as follows:

$$\frac{d\ln Y}{dt} = \frac{\dot{Y}}{Y} = y \tag{6.3n}$$

Deriving all the variables of equation (6.1n) over time, we obtain equation (6.2).

6 See Smith, 1975, Miyao, 1987a. These curves are obtained by setting equations (6.4) and (6.5) equal to zero. As regards equation (6.1), if we substitute equation (6.1) in (6.4), we obtain:

$$k = sL^{\beta}\left(\frac{K}{L}\right)^{\alpha-1} + \mu(i_r - i_w) = 0 \tag{6.4n}$$

i_r is by definition equal to the marginal productivity of capital, i.e. it is equal to:

$$i_r = \alpha L^{\beta} K^{\alpha-1} L^{1-\alpha} \tag{6.5n}$$

which when substituted in equation (6.4n) yields:

$$sL^{\beta}\left(\frac{K}{L}\right)^{\alpha-1} + \mu(\alpha L^{\beta} K^{\alpha-1} L^{1-\alpha} - i_w) = 0 \tag{6.6n}$$

which can be rewritten as:

$$sL^{\beta}K^{\alpha-1}L^{1-\alpha} + \mu\alpha L^{1-\alpha+\beta}K^{\alpha-1} - \mu i_w = 0 \tag{6.7n}$$

In its turn, equation (6.7n) can be written as:

$$\left(\frac{s+\mu L}{\mu i_w}\right)^{\frac{1}{1-\alpha}} = KL^{\frac{\alpha-1+\beta}{1-\alpha}} \tag{6.8n}$$

Solution of equation (6.8n) for K produces:

$$K = \left(\frac{s+\mu L}{\mu i_w}\right)^{\frac{1}{1-\alpha}} L^{\frac{1-\alpha+\beta}{1-\alpha}} \tag{6.9n}$$

Likewise, as regards equation (6.5) we have:

$$l = n + \lambda(w_r - w_w) = 0 \tag{6.10n}$$

Setting w_r equal to marginal productivity, we obtain:

$$n + \lambda((1-\alpha)L^{\beta}K^{\alpha}L^{-\alpha} - w_w) = 0 \tag{6.11n}$$

which with simple steps leads to:

$$K = \left(\frac{w_w\lambda - n}{\lambda(1-\alpha)}\right)^{\frac{1}{\alpha}} L^{\frac{\alpha-\beta}{\alpha}} \tag{6.12n}$$

7 The trend of equation (6.3) depends on the time trend of the growth rate of per capita capital $(k - l)$, which for convenience may be called $(k - l)$. From equations (6.4) and (6.5) we obtain:

$$H(k-l) = sA(k-l)^{\alpha-1} + \mu(\alpha A(k-l)^{\alpha-1} - i_w) - n - \lambda((1-\alpha)A(k-l)^{\alpha} - w_w) \tag{6.13n}$$

when $\lim_{(k-l)\to\infty} H(k-l) = \mu i_w - n + \lambda w_w$ and $\lim_{(k-l)\to 0} H(k-l) = \infty$ and $H'(k-l) < 0$. These relations explain the trend of the per-capita growth rate curve in Figure 6.3 and imply, for the differential equation (6.9n), the existence, uniqueness and stability of equilibrium growth. See Miyao, 1987b.

8 One implication of the model set out in Section 6.2.1 is that low-wage regions should exhibit higher rates of growth of both capital and the capital/output ratio, and consequently of per capita income. Moreover, low-wage regions should also record a higher rate of wages growth because of the increase in the capital/output ratio. However, statistical tests conducted by the authors of the theoretical model showed that this was not the case. On examining the American regions, Borts and Stein found that in the periods 1919–1929 and 1948–1953 capital flowed to high-wage regions – regions, moreover, with higher rates of wages growth. Only in one period, between 1929 and 1949, did the empirical reality seem to support the theoretical hypotheses; a result that, as Borts and Stein themselves acknowledged, was too weak to give empirical validity to their model. See Borts, 1960; Borts and Stein, 1964 and 1968 (original edition, 1962). Numerous neoclassical authors, convinced that the original model was fundamentally sound, blamed its contradiction by the empirical evidence on erroneous methodology: Smith argued that the problem was due to the excessive sectoral disaggregation of the data used by Borts and Stein, while Coelho et al. pointed to the erroneous use of nominal wages as a proxy for individual well-being, given substantial differences in the costs of living among regions. See Coelho and Ghali, 1971; Smith, 1974; Coelho and Shepherd, 1979.

1/12 of a unit of good B (1/3 – 1/4 = 1/12) from the trade. Given that each unit of B is produced in four hours, 1/12 of B is produced in 1/3 of an hour, which represents the advantage in terms of labour-hours saved obtained from the trade.

22 'For Ricardo, the main benefit deriving from imports is that wage goods can be obtained at lower prices.' See Onida, 1984, p. 65.

23 Armstrong and Taylor's observation (2000, p. 123) is emblematic: 'that trade is based on comparative advantage and not absolute advantage is universally accepted.'

24 See Camagni, 2002.

25 See Heckscher, 1919.

26 See Leontief, 1953 and 1956.

27 See Moroney and Walker, 1966.

28 See Camagni, 1999a.

29 See Keesing, 1966.

30 See Posner, 1961, Vernon, 1966; Nelson and Norman, 1977.

31 In a recent work, Camagni responds to the provocative thesis of the noted international economist Paul Krugman that territories, unlike firms, can only compete on the basis of comparative advantage: see Camagni, 2002, on Krugman, 1996b and 1998. For broader treatment of the arguments presented in the main text see Camagni, 2002.

32 Ricardo's model was developed from a normative perspective; it concludes that it is economically convenient for regions (or countries) to trade, but nothing in the model is able to determine whether or not trade actually takes place.

33 The Ricardo model, which demonstrates the existence of comparative advantages in the economies of two nations, is based on a barter economy in which 'units of goods' are exchanged. But the model lacks a theory of wages, prices and money necessary to understand real trade decisions. 'The monetary conditions of trade are an advantage in absolute costs.' See Onida, 1984, p. 81.

34 'Specie' denotes commodity-money or international means of payment, whose net flow reflects the deficits-surpluses in the trade balances of countries. See Onida, 1984, p. 85.

35 See Section 5.2.

36 On the effects of creation of the European Common Market see Scitovski, 1958; Thirlwall, 1974; Balassa, 1975. On those of the creation of the Single Market see Cecchini, 1988; Camagni, 1992b; Quévit, 1992.

37 In 2004, eight countries in Eastern Europe and two from the Mediterranean area (Malta and Cyprus) joined the European Union. In 2007 it was the time of Bulgaria and Romania; with the entry into the European Union of Croatia in 2013, the European Union changed from a Europe of 15 to a Europe of 28.

38 See Bachtler and Downes, 1999; Petrakos, 2000; Traistaru et al., 2003; Resmini, 2007.

39 See Camagni and Pompili, 1990; Capello, 2002a.

40 See Linder, 1961, Lancaster, 1980. For comments on the models see Barker, 1977.

Part III

Theories of local development

Diversified-relational space

9 In Borts and Stein's words: 'The forces we have outlined produce a *permanent divergence of regional growth rates*. The only way in which growth rates might converge is through the role played by other autonomous forces operating within the framework of such a set of economic relations.' Borts and Stein, 1968, p. 184 (emphasis added); original edition 1962.

10 The value of marginal product is the price of the good produced multiplied by the marginal productivity of the factors used to produce that good. It represents the revenue obtained by the firm from the use of the extra unit of the production factor. The value of marginal product has to be equal to the cost of the production factor in order for a firm to maximize its profit. In the case of labour, therefore:

$$VPMa_l = P_x * PMa_l = w \tag{6.14n}$$

where $VPMa_l$ is the value of marginal product (i.e. the marginal revenue), P_x the price of good x, PMa_l the marginal productivity of the labour factor, and w the wage. Likewise, for capital:

$$VPMa_k = P_x * PMa_k = i \tag{6.15n}$$

where i represents the remuneration of capital. When marginal revenue equals marginal costs, i.e. when equations (6.14n) and (6.15n) hold, the firm achieves profit maximization.

11 With this assumption, the model resembles the export-base model: external demand – which is not explained but assumed exogenously – is the source of growth.

12 The migration of labour from the agricultural to the manufacturing sector induced by industrialization is a good example of this source of local growth. There is a version of the model that assumes the existence of a single region with two sectors: one with low, the other with high factor productivity. Under standard neoclassical assumptions, the migration of workers to the higher-wage sector is followed by higher income levels produced by the reallocation of resources which guarantee conditions of optimal intra-regional allocation of resources. See Borts and Stein, 1964, Chapter 7; McCombie, 1988.

13 Recall the item 'Compensation of employees and property income owned by non-residents in the region' in the current account of the balance of payments (see Table 5.1). These incomes do not enter the formation of disposable income; in Table 5.2, in fact, the remuneration of employees and property incomes entering the formation of gross disposable income are net incomes, i.e. the only ones earned by residents outside the region; see Chapter 5.

14 Borts writes: 'Migration does not appear sufficient to produce convergence. It clearly produces less divergence than would occur were migration to halt.' See Borts, 1960, p. 346.

15 A case in point is Northeastern and Central Italy in the 1970s, where growth was driven by traditional labour-intensive light manufacturing industries characterized by a high level of productivity and a low labour cost, compared with the Northwestern (Lombardy, Piedmont and Liguria) and Southern regions of the country. See Camagni and Capello, 1990, and Chapter 8 in this book.

16 See Holland, 1977; Camagni, 1999a.

17 See Okun and Richardson, 1961.

18 See Lutz, 1962. For discussion of the Italian dual economy see Spaventa, 1959, Ackley and Spaventa, 1962, Graziani, 1983. For a critical examination of Lutz's theory see Holland, 1971.

19 Interesting in this regard is Ohlin's observation that the theory of international trade is a particular case of the theory of interregional trade with the exclusion of production factor mobility. See Ohlin, 1933.

20 See Ricardo, 1971, original edition, 1817. Historians of economic thought are still unsure as to who originally developed the notion of comparative advantage, which is present in the works of both Torrens (1815) and Ricardo (1817). However, there is no disputing that both authors made a crucial contribution to the development of the theory of comparative advantages.

21 The advantage in terms of labour-hours is calculated as follows: by obtaining from the trade one-third of good B, when it could itself produce one-quarter, the region obtains an extra

7 Territorial competitiveness and exogenous development

7.1 Diversified space: the components of territorial competitiveness

This part of the book examines an approach to the study of regional development that runs counter to those treated thus far. It differs from them in its conception of space. Whilst the theories discussed in previous chapters use the term 'space' to denote territorial areas assumed to be internally homogeneous and uniform, the theories now considered conceive 'space' as *diversified*. This change of perspective allows economic activities and production factors, demand and sectoral structure, to be treated as spatially heterogeneous within a region, so that territorial relations are cast in new light.

This new conception of space enables identification of highly distinct polarities in a territory. Activities, resources, economic and market relations structure themselves around these polarities to generate a cumulative process of territorial agglomeration and a virtuous circle of development. This conception of space restores one of the inspiring principles of location theories – that of agglomeration economies as the source of local development – to theories of regional development. It is evident that thus severed is any connection with geographical space, abstract or administrative. A more complex conception of space takes over, one based on the economic and social relations that arise in a territorial area. Whence derives the expression *diversified-relational space*.

When space is conceived as 'diversified-relational', theories radically change in their nature. A macroeconomic and macro-territorial approach gives way to a micro-territorial and micro-behavioural one. Abandoned is the notion of a region as a portion of a national system acting and reacting economically as a single, internally homogeneous system. Its place is taken by individual economic actors (large or small, public or private, multinational or local) whose behaviour is studied in terms of location choices, productive and innovative capacity, competitiveness and relations with the local system and the rest of the world.

The qualitative nature of theories – only in recent years superseded thanks to the more advanced and sophisticated modelling techniques examined in the next part of the book[1] – led in the mid-1970s to the distinction in the literature between '"pure and exact" regional theory without agglomeration economies, on the one hand, and "applied regional theory" which is inexact but takes agglomeration factors into account, on the other hand' drawn by Edwin von Böventer.[2]

The theories analysed in this part of the book resemble those discussed in the previous chapter in that they conceive development as a process generated and sustained by supply-side elements. But we shall see that they embrace a conception of development that has little to do with that of the theories previously examined. They abandon the short-run view of development as a simple increase in income and employment, and also that of individual well-being, and assume a longer-term perspective. They identify all the tangible and intangible elements in a local area that determine its *long-term competitiveness* and enable it to maintain that competitiveness over time. To reprise the distinction between development and growth, this part of the book deals with theories of local development, whilst the endeavour to identify the (short- or long-period) growth path pursued thus far is abandoned.

The theories analysed here therefore seek to identify the factors that render the costs and prices of production processes lower than they are elsewhere. These factors are (i) elements *exogenous* to the local context, which originate externally to the area and are transferred into it either fortuitously or deliberately, and (ii) *endogenous* elements, which arise and develop within the area and enable it to initiate a process of self-propelling development.

Exogenous elements comprise the following: the fortuitous local presence of a dominant firm or a multinational company; the diffusion in the area of an innovation produced elsewhere; or the installation of new infrastructures decided by external authorities. Although these elements have nothing to do with local features and productive capacities, once they are present in an area they may catalyse new economic activities and development. Endogenous elements are entrepreneurial ability and local resources for production (labour and capital); and in particular the decision-making capacity of local economic and social actors able to control the development process, support it during phases of transformation and innovation, and enrich it with external knowledge and information. All these are factors strengthened and enhanced by a concentrated territorial organization that generates local processes of knowledge-acquisition and learning; networks of economic and social relations that support more efficient and less costly transactions;[3] and advantages of economic and physical proximity among economic actors.

The assumption of diversified space entails definitive abandonment of the notion that regional development consists solely in the allocation of resources among regions. Instead, regional development must be conceived as stemming from local productive capacity, competitiveness and innovativeness. The neoclassical model of interregional growth (Borts and Stein's one-sector model) presumed that the national growth rate is exogenously determined, and that the problem for regional development theory is explaining how the national growth rate is distributed among regions. According to this logic of *competitive development*, the growth of one region can only be to the detriment of the growth of another region, in a zero sum game.[4] The theories examined here adopt a notion of *generative development* whereby the national growth rate is the sum of the growth rates achieved by individual regions. National economic development may well increase because of growth achieved by a particular territorial area, and this growth may also come about – in the presence of increasing returns (as for the theories discussed in the next chapter) – with the same resources.

This chapter examines theories that identify the elements exogenous to the system which determine long-term competitiveness: the presence of a dominant firm (Section 7.2), of a multinational company (Section 7.3), the diffusion of an innovation originating in another area (Section 7.4), the construction of transport and social infrastructures (Section 7.5), and finally the adoption of advanced communication technologies (Section 7.6). Left for treatment in the next two chapters are theories that, with intriguing and impressive insight, seek to identify the endogenous elements that determine local competitiveness. These theories hypothesize the existence of increasing returns generated by territorial agglomerations; they in fact conceive the development path as dependent on the efficiency of a territorially concentrated organization of production, not on extra economic resources or on their more efficient spatial allocation.

7.2 The growth-pole theory

7.2.1 The economic approach: Perroux's contribution

The first theory that abandons the notion of uniform-abstract space to conceive of a diversified-relational space is the 'growth-pole theory' first formulated in 1955 by the French economist François Perroux.

The basis of Perroux's theory is encapsulated in his celebrated statement – which despite its simplicity has been important in its consequences: 'Development does not appear everywhere at the same time: it becomes manifest at points or poles of development, with variable intensity; it spreads through different channels, with various final effects on the whole of the economy.'[5]

Thus, in the same period when the principal models of interregional growth were being produced, Perroux formulated a theory of local development that envisaged selective growth at certain points in space where a 'propulsive unit' triggered the development process. Perroux identified this element as the fortuitous presence in the area of a dominant firm, which he called '*l'industrie motrice*' owing to its capacity to influence through its investment decisions the levels of investment undertaken by the firms connected with it.[6] Because of its dynamics and technological dynamism, the dominant firm responds to the needs of an external market (and here the influence of the export-base model is evident). And thanks to its dominant position in the sector and in the economy, it generates a series of positive effects on the sector to which it belongs, and on the economy as a whole.

A technological innovation by the dominant firm that reduces the price of a good, or enhances its quality, increases external demand for that good. This stimulates greater production of the good which in its turn generates a growth-pole through a series of positive effects:

- *a Keynesian multiplying effect on income* that horizontally pervades the entire economy. Increased production by the dominant firm augments employment in both the firm itself and in those connected with it, with a consequent increase in incomes and consumption;
- *a multiplying effect à la Lentief*, connected with intersectoral input–output effects, which vertically pervades the dominant firm's *filière*. Firms and sectors upstream from the dominant firm see their production and outlet markets expand. Relations among firms act as channels transmitting the development without which the growth-pole could not exist (the theory thus closely reflects a conception of diversified-*relational* space);
- *an acceleration effect* on firms' investments. Growth of demand for the dominant firm's goods and those of the firms connected with it stimulates investments

(there is an evident reference here to the Harrod–Domar model). These investments are facilitated by higher profits which generate higher levels of reinvestment of those same profits. Like the input–output effect, this acceleration effect operates vertically along the dominant firm's *filière*. This gives rise to selective development because, especially in its direct effects, development may be confined only to the sector to which the dominant firm belongs, and to the sectors connected with it;[7]

- *a polarization effect* that produces what Perroux calls a 'growth-pole'. Increased demand for intermediate goods and services generated by the dominant firm induces other firms to locate close to it in order to (a) minimize their transportation costs in serving the propulsive firm, (b) exploit the infrastructures and fixed social capital activated by the pole, (c) improve the local managerial or entrepreneurial skills produced by the economic activities generated by the dominant firm, and (d) exploit the greater demand produced by higher employment.

This theory comprises a number of key features for the interpretation of development already put forward by previous theories: the importance for the development process of infrastructures, services and input–output relations among firms and sectors that balanced development models had already emphasized; the positive effects of growth in demand (real and expected) on the level of investments already highlighted by the Harrod–Domar model; and the Keynesian income multiplying mechanisms already present in the export-base model.

The difference resides in the way in which these factors are conceived – no longer in macroeconomic and macro-territorial terms, but rather in microeconomic and micro-behavioural ones.[8] Development is generated by the dynamism of a firm and by its links with other firms, and the cumulative growth process is the result of rational behavioural reactions by the various actors involved in the dominant firm's activities.

Perroux thus for the first time incorporated into a theory of local development the possibility of selective development; that is, development confined to particular sectors or particular areas of a region by cumulative processes that work to the advantage of specific sectors and areas. In Perroux's approach, therefore, growth does not necessarily and automatically spread through all the economy's sectors and through the national and regional territory.

7.2.2 *The territorial approach: Boudeville's contribution*

Although the growth-pole theory aimed to interpret local development, it lacked a clear local dimension. According to Perroux's theory, the channels through which development spreads are input–output relations, but it gives these relations no concrete spatial location. Not surprisingly, therefore, it has been argued in the literature that economic space and geographic space do not coincide in Perroux.[9]

In 1964, Jacques-R. Boudeville endeavoured to emphasize precisely this spatial/territorial component of the growth-pole theory, by imposing clear geographic boundaries on the positive development effects generated by the propulsive industry. By constructing a simple extension of Perroux's theory, Boudeville identified three ways to define the geographic boundaries of polarization effects. For this purpose, he used the following three hypotheses on the geographic location of the actors involved in the development process, or on the geography of positive spill-over effects:

- the propulsive industry and the firms connected with it are geographically clustered;[10]
- the propulsive firm is located in a city. Hence – in keeping with the classical tradition of urban economics – the input–output relations that generate development can be hypothesized as operating within that same urban area;
- the positive effects generated by the dominant firm impact only upon the local area. This amounts to hypothesizing the absence of leakages in the income-multiplying effects evidenced by the export-base theory, and to arguing that a growth-pole comes into being when the positive effects of a dominant firm are confined to the local area.

These three interpretations have an important feature in common: for all three of them the key factor in development is no longer, as in Perroux, sectoral interdependence alone. For local economic development to come about, there must be a *spatial concentration* of production activities that determines the positive final effect exerted by the dominant firm on local development.

This last point has an important implication, which represents a watershed with respect to the logic used by previous interregional growth models to interpret regional development: the spatial concentration of economic activity is a territorial organization of production that generates development more efficiently than does spatial dispersion. The growth-pole theory was therefore the first step towards conceiving space as an active factor in development. It opened the way for the analyses of endogenous development presented in the next chapter. For these analyses, the spatial concentration of activities is the source of increasing returns in the form of agglomeration economies, localization economies, technological externalities and localized learning processes – all of which are elements that enhance the competitiveness of local firms and foster local development.

7.2.3 Critical assessment of the theory

The merits of the growth-pole theory have already been pointed out. It suggested, for the first time, the existence of selective local development that works in favour of some sectors and some specific local areas but does not necessarily benefit the region as a whole. The real world is constellated by strong areas (with greater densities of manufacturing activity, and a greater capacity for economic growth) and weak areas, even within the same region; for the first time, the growth-pole theory is able to explain phenomena of concentrated settlement.

Moreover, the theory has the outstanding merit of recognizing input–output relations among sectors on the one hand, and the spatial concentration of productive activities on the other, as the crucial factors in development. As regards sectoral relations, it is the first theory of regional development to have emphasized the competitiveness of certain sectors and industrial dynamics in explanation of local development.[11] As for agglomeration economies, this is the first time that these become an essential component of theories of local development as well.

Finally, it is a theory that brings together notions developed by other, and apparently entirely distinct, theories. The ideas of the central place theory[12] reappear in Perroux's argument that the pole furnishes higher-level services (services to businesses, infrastructural and health services, educational, recreational and cultural ones) for a broader area, and that the availability of these services attracts new businesses into the area around the pole. Once again, firms take their location decisions in light of the two crucial features emphasized by location theory: transportation costs (in this case of intermediate goods delivered to the dominant firm) on the one hand, and agglomeration economies on the other, generate a polarization effect and give rise to a theory of local development.

The experiences of the European countries where public intervention has been inspired by this theory – intervention either through the creation of state-controlled enterprises (as in Italy) or through policies to attract foreign companies (as in the United Kingdom and Ireland) – have revealed many of the flaws in growth-pole theory.

A first defect of the theory is its failure to explain the reasons for the initial presence of a propulsive industry in an area: this presence the theory assumes to be exogenous. Because the growth-pole theory does not explain why the propulsive firm has located in that particular area, it is unable to distinguish the effects of a natural pole from those of a planned pole.

Incentivizing the location of large firms in weak areas through government growth policies is an excessively banal normative interpretation of Perroux's and Boudeville's theory. For a pole to come into being, the large firm, or the industrial complex, must be embedded in a broad production *filière* that subcontracts and outsources numerous activities; investment by the dominant firm thus generates very strong multiplying effects, and it is in these that the definition of a pole resides. In order to generate the effects of a natural pole, a planned pole must necessarily be able to create a local network of intersectoral relations if the positive effects of the large firm's dynamism are to remain *in loco*. And this is an aspect which the majority of public intervention schemes guided by this theory have greatly undervalued.[13] There are, in fact, numerous examples of the creation of large local industrial complexes (e.g. steel and petrochemicals) in the Italian *Mezzogiorno* which have established intersectoral relations outside the region and transmitted the beneficial effects of development beyond the confines of the local economy.

A second shortcoming of the growth-pole theory is that it has deliberately ignored the negative effects (Albert Hirschman's 'backwash effects'[14]) accompanying the formation of a pole, emphasizing only its positive ones (Hirschman's 'spread effects'), and stressing expectations of success in creation of a pole. But it may easily happen that the location of a large firm in an area has an initial crowding-out effect on local firms – especially crafts businesses – resulting from the shock on prices and wages generated by the advent of the large firm, with a markedly negative impact on local employment.[15]

Figure 7.1 illustrates the evolution over time of the negative and positive effects exerted on an area's economy by the formation of a pole. The positive effects tend to appear in the long period, after initial resistance has been overcome and relations have been established between the dominant firm and local businesses. The negative effects, which are very pronounced in the first period, subsequently attenuate, when the local economy has reorganized itself around the large firm. The result is a pattern

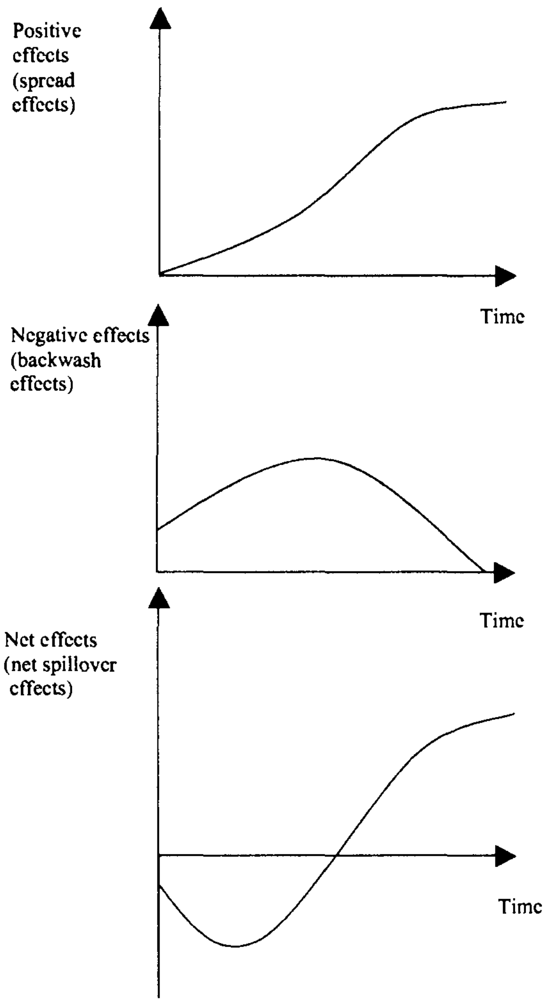


Figure 7.1 Temporal evolution of the positive and negative effects of a pole

of net effects (so-called 'net spillover effects') which is initially markedly negative but then becomes positive when the development-generating effects absorb the negative ones. It has been estimated that the period of net negative impacts may last even for decades.

Moreover, close inspection of the growth-pole theory reveals a contradiction in its logic when applied for normative ends. If the aim is to develop a weak area, the dominant firm must require only few local inputs; but for this reason it is unable to generate large-scale spillover effects on the local economy. In order to obviate this difficulty, Italy switched in past decades from sectors like steel and petrochemicals (1960s) to mechanical engineering and car production (1970s and 1980s) – sectors requiring greater inputs and therefore presumed to generate spillovers. Neither in this case, however, did the growth-pole policy produce the results expected. This demonstrates that the process is a highly complex one, and requires interventions of another kind;

not only incentives to industry but also support for the development of intangible elements like human capital, knowledge and learning.

The growth-pole theory has indubitably played an important part in the history of economic analysis of local development. It was the first theory able to interpret development using a microeconomic approach. It showed that factors like sectoral competitiveness and territorial concentration are crucial for development, and it introduced the notion of selective development in some areas and some sectors. On the theoretical front, however, it was unable to explain why an engine of development (a dominant firm) should be present in some areas and not in others. Because of this shortcoming, normative application of the theory has resulted in sometimes severe policy failures.

7.3 Multinational companies and regional growth

7.3.1 The role of multinationals in regional growth

On the basis of criticisms of the neoclassical model for its optimistic view of capital mobility, and of Perroux's equally optimistic view of the role in local development of large-sized firms, in the 1970s a theory was put forward that interpreted regional development in terms of the impact of large multinational firms on local growth. The micro-behavioural nature of this model's reasoning justifies its inclusion in this chapter.

Adopting a 'radical' approach, this theory focuses on the impact of the location choices of multinational firms on regional development. Its overall thesis is that the location decisions of multinational firms are driven by the profit motive. This, the theory argues, is evident in their choice of areas with low labour costs for the location of labour-intensive unskilled production activities.[16] On this logic, weak regions are the preferred sites for these kinds of low value-added functions, and the development (crisis) of these areas is strictly linked to their success (failure). Breaking the production cycle down into its various functions and finding an appropriate location for them is, according to the theory, the winning strategy for multinational firms. However, this strategy tends to consolidate the division of labour between rich regions as the centres of advanced managerial high-value-added functions, and poor ones destined to receive lower-level activities. There is therefore the risk that what Liepitz calls an 'integration/domination' relationship with the advanced regions may become permanent.[17]

On this radical view, it is capitalist accumulation that causes spatially uneven development. The workings of the capitalist system reproduce and exacerbate regional disparities over time, widening the gap between rich and poor areas within the country.

Accentuating this tendency are further risks for weak areas inherent in a capitalist economy: principally the hyper-mobility of capital, which entails constant changes in the physical location of multinationals. This exposes weak regions to expansion followed by decline of their economies, with rapidly alternating waves of growth and recession. The rapidity itself with which growth comes about generates crises: the scarcity of infrastructures, labour and productive capacity in periods of expansion induces a rise in local wages and prices which impoverishes the entire regional economy.[18]

The persistence of regional imbalances despite public interventions in favour of large firms in the North of England, the South of Italy and Ireland has empirically belied the claims of this 'radical' current of thought.

The 1980s saw the advent of a more balanced school of thought that emphasized also the positive processes engendered in local economies by the presence of multinationals.[19] The development of this school of thought was boosted in the 1980s by numerous technological innovations that altered the standard functional division of labour. The reprogrammable systems made possible by CAM/CAD applications, and the computerization of numerous administrative and managerial procedures, gave rise to a new organization of industry characterized by closer functional integration (production, design, research, marketing and strategy), deverticalization and reorganization of the production cycle.[20]

In general, the presence of multinationals has been associated with a higher static efficiency of the entire area made possible by technological transfer and increased productivity. Advantages stemming from the presence of multinationals in an area have been classified as direct and indirect. The direct ones refer to:

- job creation;
- increased managerial and technological expertise;
- increased wage levels thanks to the formation of higher quality occupations;
- change in the industrial and functional mix;
- increase in the productive capacity.

Added to these benefits have been indirect advantages in the form of foreign direct investment (FDI) induced spillovers; these refer to the pecuniary and technological externalities that the local productive system receives from the presence of large multinationals. The latter are in fact unable to internalize all benefits stemming from the knowledge and the technology that they produce. In particular, the indirect advantages have been identified as rather numerous, namely:

- a strengthening of the productive system in areas with scant entrepreneurship;
- enhancement of industrial agglomeration effects;
- stimulus for new industrial investments upstream and downstream from the multinational firm;
- the creation of new firms upstream and downstream from the multinational;
- localized technological spillovers;
- the increase of managerial and technological know-how in the local area;
- cross-fertilization between firms and local institutions in the provision of vocational training.

These advantages, or spillovers, may be diffused in the area that hosts the multinational through technological externalities; that is, through imitation or reverse-engineering processes, labour-force training, and an increase in the quality of the production processes of local firms due to the greater competition that the large multinational generates among potential suppliers. Moreover, the advantages may be diffused through input–output linkages between the multinational and the local firms. In this case, they take the form of pecuniary externalities, since they are mediated by the market.[21]

The intensity with which these processes arise in the local economy depends on FDI characteristics and on the characteristics of the area itself. The degree of the group's vertical integration, the technological intensity of its production process, the size of its *filière*, its position in that *filière*, the type of investment (greenfield or the purchase of already existing firms) and the extent to which production is outsourced are all aspects of the *modus operandi* of multinationals that affect the benefits that they generate for the local economy.

As regards an area's characteristics, its already existing productive system (assessed quantitatively and qualitatively), human capital and technological knowledge determine the extent to which location in the area by a multinational company affects the local economy. A recent body of literature stresses that a multinational's input–output relations with local firms are crucially important for local development – which recalls the growth-pole theory and its emphasis on the importance of intersectoral relations for local economic development. This more recent school of thought mathematically models the effects of input–output relations on economic growth as increasing returns in an aggregate production function that are entirely absent from Perroux and Boudeville's theory.[22]

The intensity of the effects has been highlighted as depending on a combination of FDI characteristics and the characteristics of the area in which multinationals locate. In its turn, this combination has been associated with the strategic reasons that have induced multinationals to produce abroad. When the reasons to produce abroad have been efficiency seeking, the advantages for the host areas have always been expected to be limited, at least in the short run, due to the low quality/low technology functions located abroad by the multinational, and to the choice of a host area characterized by low labour costs and low-activity production functions. By contrast, when the decision to produce abroad has been taken in order to obtain strategic resources, the functions involved are high-level ones (e.g. research and development activities), and the choice of where to locate is driven by the presence of a highly specialized human capital, a competitive industrial tissue and a rich creative context. All this implies more favourable conditions for the host area, since it is expected to be able to develop co-operation strategies between the multinational and the local firms for the exchange of knowledge and technologies, and to prevent pure exploitation strategies by the multinational.[23]

Finally, the literature on the role of multinationals in local development has more recently focused on aspects of technology transfer (or technological spillovers[24]) from multinational firms to local economies. These spillovers are more frequent in advanced areas with high levels of innovation. They also occur most frequently in industrial areas similar to the sector in which the multinational operates and with already-existing specific knowledge that the additional knowledge generated by the multinational enriches.[25] Moreover, the innovativeness already possessed by a local system is important if the latter is to exploit the technological externalities produced by a large multinational. It is accordingly useful to examine two features: the endogenous technological potential of an area, and the mechanisms by which innovation spreads through space.[26]

Despite the plethora of FDI-regional growth studies, the long-run relationship between FDI and growth is not yet clear, and the empirical results are far from being unanimous; the conditions for a long-term FDI-regional growth nexus are complex and numerous, and they result from a combination of FDI-characteristics and local factors. A generalization of the phenomenon is consequently rather complex, and for this reason it still captures the interest of a large number of scholars.

7.3.2 *The determinants of regional attractiveness for multinationals*

On the basis of the indisputable short-term direct advantages of greenfield investments for employment and growth of the host areas, and of the existence of success stories of FDI-driven growth patterns (like some Eastern European country regions – mostly capital regions – and Ireland), a large number of studies have been devoted to identification of the sources of local FDI attractiveness, with the aim of providing an explanation for their unequal spatial distribution.

Generic reasons for the FDI location choices have been related to their strategies to become multinational. In particular:

- the presence of large final markets when the reason to produce abroad is a market-seeking strategy;
- the presence of low-cost labour, of a limited distance to the final goods market and of limited international trade barriers when the reason of the firm to become a multinational is a cost minimization-seeking strategy;
- the presence of high-quality assets and of human capital when the choice to produce abroad is a strategic-seeking strategy;
- the presence of natural resources and of raw materials when the reason for the firm's international production is oriented to a resource-seeking strategy.

The theory of FDI developed along the above-mentioned lines has been criticized for its a-spatial nature, since it has considered geography as 'highly stylized and unspecific',[27] suggesting location factors applicable to a high number of regions, and therefore being unable to explain real FDI location choices.

During the 2000s, theories have started to look for the determinants of specific local assets attracting FDI. They have done so by considering specific territorial elements of the region on the one hand, and spatial interdependences in the location determinants on the other.

As for territorial characteristics, the presence of large cities has been highlighted as an important aspect for FDI attractiveness: cities are the loci of qualified labour, of amenities, of knowledge, of advanced and specialised services for industrial activities, of 'urban atmosphere'. More recently, 'soft' assets like social and relational capital, quality of local institutions, quality of life, quality of transport infrastructure and accessibility have been mentioned as important sources of FDI attractiveness. In this stream of studies a crucial question has become whether the location choices of FDIs respond first to macroeconomic and national elements (stability of the national banking system and of the macroeconomic conditions; quality of national institutions), and then to regional aspects, or whether the regional elements determine the firm's location choice notwithstanding the national socio-economic conditions.[28] Recent empirical analyses have shown that the location choices among countries and regions are strongly interrelated, demonstrating that multinationals take account of both aspects at the same time.[29]

As for spatial interdependences in the location determinants, the literature has recently highlighted the role of local assets of nearby regions in influencing FDI location choices. From the empirical perspective, this has been made possible by the development of spatial econometric techniques that allow, in the estimation of location choices, inclusion of the characteristics of nearby regions, weighted by distance.

Like the FDI/regional growth nexus, also identification of the sources of FDI attractiveness is highly complex, due to the numerous intertwined aspects that are at work and influence the location choice. Despite the numerous efforts already made in the literature, this is a field in which further conceptual and empirical reflections are necessary.

7.4 The spatial diffusion of innovation

7.4.1 Hägerstrand's model: geographical distance

We have not yet examined the role of innovation in local development. The interregional theories and models of local growth presented in Part II of this book conceived – in a typically neoclassical manner – innovation as 'manna from heaven' available to all economic actors at no cost, and as such with no influence on the growth capacity of systems; technological progress, these theories and models believed, came about at the same pace and simultaneously in all sectors, among all economic actors, and in all territorial contexts.

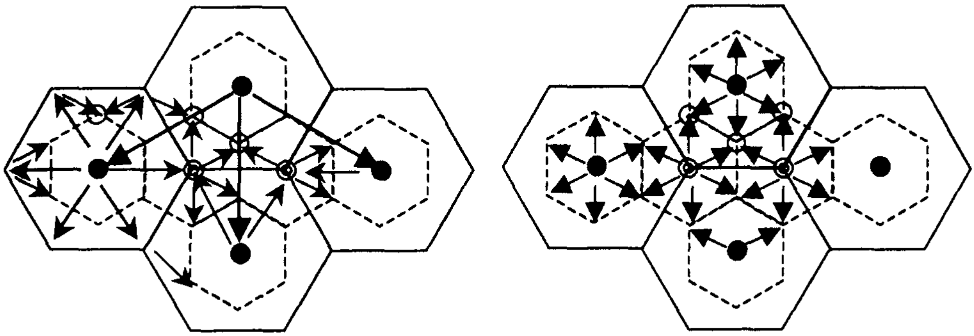
However, if the assumption of perfect information is abandoned for the more modern notion adopted in the 1980s of information asymmetry, the simplistic framework assumed thus far no longer holds. Innovation comes about in entirely different ways in different areas and is thus a key factor in explanation of the differing capacities of regions to grow. It accounts for the process of output growth, which cannot be directly attributed to an increase in the production factors, in equilibrium and with constant returns to scale.[30]

Innovation is therefore of key importance for explanation of why local systems grow; and any thoroughgoing theory of regional development must be able to specify the sources of innovation and the factors that give a local system innovative capacity.

An early approach to analysis of these matters conceived innovation as an exogenous factor in development; innovation, this approach maintained, propagates through specific territorial channels to generate positive impacts on a local area from outside. Analysis should therefore examine the territorial routes whereby innovation reaches a particular area; routes formalized in models of the spatial diffusion of innovation.

The best-known, and probably earliest, of these models was developed by the Swedish geographer Torsten Hägerstrand.[31] His pioneering work on innovation diffusion, on which numerous subsequent studies were based, maintained that the temporal development of an innovation displays an S-shaped pattern represented by a logistic function, and that temporal phases of the cycle must be combined with spatial ones to depict a spatial-temporal diffusion of innovation moving through the following three phases:

- the first 'adoption' stage, when the urban hierarchy canalises the course of diffusion: the innovation centre is often the primary city or some other metropolitan centre; the centres next in rank then follow (Figure 7.2a);



● Cities of order n

◎ Cities of order n - 1

○ Cities of order n – 2

Figure 7.2 Hägerstrand's channels of innovation diffusion

- the second 'diffusion' stage, when the hierarchical effect and the 'neighbourhood effect' (the latter illustrated by Figure 7.2b) act simultaneously, with different weights according to the moment in time. Initially, the hierarchical effect still predominates; however, as time passes, the friction of space puts outlying large centres out of the immediate range of the diffusion's influence, and the neighbourhood effect predominates;[32]
- the third 'saturation' stage, when the spatial diffusion of the innovation becomes random. Saturation may be reached around the innovation centre while the rate of diffusion is still low in distant areas. The overall deceleration in the diffusion pattern may conceal a catching-up process in which the innovation is still spreading through distant areas although adoption at the centre has halted.

In Hägerstrand's model, an innovation is diffused by an epidemic process: the pure likelihood of contact between people who have already adopted an innovation and its potential adopters explains innovation diffusion in this model, which implicitly assumes that every potential adopter has the same opportunity to adopt, and that spatial variations in adoption are due solely to information flows that spread territorially at different times.

The model assumes that information about innovation automatically entails its adoption; and when Hägerstrand seeks to include the uneven distribution of receptiveness among an innovation's potential adopters, he once again resorts to a measure of information intensity.[33]

Moreover, it should be borne in mind that the use of a logistic function is only acceptable if it can be assumed that potential adopters are all equally likely to adopt a particular innovation. The element that makes this assumption unacceptable is space; it is very difficult, in fact, to hypothesize that potential adopters located in two territorial areas different in structure and productive performance will have the same receptiveness to innovative processes. Hence space does not perform a significant role in Hägerstrand's model except in its form as pure geographical distance between actual and potential adopters.

7.4.2 The contribution of Griliches and Mansfield: economic distance

The economists Zvi Griliches and Edwin Mansfield have examined the spatial characteristics that condition the innovation adoption process. They introduce into Hägerstrand's model the idea that the spatial diffusion of innovation is influenced less by geographic distance among adopters than by economic distance; the amount of productive activity in an area, and its levels of income, consumption and investment, can straightforwardly explain the greater receptiveness of an adoption area.[34]

Griliches and Mansfield formulate a two-stage methodology for the empirical analysis of innovation diffusion. The first stage involves estimation of a logistic function taking the following form:

$$D = K / (1 + e^{-(a+bt)}) \tag{7.1}$$

where D is the density of adoption, or the cumulated number of adopters, a the moment in time when the first adoption occurs (origin of the logistic curve), b the speed of adoption (the slope of the logistic curve) and K the asymptote towards which the curve tends (the ceiling of the logistic curve); that is, the maximum number of potential adopters that the innovation can reach (Figure 7.3). The analytical properties of the logistic function enable simple estimation of its parameters.[35]

Once the parameters of the logistic function are known, the next stage is estimation, by means of interregional cross-section regressions, of the incidence of the local economy's main characteristics on the historical moment of the innovation's adoption, on

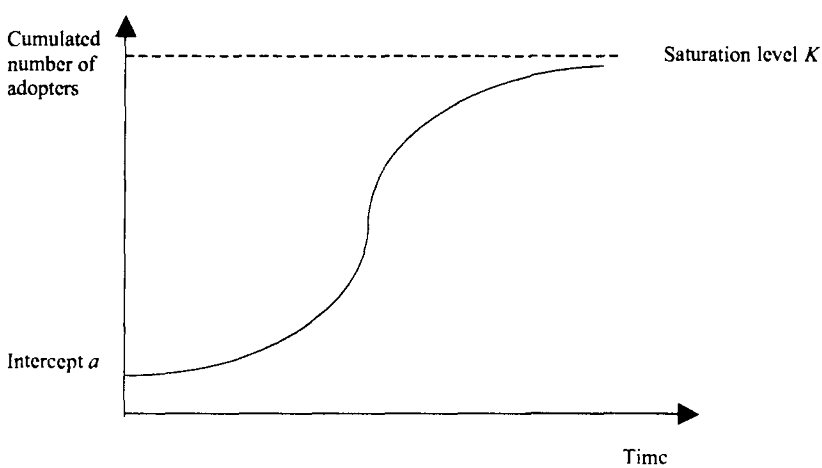


Figure 7.3 The logistic function

the speed of its penetration and on its level of saturation. This furnishes a clear 'snapshot' of the various spatial patterns of the innovation's adoption.[36]

Griliches applied this methodology to explain the wide cross-sectional differences in the rates of use of hybrid seed corn in the United States. He found that the lag in the development of adaptable hybrids for particular areas and the lag in the entry of seed producers into these areas (differences in origins) were explained by varying profitability of entry, 'profitability' being a function of market density, and innovation and marketing costs. Differences in the long-run equilibrium use of hybrid corn (ceilings) and in the rates of approach to that equilibrium (slopes) were explained, at least in part, by differences in the profitability of the shift from open pollinated to hybrid varieties in different parts of the country.[37]

When the same methodology was applied to interpret the diffusion of fixed telephone services in the 20 Italian regions, it provided clear evidence that adoption of the innovation (origins) differed markedly among regions because of socio-economic factors: the industrial performance of regions and the educational level of the local population amply explained both the level of saturation and the speed of adoption of the service (i.e. the ceilings and the slopes). The urban structure of a region was instead important in determining the historical moment when the innovation was adopted (expressed by the value of the *a* parameter) – thus highlighting the role of cities in the birth of innovations.[38]

In the case of process innovation adoptions (industrial automation, robotization, etc.), one may logically expect the adoption rate to be lower in weak areas. At macro level, an important obstacle against adoption in weak regions is the lower cost of labour, which reduces the relative profitability of new labour-saving technologies. From a microeconomic point of view, there are factors of cultural and organizational backwardness that generate high adjustment costs in the switch from an old to a new technology. These costs are such that the latter is much less profitable in weak than in advanced regions.

From a dynamic perspective, there are two further critical elements in explaining the late adoption of innovations in weak regions. The first is the irreversibility of non-adoption choices, which may condemn an area to a permanent state of technological backwardness. Irreversibility springs from complex, irreversible processes of cumulative learning and investment in knowledge that accompany transition to a new technological trajectory. These processes influence the costs of, and revenues from, adoption of the new technology, and they alter its relative profitability. In the first period, when the innovation is adoptable, it may well be the case that the costs of adopting a new technology exceed the revenues, justifying, from a static point of view, the non-adoption decision (Figure 7.4). But the time trend of the costs and revenues of adoption shows that there is a temporal interval within which adoption of the new technology is profitable. However, beyond a certain point (time 2 in Figure 7.4), the non-adoption decision becomes irreversible because the costs of adoption are greater than the revenues, and the trend increases in time. As a result, the weak region is condemned to competing against the advanced regions with a more limited and obsolete technological endowment.[39]

The second critical element concerns the need for anticipatory and far-sighted public policies; in fact, if incentives for new technologies are allocated in the first phases of adoption, they may be less than those that are subsequently necessary when the technologies are being developed. As Figure 7.5 shows, if a new technology starts

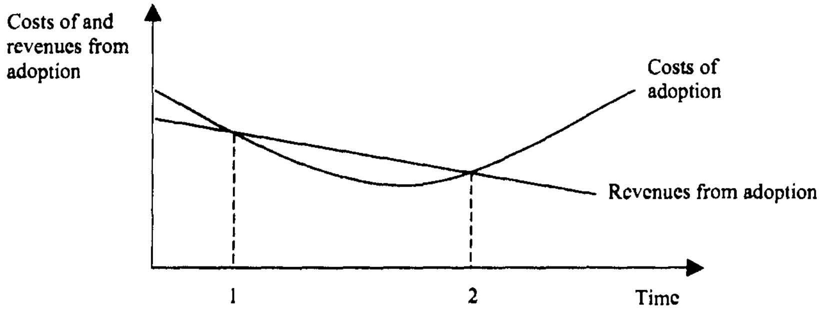
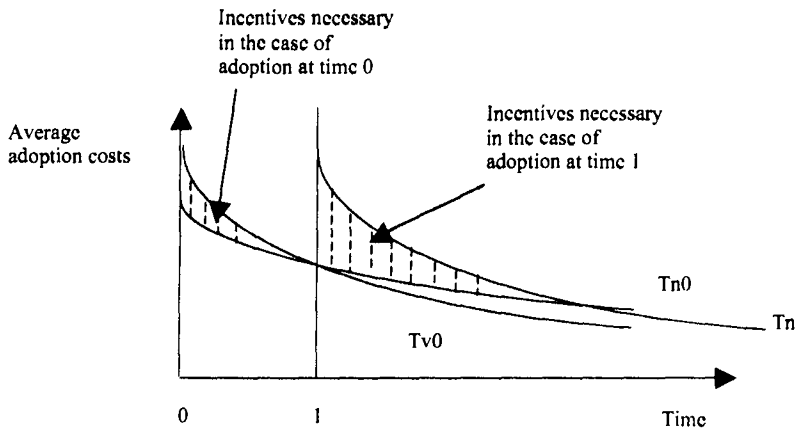


Figure 7.4 Time trend of the costs of and revenues from adoption
Source: Camagni and Capello (1998)



Tv0 = Average adoption cost of the old technology at time 0

Tn0 = Average adoption cost of the new technology at time 0

Tn1 = Average adoption cost of the new technology at time 1

Figure 7.5 Temporal evolution of the incentives necessary for the adoption of a new technology
Source: author's elaboration on Camagni (1998)

to be developed at time 0, its development may involve a limited extra cost with respect to the old technology. But if the new technology is adopted at time 3, the public costs required to support the late adoption are higher. Consequently, anticipatory and far-sighted policies supporting the introduction of new technologies are all the more necessary and desirable, but generally more difficult to put in place in backward regions.

7.4.3 *The limitations of the logistic-epidemic model*

Although the logistic model is well known in the literature on the spatial diffusion of innovation, it has not been immune to criticisms centred on its evident interpretative shortcomings.[40]

First, the logic of the model does not envisage the technological evolution of an innovation – that is a post-innovation improvement. Even less does it contemplate paradigmatic shifts to other technological trajectories brought about by radical innovations. Yet changes in technological knowledge may induce product innovations, with the consequence that they change or even halt the development predicted by the logistic pattern on the old product. Generated at the same time is a process of new product development, represented by a new logistic function, which may co-exist with the old technology for some time, giving rise to a technological pluralism inconceivable on the logic of the logistic model.[41]

A second shortcoming of the logistic approach is that it conceives technological development as resulting from the behaviour of potential adopters and from demand for technology, while taking for granted the existence of a completely flexible supply able to satisfy that demand. However, it is now recognized that innovation is the result of a virtuous 'demand/supply' circle in which both components interact to influence the time scale of the innovative process and the ways in which it comes about.

Finally, the logistic model defines the number of potential adopters *ex ante*, exogenously; an aspect that has a considerable negative bearing on the model's interpretative capacity.

7.4.4 *The product life-cycle and the life-cycle of regions*

The notion that the spatial diffusion of innovation is a continuing process in time represents the central component of the logistic model used by the regional life-cycle theory proposed by Norton and Rees and based on Hirsch and Vernon's well-known product life-cycle theory.[42]

Norton and Rees's life-cycle theory interprets regional differences in technological capacity as stemming from physiological processes due to technological ageing. Technological development has three stages associated – through analysis of demand, production and innovative processes – with three specific locations of innovation (Figure 7.6a), as follows:

- the first stage is the take-off of a new product. When incremental innovations in the characteristics of a product are frequent, and production processes have not yet been standardized, the strategic factors necessary for the innovation are research and innovation capabilities, the quality of labour and ready access to specific information. The natural location is an urban and metropolitan area where demand is more inelastic (rigid) to prices and there is a greater receptiveness to innovation;
- the second stage is product maturity. Here incremental process innovations predominate, and the strategic factors for innovation are managerial ability and the availability of capital. Production processes require large-scale plants because they are now highly capital-intensive. The more peripheral areas of advanced countries, where land costs less, are the best locations for manufacture of the innovative product;
- the third stage is the standardized production of the innovative good. The strategic factor is now the cost of labour, and the optimal location is a developing country.

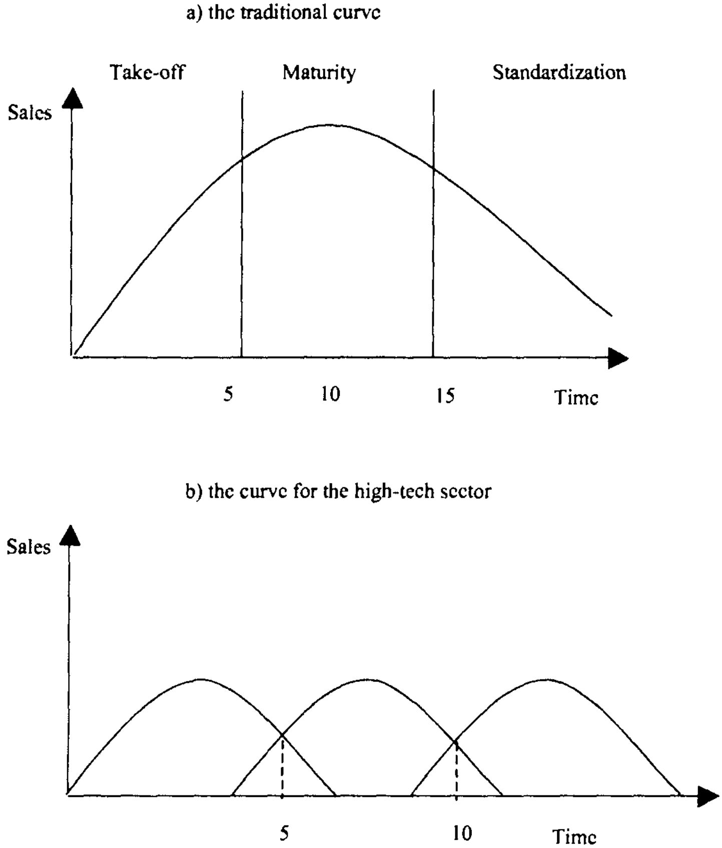


Figure 7.6 The product life-cycle

The result of the above process is the progressive diffusion of the product innovation from more central metropolitan areas to the periphery. This comes about through a mechanism whereby the innovation 'filters down' from stronger areas to weaker ones.

A strength of this theory – one which distinguishes it from the logistic-epidemic model – is its ability to conceive of an *interregional technological pluralism*. This results from a 'snapshot' taken in a particular time-frame of the interregional movement of technologies – a movement driven by the simple process of the physiological ageing of technologies. It may happen, in fact, that whilst central areas stagnate, the periphery experiences notable innovation characterized by 'creative imitation'.[43]

However, this reasoning has a number of flaws. The first is that, whilst the life-cycle theory was well suited to interpreting the spatial diffusion of innovation in the 1950s and 1960s, when technological change in products took the form of long waves comprising the traditional stages of take-off, maturity and decline, the same cannot be said of the 1980s, when the new technological paradigm associated with high-tech industry imposed an extremely rapid pace on product evolution. The life-cycles of products were therefore drastically reduced; and the patterns of spatial adoption changed as a consequence (Figure 7.6b).[44] The pressure exerted by the shorter life-cycles of products was not in fact accompanied by a more rapid spatial diffusion of innovation from the centre to the periphery. Moreover, the nature of the new technological paradigm in the 1980s was such that it revitalized 'traditional' production, inducing it to relocate from the periphery to the centre. The new electronic technologies introduced in traditional sectors like textiles, cars, clothing and precision instruments enabled them to renovate production by improving products and processes; and this gave central regions, rather than peripheral ones, ample opportunities for economic recovery.

The second weakness of the life-cycle model is that it entails, with no apparent exceptions, a view of interregional technological development as a simple and linear process of technology transfer between territorial areas. By so doing, it rules out the possibility that there might be some subjective element in the diffusion process, such as the interest, ability and receptive capacity of one area compared to another.[45] It thus makes the same mistake as committed by Hägerstrand when he assumed that information necessary entails adoption.

It will be evident from the foregoing discussion that the spatial diffusion of innovation is a highly complex process in which demand components (characteristics of the potential adopters) interact with supply components (characteristics of the prevailing technological paradigm) and, not least, with elements specific to the context in which the innovation diffuses (structural characteristics of the area). All this is very difficult to incorporate into models for which the diffusion mechanism is an epidemic process. Excluded by definition is the existence of areas able to absorb innovation in different manner. Yet this to overlook one of the key elements in explanation of the spatial diffusion of innovation. Nonetheless, to be commended are the efforts made by the various models to include this element in their analyses; either in the form of 'economic distance', as in Griliches and Mansfield, or (and this will be examined in detail in Chapter 9) local endogenous factors – most notably the greater presence of the knowledge and learning that support adoption processes and account for an area's greater capacity to adopt and exploit innovation.

7.5 Infrastructures and regional development

Many of the growth theories discussed agree on the importance of infrastructures for regional development. The theories of balanced development, stages of development, the export base and growth-poles underline the role of infrastructures in determining a local system's growth and development path. According to these theories, export capacity, the production system's competitiveness and an area's capacity to attract new activities result, *inter alia*, from a developed infrastructural endowment.

In light of these theories, more recent analyses have paid closer attention to infrastructures, seeking to identify, among the many possible determinants of an area's growth, the real contribution made thereto by infrastructures.

These analyses consider infrastructural endowment to be one of the factors that – together with geographical location and an agglomerative sectoral structure – determine a regional development potential.[46] A better infrastructural endowment attracts new firms into an area, and it is a source of competitiveness for the firms already operating in that area. It heightens the productivity of the production factors, and by increasing accessibility, reduces their purchase costs – thus generating positive externalities on local development.[47]

Because the production of infrastructures is indivisible, they are often produced by the public sector; for which reason they are termed 'public capital' or 'social fixed capital'.

The theoretical analysis described has been accompanied by a large number of empirical studies intended to measure the contribution of 'social fixed capital' to factor productivity.[48] The most common method is to estimate an aggregate production function (at the regional or provincial level) in order to verify the existence of scale economies – or (which is easier from an econometric point of view) a multiplying coefficient connected with the infrastructural endowment.[49]

The most important contribution to such analysis has been David Aschauer's study[50] which estimated a production function to show that public capital exerted a strong positive influence on total factor productivity in the United States between 1945 and 1985. Since Aschauer's study, a large number of other empirical surveys have shown – for different geographical areas and different time periods – that greater public capital intensity has positive and significant effects on output elasticities, meaning that a greater provision of public capital increases the magnitude of the impact on regional output. Table 7.1 summarizes the range of alternative estimates of the impact in the recent literature.[51]

The discovery by these empirical analyses of a correlation between infrastructural endowment and economic growth suggests that 'social fixed capital' is a determinant of local competitiveness and factor productivity. The closeness of the correlation largely depends on the type of public capital considered: 'economic' infrastructures (transport facilities, roads, motorways, railways, airports and electricity generating stations) – these being directly functional to firms – give rise to greater increases in productivity compared with 'social and civil infrastructures' (hospitals, schools, universities, public housing projects and sewerage systems). Although the latter infrastructures directly affect the quality of life and human capital, they influence production only in the longer run – and with effects not necessarily restricted to the area in which the infrastructures are installed. Moreover, the numerous estimation methods and the diverse territorial disaggregations used by studies on the matter account for the marked variability in the values of income elasticity to social fixed capital (Table 7.1).

Yet the infrastructural policies implemented over the years have had only very limited positive impacts on regional disparities. Besides national policies, 80 per cent of the structural funds allocated by the European Union in the 1970s and 1980s to the development of infrastructures in the Objective 1 regions did not reduce regional disparities: these remained constant in the 1970s, indeed worsened in the 1980s, and improved in the 1990s; lastly, a process of 'reverse convergence' took place during the economic crisis of the end of the 2000s.[52]

These results show that it is necessary to proceed with caution if infrastructural investment is to generate economic development.[53] Infrastructural development must necessarily match the needs expressed by the industrial specialization of the area in which the infrastructures are to be installed – as already amply evidenced by the theory of balanced development.[54] Even more misguided is the idea that the creation of infrastructures alone in a weak economic region can engender economic growth, if there is not a 'fertile' productive context on which to graft development. Moreover, the building of a transport infrastructure may increase competition in the area because it makes the local market accessible to external firms. Finally, in the presence of an already

Table 7.1 Alternative estimates of the impact of infrastructure on output

Studies	*Output elasticity*	*Levels of analysis*	*Infrastructure variable*	*Output variable*
Aschauer (1989)	0.39	National	Public capital	National output
Munnell (1990)	0.33	National	Public capital	National output
Aschauer (1989)	0.24	National	Core public capital	National output
Hulten and Schwab (1995)	0.39	National	Public capital	National output
Moomaw et al. (1995)	0.07–0.26	State	Public capital	Gross state product
Moomaw and Williams (1991)	0.25	State	Highway density	Total factor productivity
Costa et al. (1987)	0.20	State	Public capital	Ouput
Munnell (1990)	0.15	State	Public capital	Gross state product
Aschauer (1990)	0.11	State	Core public capital	Per capita output
Munnell (1990)	0.06	State	Highway capital	Gross state product
Deno (1988)	0.31	Metropolitan	Highway capital	Manufacturing output
Duffy-Deno and Eberts (1991)	0.08	Metropolitan	Public capital	Personal income
Eberts (1986)	0.19–0.26	Metropolitan	Core public capital	Manufacturing value added

Source: Guild (1998)

well-developed infrastructural endowment, further investment in fixed social capital produces – as in the case of any intensively used factor – a very small increase in local production.[55]

In sum, whilst it can be argued that an endowment of fixed social capital is a necessary condition for local development (as also shown by the results of empirical surveys), it is wrong to believe that it is a sufficient one. A series of other, equally necessary, factors must exist – among them entrepreneurship, specialization and innovative capacity – if an infrastructural policy is be truly effective.

7.6 New communication technologies and regional development

New opportunities, but also new threats, for the development of local systems arise from the advent of the technological paradigm of the information and communication technologies (ICTs). The adoption and use of ICTs (or 'computer networks') open up broad avenues for innovation that encourage local development. Product innovations (e-business, e-commerce), innovations in product distribution (online marketing) and process innovations (just-in-time production, functional integration) spring from the presence and exploitation of these technologies, giving the local production system greater competitiveness and efficiency.

The opportunities for development afforded by the new communication technologies depend on the strategic use made of those technologies themselves – a use, that is, which combines new technological potentialities, new organizational methods (required by the innovations themselves) and new ways to penetrate markets with more innovative and qualitatively better products (business ideas). Appropriation of the potential profits and higher levels of competitiveness offered by these technologies therefore requires knowledge and innovative and creative skills that are certainly not distributed uniformly in space. Knowledge results from slow processes of learning fuelled by information, knowledge and investments in research and training stemming from local experience and expertise embodied in human capital and local relational networks, and in the local labour market. Knowledge is increasingly embedded in the local production system, so that learning processes are highly selective, and they determine capacities for technology use which differ markedly across regions.[56]

This last point has an important implication that relates to what was said earlier about infrastructures in general: the mere adoption of these technologies is a necessary but not sufficient condition for local development. Technologies in the strict sense exert a neutral influence on the growth patterns of local production systems, in fact; they represent a set of opportunities available at a certain cost, a 'quasi-public good'. The discriminating factor – and this is by no means a 'public good' – is the cultural and organizational capacity to exploit their potential with a creative array of technologies, organizational styles and business ideas. These capacities are very likely to be present in central areas, especially in the early phases of adoption and, at least initially, they boost the centripetal forces of development.

This point reminds us that the effect of ICTs on regional disparities is still dubious; in the past 15 years two currents of thought have interpreted the impact of ICTs on regional disparities:[57]

- the first maintains that new ICTs are able to resolve the problem of peripherality: greater access to information, knowledge and specific services to production reduce the disadvantages of a peripheral location – the disadvantages emphasized by the 'centrality–peripherality' approach described in Chapter 4;
- the second, and contrary, current of thought argues that the diffusion of these technologies takes the form of a centripetal process driven by the presence in stronger areas of greater potential demand, and of more knowledge about these technologies and the ability to exploit them. On this view, the centripetal adoption process gives the centre broad opportunities to maintain and widen the gap between it and the weak regions.

Empirical studies demonstrate that in the first phases of adoption – or in all the phases of technological progress that require new strategies of technology adoption – the diffusion of technologies is a centripetal process. A study carried out in the early 1990s on the impact of ICTs on regional development in Northern and Southern Italy showed the existence of an 'ICT adoption/local development' nexus in the North. But this appeared to be entirely absent in the South, owing to a lack of the knowledge necessary for its strategic use. A similar conclusion was reached by an empirical study on the USA: a state-by-state regression analysis showed the variation in returns on telecommunications investments across states. This variation may have been due to the inefficient utilization of telecommunications infrastructure as a factor of production. The same study found that the states obtaining significantly positive benefits were those in which firms used the telecommunications infrastructure more efficiently.[58]

More recent studies have again stressed that if knowledge on how to use these technologies is not sedimented in the local labour market and fuelled by strong relations among local firms, the adoption of new technologies will not generate processes of local development.[59]

7.7 Conclusions

This chapter has surveyed the first theories based on a notion of *diversified space* which enables production activities and factors, demand and sectoral structure to be conceived as unevenly distributed within a region. According to this approach, the economic and social relations that arise in a geographical area perform an important role in explaining a local system's development; hence the expression *diversified-relational space.*

Of the numerous approaches to diversified-relational space, the chapter has described those that identify elements exogenous to the system as determining long-period competitiveness: the presence of a dominant firm or a multinational, the diffusion of an externally originating innovation, the creation of transport and social infrastructures and the adoption of new communication technologies. The next chapter will instead examine the theories that, with intriguing and impressive insight, have sought to identify the endogenous factors that determine local competitiveness.

Review questions

1 What is meant by a diversified-relational space? How does this definition of space change theories of regional development?
2 What is the conception of growth behind the theories based on a diversified-relational space?
3 What is meant by exogenous development?
4 What are the main novelties of the growth-pole theory? What are the limits of this theory?
5 What are the main determinants in the location choices of multinationals?
6 What are the local conditions for multinationals to generate economic development?
7 What is the aim of Hägerstrand's model? What are the limits of this model? What is added to this model by Grilisches and Mansfield's contributions?
8 Is the statement 'infrastructures generate growth' true? Explain why.
9 Is the statement 'ICTs decrease regional disparities' true? Explain why.

Selected reading on empirical findings

About infrastructure and regional development

Biehl D. (1986), *The Contribution of Infrastructure to Regional Development, Regional Policy Division*, European Community, Brussels.

Bröcker J. and Schneekloth N. (2006), 'European Transport Policy and Cohesion: An Assessment by CGE Analysis', *Scienze regionali – Italian Journal of Regional Science*, vol. 5, no. 2, pp. 71–87.

Fabiani S. and Pellegrini G. (1997), 'Education, Infrastructure, Geography and Growth: An Empirical Analysis of the Development of Italian Provinces', Banca d'Italia, *Temi di Discussione*, no. 323.

Spiekermann K. and Wegener M. (2006), 'Accessibility and Spatial Development in Europe', *Scienze regionali – Italian Journal of Regional Science*, vol. 5, no. 2, pp. 15–46.

About ICTs and regional development

Butler J., Gaspar J.M.B. and Jeppesen E. (1986), 'Telecommunications and Regional Development in Portugal', *Arbejdsrapport*, vol. 16, Aarhus University.

Capello R. and Nijkamp P. (1996), 'Telecommunications Policy for Regional Development: Theoretical Considerations and Empirical Evidence', *The Annals of Regional Science*, vol. 30, no. 1, pp. 7–30.

Ding L. and Haynes K.E. (2004), 'The Role of Telecommunications Infrastructure in Regional Economic Growth of China', paper presented at the Telecommunications Policy Research Conference, Washington, DC, 1–3 October.

Yilmaz S. and Dinc M. (2002), 'Telecommunications and Regional Development: Evidence from the U.S. States', *Economic Development Quarterly*, vol. 16, no. 3, pp. 211–228.

About multinationals and regional development

Cantwell J. and Piscitello L. (2002), 'The Location of Tecnological Activities of MNCs in European Regions: The Role of Spillovers and Local Competencies', *Journal of International Management*, vol. 8, no. 1, pp. 69–96.

Crescenzi R., Pietrobelli C. and Rabellotti R. (2012), 'Innovation Drivers, Value Chains and the Geography of Multinational Firms in European Regions', LSE 'Europe in Question' Discussion Paper Series No. 3/2012.

Iammarino S. and McCann Ph. (2013), *Multinationals and Economic Geography*, Edward Elgar, Cheltenham.

Li S. and Ho Park S. (2006), 'Determinants of Locations of Foreign Direct Investment in China', *Management and Organization Review*, vol. 2, no. 1, pp. 95–119.

Nachum L. and Wymbs C. (2002), 'Firm-specific Attrbutes and MNE Location Choices: Financial and Professional Service FDI to New York and London', ESRC Centre for Business Research, University of Cambridge, Working Paper No. 223.

Pusterla F. and Resmini L. (2007), 'Where Do Foreign Firms Locate in Transition Countries? An Empirical Investigation', *Annals of Regional Science*, vol. 41, no. 4, pp. 835–856.

Resmini L. (2007), 'Regional Patterns of Industry Location in Transition Countries: Does the Economic Integration with the EU Matter?', *Regional Studies*, vol. 41, no. 6, pp. 747–764.

Urata S. and Kawai H. (2000), 'The Determinants of the Location of Foreign Direct Investment by Japanese Small and Medium-sized Enterprises', *Small Business Economics*, vol. 15, no. 2, pp. 79–103.

Further reading

Antonelli C. (eds) (1992), *The Economics of Information Networks*, North Holland, Amsterdam.

Biehl D. (1986), *The Contribution of Infrastructure to Regional Development, Regional Policy Division*, European Community, Brussels.

Boudeville J.-R. (1966), *Problems of Regional Economic Planning*, Edinburgh University Press, Edinburgh.

Capello R. (1994), *Spatial Economic Analysis of Telecommunications Network Externalities*, Avebury, Aldershot.

Casi L. and Resmini L. (2014a), 'Spatial Complexity and Interactions in the FDI Attractiveness of Regions', *Papers in Regional Science*, vol. 93, no. S1, S51–S78, DOI: 10.1111/pirs.12100.

Ding L. and Haynes K.E. (2004), 'The Role of Telecommunications Infrastructure in Regional Economic Growth of China', paper presented at the Telecommunications Policy Research Conference, Washington, DC, 1–3 October.
Gillespie A. and Williams H. (1988), 'Telecommunications and the Reconstruction of Regional Comparative Advantage', *Environment and Planning A*, vol. 20, no. 10, pp. 1311–1321.
Holland S. (1971), 'Regional Underdevelopment in a Developed Economy: The Italian Case', *Regional Studies*, vol. 5, no. 2, pp. 71–90.
Iammarino S. and McCann Ph. (2013), *Multinationals and Economic Geography*, Edward Elgar, Cheltenham.
Vickerman R. (ed.) (1991), *Infrastructure and Regional Development*, Pion, London.

Notes

1 The reference is, for example, to formalization of equilibrium in non-linearity conditions and equilibrium under monopolistic competition. The latter was proposed towards the end of the 1970s by Dixit and Stiglitz, and it provides the basis for some of the models presented in Part IV.
2 See von Böventer, 1975, p. 3. When von Böventer refers to '"pure and exact" regional theory without agglomeration economies', he means the theories presented in Part II of this book; when he refers to '"applied regional theory" which is inexact but takes agglomeration factors into account', he means theories expounded in more qualitative form, which will be the ones developed in this part of the book.
3 'Transaction costs' are the costs that arise from the exchange of information and documents relative to commercial transactions, for which reason they are also called 'costs of market use'. See Williamson, 1975.
4 This is the case of the weak region achieving greater growth than the rich region in Borts and Stein's one-sector model. To be stressed is that the view of development adopted by other neoclassical models, like the Heckscher–Ohlin model, is one of generative development, not of competitive development. On the distinction between competitive and generative development see Richardson, 1973 and 1978.
5 Perroux, 1955, p. 308, my translation. The same ideas are set out in embryonic form in Perroux, 1950.
6 Higgins proposes the following definition: a firm *A* can be called a 'dominant firm' if its investment decisions influence the investment decisions of a group of firms (*B*) connected with it, and therefore if the following holds:

$$I_B = f(I_A) \quad \textit{with} \quad \Delta I_B / \Delta I_A > 0 \tag{7.1n}$$

where *I* is the level of investments. See Higgins, 1977.
7 See Boudeville, 1964, English translation, 1966; Paelink, 1965.
8 Hansen, 1967.
9 Paelink, 1965.
10 See also Chapter 8.
11 Note that, chronologically, the growth-pole theory preceded shift-share analysis, the technique already discussed and that sought to describe the productive structure and sectoral dynamic of an area in order to explain its development.
12 See Chapter 3.
13 For a critical survey of the normative application of growth-pole theory see Parr 1999a and 1999b.
14 Hirschman, 1958.
15 An example is provided by the steelworks constructed in Taranto (Southern Italy) during the 1960s, the direct consequence of which was the closure of small crafts firms, which were forced to leave the urban area by the increases in living costs, land rents and wages caused by the external intervention.
16 On this see Holland, 1977; Massey and Meegan, 1978; Carney et al., 1980; Damette, 1980; Lipietz, 1980.

17 Lipietz, 1980.
18 On this see Holland, 1977; Carney, 1980; Damette, 1980. In particular, Damette introduced the idea of the hyper-mobility of capital.
19 For this interpretation see, e.g., Nauwelaers et al., 1988; Young et al., 1988. In the 1990s the extraordinary economic growth achieved by Ireland thanks to its ability to attract FDIs (the so-called 'Irish miracle') amply demonstrated that the relationship between FDIs and a country's growth is not always negative, and that it is more complex than the 'radical' school envisaged.
20 See Camagni, 1988.
21 For the literature on the transmission channels of the advantages of a multinational in the host area, see Blomstrom and Kokko, 1998; Blonigen, 2005; Johnson, 2006; Kugler, 2006.
22 See Rodriguez-Clare, 1996; Borensztein et al., 1998; Markusen and Venables, 1999, and Chapter 10 in this book.
23 There is a risk in this case that the host area may be a donor of knowledge without being able to receive advantages from the presence of multinationals. On this issue, see Blomström and Kokko, 1998.
24 On multinationals and technological spillovers see Blomstrom and Kokko, 1988; Cantwell, 1989; Cantwell and Iammarino, 1998; Cantwell and Piscitello, 2002.
25 On this see Holland, 1977; Massey and Meegan, 1978; Lipietz, 1980; Carney et al., 1980; Damette, 1980.
26 For the literature on technological spillovers from multinationals, see Iammarino and McCann, 2013.
27 See Iammarino and McCann, 2013, p. 61. For a critique on the a-spatial nature of the traditional determinants of location choices, see also Barba Navaretti and Venables, 2004.
28 On these studies, see Pusterla and Resmini, 2007; Casi and Resmini, 2014b.
29 See Casi and Resmini, 2014a.
30 In a celebrated study, Solow showed that more than 40 per cent of US growth between 1900 and 1949 was due to a factor (Solow's famous 'residue') that could not be explained in terms of factor growth; see Solow, 1957. In subsequent years, empirical surveys carried out at national, regional and local level showed that more than one-third of output growth in the second half of the twentieth century was due to the growth of factor productivity achieved through technological progress. It is debatable whether the assumption of constant returns made by these studies is necessary. Nevertheless, it proves useful in several respects: in the presence of increasing returns to scale, output increases more than input, and it is conceptually and empirically difficult to determine the specific contribution of technological change to economic growth. Instead, on the hypothesis of constant returns, the entire increase in input not obtained from increases in the production factors is due to technological progress. For discussion of the role of the spatial diffusion of innovation in the regional development process see, among others, McCombie, 1982.
31 Hägerstrand, 1966. For the preliminary reflections that gave rise to the model see Hägerstrand, 1952.
32 'This accounts for the fact that though most innovations are adopted at a higher rate than the economy's mean rate in cities and large towns (obvious exceptions are innovations in the agricultural sector, such as new agricultural machines), certain rural districts surrounding the innovation centre may also apply the innovation at a rate above the mean.' See Richardson, 1972, p. 313.
33 In Hägerstrand's words: 'a person becomes more and more inclined to accept an innovation the more often he comes into contact with other persons who have already accepted it'. See Hägerstrand, 1967, p. 264.
34 Griliches, 1957 and Mansfield, 1961 and 1968.
35 The linearity of the function obtained with a transformation into logarithms allows estimation of its parameters with ordinary least squares regression – an easily applicable econometric technique. In Griliches' words, 'the logistic was chosen because it is simpler to fit and in our context easier to interpret'. See Griliches, 1957, p. 503.
36 In this case, too, the determinants of the values assumed by the three parameters are easily identified by means of multiple linear regression models based on ordinary least squares.
37 See Griliches, 1957, p. 501.

38 See Capello, 1988. The same analysis has been conducted on the adoption of industrial automation technologies. See Camagni, 1985.
39 The pattern of adoption costs is explained as follows: adoption costs are at first decreasing because of the creation of local knowledge in the new technology and because of the reduction of sunk costs on the old technology, made possible by increasing depreciation of the old technology. They then increase because of the constant accumulation of knowledge in the old technology. In their turn, the revenues (net of the opportunity-cost of still using the old technology) decrease over time because of learning processes centred on the old technology. Figure 7.4 shows that if the new technological trajectory is not chosen immediately, it may never be introduced; in fact, the more the adoption is postponed, the greater the risk of lock-in to the knowledge developed for the old technology, and the more costly the transition from the old to the new technology. See Camagni and Capello, 1998. The pattern of adoption costs and revenues is also affected by the structure of the market in which the innovation is produced. See Capello et al., 1999, chaps 5–6.
40 See Davies, 1979; Brown, 1981; Stoneman, 1986. For a critical analysis of epidemic models see Hagget et al., 1977, chap. 7.
41 The logistic model used to analyse the telephone service would find it very difficult today to describe its real diffusion *ex post*. The model was developed in the mid-1980s and did not foresee the advent of the cell-phone, which has radically changed the trend in adoption of the fixed telephone service.
42 See Vernon, 1966; Hirsch, 1967; Norton and Rees, 1979.
43 See Davelaar and Nijkamp, 1990.
44 In the words of Abernathy and Utterback, the curve of the product life-cycle 'lost its tail'. See Abernathy and Utterbach, 1978.
45 As Davelaar puts it, 'the "swarming" processes are "creative" processes and not simple "carbon copy" processes of imitation'. See Davelaar, 1991, p. 29. On this statement, see also Davelaar and Nijkamp, 1990.
46 The 'regional development potential' model was first formulated by Biehl, who maintained that 'an essential requirement for a theory which claims to establish the contribution of infrastructures to regional economic development is that it must not restrict itself to the infrastructures themselves but also considers other possible determinants of regional economic development; and this is precisely the case of the regional development potential approach'. See Biehl, 1986.
47 For analysis of infrastructures as the source of endogenous growth see Barro, 1990.
48 Infrastructural endowment is estimated using two methods. The most frequent one measures public capital in monetary terms as a proportion of the total stock of capital. The method less frequently used – mainly because of the difficulty of obtaining suitable data – calculates it on the basis of 'physical' indicators of infrastructural endowment.
49 Besides infrastructural endowment, often included in the production function are further factors deemed decisive in the overall competitiveness of the local system considered. For example, Biehl emphasizes location, sectoral composition and the agglomerative structure; Ferri and Mattesini highlight the role of human capital; Fabiani and Pellegrini stress the importance of geographical factors, the environment and the sectoral structure. See Biehl, 1986; Fabiani and Pellegrini, 1997; Ferri and Mattesini, 1997. See Elhance and Lakshmanan, 1988, for the solution of some methodological problems in estimates of a quasi-production function.
50 See Aschauer, 1989 and 1990.
51 On the theoretical and empirical aspects of the subject, see, among others, Hansen, 1965a and 1965b; Eberts, 1986; Costa et al., 1987; Deno, 1988; Aschauer, 1989 and 1990; Rietveld, 1989; Munnell, 1990; Duffy-Deno and Eberts, 1991; Moomaw and Williams 1991; Hulten and Schwab, 1995; Moomaw et al., 1995; Guild, 1998.
52 See European Commission, 2004 and 2011; Camagni and Capello, 2015.
53 On this see Bruinsma et al., 1990; Vickerman, 1991.
54 See Rosenstein-Rodan, 1943, and Chapter 4 of this book.
55 Hansen, a pioneer of these studies, showed that development in Belgium varied by region according to the classes of investment. Lagging regions benefited more from increased social overhead capital (e.g. social services including health and education), while intermediate

regions benefited more from increased spending on economic overhead capital (e.g. roads and power supplies). Congested regions were less affected by changes in both types of social capital. See Hansen, 1965a and 1965b. Paci and Saddi have obtained different levels of income elasticities to infrastructures for the North and the South of Italy: 0.14 for Northern Italy and 0.20 for Southern Italy. See Paci and Saddi, 2002.

56 See Camagni, 1991; Lundvall and Johnson, 1994; Asheim, 1996; Amin and Wilkinson, 1999; Keeble and Wilkinson, 2000. For a definition of collective learning see Chapter 9.

57 On the debate concerning convergence or divergence in development rates following the advent of ICTs see Goddard and Pye, 1977; Gillespie and Williams, 1988; Hepworth and Waterson, 1988; Gillespie et al., 1989; Capello, 1994.

58 See Yilmaz and Dinc, 2002. For empirical analyses on the effects of telecommunications on regional development in China see Ding and Haynes, 2004; for Portugal, see Butler et al., 1986.

59 See Capello, 1994; Capello and Spairani, 2004.

8 Territorial competitiveness and endogenous development

Agglomeration economies

8.1 The endogenous sources of competitiveness: agglomeration economies

Throughout this book thus far, space has performed two distinct roles in models and theories: (i) the role of a physical barrier – or of a spatial friction – against economic activity, taking the form of the physical distance between input and output markets conceptualized by models as a generic transportation cost;[1] (ii) that of a 'physical container' of development, a simple geographical area often associated with the administrative region by aggregate macroeconomic theories – but also with smaller local areas (simple geographic agglomerations within a region, as envisaged by the more microeconomic theories examined in the previous chapter). In both cases, space plays no part in determining the development path of a local economy. The same economic logic explains the development of regions, metropolitan areas or, more generally, densely populated industrial areas. The export-base theory can be applied just as well to a region as to a country, with no change in the logic of its underlying reasoning. The Harrod–Domar model, too, and likewise the neoclassical growth models, fit both regional cases and national ones, which testifies to their aspatiality.

In this (and in the next) chapter, a radical change in the conceptualization of space gives it a very different role in development. No longer a simple geographical container, space is conceived as an economic resource, as an independent production factor. It is the generator of static and dynamic advantages for firms, and a key determinant of a local production system's competitiveness. According to the theories examined in this (and in the next) chapter, space is a source of increasing returns, and of positive externalities taking the form of agglomeration and localization economies. Higher growth rates are achieved by local production systems where increasing returns act upon local productive efficiency to reduce production and transaction costs, enhance the efficiency of the production factors and increase innovative capacity. Regional development consequently depends upon the efficiency of a concentrated territorial organization of production, not on the availability of economic resources or their more efficient spatial allocation.

This new conception of space has several implications. Space can only be *diversified* space in which it is easy to distinguish (even internally to a region) the uneven distribution of activities. Development comes about selectively in areas where the concentrated organization of production exerts its positive effects on the parameters of static and dynamic efficiency. At the same time, space is *relational*, in that the economic and social relations that arise in an area perform crucial functions in various respects. They ensure the smoother operation of market mechanisms, more efficient and less costly production processes, the accumulation of knowledge in the local market and a more rapid pace of innovation – all of which are factors that foster local development.

Second, on adopting this new notion of space it is no longer possible to treat development as exogenous in origin. Development is now by definition *endogenous*. It is fundamentally dependent on a concentrated organization of the territory, embedded in which is a socio-economic and cultural system whose components determine the success of the local economy: entrepreneurial ability, local production factors (labour and capital), relational skills of local actors generating cumulative knowledge-acquisition – and, moreover, a decision-making capacity that enables local economic and social actors to guide the development process, support it when undergoing change and innovation, and enrich it with the external information and knowledge required to harness it to the general process of growth, and to the social, technological and cultural transformation of the world economy. The theories presented in this chapter accordingly endeavour to identify the genetic local conditions that determine the competitiveness of a local production system and ensure its persistence over time. They seek out the local factors that enable areas, and the firms located in them, to produce goods demanded internationally with an (absolute) competitive advantage, to maintain that advantage over time by innovating, and to attract new resources from outside.

As we shall see, theories of local endogenous development divide into two broad strands. On the one hand neo-Marshallian inquiry, which views local growth as resulting from externalities acting upon the static efficiency of firms, has been expanding and consolidating for years. On the other, the neo-Schumpeterian literature, which has arisen more recently, interprets development as resulting from the impact of local externalities on the innovative capacity of firms.

The logical leap of interpreting space as an active factor in development forcefully imposed itself upon the history of economic thought in the early 1970s, when unprecedented patterns of local development in Italy surprised theoreticians by resisting explanation based on conventional models. During the early 1970s, the sudden and rapid growth achieved by certain Italian regions – those of the Northeast and the Centre in particular[2] – when the country's industrialized areas[3] were showing evident signs of economic crisis could be explained neither by a neoclassical paradigm of interregional mobility of production factors (which greatly decreased in those years), nor by a paradigm centred on large firm efficiency (à la Perroux), nor by a Keynesian paradigm of development driven by external demand.

Numerous neo-Marshallian theorists around the world pursued very similar lines of theoretical inquiry during the 1970s and 1980s (still today there is no lack of theory on the matter): Walter Stöhr developed the concept of 'bottom-up development', Enrico Ciciotti and Reinhart Wettmann that of 'indigenous potential', Bengt Johannison of 'local context', Bernardo Secchi and Gioacchino Garofoli of 'system areas', and Claude Courlet-Bernard Pecqueur and Bernard Ganne of 'localized industrial system'.[4] But the first systematic theory of endogenous development was produced in Italy by Giacomo Becattini with his seminal study on the 'Marshallian industrial district' published in the mid-1970s.[5] The theory of the industrial district – which originated in the work of the great neoclassical economist Alfred Marshall[6] – was the first to conceptualize external economies (of agglomeration) as sources of territorial competitiveness. It did so with a model in which the economic aspects of development are reinforced by a socio-cultural system that fuels increasing returns and self-reinforcing mechanisms of development (Section 8.2).

In what follows, no criticism will be made of the qualitative nature of the theories examined, even though this is an aspect to which orthodox economists have often objected. On the contrary, it will be argued that these theories have enriched economic analysis by identifying the intangible elements (knowledge, learning, relationality, social capital) which come together to form local competitiveness. Far from being of scant economic significance, the chapter maintains, these elements should be valued and appreciated for their contribution to knowledge on local development processes.

Finally, when space is viewed as generating advantages for firms, and therefore as an active component in the development process, scholars of local development shift their attention to the role of the urban space (the city) as the place where agglomeration economies are generated – be these localization or urbanization economies – and therefore as the place where the economic development of the entire region is rooted and structured. Hence, as the models of Christaller and Lösch show, the existence of an advanced and efficient city, and of an urban system organized into a network of vertical and horizontal relationships reflecting an efficient division of labour, may determine the success and development of a region (Section 8.3).

In this stream of thought, the theory of the optimal city size represents the first step to identify the urban size that allows the full exploitation of urbanization economies. As we will see, further reflections have associated urban efficiency increases not only with urban size, but with an increase in the endowment of high-quality urban production factors, high-level functions, high-quality services offered, in a dynamic perspective. In this way, these theories are able to explain not only the physical growth of cities, but also their structural dynamics, and to identify under which conditions (not only related to physical size), cities are able to grow, and why (Section 8.4).

8.2 The Marshallian industrial district

8.2.1 The genetic conditions of an industrial district

The 1970s witnessed the miracle of the 'Third Italy'[7]: the northeastern and central regions of the country that recorded surprisingly high growth rates in a period of general economic crisis provoked by severely adverse macroeconomic conditions (oil shocks, inflation, unemployment, stagnating consumption and investment, devaluation of the national currency, the lira). The miracle was interpreted first as a short-term phenomenon due to industrial conflict in the large companies of the industrialized areas, and then as the territorial effect of production decentralization (and therefore again as a process dependent on the centre). Finally, when empirical analyses demonstrated the autonomy and originality of the Third Italy development model, it was hailed as a new form of capitalist economic development.[8]

Numerous case studies on 'success stories' in non-metropolitan, diffused development prompted analysis of the factors responsible for the economic success of Third Italy areas, the distinctive features of which were close concentrations of small-sized firms and a form of entrepreneurship that seemingly stemmed from the historical structure of local agricultural systems. These studies on success factors were flanked and enriched by surveys on the flexibility of local labour markets – which permitted part-time work in the agro-food industry and rapid and easy labour mobility among firms – and by sociological analyses of a cultural, social and political homogeneity that underpinned a long tradition of co-operation in agriculture and trade.[9]

The seminal theory which derived from the empirical results of these studies – and which subsequently bred a large and ramified body of literature – was that of the *Marshallian industrial district.* This term denotes a local area with a strong concentration of small and medium-sized firms, each specialized in one or a few phases of the production process (or activities subsidiary to it) serving the needs of the area's principal sector.

A spatial concentration of small firms is the first genetic element necessary, though not sufficient, for an area to be an industrial district. A district's economic-productive organization is rooted in a social and cultural system of shared values that penetrates the market and structures its workings.[10] It is this relationship between economy and social structure that drives development: the symbiosis between market and society produces the synergy, co-operation and interaction that give rise to the increasing returns and location advantages of district firms. The genetic conditions that must be in place for a geographical area to be an industrial district are the following:

- spatial proximity, or geographical contiguity among firms;
- social proximity: a system of institutions, codes and rules shared by the entire community regulates the market; this system induces firms to co-operate and, in general, to resort to the local market when activities, phases or services prove too costly for them to produce internally;
- a concentration of small firms, the main features of which are productive flexibility and rapid adjustment to market volatility;
- marked industrial specialization of the area, in which all phases of the production chain are undertaken: from design of the product, through production of all intermediate goods necessary for the product's manufacture, to its marketing world-wide.

8.2.2 *District economies*

The combined presence of the above economic-territorial conditions gives rise to the competitive advantages which make firms successful. Put in purely economic terms, these conditions generate increasing returns in the form of agglomeration economies; or more precisely, localization economies, or again 'district economies', which are the advantages (in terms of lower costs or increased productive efficiency) deriving to firms from proximity with other firms operating in the same sector.[11] They enable small firms to overcome the obstacles due to their small size, without their having to forgo the advantages which that same small size gives them.[12]

District economies derive from the factors now described (and set out in Table 8.1):

- *lower production costs.* The existence in districts of numerous, highly specialized local suppliers reduces the costs of transporting intermediate goods. A local labour market with high levels of elasticity – in the sense of rapid and smooth adjustment of the labour force to quantitative changes in demand – also reduces production costs. At the same time, easy recourse to the labour market, made possible by shared social rules and social sanctions on opportunistic behaviour, enables firms to draw on external labour and, above all, to outsource more complex and costly production phases. These are further factors which reduce production costs;

Table 8.1 A district's genetic conditions and advantages: a taxonomy

Genetic conditions (sources) / *Advantages (effects)*	*Spatial proximity*	*Social and cultural proximity*	*Concentration of small firms*	*Industrial specialization*
Reduction of production costs	Lower transport costs for intermediate goods	System of local agents Recourse to external labour (home work) Outsourcing of production phases	Production flexibility	Availability of skilled labour Inter-firm division of labour
Reduction of transaction costs	Labour demand/ supply match Broad local market upstream and downstream	Networks of interpersonal relations System of shared rules and institutions Common code of behaviour Sense of belonging Explicit capacity for inter-actor cooperation Informal contracts	Flexible, non-bureaucratized relationships among firms	Adequate technical knowledge for choice of suppliers
Increase in the efficiency of the production factors	Existence of a critical mass for specialized and infrastructural services Broad market for specialized inputs	Widespread industrial culture Mobility of tacit information Widespread entrepreneurial expertise	Flexibility in the quantity and quality of inputs to the production process	Information services for specialization sectors
Increase in innovative capacity (dynamic efficiency)	Localized accumulation of knowledge	Socialization to the risk associated with innovative activity Accumulation of shared knowledge	Competition-driven stimulus to innovation	Accumulation of specific knowledge

- *reduced transaction costs* (i.e. the costs of economic transactions). The geographical proximity that characterizes a district market facilitates the matching of labour supply and demand through a close-knit local information network. But even more important is social proximity: the system of shared rules of behaviour, the code of conduct internalized through socialization and the sense of belonging to a community partly inherited from the farm-management system characterizing the history of industrial districts.[13] Social proximity generates governance mechanisms which discourage opportunism or dishonesty in transactions; it thus substantially reduces costs and recourse to the market. The sense of belonging to a specific community and the social identity pervade local society and underpin trust relations, which foster inter-firm co-operation in the form of informal, non-bureaucratized and flexible contracts; to use Becattini's apt description, the Marshallian industrial district is a 'localized thickening of inter-industrial relations'.[14] Finally, the strict specialization of firms gives them the technical expertise they need to assess the quality of the large number of suppliers in the area efficiently. Consequently, recourse to the market is again less costly than it is in a differentiated production system;

- *increased efficiency of the production factors.* External economies do not have positive effects on costs alone. Production resources remaining equal, the system of shared social values, the spatial concentration of specialized firms, and their small size, act upon the production capacity of firms to increase the efficiency of the production factors. Upstream and downstream from the production process, the presence of a critical mass of firms generates a series of services that enable better use to be made of local production, and that also have synergic effects on the market image of the local economy. Social proximity engenders what Marshall called 'industrial atmosphere', by which he meant an industrial culture consisting in the indivisible 'intangible assets' of the production system as a whole: an entrepreneurial mentality, a spirit of co-operation, local technical knowledge about the production cycle and the socialization of knowledge make firms – other conditions remaining equal – more productive. Specialization of firms in different stages of the *filière*, the vertical and horizontal division of labour and close purchase/sale relations among firms give rise to a greater overall efficiency manifest in increased earnings and profits (and the area's greater attractiveness for the location or creation of new firms);
- *increased dynamic efficiency*, in the sense of the innovative capacity possessed by firms operating in the district. Industrial district theory has adopted Marshall's view of a specialized area – 'where so great are the advantages which people following the same skilled trade get from near neighbourhood to one another. The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously'[15] – to underline the decisive importance of locally accumulated knowledge for the innovative capacity of firms. This aspect has been subsequently reprised and amplified by the theories examined in the next chapter.

The high level of collective efficiency achieved by firms in industrial districts is therefore explained mainly by district economies. Accordingly, these theories conceive the territory, with its networks of inter-firm and social relations, as the source of economic growth – that is, as a factor that actively contributes to determining the development path and productive capacity.

8.2.3 *Beyond district economies*

Although district economies are the most evident economic advantages springing from the co-presence in an area of small firms operating within the same sector, they are generated and reinforced by factors present in the economic and social context.

A first factor is the inextricable interweaving of economic, geographical and social elements. It has often been stressed, regarding the concept of 'industrial district', that a simple clustering of small firms in a particular area does not in itself constitute an industrial district. Social proximity – defined as a shared code of behaviour and a set of common values penalizing opportunistic behaviour – is a typical feature of a district. Social proximity penetrates the market, structures it around clearly defined rules, and gives it efficiency. The strength of this organizational model is the close relationship between the economy and the social structure. In this regard, analysts have formulated the notion of a 'community market' – this being the level of transactions governance that lies between the market and the community – because the information that transactions require resides both in the prices system and in an implicit code of behaviour that economic agents internalize through socialization.[16]

A second factor enhancing the efficiency of district firms is the integration between co-operation and competition; indeed, striking an appropriate balance between these two processes determines the survival itself of the district organizational model.[17] Notwithstanding the impression that might be gained from industrial district theory's constant emphasis on co-operation, the firms operating in a district engage in aggressive competition with each other, being obliged to do so by the ready substitutability of the goods which they produce. Competition is the driving force behind district firms obliged to maintain their goods at high-quality levels and to innovate their production techniques (even if only by imitation). Simultaneously, forms of explicit co-operation characterize a market regulated by social norms and sanctions that punish opportunistic behaviour; in 'repeated games' (i.e. transactions which take place several times sequentially between the same economic actors) 'reputation' is an intangible asset that ensures a firm's survival in the market.[18]

Finally, the presence of a governance structure (local agents and institutions) that buttresses the transactions regulation system ensures the efficient operation of the 'community market' by explicitly supporting forms of competition and co-operation. In order that competition does not degenerate into aggression damaging to firms, the district's industry associations impose price controls in the form of agreed and indicative tariffs, modifiable according to the manufacturing process. The risk of information asymmetry is thus abated, and transaction costs are reduced. The local governance system likewise prevents co-operation from degenerating into financial agreements or protectionist cartels that cancel out the positive effects of competition.[19]

Forty years since its first conceptualization, the Marshallian notion of the industrial district is still largely unsurpassed as a tool for the study of local systems. Although its appeal may have been diminished by a plethora of redundant and repetitive empirical studies, the Italian industrial district model has recently been used by development economists to interpret the small-firm systems now arising in the developing countries.[20] However, as often happens when conclusions drawn from specific empirical cases are applied to other contexts, analyses conducted in Latin America suggest that there is a substantial difference between the Italian industrial district and the clusters of small firms now emerging in Brazil, Mexico and Argentina – especially because of the lesser degree of co-operation in those countries.[21]

Identifying the determinants of the success or crisis of industrial districts – the 'catastrophic' patterns in their evolution – is still today the most interesting and fruitful aspect of research in this field, since the dynamics of industrial districts still lack sound conceptual interpretation.

8.2.4 *Some critical remarks*

As said, industrial district theory has had the outstanding merit of being the first theory to give space an active role in economic development, thereby enriching the concept of agglomeration economies with social, psychological and cultural dimensions.

A second merit of the theory is that it highlights the endogenous factors in development: entrepreneurship, production flexibility, district economies, and the presence of a social and cultural context and an institutional structure able to catalyse 'indigenous potential'. The theory of bottom-up development indirectly entails the immobility of certain factors, such as skilled labour, specific knowledge and expertise, but also of intangible elements like a socio-cultural system that supports transaction and market mechanisms. Thus explained is the selectivity of territorial development, and the difficulty of generating development processes artificially.

Industrial district theory has had the further virtues of producing a conceptual model able to explain what was 'inexplicable' at the time of the theory's formulation, and of opening analysis of regional development to consideration of genuinely territorial elements.

However, having acknowledged these various virtues of district theory, mention should also be made of weaknesses in its logical-conceptual structure. First, as often happens at the moment of a 'catastrophic' break with already-existing theories, the approach has an evident tendency to emphasize the novel and to undervalue the findings of previous theories. It places pronounced emphasis on endogenous aspects and tends *entirely to ignore the exogenous and objective elements that accompany a development path*, in particular the macroeconomic and macro-territorial conditions that act upon the economies of individual areas. The influence exerted by these elements on the birth, development and crisis of district areas is undeniable. In the early 1970s, for example, the manufacturing and exporting difficulties of the large industrial areas in Italy led to general medium-period exchange rate weakness, and to a decrease in the cost of labour (expressed in international currency), which worked mainly to the advantage of the NEC regions because of their specialization in labour-intensive 'tradeable' manufactures with greater elasticity to price.[22] The same situation arose in 1992, when the general weakness of the Italian economy, together with instability in the European financial markets, induced the Italian economic policy authorities to heavily devalue the lira outside the 'monetary snake' bands, taking Italy out of the European monetary system.

It is therefore necessary to analyse not only the internal elements of dynamism and entrepreneurial ability present in individual regions but also the interdependence among regional economic systems and the feedbacks that occur over time.[23] A useful way to consider these connections is to apply the concept of a region's 'relative locational advantage'. This is measured by means of two indicators – productivity defined in the broad sense as the overall efficiency of the local social-productive system, and the cost of labour, also defined in the broad sense as the cost of 'labour force reproduction' – which are used to determine all the socio-environmental factors that affect the real purchasing power of wages in each region. The relative locational advantages of the three Italian macro-regions highlight very clearly the favourable conditions enjoyed by the NEC regions during the 1970s, and the contemporaneous loss of competitiveness by the Northwest (Figure 8.1a).[24]

Finally, comparison between productivity and cost of labour evidences the economic revival of the 'central' regions in the 1980s and – more interestingly – the crisis of relative

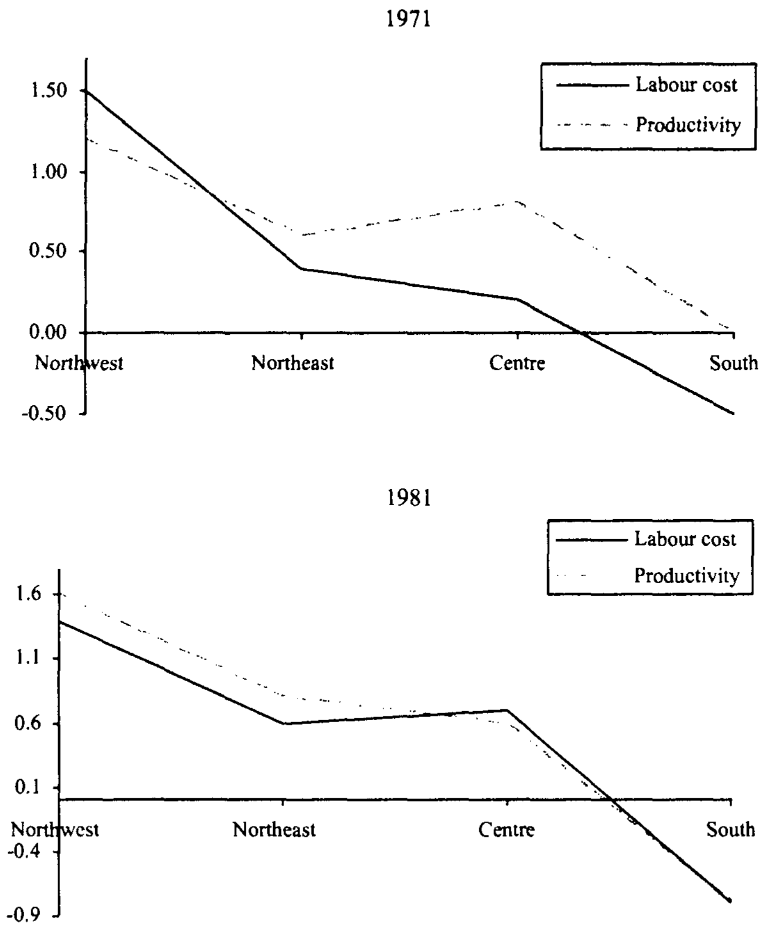


Figure 8.1 Regional locational advantages in Italy, 1971 and 1981
Source: Camagni and Capello (1990)

competitiveness that hit some regions, especially those of central Italy: a crisis that was neither foreseen nor explained by industrial district theory (Figure 8.1b).

This last point introduces a second shortcoming of district theory: *its static theoretical framework and its tendency towards ex-post descriptivism of spatial phenomena.* The theory is able to quantify the relative advantage of the Third Italy entirely satisfactorily. But it is less successful in identifying the determinants of the growth and dynamics of the Third Italy areas, their ability to respond to increasing worldwide competition, rapid technological change, negative feedbacks in the form of manpower shortages, increased costs of labour and production factors, and the physical and infrastructural congestion generated by economic success. These elements may cancel out some of the locational advantages on which the success of these areas was initially based.

A further weakness (one particularly apparent in subsequent schools of thought) is the theory's excessive emphasis on specialization and flexibility. Characteristic of the small firm and of a 'post-Fordist' model of production organization,[25] flexibility is today also a distinctive feature of large firms if they utilize modern flexible production technologies (CAD/CAM) and the new forms of production organization (just-in-time). Likewise, external economies, which lend themselves well to interpretation of the increasing returns achieved by district firms, also arise and develop in metropolitan areas where large firms are located. Once again, therefore, consideration of subjective, endogenous and local factors alone fails to account for the competitiveness of more recently industrialized areas, and even less for their ability to maintain this competitiveness over time.

Finally, the importance, richness and strength of industrial district theory are evident. But equally obvious is the extreme difficulty of measuring the economic advantages that it theorizes. This difficulty may in part explain the large number of qualitative studies produced with the purpose of determining the presence of genetic elements and success factors in individual districts. The problem with this approach is that it may lapse into mere anecdotal description, thus clogging the literature without adding a great deal to the conceptual framework already developed. Consequently to be welcomed are those studies which have sought to make quantitative measurement – using statistical tools and cross-section econometric analyses – of a firm's external economies, using either municipal-level data (for inter-district analyses) or data disaggregated at the level of the individual district firm (for intra-district analyses). These studies have the indubitable merit of removing the anecdotal content from empirical analysis and of furnishing quantitative measures of phenomena difficult to gauge.[26]

We may therefore conclude that the theoretical/conceptual contribution of industrial district theory has been very fruitful. Nevertheless, still today, 40 years after its first formulation, the theory is better suited to describing spatial phenomena than to interpreting their dynamics.

8.3 The urban structure and regional development

Especial emphasis has recently been given to the idea (also mooted by authors in the past) that an efficient, modern and advanced urban structure able to grow in balanced manner, and pursuing goals of equity, competitiveness and sustainability, determines a region's economic success.[27]

On this view, regional development springs substantially from the balanced growth of individual cities – these being sources of increasing returns for the people who live and work in them – and of the city system in which each individual city is embedded.[28] Each individual city must therefore find a growth path that reinforces the elements from which its static and dynamic efficiency derive. Such growth, moreover, is enhanced when the individual city is embedded in a system of cities in its turn able to develop in a harmonious and balanced manner, with an even mix of well-connected and integrated urban centres.

In economic terms, these assertions are borne out by numerous theoretical findings. Maintaining the distinction between analysis of the efficiency of the individual city and that of the city system, the theoretical underpinnings of these statements emerge very clearly, from both the static and dynamic points of view (Table 8.2).

First, a city is a spatial cluster of productive and residential activities. The concentration of activities, the density of the contacts that develop within them, and easy access to advanced information and knowledge are evident advantages springing from an

Table 8.2 Sources of static and dynamic urban economies

Spatial dimension / *Temporal dimension*	*Cities*	*City systems*
Static	The city as a cluster Presence of pronounced sectoral mix Density of proximity contacts Reduction of transaction costs	The city system as an efficient production system Balanced urban structure Efficient networks of interconnection Specialization economies Network externalities from complementary activities
Dynamic	The city as a *milieu* Reducer of uncertainty: – information transcoder – *ex-ante* coordinator of collective action – substrate for collective learning	The city system as an efficient system of innovative cooperation Generator of dynamic network externalities

urban location. They are opposed to localization (or district) economies because they stem from the presence of a mix of sectors.[29] A broad and diversified labour market, the availability of typically central and urban services (advanced, financial, insurance, managerial, etc.), a supply of managerial and executive skills, communication and information structures, characterize an urban location, and they affect the factor productivity of the firms situated therein.

But the importance of the city as the engine of development also resides in its ability to generate dynamic economies and become the preferred location for new high-tech companies, and, in general, for innovative functions. Besides the well-known role attributed in the 1950s and 1960s to cities as 'incubators' of the new – or as the 'nurseries' of small firms – and supported by empirical data,[30] a new interpretative factor has recently been adduced in explanation of the dynamic efficiency of cities. Like the *milieu*, the city performs the crucial function of reducing dynamic uncertainty and creating processes of collective learning to the advantage of local actors. The urban environment comprises the co-operation, synergy and relational proximity that in *milieu* theory determine the dynamic efficiency of firms. Shared values, common codes of behaviour, a sense of belonging and mutual trust are features which the urban system shares with the *milieu*. And they account for the ability of the urban system, as well, to reduce uncertainty and generate processes of the socialization of knowledge and collective learning.[31]

However, it is not solely in the efficiency of the individual city that one grasps the effect of the urban system on regional economic development. As shown by the first theorists of general spatial equilibrium and the structure of city systems, Christaller and Lösch, a well-balanced urban system with an even mix of large, medium and small cities and towns, endowed with efficient transport networks, is the ideal territorial system in terms of efficiency and well-being. A city system of this kind, in fact, makes it possible to exploit the geographical, historical and cultural specificities of each individual city to provide a broad and diversified range of possible locations for firms and households, and to avoid the hyper-concentration of production and residential activities in a few large-sized cities, where the advantages of scale economies are easily eroded by the high social and environmental costs associated with large urban size.

Again in regard to city systems, an important theoretical contribution to understanding of their dynamic efficiency has been made by the theory of city networks discussed earlier.[32] A network organization of urban centres, hierarchically ordered or of similar size, gives rise to evident advantages associated *inter alia* with the innovative co-operation necessary for the undertaking of innovative projects (infrastructures, or service provision, or even large-scale urban planning).

The sources of an urban system's static and dynamic efficiency are evident in the theories just discussed. But it should be pointed out that the theory that associates regional development with the dynamism of the urban structure rightly gives great importance to social aspects (cohesion) and environmental ones (sustainability) as well. These too are sources of static and dynamic advantages (disadvantages), and they are decisive factors in the balanced growth of an urban system in that they act upon the location choices of firms, and upon the ability of a city to retain businesses already operating in it, and to attract new ones. In the age of globalization, success in the fierce competition among the large European cities for direct foreign investments hinges on their economic efficiency and environmental and social quality.

8.4 Agglomeration economies: size, productivity and urban growth

8.4.1 The optimal city size

As mentioned before, when space assumes an active role in the development process, agglomeration economies return to the centre of the interpretation of regional and urban growth patterns.

Cities are by definition sources of agglomeration economies for firms and individuals in the form of urbanization economies. In cities, manufacturing activities receive advantages from the presence of fixed social capital (transport infrastructure, advanced telecommunications networks and services), of advanced services for firms, of a large and diversified market for intermediate and final goods; advantages that increase with the size of the city. Individuals obtain advantages from the presence of public services (education, health, transport and social infrastructure in general), and of private services (cultural and recreational activities), but they also benefit from a 'variety' of options in diverse fields: work and residential opportunities, leisure time, lifestyles in general. Also in this case, the advantages increase with city size. Thanks to the presence of these externalities, production factors like capital and labour register a higher productivity, for the same K/L ratio, in large cities compared with small ones (Figure 8.2).

For this reason, the presence of cities able to maximize static efficiency, to exploit their agglomeration economies and to grow along virtuous and cumulative growth paths is of great importance for the dynamics of the region itself. In its theoretical justification of the European cohesion policies, the European Union has always been concerned about the trade-off between policies in support of weak areas (equity aim) and policies able to exploit the efficiency of strong areas (in general, large urban regions) (efficiency aim) leading to a higher wealth, and consequently to the creation of a larger amount of resources to be distributed to weak areas, with respect to a pure equity policy.[33]

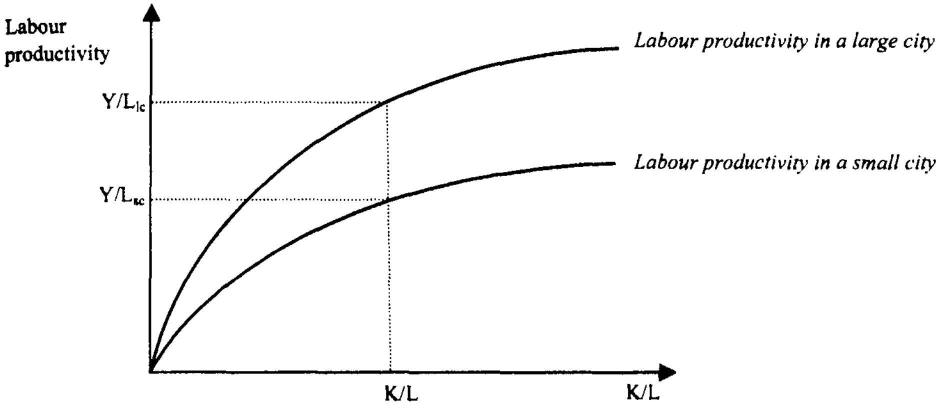


Figure 8.2 Labour productivity: static urban agglomeration economies

In the 1970s, William Alonso formulated the 'optimal city size theory', which envisages the size above which an increase in the physical dimension of the city decreases the advantages of agglomeration. The theory states that the indivisibility and synergy mechanisms that are at the basis of economies of scale in cities apply up to a certain urban size, after which diseconomies of scale due to congestion effects take place and decrease the average advantage curve of an urban location. Defined by Alonso as the land rent costs associated with urban size, average location costs, in their turn, start to increase above a certain threshold. Prohibitive urban land costs and, in a modern version, increasing environmental costs are all elements that explain the U-shape of the average location costs; and the city, like every intensively exploited economic resource, shows decreasing returns above a certain threshold level.

According to this theory, therefore, the average location benefits and costs curves, which represent the benefits and the costs of people already located in the city, respectively assume an inverted U-shaped form and a U-shaped one; the former increase and then decrease, while the latter decrease and then increase. The same applies to the marginal location benefits and costs curves; these represent the location advantages and costs for a person choosing the city as his/her new location.

On the basis of these curves, it is possible to identify different 'critical' sizes of the city, among which the most interesting are (Figure 8.3):

- a minimum size for the existence of the city (A), when the average costs equal the average benefits;
- an optimal size for people already living in the city, when the distance between the per-capita location benefits and costs are maximum (B);
- an optimal size for the entire community (C), when the marginal location benefits equal marginal location costs, so that the advantages of a marginal expansion are perfectly compensated by the new costs; it therefore represents the optimal size in a national planning perspective, and not only for the single city;
- a maximum size of the city (D), after which per-capita location benefits are smaller than per-capita costs.[34]

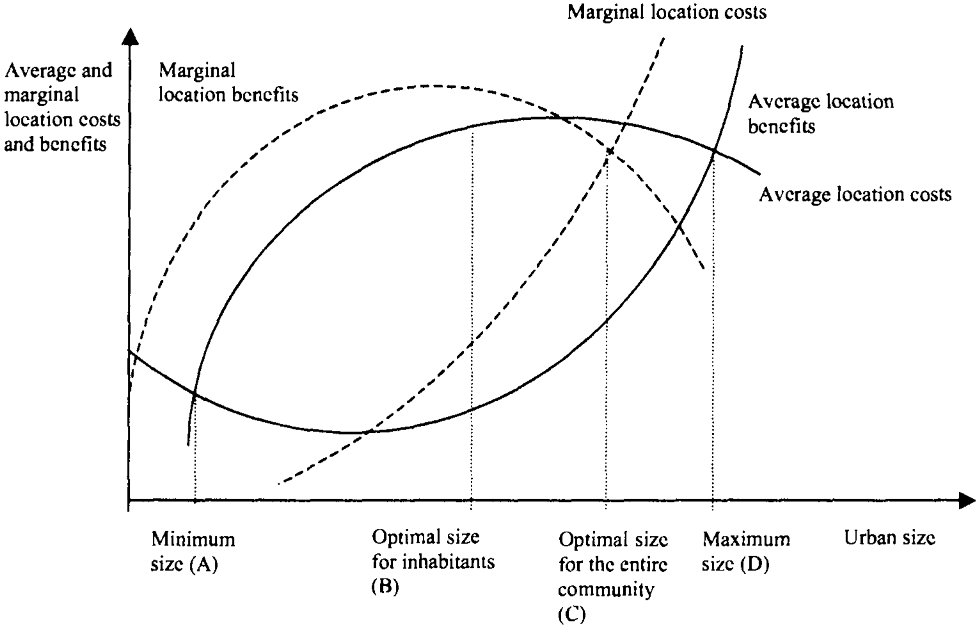


Figure 8.3 Average and marginal location costs and benefits: the critical sizes of the city
Source: Richardson (1978)

On the basis of different methods, each of them with intrinsic limits, a large number of empirical analyses have been conducted to test both the existence of agglomeration economies in an urban environment, and the shape of the benefits and costs curves, the purpose being to determine an optimal city size. Three main methods have been used, namely:[35]

a) estimation of an aggregate urban production function, which allows verification of the existence of economies of scale through the presence of a multiplicative parameter linked to urban size.[36] The underlying assumption is that all cities have the same production function; a hypothesis inevitably at odds with the reality, where cities are specialized in different functions, at different complexity levels;
b) estimation of a sectoral aggregate production function, which allows verification of the existence of higher factor productivity in industries localized in cities with greater industrial concentration.[37] This method has been conceived in order to overcome the limits encountered by the assumption that cities all have the same production function; this method is able to estimate economies of scale at the industry level (defined by Hoover as localization economies generated internally to an industry and externally to the firm).[38] However, also this method has limitations because it is unable to measure the advantages that stem from the mix of productive activities and industries that are present in large cities with respect to small ones; a mix generally formed by service activities, which are not contained in these kinds of analyses due to the well-known problem of measuring the output of service industries;

c) direct analysis of the difference in income and wages, after controlling for the different living costs between large and small cities. On this logic, large cities should show higher real wages, as an effect of higher productivity. Results confirm this expected outcome, although higher wages could also be interpreted as higher monetary compensations for the disamenities (high land values, social and environmental problems) that workers have to face in metropolises, rather than higher remunerations of higher productivity levels.[39]

Many criticisms have been made of the neoclassical approach to optimal city size theory. They include the observations that:

- cities are different from one another. They are characterized by different functions, and they perform different specializations. The use of the same urban production function for all cities in econometric analyses estimating optimal city size is extremely restrictive. In the words of Richardson: 'we may expect the efficient range of city sizes to vary, possibly dramatically, according to the functions and the structure of the cities in question;[40]
- if cities are different from one another, the optimal city size may also be different, depending on the specific characteristics. Richardson elegantly compares the 'optimal city size' theory with the theory of the behaviour of firms. We would never expect the optimal position for each and every firm to occur at the same level of output, so why should we expect the optimal point in each city to be located at the same population level?
- cities exist in an interurban environment, which influences their efficiency. In fact, it may happen that a small city efficiently linked to a well-interconnected metropolitan and regional system, with a clear division of labour not only vertical (hierarchical) but also horizontal (functional) among centres, is more efficient than a similar isolated city. The optimal city-size theory, on the contrary, does not consider the spatial context in which cities operate;
- cities generate a large variety of externalities as a result of the qualitative characteristics of the urban production environment. Already in 1961, Benjamin Chinitz expressed some doubts about the assumption that urban factor productivity depends mainly on the physical size of cities. He emphasized, on the contrary, the importance of a diversified and competitive urban production system as a source of urban productivity. Such a system is able to provide a far larger variety of externalities for small firms than an oligopolistic and specialized urban structure.[41]

These criticisms, linked to the lack in the real world of a performance of cities coherent with the theory of the optimal city size, have over time induced interpretation of the existence of agglomeration economies from different perspectives, in an attempt to overcome the idea that they merely depend on urban size.

8.4.2 *Beyond urban size: the SOUDY model*

In the mid-1980s a new model, named SOUDY (supply-oriented urban dynamics), was conceived with the aim of superseding the assumption of the 'optimal city size theory' that cities are all similar. It introduced into the conceptual reasoning the assumption that cities are of different types defined on the basis of the range of functions that they host.[42]

As shown in Figure 8.4, for each economic function and an urban range associated with it, the model assumes an average (aggregate) benefit curve that increases for higher-order functions due to a) growing entry barriers, b) decreasing elasticity of demand that allows extra profits to be gained in all market conditions, and c) increasing possibility of obtaining monopolistic revenues due to the use of scarce, qualified factors.

Higher-order functions are characterized by higher thresholds for the level of appearance in the city (in terms of urban population) (d_1, d_2, d_3 . . .). Once a location cost curve with a traditional U-shaped form as suggested by Alonso is added in the figure, one can define a minimum and a maximum efficient city size (d_1–d_3 for the function – and centre – of rank 1; d_2–d_4 for the function – and centre – of rank 2 . . . in Figure 8.4) obtained as the difference between the average location benefit and the average location costs.

The model assumes that an 'efficient' city-size interval exists separately for each hierarchical rank, and that it is associated with rank-specific economic functions. In other words, for each economic function characterized by a specific demand threshold and a minimum production size, a minimum and a maximum city size exists beyond which urban location diseconomies overcome production benefits typical of that function.

As each centre grows, approaching the maximum size compatible with its rank ('constrained dynamics'), it enters an instability area (e.g. in d_2–d_3 in Figure 8.4) where it becomes a potentially suitable location for higher-order functions, thanks to the achievement of a critical demand size for them. In dynamic terms, each city's long-term growth possibilities depend on its ability to move to higher urban ranks, developing or attracting new and higher-order functions ('structural dynamics'). This 'jump' is

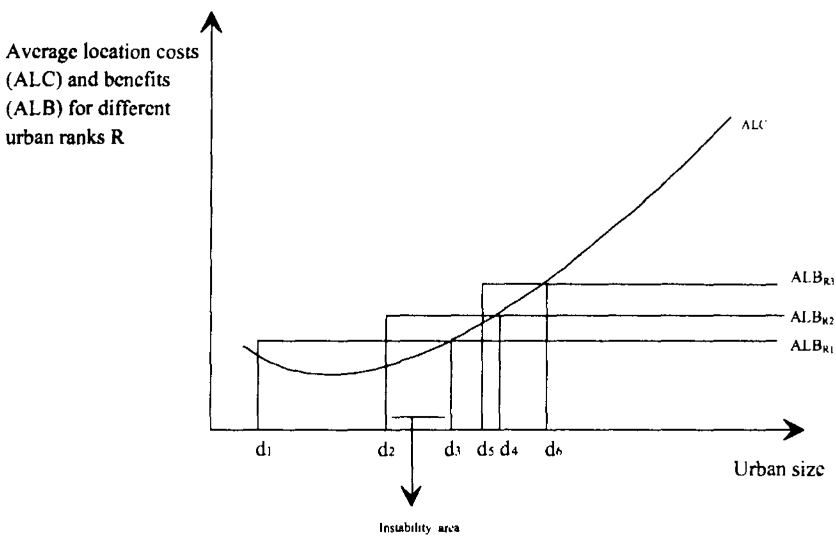


Figure 8.4 Efficient urban size for different urban functions
Source: adapted from Camagni et al. (1986)

not attained mechanically: it represents a true urban innovation and is treated as a stochastic process in the dynamic model. The city can stop growing if it does not innovate, and it can go on growing if it innovates in the function that it hosts.

With its logic, the SOUDY model is able to take into consideration many aspects that are not analysed in the 'optimal city size theory', and to interpret the real world more deeply, by suggesting:

- the need to replace the 'optimal size' by an 'interval' within which the city size is 'efficient',[43] i.e. where average production benefits exceed average location costs;
- the need to allow for different 'efficient' urban intervals according to the functions actually performed by the cities;
- the possibility of decoupling urban ranks from urban size. Differently from Christaller's approach, two cities of the same size (for example, size d_2 in Figure 8.4) can belong to two ranks (1 and 2 in the example), depending on their capacity to attract/develop higher functions.[44]

Thanks to its logic, the SOUDY model is able to interpret urban dynamics both as a process of a city's physical growth and as a process of structural evolution that characterizes the city in the long run. However, it was not until the 2000s that agglomeration economies returned to the centre of the explanation of urban dynamics, and the SOUDY model was recognized as an original and forward-looking model of urban dynamics.

8.4.3 Agglomeration economies and space: the geographical approach

The interpretation of agglomeration economies as sources of urban dynamics returned to the scene when the relatively better growth rates of smaller cities with respect to larger ones, identified at the end of the 2000s, could not find plausible theoretical explanations. In fact, when one assumes that larger cities are more productive and efficient (and therefore expects them to be more attractive and more dynamic[45]), the relatively higher performance of small cities with respect to large ones does not receive any explanation. When instead the higher performance of small and medium-sized cities is interpreted through the presence of diseconomies of scale in large cities, as explained by the 'optimal city-size theory', further doubts arise as to the lack of a theoretical explanation of why, at a specific moment in time, large cities enter a phase of decreasing returns.

One way to resolve this apparent contradiction between theory and reality is proposed by those scholars interested in the geographical foundations of agglomeration theories,[46] who build upon and enrich the concept of 'borrowed size' developed by Alonso: '*a small city or a metropolitan area exhibits some of the characteristics of a larger one if it is near other population concentrations*'.[47] Behind this statement lies the claim that smaller places can 'borrow' some of the agglomeration benefits of their neighbours while avoiding agglomeration costs.[48]

This approach highlights the fact that urban agglomeration effects are not necessarily limited to the physical boundaries of a city, but can spill over to surrounding areas. The physical distance at which agglomeration economies are able to exert their effects is the main element in this approach, which explains why smaller cities can sometimes grow thanks to (and at the expense of) other nearby cities. This approach can explain

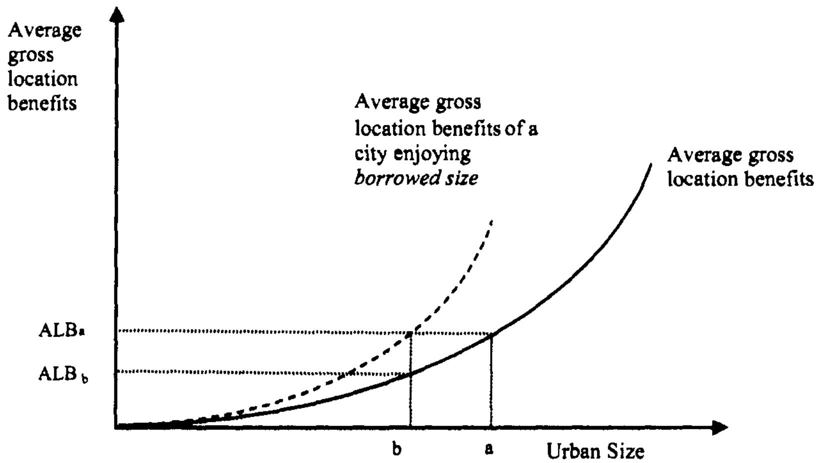


Figure 8.5 Agglomeration economies and borrowed size
Source: Camagni et al. (2015)

why smaller cities can be more efficient than larger cities, as well as why efficient polycentric urban structures at local (regional) level exist where agglomeration economies are 'borrowed' from the entire urban system.

Figure 8.5 represents the average gross urban benefits for different levels of urban size. A city enjoying *borrowed size* obtains average benefits (ALB_a) from its size (b) which are typical of a larger city (a). This means that size and proximity generate technological externalities that boost the productivity of small cities to the level of larger ones.

According to these scholars, the opposite case is also possible: because of physical proximity to large cities, cities of smaller size are limited in their capacity to create new functions, and therefore to grow, and they lose (instead of acquiring) advantages. Called 'agglomeration shadows', the proximity disadvantages are interpreted as the regulatory elements of the formation of urban systems; new urban areas are formed at a distance where they are able to avoid competition with already-existing cities, of similar or greater size, as in the Christaller model.[49]

Notwithstanding the stimulating intuition that agglomeration economies are not confined to the physical boundaries of the city, and that higher urban productivity levels may be achieved if the city is located in a large urban system, this approach has the same limitation as the mainstream approaches that interpret urban growth as a phenomenon of large cities because of their greater (static) productivity. The higher productivity of larger cities is a source of urban dynamics since it is associated with their greater attractiveness: however, reasoning in this way implies an indirect link between agglomeration economies and urban growth that must be empirically tested. Furthermore, the same approach suggests that if the city grows, its productivity increases, and so does its attractiveness, and its size. Also this way of reasoning is unsatisfactory, since the starting point of the reasoning ('if the city grows') is in reality exactly what an interpretative theory on urban dynamics should be able to explain; leaving this aspect to a probability (an *if*) means renouncing explanation of the determinants of urban dynamics.[50]

Moreover, none of these approaches provides a theoretical explanation of the conditions under which 'borrowed size' or 'agglomeration shadows' may take place; nor are they able to identify empirical regularities that identify when one prevails over the other: a major limitation that calls for additional theoretical reflections and empirical studies.

Finally, the geographical approach, as well as the mainstream approach that will be presented in Chapters 10 and 11, assumes that no thresholds exist to increasing returns; whatever physical size the city achieves, it exploits agglomeration economies. However, like any resource exploited intensively, the city is subject to decreasing returns: a situation that finds explanation when the true source of urban dynamics is identified in the net urban benefits, not in gross urban benefits.[51] In fact, firms and individuals base their location choices, and the benevolent planner his/her normative choices, on the location benefits once the higher labour and land costs and general size disamenities have been subtracted. As Figure 8.6 shows, when urban net benefit is the determinant of location choice, a maximum size of the city exists, even in the presence of permanent increasing gross benefits.

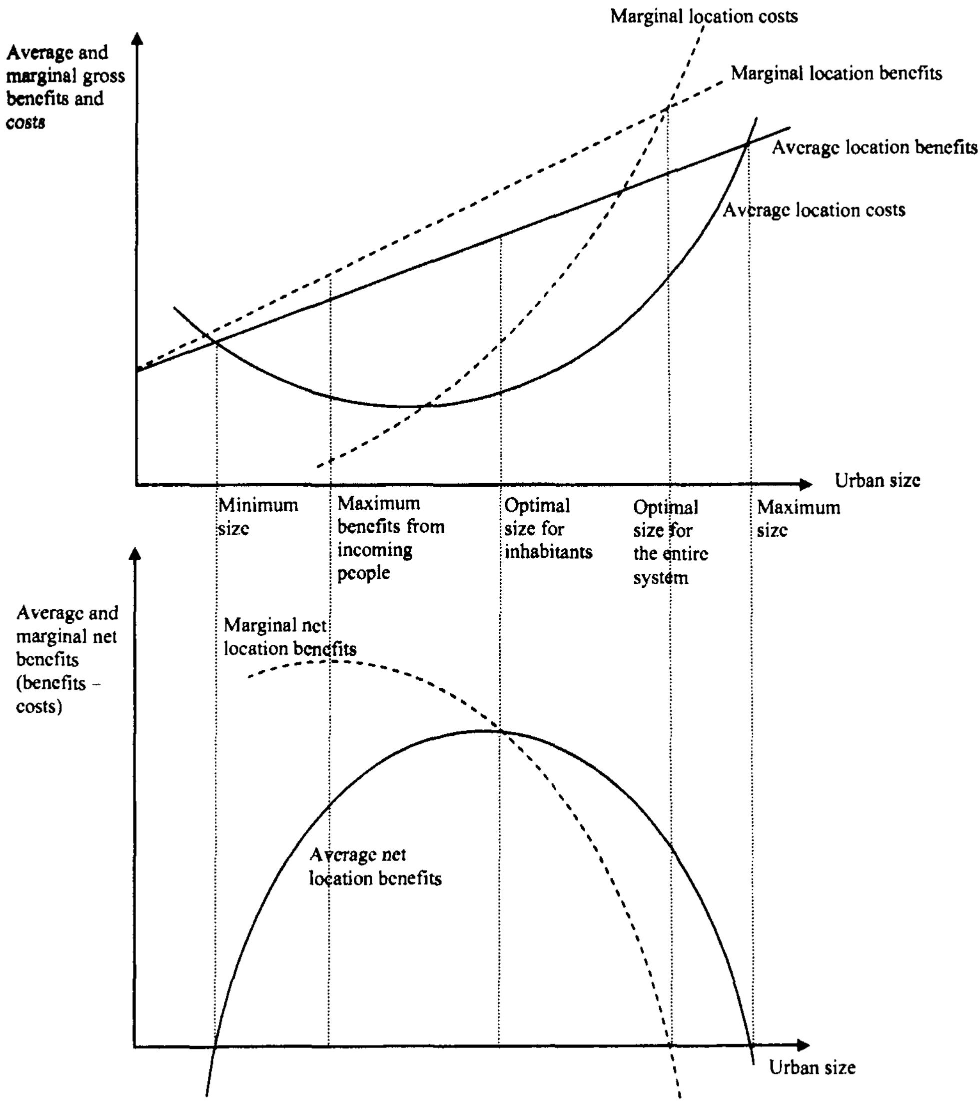


Figure 8.6 The *primum mobile* of urban dynamics: net urban benefits
Source: Alonso (1971)

8.4.4 *Dynamic agglomeration economies and urban growth: the macro-territorial approach*

A recent approach to agglomeration economies investigates the macro-territorial foundations of agglomeration economies.[52] It takes the city as the main unit of analysis, as do the large number of empirical studies measuring the scale effects of urban size.

The main idea in the macro-territorial approach is that explanation of urban growth requires true consideration of the time dimension, and that for this reason agglomeration economies in their static version must be replaced by a concept of dynamic agglomeration economies. These latter are defined as changes over time in productivity advantages associated with urban size; once the determinants of, and the preconditions for, dynamic agglomeration economies have been identified, urban growth finds a direct explanation.

This approach starts by acknowledging two bridging elements between a static and a dynamic interpretation of agglomeration economies. First, if agglomeration economies are assumed as the driving forces behind attractiveness for new activities and population, they must be conceived as net and not gross urban benefits, at a macro-urban and not micro-pecuniary level. Second, other factors contribute, together with pure size, to explaining urban efficiency levels; and changes in the intensity of these factors cause increases in agglomeration economies *irrespective of* the size of the city.

On a simplified view, in fact, efficiency increases may be taken for granted on passing from small to medium and large cities; only in very large cities should the problem of a downturn in urban returns to scale eventually emerge. Assuming a more complex view, the new theoretical conjecture claims that the exploitation of agglomeration economies is relatively straightforward within each of the three/four traditional size classes (small, medium, large, mega-cities), but it implies the presence of specific limiting/enabling factors when cities approach some critical instability point. Therefore, cities may experience a halt in their growth path, and even a decline, irrespective of their size, in the absence of these conditioning factors. These factors are not really quantitative in nature. Rather, they are qualitative; and quantum leaps in their endowment are needed at specific intervals if agglomeration economies are fully to exert their beneficial effects. The quality of the activities hosted, the quality of production factors, the density of external linkages and co-operation networks, the quality of urban infrastructure – for internal and external mobility, education, public services – are all factors enabling an increase in productivity advantages, and a long-term 'structural dynamics' process (in the language of dynamic ecological models) via what can be called a process of urban evolution and transformation.

In this sense, the explanation of a relatively good urban economic performance is not mechanically linked to the existence of static agglomeration economies. Instead, this approach highlights the conditions under which agglomeration economies manifest themselves and may be fully exploited within each urban size class.

This approach confirms the existence of agglomeration economies, as well as the risk of agglomeration diseconomies, but this *general law works within each class of cities*. Some large cities are able to escape agglomeration diseconomies, despite their large size, while small ones can experience decreasing returns despite their small size (Figure 8.7). The explanation of this apparent contradiction lies in the capacity of cities to overcome agglomeration diseconomies either by innovating the functions that they host or by stimulating network co-operation and interaction with other cities. All

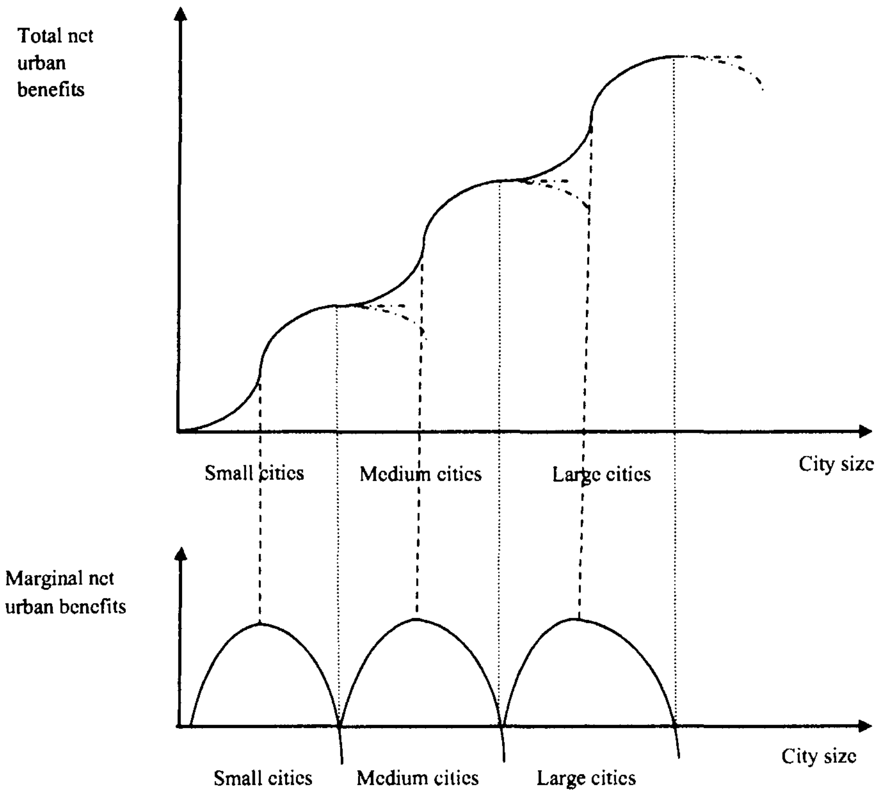


Figure 8.7 Urban evolution in a simplified urban hierarchy (total and marginal net urban benefits by city size classes)

Source: Camagni et al. (2015)

this is obtained, however, without it being possible to precisely indicate the width of the interval (the size at which the city in each class reaches its maximum performance), as well as the speed with which the maximum size is achieved (the slope of the logistic curve of total net urban benefits).

To summarize, this approach highlights that there exists a ***unique law of agglomeration economies that applies to all cities of any size and showing marked specificities within each size class. Within each city class, the quality of territorial capital assets – the presence of high-value functions or networking and co-operation capabilities – is the condition sine qua non to avoid entering a phase of decreasing returns.*** From this perspective, smaller cities, especially, have high potential for growth if they enter a virtuous and cumulative path of transformation and innovation through the exploitation of high-quality territorial assets despite their limited size.

This approach is not only able to explain the physical growth of the city; it potentially interprets a structural evolution of the city, which represents an increasingly important stylized fact. Indeed, innovation has for a long time shaped relative urban growth, mainly through the creation of new producer and consumer services, the increasing sophistication of existing services, the improvement of service functions within industries and their selective decentralization along the urban hierarchy, and the adoption of new technologies in internal mobility and communication, thereby changing the nature of cities, their specialization patterns and their growth opportunities.

8.4.5 *Urbanization vs. localization economies*

It has been argued in previous sections that cities are the natural sources of economies that stem from the variety and the richness of large markets of goods and services, of job opportunities, of different lifestyles, of advanced services for firms and families: all advantages defined as urbanization economies. However, since the 1970s, a rich debate has focused on identification of the relative importance of industrial diversification with respect to specialization as sources of increasing returns in cities.

A plethora of empirical studies have been conducted to highlight which cities, and within each city which industries, produce higher agglomeration economies. The advantages arising from highly specialized cities, in which the 'base sector', as Hoyt called it in his theory, is easily identifiable and a source of localization advantages, of Marshallian economies (as mentioned in the first sections of this chapter), are compared to the advantages that stem from diversified cities. In diversified cities, firms can enjoy the presence of a mix of industries, of a large market of final goods and of inputs; in general, of the so-called urbanization economies, or 'Jacobian externalities', which take the name of Jane Jacobs, the first scholar to highlight the importance of this concept in her famous book *The Economies of Cities*.[53]

According to the Marshallian tradition, firms located in specialized cities may draw advantage from the reduction of production costs when the size of the industry increases; the sources of advantages are embedded in the presence of specialized inputs, of a specialized market for final goods, in a labour market whose specialization guarantees an easy match between firms' needs and labour force expertise, in a specialization of firms within the entire product life cycle, and in the strong input–output linkages among local firms, the latter known as 'pecuniary externalities' since they are mediated by the market.

The Jacobian tradition calls for the importance of advantages that stem from the presence of a rich mix of industries in the area; these advantages increase with the size of the city. New York, Los Angeles and San Francisco are all cities that present a larger and richer mix of industries than cities of smaller size; in particular, the variety of industries is a source of externalities in the form of new, complementary and diversified knowledge which enrich the innovative ability of local industries, and therefore their dynamics.

The methodology applied by the empirical analyses is based on the introduction of specialization and diversification indicators in regional or urban production or growth functions, in order to test their role after controlling for the city size.[54]

Despite the richness of the studies, however, and the sophisticated econometric techniques and geo-referenced data that have been developed over time, the empirical evidence is mixed, and a clear and indisputable confirmation of whether industrial diversification or specialization determines urban productivity (and urban growth) is still lacking[55] – to the point that it has been rightly underlined that the question is misplaced. The reasons for this statement are manifold. First, it is not really pure diversification or specialization which is important, but the mix of industries present in the city. The cross-fertilization of ideas and knowledge among industries at the basis of urbanization economies takes place when the local industries are different but characterized by a large common knowledge base that guarantees the complementary knowledge to be exchanged.[56] Second, an industrial heterogeneity exists in exploiting the different sources of advantages. High-tech industries are more interested in exploiting knowledge exchange with respect to traditional industries. The latter probably gain more advantages from the input–output linkages, in the form of stable and long-term relationships with local suppliers and of a specialized labour market.[57] Finally, by abandoning the deterministic logic of the first two considerations that the mere physical proximity among industries or individuals inevitably brings a knowledge exchange, an additional explanation emerges concerning the lack of an indisputable confirmation that industrial diversification or specialization determines urban productivity (and urban growth). As will become clear in the next chapter, on assuming a relational logic whose aim is to highlight the conditions under which co-operation among agents or firms takes place, stochastic and probabilistic elements emerge in the capacity to exploit the advantages present in the area, linked to the variety and richness of individual behaviours, that can easily explain the volatility in the aggregate result.

8.5 Conclusions

The chapter has surveyed the theories that, since the mid-1970s, have endeavoured to make economic analysis incorporate a concept of space that performs an active role in defining the patterns of a local area's economic development. With these theories, the concept of 'space' gives way to that of 'territory' as a factor generating economic advantages for the activities located in it. For these theories, economic development depends on a concentrated spatial organization of production activities, rather than on a greater endowment of economic resources, or on their more efficient spatial allocation.

These theories are widely studied still today, and they have been enriched with numerous conceptual and empirical studies. They represent the maximum of cross-fertilization between location theory and local development theory, where locational features explain and are explained by economic development processes inextricably bound up with each other. All this comes about within a microeconomic and micro-behavioural framework, and therefore without the rigorous mathematical formulation of macroeconomic growth models that, until the end of the 1980s, were only possible on the assumption of perfect competition and constant returns. In the next chapters we shall see that theories have been developed in recent years that include spatial aspects and the increasing returns that derive from them in macroeconomic growth models but are only able to do so because they adopt a different conception of space as 'diversified-stylized'. However, before moving in this direction, the next chapter shall present theories that, in an elegant way, have used the concept of territory as a source of dynamic advantages; as we shall see, in these theories the innovation capacity of local systems is made dependent on local socio-economic conditions strongly embedded in the local area. Different concepts of proximities (physical, relational, institutional, cognitive) are brought to the fore as explanatory elements of innovation capacities of local firms, within a rigorous microeconomic and micro-behavioural framework.

within industries and their selective decentralization along the urban hierarchy, and the adoption of new technologies in internal mobility and communication, thereby changing the nature of cities, their specialization patterns and their growth opportunities.

8.4.5 *Urbanization vs. localization economies*

It has been argued in previous sections that cities are the natural sources of economies that stem from the variety and the richness of large markets of goods and services, of job opportunities, of different lifestyles, of advanced services for firms and families: all advantages defined as urbanization economies. However, since the 1970s, a rich debate has focused on identification of the relative importance of industrial diversification with respect to specialization as sources of increasing returns in cities.

A plethora of empirical studies have been conducted to highlight which cities, and within each city which industries, produce higher agglomeration economies. The advantages arising from highly specialized cities, in which the 'base sector', as Hoyt called it in his theory, is easily identifiable and a source of localization advantages, of Marshallian economies (as mentioned in the first sections of this chapter), are compared to the advantages that stem from diversified cities. In diversified cities, firms can enjoy the presence of a mix of industries, of a large market of final goods and of inputs; in general, of the so-called urbanization economies, or 'Jacobian externalities', which take the name of Jane Jacobs, the first scholar to highlight the importance of this concept in her famous book *The Economies of Cities*.[53]

According to the Marshallian tradition, firms located in specialized cities may draw advantage from the reduction of production costs when the size of the industry increases; the sources of advantages are embedded in the presence of specialized inputs, of a specialized market for final goods, in a labour market whose specialization guarantees an easy match between firms' needs and labour force expertise, in a specialization of firms within the entire product life cycle, and in the strong input–output linkages among local firms, the latter known as 'pecuniary externalities' since they are mediated by the market.

The Jacobian tradition calls for the importance of advantages that stem from the presence of a rich mix of industries in the area; these advantages increase with the size of the city. New York, Los Angeles and San Francisco are all cities that present a larger and richer mix of industries than cities of smaller size; in particular, the variety of industries is a source of externalities in the form of new, complementary and diversified knowledge which enrich the innovative ability of local industries, and therefore their dynamics.

The methodology applied by the empirical analyses is based on the introduction of specialization and diversification indicators in regional or urban production or growth functions, in order to test their role after controlling for the city size.[54]

Despite the richness of the studies, however, and the sophisticated econometric techniques and geo-referenced data that have been developed over time, the empirical evidence is mixed, and a clear and indisputable confirmation of whether industrial diversification or specialization determines urban productivity (and urban growth) is still lacking[55] – to the point that it has been rightly underlined that the question is misplaced. The reasons for this statement are manifold. First, it is not really pure diversification or specialization which is important, but the mix of industries present in the city. The cross-fertilization of ideas and knowledge among industries at the basis of urbanization economies takes place when the local industries are different but characterized by a large common knowledge base that guarantees the complementary knowledge to be exchanged.[56] Second, an industrial heterogeneity exists in exploiting the different sources of advantages. High-tech industries are more interested in exploiting knowledge exchange with respect to traditional industries. The latter probably gain more advantages from the input–output linkages, in the form of stable and long-term relationships with local suppliers and of a specialized labour market.[57] Finally, by abandoning the deterministic logic of the first two considerations that the mere physical proximity among industries or individuals inevitably brings a knowledge exchange, an additional explanation emerges concerning the lack of an indisputable confirmation that industrial diversification or specialization determines urban productivity (and urban growth). As will become clear in the next chapter, on assuming a relational logic whose aim is to highlight the conditions under which co-operation among agents or firms takes place, stochastic and probabilistic elements emerge in the capacity to exploit the advantages present in the area, linked to the variety and richness of individual behaviours, that can easily explain the volatility in the aggregate result.

8.5 Conclusions

The chapter has surveyed the theories that, since the mid-1970s, have endeavoured to make economic analysis incorporate a concept of space that performs an active role in defining the patterns of a local area's economic development. With these theories, the concept of 'space' gives way to that of 'territory' as a factor generating economic advantages for the activities located in it. For these theories, economic development depends on a concentrated spatial organization of production activities, rather than on a greater endowment of economic resources, or on their more efficient spatial allocation.

These theories are widely studied still today, and they have been enriched with numerous conceptual and empirical studies. They represent the maximum of cross-fertilization between location theory and local development theory, where locational features explain and are explained by economic development processes inextricably bound up with each other. All this comes about within a microeconomic and micro-behavioural framework, and therefore without the rigorous mathematical formulation of macroeconomic growth models that, until the end of the 1980s, were only possible on the assumption of perfect competition and constant returns. In the next chapters we shall see that theories have been developed in recent years that include spatial aspects and the increasing returns that derive from them in macroeconomic growth models but are only able to do so because they adopt a different conception of space as 'diversified-stylized'. However, before moving in this direction, the next chapter shall present theories that, in an elegant way, have used the concept of territory as a source of dynamic advantages; as we shall see, in these theories the innovation capacity of local systems is made dependent on local socio-economic conditions strongly embedded in the local area. Different concepts of proximities (physical, relational, institutional, cognitive) are brought to the fore as explanatory elements of innovation capacities of local firms, within a rigorous microeconomic and micro-behavioural framework.

Appendix: specialization and diversification indicators

Empirical analysis of the sectoral specialization and diversification, and of the static and dynamic efficiency of cities, requires identification of indicators able to measure the productive diversification or specialization of an area.[58]

The most commonly used indicator for the productive specialization of an area is the so-called Hirshmann–Herfindahl (HH) index.[59] With s_i denoting the share of employees in the i-sector in the region, the productive specialization of the region is measured as the sum of the square of all k-industry shares:

$$HH = \sum_{i=1}^{k} (s_i)^2$$

The index assumes values from $1/k$ (equal distribution of employment in all industries) to 1 (concentration of employment in one sector). Therefore, the higher the indicator, the higher the area's degree of productive specialization.

The way in which the HH index is conceived does not allow understanding of whether the specialization in one sector is the result of one large firm or the presence of many small firms. In the former case, specialization advantages stem from economies of scale (a large plant in an industry); in the latter case, they are the result of localization economies (presence of many firms in the same sector). In order to overcome this limitation, the indicator may be revised as follows:

$$LOCECON = \sum_{i=1}^{k} \left(s_i - \frac{1}{k*} \right)^2 - \sum_{j=1}^{n} \left(z_j - \frac{1}{n} \right)^2$$

where z is the share of employees in the n firms present in the region and k^* is the maximum number of sectors to which the n firms can belong. The first part of the index measures the extent to which the industrial employment differs from an equi-distribution ($1/k^*$).[60] The second part of the index indicates the extent to which the distribution of employees in the region's firms differs from an equi-distribution ($1/n$). The indicator assumes values from –1 to 1; the closer to –1, the more the specialization depends on plant size rather than on localization economies. In this case, in fact, the index represents the situation in which the employment in firms differs from an equi-distribution.

It is evident that the complement to 1 of the HH index measures diversity. A variant of this indicator consists of an index of diversity calculated on the first five largest local industries. In this case, the productive diversity of an area is obtained as the complement to 1 of the sum of the shares of employees in the different i sectors (indicated with s_i) in the first five largest sectors, excluding the first one, as follows:[61]

$$DIV = 1 - \sum_{i=1}^{5} (s_i)^2$$

The index assumes values from 0 to 1. Close to 1 the index signals high productive diversity.

The indicators described thus far measure the industrial specialization of a region. However, it might be interesting to measure how specialization in a region changes over time. For this purpose, the Lawrence index is an appropriate indicator. It is calculated as follows:

$$LAWRENCE = \frac{1}{2}\sum_{i=1}^{k}\left|s_i^t - s_i^{t-1}\right|$$

where *i* represents a generic sector, and *k* the maximum number of sectors. The multiplicative term ½ allows the indicator to go from 0 to 1. Values close to 1 indicate an industrial specialization reconversion, while a value close to 0 shows a stability over time of the region in its specialization patterns.

Another well-known indicator is the so-called Krugman dissimilarity index.[62] The latter is built with the intention to measure the degree of dissimilarity of a region's sectoral specialization with respect to a reference area, generally its nation. It is obtained as the sum of the absolute value of the difference in the shares of employees in sector *i* in the region *r* (s_{ir}) and the same share calculated at national level (s_{in}):

$$DISS_r = \sum_{i=1}^{k}\left|s_{ir} - s_{in}\right|$$

The indicator assumes values from 0 to 2. When it is close to 0, dissimilarity is small, and therefore specialization in region *r* is similar to that of its nation *n*. Conversely, values close to 2 signal a situation of total dissimilarity in specialization between the region and its nation. Because of the way in which it is built, the indicator is unable to measure the degree of regional specialization; it is instead able to measure how the region's specialization differs from that of its nation. Therefore, it indicates a *relative specialization*.

The difficulty in obtaining a common result from the studies on diversification or specialization of urban productivity and growth has in recent years pushed scholars towards a different definition of diversity based not merely on the presence of a mix of sectors, but of complementary sectors sharing a common knowledge base.[63] As we will see in the next chapter, this reflection opens the way to new indicators and new attempts to identify the role of industrial specialization or diversity in regional dynamics.

Review questions

1 What is meant by endogenous development?
2 What role is played by space in local endogenous development theories?
3 What conception of growth is behind the theory of endogenous development?
4 What stimulated the conceptualization of the endogenous development approach?
5 How would you define an industrial district? How would you define district economies? What advantages exist in industrial districts that go beyond district economies? What are the strengths and weaknesses of this theory?
6 What theoretical elements support the idea that regional development depends on an efficient urban system within the region?

7 What does the 'optimal city size' theorize? What are the critiques of this approach?
8 What does the SOUDY model theorize? What is new with respect to the previous approach?
9 What does the geographical approach add to the interpretation of urban growth? What does 'borrowed size' and 'agglomeration shadow' mean? What are the main limits of the geographical approach?
10 What theoretical novelties are introduced by the macro-territorial approach to urban growth? What elements explain the structural dynamics of cities?
11 How is a regional productive specialization index built?
12 How is the Lawrence index built?
13 How is the Krugman dissimilarity index built?

Selected reading on empirical findings

About industrial districts

Alam G. (1994), 'Industrial Districts and Technological Change: A Study of the Garment Industry in Dehli', in *Technological Dynamisms in Industrial Districts: An Alternative Approach to Industrialisation in Developing Countries?*, UNCTAD, New York–Geneva, pp. 257–266

Dawson J. (1988), 'Flexibility Together: Surviving and Growing in a Garment Cluster, Ahmedabad, India', *The Journal of Entrepreneurship*, vol. V, no. 2, pp. 153–177.

Dei Ottati G. (1996), 'Economic Changes in the District of Prato in the 1980s: Towards a More Conscious and Organised Industrial District', *European Planning Studies*, vol. IV, no. 1, pp. 35–52.

Nadvi K. and Schmidz H. (1994), 'Industrial Clusters in Less Developed Countries: Review of Experiences and Research Agenda', Discussion Paper no. 339, IDS, University of Sussex.

Rabellotti R. (1997), *External Economies and Cooperation in Industrial Districts: A Comparison of Italy and Mexico*, Macmillan, London.

Rabellotti R. and Schmidz H. (1997), 'The Internal Heterogeneity of Industrial Districts in Italy, Brazil and Mexico', *Regional Studies*, vol. 33, no. 2, pp. 97–108.

Saxenian A.L. (1994), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Harvard University Press, Cambridge, MA.

Scott A.J. (1992), 'The Role of Large Producers in Industrial Districts: a Case Study of High Technology Systems Houses in Southern California', *Regional Studies*, vol. 26, no. 3, pp. 265–275.

About agglomeration economies and urban growth

Camagni R., Capello R. and Caragliu A. (2013), 'One or Infinite Optimal City Sizes? In Search of an Equilibrium Size for Cities', *The Annals of Regional Science*, vol. 51, no. 2, pp. 309–341.

Camagni R., Capello R. and Caragliu A. (2014), 'The Rise of Second Rank Cities: What Role for Agglomeration Economies?', *European Planning Studies*, DOI 10.1080/09654313.2014.904999.

Camagni R., Capello R. and Caragliu A. (2015), 'Agglomeration Economies and Urban Growth: Spatial Context and Structural Evolution behind Urban Growth', in *Papers in Regional Science*, doi:10.1111/pirs.12182.

Camagni R., Diappi L. and Leonardi G. (1986), 'Urban Growth and Decline in a Hierarchical System: A Supply-oriented Dynamic Approach', *Regional Science and Urban Economics*, vol. 16, no. 1, pp. 145–160.

Carlino G. (1980), 'Contrast in Agglomeration: New York and Pittsburgh Reconsidered', *Urban Studies*, vol. 17, no. 3, pp. 343–351.

Combes, P.P., Duranton, G. and Gobillon, L. (2008), 'Spatial Wage Disparities: Sorting Matters!', *Journal of Urban Economics*, vol. 63, no. 2, pp. 723–742.

Ellison G. and Glaeser E.L. (1997), 'Geographic Concentration in US Manufacturing Industries: A Dartboard Approach', *Journal of Political Economy*, vol. 105, no. 5, pp. 889–927.

Meijers E. (2013), 'Cities Borrowing Size: An Exploration of the Spread of Metropolitan Amenities across European Cities', paper presented at the Association of American Geographers annual meeting, Los Angeles, 9–13 April.

Further reading

Becattini G. (1990), 'The Marshallian Industrial District as a Socio-economic Notion', in F. Pyke, G. Becattini and W. Sengenberger (eds), *Industrial Districts and Interfirm Cooperation in Italy*, International Institute of Labour Studies, Geneva, pp. 37–51.

Becattini G. (ed.) (2004), *Industrial Districts: a New Approach to Industrial Change*, Edward Elgar, Cheltenham.

Camagni R., Capello R. and Caragliu A. (2015), 'Agglomeration Economies and Urban Growth: Spatial Context and Structural Evolution behind Urban Growth', in *Papers in Regional Science*, doi:10.1111/pirs.12182.

Camagni R., Diappi L. and Leonardi G. (1986), 'Urban Growth and Decline in a Hierarchical System: A Supply-oriented Dynamic Approach", *Regional Science and Urban Economics*, vol. 16, no. 1, 145–160.

Capello R. (1999), 'Spatial Transfer of Knowledge in High-technology Milieux: Learning vs. Collective Learning Processes', *Regional Studies*, vol. 33, no. 4, pp. 353–365.

Dei Ottati G. (2003), 'The Governance of Transactions in the Industrial District: The "Community Market"', in G. Becattini, M. Bellandi, G. Dei Ottati and F. Sforzi (eds), *From Industrial Districts to Local Development*, Edward Elgar, Cheltenham, chap. 4.

Meijers E. (2013), 'Cities Borrowing Size: An Exploration of the Spread of Metropolitan Amenities across European Cities', paper presented at the Association of American Geographers annual meeting, Los Angeles, 9–13 April.

Notes

1 The allusion is to all the location theory models examined in Part I of the book.

2 Hence the name 'NEC areas'.

3 The 'industrial triangle' comprising Lombardy, Liguria and Piedmont, i.e. the regions of northwestern Italy.

4 See Secchi, 1974; Stöhr and Tödtling, 1977; Ciciotti and Wettmann, 1981; Garofoli, 1981; Johannisson and Spilling, 1983; Stöhr, 1990; Courlet and Pecqueur, 1992; Ganne, 1992. See Vásquez-Barquero, 2002, for a well-structured survey of theories of endogenous development.

5 Becattini set out his main ideas in a study published in 1975 (see Becattini, 1975) and then developed them in a subsequent study of 1979 (see Becattini, 1979; English translation, 1989). There followed a series of works in which Becattini expanded and deepened the concept of the 'Marshallian industrial district'. A recent volume containing seminal works on the issue is Becattini, 2004.

6 See Marshall, 1920. For detailed analysis of the links between Marshall's work and the theory of industrial districts see Bellandi 1989.

7 The 'Third Italy' is the geographical macro-area comprising the regions of the Northeast and the Centre (the NEC areas). The term was coined in order to underline the inadequacy of the North/South 'dualistic' model hitherto applied to Italian economic development, and to mark the birth of a new (third) macro area of such development. See Bagnasco, 1977.

8 See Benedetti and Camagni, 1983.

9 Unfortunately, none of the pioneering studies on the subject have been translated into English and are therefore only available in Italian. See Bagnasco, 1977 and 1983 on entrepreneurship; for detailed analyses of the workings of local labour markets see Paci, 1973; Brusco, 1982 and 1990; for analysis of social and political cohesion see Bagnasco and Trigilia, 1984; Bagnasco, 1985; Trigilia, 1985.

10 Becattini's original definition of an industrial district was as follows: 'a socio-territorial entity characterised by the active presence of both a community of people and a population of firms in one naturally and historically bounded area'. See Becattini, 1990, p. 38. This definition highlights the genetic features of an industrial district: the geographical and social proximities of economic actors.

11 In his *Principles of Economics* Marshall stresses the importance of 'those very important external economies which can often be secured by the concentration of many small businesses of a similar character in particular localities: or, as is commonly said, by the "localization of industry"' See Marshall, 1920, 8th edition, p. 221.

12 Because of its limited size, in fact, a small firm cannot exploit 'economies of scale', namely the advantages deriving to a firm from production of a large and increasing volume of output in the same place. In an industrial district, this disadvantage is offset by localization economies, which Marshall calls 'external economies' because they depend on conditions external to the firm (the sector's volume of business) rather than internal to it (the volume of output by the firm). In *Principles of Economics*, Marshall argues that the advantages of large-scale may in general be achieved either by grouping a large number of firms into a single district or by constructing a few large-scale factories. See Marshall, 1920.

13 The historical structure of the agricultural systems of industrial districts has been researched by Bagnasco, who wrote in 1977: 'The family of peasant origin (although it is today different in nature) is the pivot of the production system, and we can also understand why industrialization arises in function of the local presence of this original type of family.' See Bagnasco, 1977, p. 153, my translation. To be noted is that Becattini initially used the expression 'urbanized countryside' to denote diffused industrialization, his purpose being to emphasize the agricultural origins of areas of recent industrialization. See Becattini, 1979; English version Becattini, 1989.

14 Becattini, 1987, p. 47, my translation.

15 Marshall, 1920, 8th edition, p. 225.

16 The idea of a 'community market' was first developed by Dei Ottati. See Dei Ottati, 2003 (original version, Dei Ottati, 1987). Also proposed has been the notion of 'market socialization': see Bagnasco, 1985. The importance of social aspects in industrial district theory is such that some authors have claimed that 'despite the many insights of genuinely economic character, the picture is still predominantly sociological'. See Benedetti and Camagni, 1983, p. 22, my translation.

17 On the symbiosis between competition and co-operation see Becattini, 1990; Bianchi, 1994; Dei Ottati, 1995 and 2003. On the concept of co-operation, or 'collective efficiency', see Schmitz, 1995 and 1998; Rabellotti, 1997.

18 In this regard Becattini writes: 'The dynamic and self-reproducing nature of the district consists of a continuous comparison between the cost of performing any given operation inside the firm and the cost of having it done outside. . . . It should be noted that it is not a matter of a generic comparison between doing and buying, but a specific comparison between doing and having done. . . . It should also be noted that this is almost always "doing together"': Becattini, 1990, p. 48. For detailed analysis of reputation as capital see Dei Ottati, 1995.

19 See Dei Ottati, 1995.

20 I would point out – and not just as a matter of pride – that whilst regional economists drew their initial theories (for example the theory of stages of development) from analysis conducted by development economists, today the reverse is happening: the tools developed by regional economics are being used by economists of development. This demonstrates the significant evolution and autonomization undergone by thought on regional economic development in the past 60 years.

21 See Schmitz and Musyck, 1994; Rabellotti, 1997.

22 See Benedetti and Camagni, 1983; Cappellin, 1983.

23 The reference is to the fact that, in order to understand the performance of the northeastern Italian regions in the 1970s or 1990s, analysis must be made of the performances in those years of the northeastern and southern regions as well.

24 The economic reasoning is as follows. When goods prices (P_n) are assumed equal at national level, and fixed with a mark-up ($1 + \lambda$) differentiated regionally (r) on the cost of labour per unit of output (w_r / x_r), they become:

$$P_n = (1+\lambda)_r (w_r / x_r) \tag{8.1n}$$

where w denotes wages and x productivity. On rearranging elements in (8.1n), a regional locational advantage, indicated by a positive gross profit rate (λ), is obtained in regions where:

$$x_r > (w_r / P_n) \tag{8.2n}$$

and, in relative terms, above-average profit is obtained in regions where:

$$x_r / w_r > x_n / w_n \tag{8.3n}$$

For economic treatment see Benedetti and Camagni, 1983; for application to the Italian case see Camagni and Capello, 1990.

25 See Piore and Sabel, 1984.

26 For a quantitative study of industrial districts using municipal-level data see Pietrobelli, 1998. For a methodology based on firm-level data see Rabellotti, 1997; Signorini, 2000.

27 When commenting on Lösch and Christaller's theories, Beguin observes that 'un bon réseau urbain hiérarchisé peut contribuer à favoriser un dévéloppement régional équilibré' (a good hierachical urban network can contribute to favouring a balanced regional development): Beguin, 1988, p. 242.

28 The official document on the 'European Spatial Development Plan' (ESDP) presented to the European Council of Ministers at Noordwijk in 1997, subsequently supplemented at the meeting of ministers held in Glasgow in June 1998, and definitively approved at Potsdam in June 1999, stresses the importance of an efficient urban system for regional development: 'The development of Europe's cities and the relations between them constitutes the most significant factor affecting the spatial balance of the territory of Europe': ESDP, 1998, p. 47. Again, 'regions as a whole can become competitive only if their towns and cities are motors of economic growth': ESDP, p. 51.

29 As early as 1961, Chinitz stressed that cities with more competitive and diversified structures furnishing externalities for small firms have greater growth potential than cities with oligopolized and specialized structures in which the 'internalizing' of service functions by large firms impoverishes the urban environment. See Chinitz, 1961.

30 The pioneering studies by Vernon and Hoover and Vernon in the USA on the concentration of small and medium-sized firms in the heart of cities provided clear evidence of the 'incubator role' performed by cities. See Vernon, 1960; Hoover and Vernon, 1962.

31 For the detailed explanation of the theory of the milieu innovateur see Chapter 9. On the concept of ***urban milieu*** see Camagni, 1999b. For empirical evidence on the existence of '***milieu*** effects' in five European metropolises, Amsterdam, Milan, Paris, London and Stuttgart, see Capello, 2001.

32 See Chapter 3.

33 This debate refers to two important institutional reports: the Sapir Report (Sapir, 2003) and the World Bank Report (World Bank, 2009). On this debate, see Barca, 2011; Farole et al., 2011; Kim, 2011; McCann and Rodriguez-Pose, 2011; Camagni and Capello, 2015.

34 Alonso (1971) stressed the mistaken tendency of many authors to look for 'optimal city size' only by minimising the location cost function. As he argued, this would be sensible only if output per capita were constant (Alonso, 1971, p. 70).

35 See Camagni, 1992a.

36 For studies based on this method, see, among others Segal, 1976; Marelli, 1981.

37 For studies based on this method, see, among others, Shefer, 1973; Sveiskauskas, 1975; Carlino, 1980; Henderson, 1985.
38 See Chapter 1 for the definition of the different forms of agglomeration economies.
39 For studies based on this method, see, among others, Fuchs, 1967; Hoch, 1972.
40 Richardson, 1972 p. 30. See also Henderson, 1985 and 1996.
41 Carlino (1980) provides a criticism of Chinitz's analysis, and demonstrates on a sample of 65 American towns that economies of scale, both internal and external to the firm, play a role in the definition of urban productivity.
42 See Camagni et al., 1986.
43 Richardson (1972) had suggested replacing the concept of optimal city size with an efficient interval of urban size in which urban marginal benefits are greater than marginal location costs.
44 The two cities will differ, though, in dynamic terms: the one belonging to the lower rank (R_1) will not grow further, having reached the maximum size of its interval, while the one having developed the higher functions (linked to rank 2) will grow, due to the presence of new and wide net urban benefits (profits).
45 See Krugman, 1991b.
46 On the concept of 'missing link', see Meijers, 2013; Burger et al., 2014.
47 See Alonso, 1973, p. 200; italics added.
48 Parr (2002) argues that agglomeration costs are more confined to city boundaries than agglomeration benefits.
49 See Krugman, 1993; Fujita and Mori, 1996 and 1997; Dobkins and Ionidas, 2001.
50 See Krugman, 1991a; World Bank, 2009.
51 See Alonso, 1973; Camagni et al., 2015.
52 For this approach, see Camagni et al., 2013, 2014 and 2015.
53 See Jacobs, 1969.
54 For a synthesis of the 'Marshall vs. Jacobs' externalities, see Beaudry and Schiffauerova, 2009. For the construction of specialization and diversification indicators, see the Appendix to this chapter.
55 On advanced methodologies and on the use of geo-referenced data, which allow account to be taken of space and industry at the same time, see Ciccone and Hall, 1996; Ellison and Glaeser, 1997; Ciccone, 2002; Rosenthal and Strange, 2001 and 2003; Henderson, 2003; Combes et al., 2008; Puga, 2010.
56 This idea is known as 'related variety', a concept developed within the evolutionary economic geography school of thought developed in the 1990s, and it will be largely presented in Chapter 9. See Frenkel et al., 2007; Boschma and Iammarino, 2009.
57 See Faggio et al., 2013.
58 We present here the basic indicators that exist in the literature. Moreover, indicators exist on industrial localization at regional level. For a review of indicators of this latter type, see Fratesi, 2008.
59 Differently from the location quotient, which measures the specialization of an area in different sectors (see Chapter 5), the HH index provides a measure of the area's aggregate specialization. The index was invented separately by the two authors (Hirschman, 1945; Herfindahl, 1950). One of the two recognized the two authors' paternity, underlining the difference with the other indicators with which it was often confused (Hirschman, 1964).
60 The use of k^* instead of k avoids distortions in the indicator when the number of firms is less than the number of industries. In this case, in fact, the maximum number of industries in which firms could be distributed, showing an equal-distribution, is n and not k.
61 In some works, the first sector of specialization has been excluded, representing in general an extreme situation.
62 See Krugman, 1991a.
63 This theory is presented in Chapter 9. For the most recent indicators, see Appendix of Chapter 9.

9 Territorial competitiveness and endogenous development

Innovation and proximity

9.1 The endogenous sources of competitiveness: innovation and proximity

Thus far, we have examined the role of space as a generator of locational advantages – lower production and transaction costs, and a more efficient use of resources – that enables firms to achieve higher levels of productivity and profit.

However, the effects of space on economic activity do not consist solely in improvements to the static efficiency of production processes (that is, an increase in firms' revenues or a decrease in their costs); they are also manifest in the innovative and creative capacity of firms. In this case, space is a source of *dynamic efficiency*. Areas with high concentrations of economic activity enjoy easy information exchange, frequent face-to-face encounters, the presence of research and development activities and advanced services, an availability of skilled labour, co-operativeness facilitated by shared rules and codes of behaviour, and local social capital, which facilitate and incentivize innovation by the firms located within those areas.[1]

These features are easily present in urban areas, which have for this reason always been recognized as the natural sites of innovative activity, the 'incubators' of new knowledge; cities are the principal centres of research, given their large pools of expertise, and the availability of advanced services (finance and insurance) ready to carry the risk of any innovative activity. Yet it is indisputably also the case that certain non-metropolitan areas of small size display an innovative capacity that persistently outstrips that of other geographical areas, and they achieve levels of innovation sometimes greatly disproportional to their manufacturing weight. They thus testify to the presence of some form of increasing returns to the concentration of innovative activity. Cases in point are Silicon Valley in California, 'Route 128' in the Boston area, Baden-Württenberg in the South of Germany, Jutland in Denmark, Småland in Sweden and Sophia-Antipolis close to Nice, to cite only some examples.

Understanding these phenomena became of particular interest in the 1980s. In those years, under the impetus of profound technological changes, innovation came to be considered the driving force of economic development, and knowledge the key factor in local economic success. Instead, the uneven spatial distribution of innovative activity was taken to be the primary cause of regional imbalances. In periods when there are evident signs of the hypermobility of labour and capital, the most immobile of factors are knowledge and the intangible elements connected with culture, skill and innovative capacity; it is on these elements that the competitiveness of local systems depends.

From what has just been said, one understands why identification of the endogenous, local conditions determining an area's innovative capacity became the most important aim of regional development theories developed in the 1980s. These theories differed sharply from the studies on the spatial diffusion of innovation discussed in Chapter 7; their primary aim was no longer to interpret innovation processes through exogenous factors, but to identify the local endogenous determinants of innovations. Their emphasis on elements endogenous to the innovative process fully justifies their inclusion in this chapter.

For these various theories, the endogenous determinants of innovation are increasing returns in the form of dynamic location advantages deriving from:

- *spatial, geographical proximity* among firms, which facilitates the exchange of tacit knowledge: this characterizes reflection by economic geographers concerned to explain the concentration of innovative activities (Section 9.2);
- *relational proximity* among firms, defined as interaction and co-operativeness among local agents, the source of collective learning processes and socialization to the risk of innovation (i.e. territorialized relations among subjects operating in geographical and social proximity): this was the approach taken by territorial economists in explaining the dynamics of local systems in terms of local innovative capacity (Section 9.3);[2]
- *institutional proximity*, taking the form of rules, codes and norms of behaviour which (i) facilitate co-operation among actors and therefore the socialization of knowledge and (ii) assist economic actors (individual people, firms and local institutions) to develop organizational forms which support interactive learning processes: this aspect was emphasized by more systemic approaches seeking to understand the evolution of complex systems like the innovative system (Section 9.4);
- *cognitive proximity* among economic agents, interpreted as the existence of a common knowledge base that guarantees mutual understanding among actors characterized by complementary knowledge, as suggested by evolutionary economic geography in its explanation of the formation of clusters of innovative firms (Section 9.5).

As we shall see, the development of these theories marks the overcoming of the simple view of pure geographical proximity as the explanatory element of knowledge exchange. During the mid-1980s with the theory of the '*milieu innovateur*', and in the 1990s with the French proximity school,[3] new and more profound analyses were added to the interpretation of local knowledge exchange, summarized in diverse concepts of proximities – relational, cognitive, organized, social, technological – without avoiding a certain confusion and an overlap among concepts. Some of these concepts were elegantly inserted into regional development theories – this is the case of relational, institutional and cognitive proximities – and it is around these concepts of proximities and on the theories on which they are based that this chapter is organized.[4]

The feature shared by all the approaches considered is that each concept of proximity is analysed in its capacity to reduce uncertainty associated with innovative activity and the solution of the co-ordination problem among actors acting individually. The existence of proximity allows the development of 'interactive learning', a learning process based on co-operation and the exchange of knowledge among actors. In most of the theories, co-operation is understood to be a result of an explicit decision by economic actors to co-operate, and the different concepts of proximities help identify the criteria on the basis of which partners are chosen. Only the *milieu innovateur* theory explicitly conceptualizes the exchange of knowledge as a spontaneous phenomenon, which takes place through the socialization of knowledge at local level, even against the will of local actors. In this respect, the process of knowledge accumulation in the *milieu innovateur* theory differs from the interactive learning concept, and is defined as a process of 'collective learning'.

Given the abundance of proximity concepts, it is necessary to understand the degree of complementarity that exists among them, and the usefulness of the concept of physical proximity for interpretation of knowledge exchanges amid the plethora of new approaches. As we shall explain (Section 9.6), the theoretical bases of the different concepts of proximity differ substantially. Moreover, as we shall see, when some approaches interpret the mechanisms behind proximity effects at local level, they refer again to a concept of physical proximity.

As for the theories of endogenous development associated with agglomeration forces, we do not criticize the qualitative nature of the approaches presented here; on the contrary, we think that they enrich the economic theory by highlighting the role of intangible elements (knowledge, learning, relational and social capital) in determining local competitiveness. Moreover, some of these theories present a new approach to the interpretation of local competitiveness. They in fact no longer embrace the traditional functional approach, characterized by deterministic cause–effect relationships whereby the presence of a certain degree of knowledge in the region mechanically leads to innovation. They instead assume a relational approach, according to which probabilistic elements – envisaged as the ways in which economic actors perceive the economic reality, react to external stimuli, and are capable of co-operative and synergic behaviours – come into play in explaining when the existence of knowledge really leads to innovation. These elements enrich the interpretation of the real world, and open the way to a more profound and sophisticated interpretation of economic development.

Therefore, from these theories it is possible to develop a new concept of 'regional innovation patterns', a concept able to interpret the different modes through which regions innovate on the basis of the existence of the local preconditions for the generation of knowledge and relationality among economic actors, both internal and external to the area (Section 9.7). It is on these 'regional innovation patterns' that modern innovation policies should be developed (Section 9.8).

9.2 Knowledge spillovers: geographical proximity

That innovative activity has a natural tendency to concentrate in space has been confirmed by numerous empirical studies. Using both input indicators (e.g. spending on research and development) and output indicators (e.g. number of patents) of innovative activity, these studies show that innovation is concentrated in central and metropolitan areas. Moreover, in all the industrialized countries, analyses of the location of high-tech firms reveal marked polarization effects due to the pronounced preference of these firms for central locations with strong sectoral specialization.[5]

Explanation of the phenomenon is straightforward: concentrated location facilitates exploitation of technological and scientific knowledge developed by research centres and universities; it gives easier access to the tacit uncodified knowledge required for imitation and reverse engineering; and it ensures the ready availability of skilled labour and advanced services.

Moreover, the complex and systemic nature of innovative processes explains their cumulative character: clusters of incremental innovations follow an initial radical innovation that marks out a 'technological trajectory' along which knowledge grows and develops within well-defined technological boundaries. At local level, demand for and the supply of innovative factors interact and mutually reinforce each other. Advanced firms enrich the surrounding environment by diffusing their technological and organizational expertise, while the surrounding environment simultaneously sustains their activity. The outcome is a cumulative polarization of research and innovation activities which reinforces the natural tendency for innovation to concentrate in space.

The role of agglomeration economies, both urban and sectoral, in explaining the concentration of innovative activity was demonstrated long ago by Marshall. But interest in dynamic agglomeration economies (the agglomerative advantages that foster innovation by firms) has grown considerably in recent years, as recognition has gained ground of the importance of innovation for the competitiveness of local systems.

The *theory of technological spillovers* developed in the 1990s linked the spatial concentration of innovative activities with the increasing returns that concentrated location generates on those innovative activities themselves. Cross-fertilizations, dynamic interactions between customers and suppliers, synergies between research centres and local production units occur within circumscribed geographical areas like highly specialized metropolitan areas. They do so as the result of the rapid exchange of information and transmission of tacit knowledge made possible by face-to-face encounters. In a concentrated location, the beneficial effects of a firm's research and development activities are not confined within the boundaries of firms; they 'spill over' into the surrounding environment, to the advantage of innovative activity by other firms.

A large number of empirical analyses, mainly econometric, have successfully measured the technological spillovers and the knowledge advantages enjoyed by spatially concentrated firms. Now briefly outlined are two of the methods employed to measure these effects:[6]

a) estimation of an aggregate knowledge production function at regional level, in order to verify the existence of technological spillovers; or in simpler econometric terms, to verify the existence of differing effects exerted by research and development (R&D) activities, conducted within and without a region, on its patenting activity.[7] The results confirm the existence of spillovers from innovative activity, in that the significance of the parameter associated with local R&D is greater than that of the parameter for external R&D;[8]
b) estimation of a disaggregated knowledge production function for individual local sectors that separately includes not only expenditure on local and external R&D, but also expenditure on R&D by the same sector and by different ones. The purpose is to determine the differing impacts on innovative activity of diversified

and specialized knowledge. Once again, the results show that expenditure on local R&D is, for the majority of sectors, more significant than expenditure on external R&D, and that diversified rather than specialized knowledge is important for local innovative capacity.

However, the theory can be criticized on various grounds. First, it should be borne in mind that research and development expenditure and number of patents are highly selective indicators of innovative capacity. Both capture only product innovations, that is, breakthroughs often associated with the innovative activity of large firms. They entirely neglect the process innovation, the creative imitation and the reverse engineering that characterize the innovative processes of small firms.

Even more dubious is the concept of space assumed by the theory. This space is purely geographical, a physical distance among actors, a pure physical container of spillover effects which come about – according to the epidemiological logic adopted – simply as a result of physical contact among actors. Important consequences ensue. First, this view is unable to explain the processes by which knowledge spreads at local level, given that it only envisages the probability of contact among potential innovators as the source of spatial diffusion. Second, it concerns itself only with the diffusion of innovation, not with the processes of knowledge creation. It thus imposes the same limitations as did Hägerstrand's pioneering model in regard to the spatial diffusion of innovation; the diffusion of knowledge means adoption, and adoption means more innovation and better performance.[9] Thus ignored, however, is the most crucial aspect of the innovation process: how people (or the context) actually learn. This is the aspect of overriding interest not only for scholars but also, and especially, for policy-makers, should they wish to explore the possibilities of normative action to promote local development.

9.3 Collective learning and the *milieu innovateur*: relational proximity

9.3.1 Local synergies and relations

In the 1980s, an international group of scholars set out to analyse the phenomenon of the spatial concentration of small firms. Their conclusion was that social interactions, interpersonal synergies, and collective action among actors – in short, what they called 'relational proximity' – are the factors that account for the greater innovative capacity of spatially concentrated small firms, and of the areas in which they are located.[10] This current of thought thus brought space as the generator of dynamic efficiency into the central focus of analysis on territorial development.

For this theory, economic and social relations among local actors condition the innovative capacity and economic success of specific local areas termed '*milieux innovateurs*'.[11] Synergies among actors are enhanced by spatial proximity and economic and cultural homogeneity, and thus produce dynamic advantages for small firms because they underpin processes of collective learning and socialization of knowledge.

Economic and social relations take two different forms in a *milieu*:

- a set of mainly informal, 'untraded' relationships – among customers and suppliers, among private and public actors – and a set of tacit knowledge transfers

that take place through job-mobility chains and inter-firm imitation processes. These informal relationships have been widely studied by the French 'proximity school', and they have recently been labelled 'untraded interdependencies';[12]
- more formalized, mainly trans-territorial co-operation agreements – among firms, among collective agents, among public institutions – in the field of technological development, vocational and on-the-job training, infrastructures and services provision.

Relationships of the former type constitute the 'glue' that creates a *milieu* effect; they are complemented by the latter, more formalized, kinds of relationship, which can be interpreted as 'network relations' proper. Both sets of relationships can be viewed as tools or 'operators' that assist the (small) firm in its competitive endeavour, enhancing its creativeness and reducing the dynamic uncertainty intrinsic to innovation processes.

The partners in trans-territorial networks are selected single economic units – enterprises, banks, research centres, training institutions or local authorities – for which location is only one co-ordinate among the many that serve to identify the unit. At first glance, therefore, these networks merely link different economic actors and have no necessary relation with space. But when the location of a unit takes on significant meaning, inasmuch as it reveals a set of relations which generate territorial development and identity (e.g. Apple at Cupertino, Silicon Valley), and when these network relations start to multiply, they do indeed become territorial. When carefully observed, the identity of the local *milieu* often prevails over the identity of the individual partner, which highlights the importance of the territorial aspect; the strategic importance of links with a company in Silicon Valley resides more in the opening of a 'technological window' in Silicon Valley than in gaining access to that specific company's know-how.[13]

9.3.2 *Collective learning and network co-operation*

'Relational capital' is defined as the set of norms and values that govern interactions among people, the institutions where they are incorporated, the relationship networks set up among various social actors and the overall cohesion of society. Relational capital is therefore explaining the intensity of social interactions, interpersonal synergies and collective action among local actors; that is, relational proximity. The latter has the same role in *milieu* theory as spatial proximity has in the knowledge spillover theory, in that it generates dynamic advantages taking the following forms (see Table 9.1):[14]

- collective learning and socialization processes;
- reduction in the risk and uncertainty associated with the innovation process;
- the *ex-ante* co-ordination of routine and strategic decisions made possible by reduced transaction costs.

These functions are performed in a large firm by its R&D department, and they are facilitated by internal diversification and complexity. A small firm finds the same functions in a highly specialized territory – as now explained.

Learning in a *milieu* takes place in spontaneous and socialized manner within the local labour market through forms of *stable and enduring* collaboration between

Table 9.1 Functions of the local milieu

Conditions / *Functions*	*Geographical proximity*	*Relational proximity*
Reduction of uncertainty	Information collection/selection Vertical integration within 'filières' Local signalling (collective marketing)	Information transcoding Selection of decision routines Risk sharing among partners
Reduction of coordination costs	Information collection Reduction of transaction costs (à la Williamson) Ex-ante co-ordination of day-to-day decisions (à la Marshall)	Reduction of control costs through trust and loyalty Social sanctions on opportunistic behaviour Ex-ante co-ordination in strategic decision-making
Durable substrate for collective learning	Labour turnover within the milieu Imitation of innovation practices	Co-operation on industrial projects Tacit transfer of knowledge Public/private partnerships in complex development schemes

Source: Camagni and Capello (2002)

customers and suppliers based on loyalty and trust. These relations produce a codified and tacit transfer of knowledge between customers and suppliers that triggers processes of incremental innovation and specific technological trajectories. Relations in the local labour market likewise perform an important role in the local production system because high turnover of skilled labour within the area and scant external mobility cross-fertilize knowledge among firms and upgrade workers' skills. Finally, firm spin-offs – independent firms created by workers previously employed by a local firm – also participate in the knowledge socialization process.

The accumulation of knowledge in large firms is ensured by the presence of R&D departments; and it is permanent because large firms are long-lived and develop their own internal capabilities and cultures. By contrast, small firms have very short life-cycles, with the consequence that they are unable to develop a solid stock of firm-specific knowledge. This difficulty is remedied by the *milieu* and by the relations within it, which guarantee continuity of knowledge through labour market stability, high people mobility within the area, and stable relations between customers and suppliers.

In *milieu innovateur* theory, therefore, collective learning is the territorial counterpart of the learning that takes place within firms. In large firms, knowledge and information are transferred via internal functional interaction among the R&D, production, marketing and strategic planning departments.[15] In *milieux*, and in local small firms systems, this function is performed by the already-mentioned high level of people mobility, by intense innovative interactions between customers and suppliers, and by firm spin-offs (Table 9.2).

Milieu theory flanks these channels of learning available to firms with a third and complementary one: learning through 'network co-operation' (Table 9.2). Through

Table 9.2 Preconditions and channels for learning processes in innovative milieux

Preconditions / *Contexts (channels)*	*Continuity*	*Dynamic synergies*	
Firms	R&D functions	Functional interaction Tacit transfer of knowledge	*INTERNAL LEARNING*
Territory	Low mobility of the labour force outside the *milieu*	High mobility of the labour force within the *milieu*	*COLLECTIVE LEARNING*
	Stable linkages with suppliers and customers	Co-operation for innovation with suppliers and customers	
		Local spin-offs	
Networking	Stability as a consequence of the complexity of strategic alliances	Transfer of knowledge via co-operation	*LEARNING THROUGH NETWORKING*

Source: Camagni and Capello (2002)

strategic alliances and/or non-equity co-operation agreements, firms acquire some of the strategic assets that they require externally, thus avoiding the costs of developing them internally. This knowledge-acquisition process stands midway between internal learning and collective learning, in that the firm comes into contact with the outside but still maintains a set of selected and targeted relationships. This form of learning assumes an important role in *milieu innovateur* theory because it permits local knowledge – which is produced by socialized and collective processes liable to isolation and lock-in – to enrich and innovate itself. Only through the co-operation with external firms that ensures an influx of new knowledge can a *milieu* avoid death by entropic uniformity. It is with this conceptual tool that the theoreticians of the *milieu innovateur* interpret the growth of small firms areas, among them the Marshallian industrial district.

9.3.3 *Beyond collective learning and network co-operation*

However, collective learning is not the only dynamic advantage generated for local firms by the *milieu*, with its assets of relational capital. A further factor facilitating firms' innovative capacity is the reduction of the uncertainty that accompanies innovative processes. In large firms, the functions of information-gathering, the codification of knowledge and the selection of decision-making routines – all of which are geared to reducing static and dynamic uncertainty – are performed by the R&D department, or by the planning unit. In the case of a *milieu innovateur*, they are undertaken in socialized and collective manner by the *milieu* itself, in which information rapidly circulates because of geographical and collective proximity.[16]

Finally, the reduction of the costs of *ex-ante* co-ordination among decision-making units, and the facilitation of 'collective action' (undertaken to furnish collective goods or simply to integrate private investment decisions), is a further element enhancing the innovative process in a *milieu*. Such co-ordination generally suffers from the availability of limited and costly information, and from the possible existence of opportunistic behaviour. The presence of the *milieu* reduces these costs because it enables information to circulate more easily; it facilitates the taking of co-ordinated decisions through proximity and social homogeneity/cohesion; while it discourages opportunistic behaviour by fostering trust and threatening social sanctions. This last social/psychological element is crucial: it derives from the sharing of common values and of similar codes of behaviour, and it acts positively by developing trust and loyalty. Conversely, it develops rapid processes of isolation and punishment for opportunistic behaviour.[17]

The influence exerted by Marshallian district theory on this approach is evident: the *milieu* theory reiterates the importance of geographical proximity, but even more so of social and cultural proximity, in guaranteeing forms of stable and enduring co-operation in small firm areas. For industrial district theory, these forms of co-operation give rise to a 'community market', the form of production organization which ensures the static efficiency of firms. For *milieu* theory, co-operation generates processes of knowledge socialization, and it reduces the risk associated with innovation, and collective learning – that is, factors of dynamic efficiency.

In recent years, econometric empirical analyses have corroborated the theory. In the case of three *milieux* in Italy, a production function was estimated using data collected at individual firm level in which efficiency parameters of the production factors were connected to:

- for labour: effects of collective learning, these being identified in the intensity of local spin-offs, and appreciation of the stability and quality of the local labour market;
- for intangible capital: effects of 'industrial atmosphere' and collective learning, these being identified in the importance to the individual firm of specialized knowledge internal to the local area, and the lesser importance of acquiring knowledge from outside.

The results showed that labour productivity is subject to increasing returns (given the small average size of firms) that are substantially reinforced by the presence of collective learning processes. Conversely, (intangible) capital productivity is subject to decreasing returns, but is greatly augmented by an increase in the appreciation and use of local specialized knowledge (Figure 9.1).[18]

In terms of economic theory, the *milieu innovateur* approach has recently been indirectly validated by stylized analytical models à la Romer and Lucas.[19] The rigidly neoclassical and aggregate form of these endogenous growth models distorts neither the hypotheses nor the intrinsic logic of the *milieu* theory – which testifies to the latter's ability to depict the endogenous economic laws underpinning the dynamic of local economic systems.

Finally, it should be pointed out that the *milieu innovateur* theory remedies the limitation intrinsic to the theory of knowledge spillovers; it explains the channels through which knowledge disseminates, not in terms of pure probability of contact, but rather in those of well-evidenced economic-territorial phenomena – supplier/customer relations, high local labour turnover, and spin-offs. The theory accordingly returns

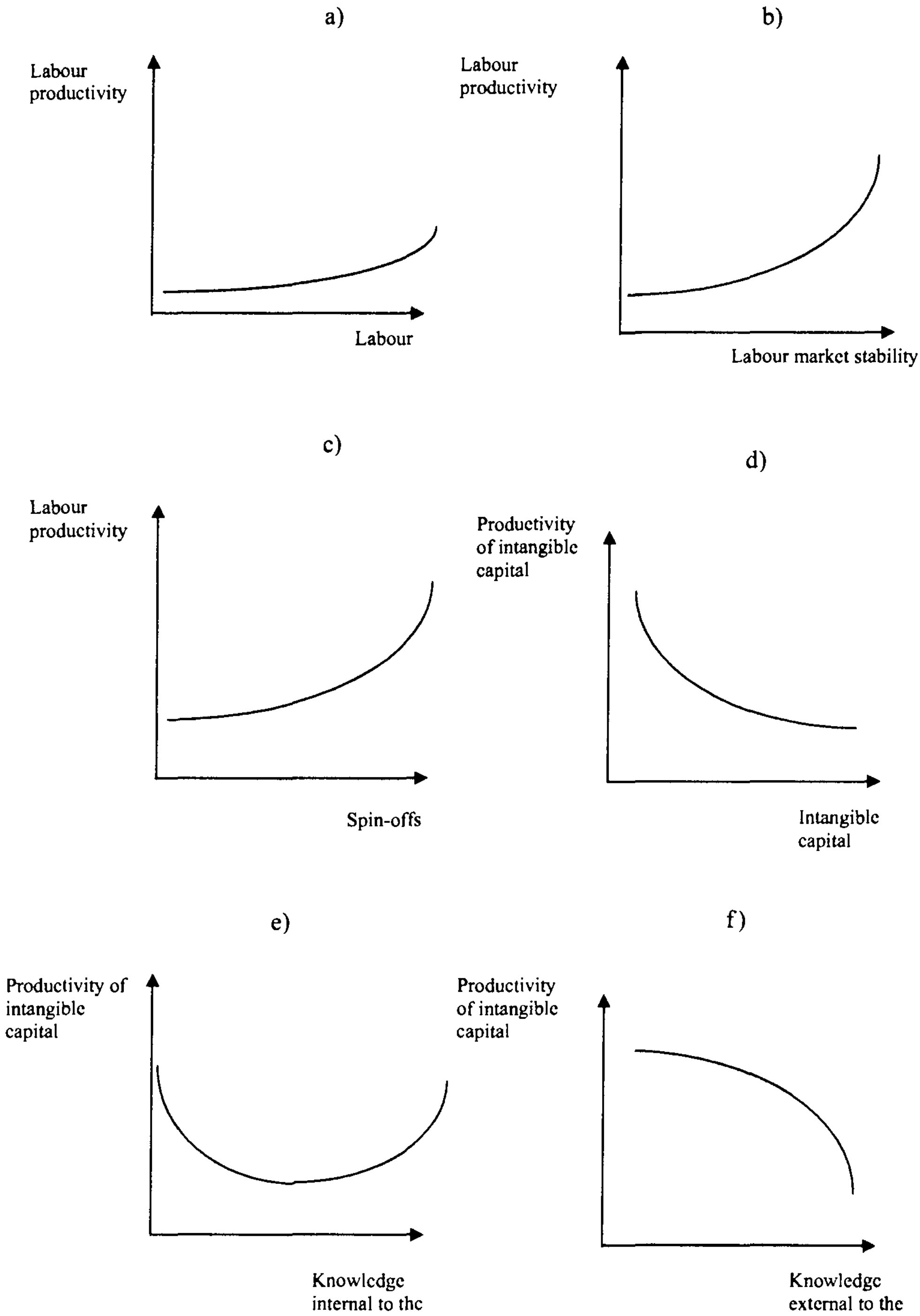


Figure 9.1 Collective learning and factor productivity
Source: Capello (1999b)

territorial factors to centre-stage in analysis of an area's endogenous innovative capacity, and it reinstates space to the active role in the economic dynamic which the theory of the Marshallian industrial district attributed to it within a framework of pure static efficiency.

9.4 The 'learning regions' and regional innovation systems: institutional proximity

9.4.1 *The 'learning region' theory*

The theory of the *milieu innovateur* has been paralleled by the international development of wide-ranging analysis of the endogenous factors at the basis of local innovative capacity. This approach has shifted its attention to institutional aspects, and specifically to the set of social, economic and cultural rules embedded in a territorial setting. It originated with the Danish school of Aalborg, and in the works of its founder, the economist Bengt-Åke Lundvall, and it has subsequently been widely adopted, mainly in the UK and the USA.[20]

The main components of this approach can be summarized as follows. The principal resource of modern economies is knowledge. Consequently, the principal processes on which an economy's competitiveness depends are learning and the acquisition of knowledge. Moreover, the complexity and systemic nature of innovation, and the brevity of the product life-cycle characteristic of technological change in recent years, entail that learning is an interactive process. Put otherwise, learning springs from co-operation and interaction between firms and the local scientific system, between different functions within the firm (between production and research and development, between marketing and research and development), between producers and customers, and between firms and the social and institutional structure. The feedbacks, interdependencies and complementarities among the various functions internal to the firm, and between the firm and external actors, required by the innovative process evince the need for co-operative and interactive forms of organizational learning. Finally, innovation is increasingly the result of an informal learning process, based on direct experience or that of others, which comes about through activities focused on finding solutions to specific technological, productive or market problems.

The consequence of these various features is that the innovative process is strongly localized: it results from the variety of traditions, norms, habits, social conventions and cultural practices that constitute what has been called 'institutional thickness'.[21] Innovation therefore cannot be understood properly unless it is examined within the socio-cultural and institutional context in which it takes place. In areas where there is 'institutional proximity' – meaning the set of norms, codes and rules of behaviour which help economic actors (people, individual firms, public and private institutions) to adopt forms of organization that facilitate interact learning – the innovative process comes about more rapidly and gives competitiveness to the economic system.[22]

A 'learning region' is in this sense:

- a region in which norms of social and institutional behaviour support interactive learning: the horizontal organization of corporate functions, co-operation and agreements among firms, and co-operation between firms and research centres, facilitate the exchange of knowledge and support the learning process;

- a region with an 'organized market' in which implicit and generally shared rules of behaviour guarantee the tacit exchange of information and the creation of knowledge. These come about through an interactive and creative learning process centred on the implementation of new products, and new technological solutions. Likewise, a 'destructive' learning process (a 'creative disruption') teaches the system to abandon obsolete and inefficient technological trajectories, and guides it through the profound transformation brought about by the technological – all the more so organizational – changes imposed by a new technological trajectory.[23]

In short, a 'learning region' is a socio-economic system able gradually to develop forms of interactive learning. It is on this 'learning ability' that a region's competitiveness depends. It is accordingly a concept that identifies the condition necessary for an economic system's competitiveness as a process (learning) more than a state (the stock of knowledge).[24]

Despite its necessary abstraction, which makes its empirical application impossible, the concept of 'learning region' has gained general consensus, not just in a particular scientific community (that of Britain and North America) but also at institutional level, given the European Union's need to devise new policy instruments with which to support regional cohesion. However, the results obtained when the concept of 'learning region' is translated into regional economic policies are rather perplexing: the interventions proposed concern the creation of education and training services, incentives for learning, the sharing of successful experiences in creating organizational forms to support interaction, and financial aid to firms undergoing corporate restructuring: all of which are interventions in support of weak regions which are well known and already applied in the past.

Also it should be stressed that there is an apparently major weakness in the theory. Although it envisages a system of homogeneous socio-economic and institutional conditions in the region, and interaction and co-operation among actors, it is nevertheless markedly aspatial. Nothing in the theory explains how and why these relations must necessarily be local; nor does it explain what territorial conditions must be in place for the 'organized market' to arise; or what territorial factors fuel the process of interactive learning.

These shortcomings are all the more evident when one considers that the concept of the 'learning region' is derived from that of the 'learning economy', and that the concept of 'learning economy' is in its turn used to denote a '*national* system of innovation' where the set of institutional rules and norms allow, strengthen and emphasize forms of interactive learning. The fact that the concept can be shifted among different territorial levels of analysis demonstrates its aspatial nature.

9.4.2 *Regional innovation systems*

Strongly rejecting the Schumpeterian idea that innovation is a linear process consisting of different and successive temporal phases of creation and transformation of knowledge into a tradable idea, the theory of regional innovation systems (RIS) embraces the opposite view: that innovation is the result of an interactive and non-linear process, of a complex system of feedbacks and retroactions among actors, and that the success of innovation resides in the capacity to merge new technological/organizational solutions with the potential demand for new products and production processes.

In fact, the RIS theory centres its analysis on the relation between two sub-systems of local actors: the system that creates and diffuses knowledge (the 'infrastructure system'), consisting of universities, public and private research laboratories, on the one hand, and the system of local firms, potential adopters (the 'business system') on the other.[25] The success of the local innovative activity lies, according to this theory, in the match and in the constant synergies between the new and creative technological solutions proposed by the 'infrastructure system' and the real technological needs of local firms. This view is based on local institutional elements, in particular norms and behavioural codes, that support co-operation, already highlighted in the 'learning region' theory.

The RIS theory brings to the fore institutional proximity as a factor that reduces the risks and uncertainty associated with the innovative process. This result is achieved without using economic-territorial elements that may emphasize the local nature of the process. Moreover, the supporters of the RIS theory strongly underline the importance of the reinforcement of both the sub-systems that compose an RIS: inefficiency and lack of development of one of the two sub-systems is interpreted as the source of weaknesses in the local innovative capacity. In this framework, innovation policies must be devoted to remedying this weakness.[26]

In regard to these normative suggestions, some words of caution are necessary: innovation may arise from external knowledge, and it may even be the result of an activity not necessarily based on research and development. The imposition of developing formal knowledge-creation activities in all regions means pushing all of them towards the same model of innovation, a strategy now widely recognized as unsuccessful.[27]

It should also be borne in mind that the 'regional innovation system' was born of the concept of 'national innovation system'.[28] The possibility of shifting the concept from one geographical level to another testifies to its necessarily aspatial nature, and the intrinsic impossibility of deducing the endogenous elements that underpin processes of territorial innovation from a theory like this one.

9.5 Evolutionary economic geography and the concept of 'related variety': cognitive proximity

Towards the end of the 1990s, a new stream of thought arose, taking the name of 'Evolutionary economic geography'. Its distinctive feature was that it centred interpretation of the dynamics of local areas on analysis of the birth and death of firms in a historical-evolutionary perspective. The innovative and locational choices of firms were analysed in a context of bounded rationality and interpreted within a theory in contrast with the assumption of perfect information of neoclassical location theories, and with the inductive approach of institutional economic geography.

The evolutionary nature of this theory leads to the description of innovation and new knowledge development as resulting from a creative process of discovery developed around existing competences, within specific technological paradigms, and along specific trajectories.[29] Limited by bounded rationality, firms are strongly influenced by their history, which influences both their innovative activities and their location choices. For this reason, new firms, usually spin-offs of other local firms, limit the uncertainty of their choice by locating in the same area; already-existing firms show an evident location inertia due to the long and stable relationships that they have established with local suppliers and customers. The result of locational decision-making processes of both new and already established firms is a concentration of firms belonging to the same industry in the local area.

For the first time in the history of spatial clusters formation, industrial concentration is not interpreted as a source of location advantages, but as a result of evolutionary processes that follow highly structured and organized routines intended to reduce the uncertainty that accompanies decisions in bounded rationality conditions.[30]

Spatial concentration (or its absence) is not only the result of a historical industrial process but also determines its future evolutionary trajectories. In fact, the probability that local firms can survive depends, in this perspective, on their capacity to exploit the information present in the area. This capacity, in its turn, depends on the existence of a common knowledge base within the industry.

It is in this way of reasoning that the concept of cognitive proximity emerges as an element crucial for explaining innovation capacity. In order to innovate through the knowledge that exists at the local level, it is necessary for firms to be endowed with the complementary knowledge necessary to be creative and generate new and innovative technological solutions. All this, however, must take place on a common knowledge base that guarantees a common language and mutual understanding among firms. In the literature, this condition is labelled 'related variety', and it is defined as a variety of interrelated technological solutions with a common knowledge basis.[31]

Although the concept of cognitive proximity was developed to explain local context formation, it has been applied to all forms of co-operation among firms, also long-distance co-operation. Interregional knowledge follows. This is generally formed through firms' networks and requires a cognitive proximity between firms to generate innovative projects in co-operation. Also in this case, the theory of cognitive proximity suggests that the greater the technological variety between two regions within a larger macro-industry, the greater the benefit that these regions obtain from the exchange of knowledge.

Through implementation of a 'related variety' indicator at the regional level based on patent activities and a disaggregation of technological classes (e.g. five-digit disaggregation) within a larger technological class (e.g. two-digit disaggregation), a number of empirical analyses have identified a positive relationship between the degree of 'related variety' and an area's growth rate.[32]

Despite the continuing success of the concept of cognitive proximity, some critical reflections on its advantages and shortcomings are necessary. The concept has certainly the great advantage of overcoming the simple idea – first propounded by Hägerstrand and then re-launched by the knowledge spillover theory – that the pure contact probability among actors can explain an exchange of knowledge. Moreover, the cognitive proximity concept enriches the concept of 'absorptive capacity' introduced into the literature to explain the differing capacities to exploit knowledge of actors localized in the same area, and interpreted rather poorly as pure technological advancement.[33] The concept of cognitive proximity encompasses more subtle cognitive elements: by simultaneously imposing a knowledge complementarity and a common knowledge basis, it identifies cognitive capacity on the basis of both the specific technological knowledge of single actors and the common aggregate knowledge of the area.

Even if we recognize the merits of this concept, it has an important intrinsic limitation: that of reducing cognitive aspects to the industry dimension. A perspective of this kind separates the learning process from the economic and territorial context, and imposes a step backward with respect to the theory of *milieu innovateur*, which highlighted the elements in the local context that make it possible to generate a collective learning process.

9.6 General remarks on the concepts of proximity

The complementary aspect of the diverse elements at the basis of knowledge transfer emerges from the different theories presented: from context conditions in the form of social and institutional rules and governance that push towards co-operation, through behavioural and subjective (relational) elements that facilitate the involuntary exchange of knowledge, to cognitive elements linked to a constructive co-operation based on a background of common knowledge. The presence of these proximities generates different positive externalities that reinforce knowledge transfer: a reduced risk of opportunistic behaviours, the limitation of uncertainty, reduced transaction costs and common understanding of technological aspects (Table 9.3).

Complementarities among the different concepts of proximity are also clear, notwithstanding some conceptual overlaps. Knowledge transfer requires at the same time relational capacity among actors, norms and rules of behaviours, and mutual trust (a condition guaranteed by the interaction between relational and institutional proximities). Relational capacity is reinforced by a cognitive map shared by actors (interaction between relational and cognitive proximities). Finally, the exchange of complementary knowledge within a common knowledge basis is facilitated by rules and social norms that punish free-rider behaviour (interaction between cognitive and institutional proximities) (Figure 9.2).

It should be borne in mind that the effects generated by the various proximities exhibit positive and negative non-linearities, as happens in all synergic processes.

Table 9.3 A comparison among the different concepts of proximity

Types of proximities	*Definition*	*Channels of knowledge transfer*	*Positive externalities associated with proximities*	*Risks associated with too much proximity*
Relational	High degree of relationality of local actors	Economic-territorial elements (suppliers-customers relationships; spin-off, specialized labour market)	In-voluntary exchange of knowledge Reduced risk of opportunistic behaviour and limited uncertainty	Risk of lock-in within local knowledge
Institutional	Rules and behavioural codes common to all local agents	Macroeconomic environment supporting co-operation	Reduced transaction costs	Institutional inertia
Cognitive	Shared knowledge	The right mix of industries	Common understanding of technological aspects	Risk of lock-in within industry knowledge

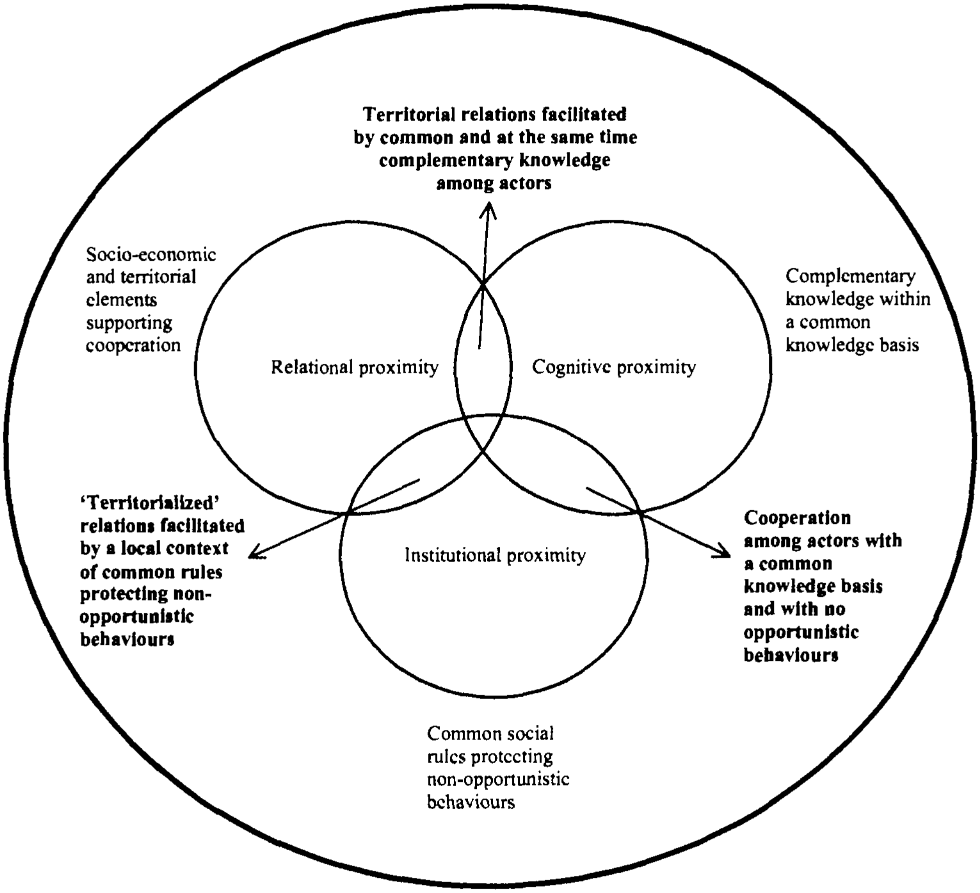


Figure 9.2 Advantages from the presence of different proximities

Whilst the simultaneous presence of the various proximities may multiply the advantages obtained, it may also multiply the risks. Accordingly, decreasing – or even negative – returns may arise. As theorized by the *milieu innovateur* approach, the presence of only local relationships risks keeping knowledge within the same technological paradigm, thus reducing local creativity and the achievement of external flows of new knowledge. Codes and rules strongly embedded in the local society may generate institutional inertia. Finally, a too high cognitive proximity risks generating lock-in within the same industry knowledge (Table 9.3).[34] Innovative processes, especially those of a radical nature, may be reduced or even become null.

Although the different definitions move towards a concept of relational space, they remain anchored to a metric and geographic vision of space; and they are far from a modern vision able to synthesize the concept of 'territory'. Only within the Marshallian theories of local districts and *milieu innovateur* can geographical space become a social space (which takes the name of 'territory') where socio-economic relationships, competences and culture, history and identity merge and thus generate competitive advantages for the local community. Only in the *milieu innovateur* theory, and in some recent hermeneutic approaches to local creativity, are physical places, identitarian relationships and collective learning processes highlighted as the sources of continuous innovative processes.[35] The natural loci for these processes are identified in urban areas. Only with an approach like this one can space be given an active role in innovation processes.

9.7 Regional patterns of innovation

Recently, the idea that the sectoral dimension is not sufficient to explain the regional innovation capacity has been strongly advocated. When applied at regional level, in fact, the sectoral logic shows two kinds of limits. The first one refers to the fact that this logic pushes towards the interpretation of formal knowledge as the main source of innovation; instead, we have previously seen how different theories have identified the source of knowledge in a variety of informal elements, like face-to-face meetings, informal co-operation, creativity, collective learning processes. The second limit refers to the idea that only knowledge stemming from local sectors is a source of local innovation; instead, a large literature has highlighted the role of knowledge coming from outside the region as a fundamental source of innovation.

In order to overcome such limits and return the local characteristics to the centre of the explanation of innovation processes, a new concept has recently been proposed, that of regional patterns of innovation. This concept interprets the different modes of innovation as the result of the presence/absence of contextual conditions necessary to create knowledge and to translate knowledge into innovation.[36]

In fact, the concept of territorial patterns of innovation is proposed and defined as a combination of territorial specificities (context conditions) that lie behind different modes of performing the different phases of the innovation process. In particular 'territorial patterns of innovation' consist in spatial breakdowns of variants of the knowledge → invention → innovation → development logical path built on the presence/absence of territorial preconditions for knowledge creation, knowledge attraction, and innovation.

The concept of territorial patterns of innovation therefore lies on a logical sequence between knowledge, innovation and economic performance; it is therefore drawn in the abstract but consistent Schumpeterian 'linear model of innovation', even if heavily criticized as unrealistic, and rooted in the idea of a rational and orderly innovation process.

The local conditions are integral part of the innovation mode, and are interpreted in this approach both as material elements, in the form of functions for the creation of knowledge (R&D laboratories and universities), and non-material, intended as the relational capacity of local actors. The pure existence of knowledge creation functions is not sufficient to guarantee an innovation process to occur; and it is not even the necessary condition, since knowledge can be acquired from outside. In order to explain this last aspect, the present theoretical framework relies on the most recent theories on proximities presented above.

For what concerns the territorial specificities (context conditions) that are behind each phase of the innovation process, this theory takes advantage of the vast and articulated literature that takes territorial elements into consideration in innovation processes, namely theories:

- concerning *knowledge creation*: human capital and education in general, universities and R&D activities, presence of an urban atmosphere have been considered, in a variety of approaches, as the territorial preconditions for endogenous knowledge creation in the vast literature that was developed during the 1980s (Section 8.3);

- concerning *knowledge diffusion*, exploiting all concepts of proximities presented above (Section 9.5);
- concerning *knowledge receptivity and the capacity to turn knowledge into innovation*: local interaction and co-operation in order to achieve reduction of uncertainty (especially concerning the behaviour of competitors and partners) and of information asymmetries (thus reducing mutual suspicion among partners); trust, sense of belonging, place-loyalty and social sanctioning in order to reduce opportunistic behaviour are all territorial elements, typical of the innovative *milieux*, that increase the capacity of a region to speed up innovation and take full advantage of collective learning processes and entrepreneurial activity (Section 9.3).[37]

Exploiting the different theories, the regional patterns of innovation approach provides a deductive framework to interpret how the different phases of the innovation process are put together at the spatial level, and why some of them take place in certain areas and others do not.

Among all possible combinations of innovation modes and territorial elements, the 'archetypal' ones may be indicated in the following, each of which reflects a specific piece of literature on knowledge and innovation in space:

- an *endogenous innovation pattern*, where local conditions fully support the creation of knowledge, its local diffusion and transformation into innovation and its widespread local adoption. Given the complex nature of knowledge creation nowadays, this pattern is expected to show a tight interplay among regions in the form of international scientific networks. From the conceptual point of view this advanced pattern is the one considered by most of the existing literature dealing with knowledge and innovation creation and diffusion (Figure 9.3);
- a *creative application pattern*, characterized by the presence of creative economic actors interested and curious enough to look for knowledge outside the region – given the scarcity of local knowledge – and creative enough to apply external knowledge to local innovation needs. This approach is conceptually built on the literature on regional innovation adoption/adaptation (Figure 9.4);
- an *imitative innovation pattern*, where the actors base their innovation capacity on imitative processes, that can take place with different degrees of adaptation of an already existing innovation. This pattern is based on the literature dealing with innovation diffusion (Figure 9.5).

Figures 9.3–9.5 show in a stylized way the three regional patterns of innovation envisaged before. As these figures show, regional modes of innovation are more complex than the simple core–periphery distinction suggested at the end of the 2000s; the latter was encouraging core regions to be the natural places for general purpose technologies, which can achieve a critical mass of scientists and knowledge able to achieve

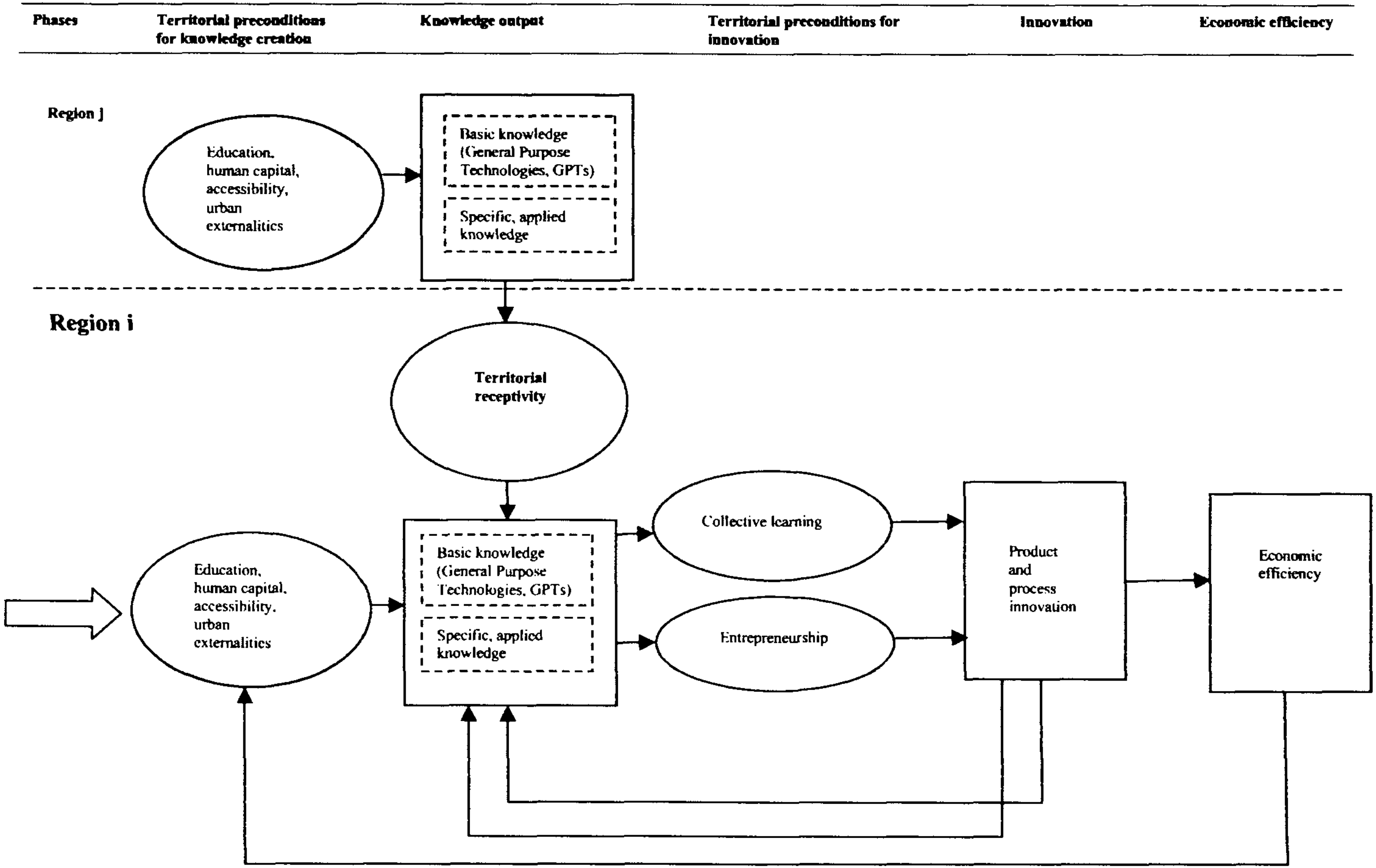


Figure 9.3 An endogenous innovation pattern

Source: Capello (2012)

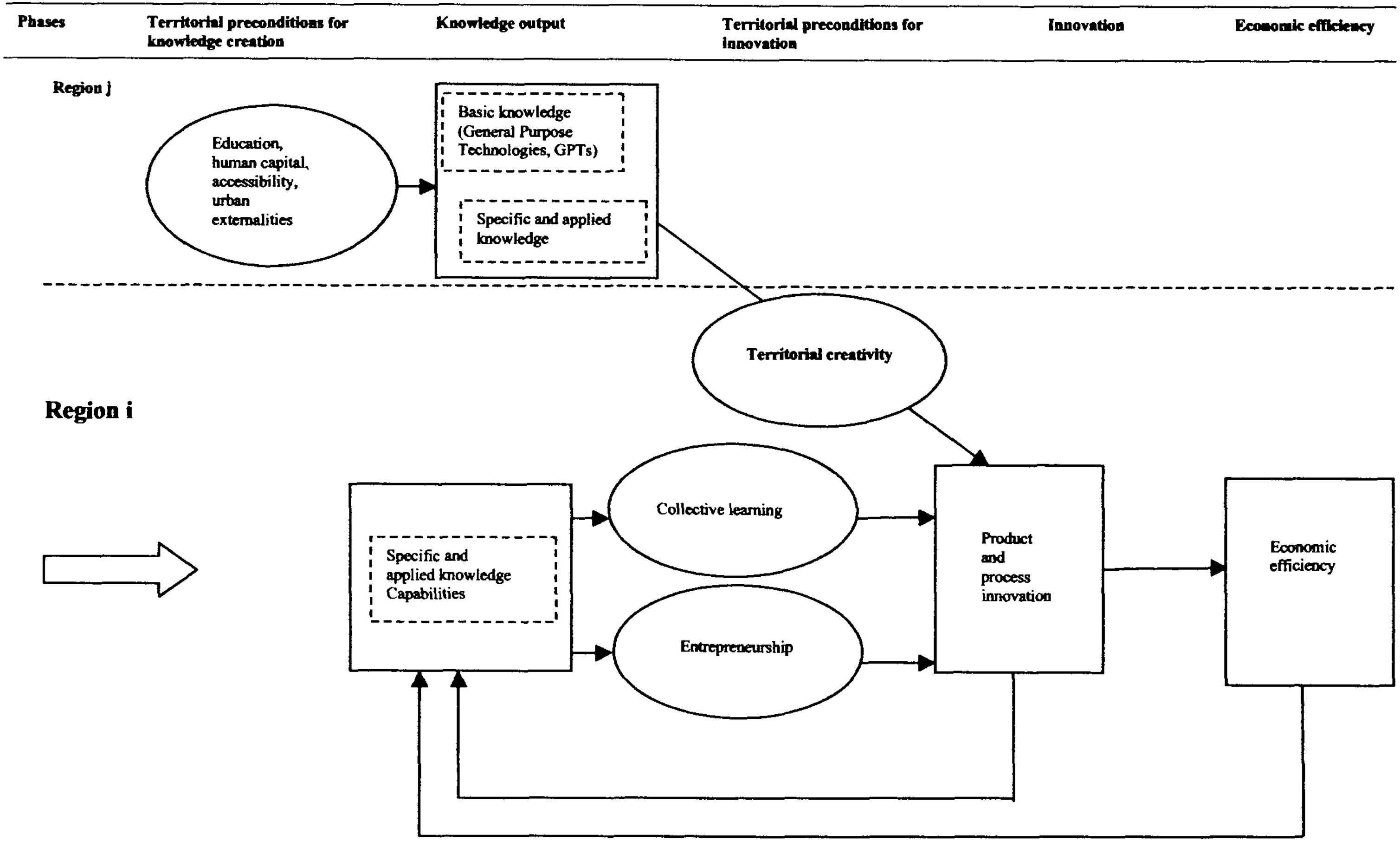


Figure 9.4 A creative application pattern
Source: Capello (2012)

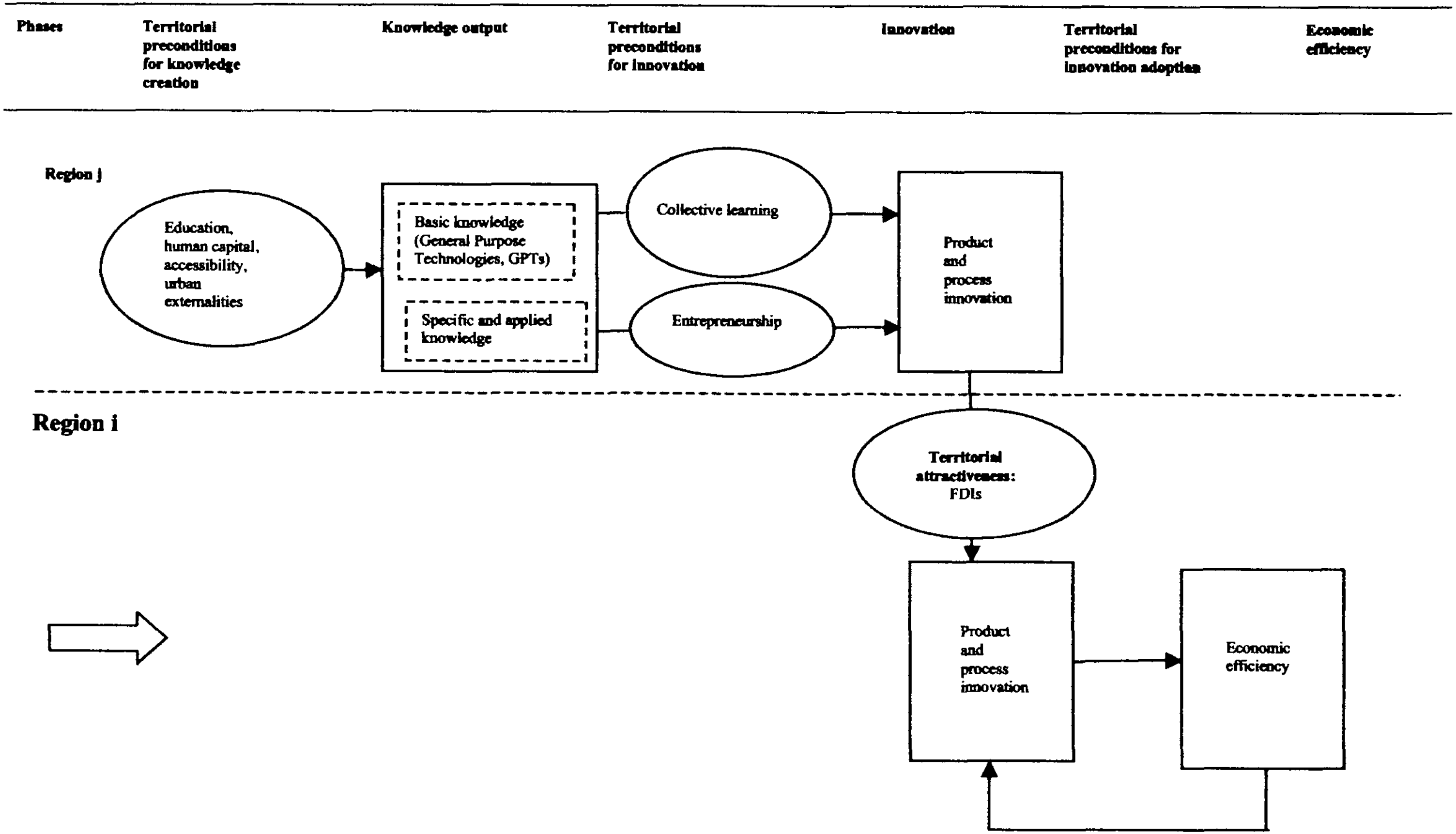


Figure 9.5 An imitative innovation pattern

Source: Capello (2012)

increasing returns to R&D, giving 'peripheral' regions the role of co-inventors of applications in their technological domain. Moreover, the approach in terms of regional patterns of innovation highlights territorial elements (context conditions) in the explanation of innovation processes.[38] These context conditions are much more articulated than the mere presence of R&D activities, which is generally applied to measure the absorptive capacity of a region.[39]

Supported by empirical evidence,[40] this approach shows that the way towards innovation and modernization is very different among regions, and it derives in a clear way from local specificities. The variety of innovative modes that the empirical analysis reports highlights how misleading is a common normative strategy for all regions in Europe, such as the achievement of the 3 per cent target of R&D investments over GDP, suggested by the official documents of the EU Lisbon and Europe 2020 Agenda. On the contrary, thanks to its taxonomy, the approach of the regional innovation mode calls for ad hoc interventions with the aim of supporting, strengthening and diversifying the virtuous aspects of each regional innovation process. It is on these local innovative profiles that the European strategy of modernization and innovation (the so called 'smart innovation strategy'[41]) finds a strong and appropriate base for its implementation.

9.8 Modern innovation policies: the smart specialization strategy

9.8.1 A new policy design and its advantages

The smart specialization approach was developed with the aim to find an explanation – and a consequent rational strategy – for the large R&D gap between Europe and some key trading partners. The most straightforward reason for the knowledge gap was outlined in the smaller share of European economy composed of high-tech, R&D intensive sectors. A second reason for the gap was pointed out in the spatial dispersion of the limited R&D efforts, generating insufficient critical mass and investment duplications, inefficient resource allocation, and consequent weak learning processes.[42]

On the basis of this diagnosis, a rational and concrete proposal was put forward by the 'Knowledge for Growth' expert group. It advocated differentiated policies for 'core' and 'periphery' regions, the former able to host laboratories and research activities on general purpose technologies (GPT), the latter oriented towards the identification of their 'knowledge domain' in which to specialize and towards co-operation with external R&D providers ('co-application of innovation').[43]

In more recent formulations, the smart specialization strategy has been translated to a regional setting – with problems, however – and extended to encompass all strategies intended to build regional competitiveness through the design and implementation of innovation; where innovation is understood in a broader sense than mere R&D (as it was in the original formulation) and as embracing creative industries, social and service innovation, new business models, and practice-based innovation.[44]

The strategic importance of the concept consists especially in the fact that implementation of the smart specialization strategy is interpreted as one of the conditions that a region must fulfil to receive funding from the European Regional Development Fund, the main funding stream within EU Cohesion Policy. Regions are required to develop their innovation strategies on the basis of their technological specialization, or – in the words of the smart specialization strategy's experts – on the basis of their *technological domains*, by which is meant the technological fields in which regions are specialized and to which regional policies should be tailored to promote local innovation processes.[45] Regional policies should be tailored to the regions' technological domains in order to promote local innovation processes in their fields of specialization. To achieve this aim, the main ingredients in translating the a-spatial smart specialization concept into a spatial setting are embeddedness of innovation policies in local knowledge-transmission linkages; relatedness supporting a process of strategic specialized diversification of sectors; and connectedness between sectors and firms which enables knowledge to diffuse.

As the experts of smart specialization strategy claim, the originality (and strength) of the concept lies in this highly innovative approach to regional innovation (competitiveness) policies. The concept in fact comprises some conceptual pillars that run counter to traditional regional innovation policies, namely:

- innovation is not merely associated with R&D, so that innovation policies should not be focused primarily on high-tech sectors, and on R&D investments. This goes against the 'one-size-fits all' innovation policy;
- smart specialization strategy does not intend to encourage a culture of 'picking winners' on a sectoral basis; rather, it pushes towards public–private partnership processes of 'entrepreneurial discovery' and learning;
- advocated in this regard is a bottom-up approach based on the self-discovery of entrepreneurial capability, thus superseding the old policy style calling for centralized planning methods to identify industrial development priorities. This policy approach is in general demand-driven because it is derived from local potentials and local needs;
- the bottom-up nature of the policy style ensures that the logic and design of the policy is appropriate for and relevant to the local regional context, rather than being imposed by a supra-regional body. It is therefore a true place-based policy, as advocated by the Barca Report;
- the endeavour to identify real potential and real development priorities is also a way to engender a policy-prioritization aimed at fostering growth which is realistic for the context concerned, and which can be explicitly applied in the regional context.

9.8.2 *The risks of the new policy design*

The above conceptual pillars distinguish this approach from the traditional old-style innovation policies. These features are innovative and modern, efficient and shareable, and they constitute a cultural leap in the design and implementation of innovation (and competitiveness) policies, moving away from an R&D-based policy that over past years has demonstrated all its fragility and inefficiency. At the same time, however, the new policy style also contains aspects and novelties that are not easy to put in place, and require additional thought for their successful implementation to overcome the risks that accompany this new policy style.

A first risk is related to the fact that local preconditions for innovation may be lacking. The main critical issue concerns the real capacity of regions, especially lagging ones, to put in place a self-discovery process; that is, a bottom-up strategy for the identification of their strengths and opportunities. Lagging regions in general lack the key elements necessary for a smart specialization strategy to be effective: they lack connectedness, entrepreneurial spirit, size in terms of market potential, industrial diversity, quality of local governance and a critical mass of capabilities to develop collective learning processes. All of these are elements that a successful entrepreneurial search process requires, and the smart specialization strategy runs the risking going against objectives of regional cohesion policy if it is not carefully monitored.

The second risk regards the difficult policy-prioritization. In the case of regions where the potential for innovation exists thanks to the presence of a variety of sectors, of entrepreneurial spirit and of enough human capital to spread the advantage of individual innovative activities around the local area so as to generate collective learning, the problem of the entrepreneurial search process lies in the criteria with which to select (and co-ordinate) among the bottom-up projects that may be proposed, and the domains in which to concentrate public resources. At the theoretical level, the concept of related variety – which refers to the variety of industries in a region that are cognitively related (see Section 9.5) – has been identified as a possible way to identify opportunities for regions to diversify into new industries. The higher the degree of related variety, the more learning opportunities are available at the local level, the more knowledge spillovers across industries occur, and the higher the regional growth. On this reasoning, policies should support discoveries that can actually build on, and are embedded in, existing related resources at regional level, and this is a basis for policy-prioritization.[46]

Another risk is that of misallocation of public resources and unlikely local strategies. A bottom-up process of strategy design carries high risks of misallocation of public resources whenever local interests and local political needs may set unfeasible industrial targets and risky innovation strategies. One suggestion for dealing with this limitation is to reason on a common policy design for regions with similar types of innovation modes. To this end, a sound taxonomy of innovative regions is required, a taxonomy which moves away from simple knowledge creation indicators (traditional patents and R&D indicators), and is able to capture the different innovation modes that regions are actually developing thanks to the presence of specific local preconditions for knowledge and innovation creation. In this regard, the regional 'innovation patterns' presented above are in my view a good way to build a useful taxonomy of innovative regions; they may be found empirically in the way knowledge and innovation are developed within individual regions according to the nature of their traditional knowledge base and productive specificities, and/or are captured from other regions via co-operation, the mobility of scientists and professionals, market procurement and trans-regional investments. The identification of territorial patterns of innovation leads to the suggestion of 'smart innovation policies'. These are defined as policies able to increase the innovation capability of an area and to enhance local expertise in knowledge production and use by acting on local specificities and on the characteristics, strengths, and weaknesses of already-established innovation patterns in each region.

Another risk associated with the smart innovation strategy is that of lock-in with respect to local historical specialization. The self-discovery process goes against technocratic approaches claiming that they can define priorities, objectives and targets on the basis of scientific techniques, and that they can identify which knowledge and inter-industry spillovers should be implemented and supported. However, this process is necessarily guided by routines and competences at the organizational level that make search behaviour localized (Nelson and Winter, 1982), not only in the industrial sense (i.e. within a specific technological paradigm), but also in the regional sense. It targets existing regional specializations and pushes towards possible future diversification processes. This strategy entails the risk of lock-in.

How can these risks be prevented? Are the proposals mentioned acceptable and shareable? The debate on these questions is still open and calls for additional insights.

9.9 The concept of territorial capital

9.9.1 *The definition and the taxonomy*

This chapter and the previous two have concentrated on the elements considered to be at the basis of local development. They are of diverse nature: material (infrastructure, presence of large firms) or non-material (knowledge, creativity, entrepreneurship, social capital), public (transport and energy infrastructure) or private (financial and productive capital), generated by endogenous (development of local creativity and knowledge) or exogenous (multinationals, investments, investments of the public sector) processes.

The variety of elements considered to generate a local development pattern has recently induced development of a synthesis concept labelled 'territorial capital', which is defined as *all local, tangible and intangible, endogenous and exogenous, assets, of public and private nature, that constitute the development potentials of an area.*

The concept of 'territorial capital' was first proposed in a regional policy context by the OECD in its Territorial Outlook, and it has been recently reiterated by D.G. Regio of the Commission of the European Union:

> Each Region has a specific 'territorial capital' that is distinct from that of other areas and generates a higher return for specific kinds of investments than for others, since these are better suited to the area and use its assets and potential more effectively. Territorial development policies (policies with a territorial approach to development) should first and foremost help areas to develop their territorial capital.[47]

Launched in a scientific context by Roberto Camagni,[48] territorial capital warrants closer inspection in order to draw up a taxonomy of all potential sources of development. The proposed taxonomy is built upon two main dimensions, chosen so as to identify the economic nature of each component of territorial capital and, consequently, the laws of accumulation and depreciation of each component (Figure 9.6):

- *rivalry*, which makes it possible to identify whether the territorial capital asset can be used only by an individual (private good) or by a specific group of people (impure public goods – available to everybody, but there is rivalry in their use since they are subject to congestion and scarcity – or club goods, available for a specific group of people that can make use of them without rivalry in their use), or available to the whole community (public goods);
- *materiality*, which makes it possible to identify a good according to its physical or intangible nature: tangible goods, intangible goods, and an intermediate class of mixed, hard–soft goods are identified.

Rivalry	Tangible goods (hard)	Mixed goods (hard + soft)	Intangible goods (soft)
High rivalry (private goods)	Private fixed capital stock Pecuniary externalities (hard) Toll goods (excludability) *c*	Relational private services operating on: - external linkages for firms - transfer of R&D results University spin-offs *i*	Human capital: - entrepreneurship - creativity - private know-how Pecuniary externalities (soft) *f*
(club goods) (impure public goods)	Proprietary networks Collective goods: - landscape - cultural heritage (private 'ensembles') *b*	Co-operation networks: - strategic alliances in R&D and knowledge - public/private partnerships in services and schemes Governance of land and cultural resources *h*	Relational capital: - co-operation capability - collective action capability - collective competencies *e*
(public goods) **Low rivalry**	Resources: - natural - cultural (punctual) Social overhead capital: - infrastructure *a*	Agencies for R&D transcoding Receptivity enhancing tools Connectivity Agglomeration and district economics *g*	Social capital: - institutions - behavioural models, values - trust, reputation - associationism *d*

Materiality

Figure 9.6 A theoretical taxonomy of the components of territorial capital
Source: Camagni (2009)

Thanks to these two dimensions, it is possible to go beyond the traditional classification of potential productive resources based on social capital, human capital, infrastructure and productive capital (fixed private capital), which are located at the four corners of the matrix. In fact, the matrix shows an intermediate class, embracing club and impure public goods, which characterizes the group of innovative elements stemming from the most recent theories. On the one side, we find networks, of material nature in the case of ICTs, but also of non-material nature in the case of co-operative networks and strategic alliances among firms for the production of new products and services. On the other, we find public goods that are subject to congestion and to depreciation in the presence of free-rider behaviour. The peculiarity of this category of goods lies in the fact that it requires new forms of governance, inclusive and in partnership, that guarantee the maximum advantage for the members of the 'club'. Moreover, a category of mixed goods emerges. This is characterized by both materiality and non-materiality, and its presence makes it possible to underline the importance of those complex territorial organizations like cities and industrial districts – as explained in detail in Chapter 8 – that are sources of agglomeration economies and generate growth for the entire community.

9.9.2 *Laws of accumulation and depreciation of territorial capital*

The taxonomy of the various components of territorial capital is not only a means to summarize all potential assets that can influence local development. The proposed taxonomy allows identification of the specific economic nature of each component, and the consequent accumulation and depreciation processes that accompany the life cycle of each asset. This is a fundamental aspect for defining the appropriate strategies for use of these resources, ensuring their protection and their valorization in the long run. Private and public, as well as tangible and intangible, goods are subject to different laws of accumulation and depreciation. The accumulation of material public goods, such as infrastructure, but also cultural and natural resources owned by the government, largely depends on the quality of their governance: the public sector is responsible for the crucial functions of control over these goods, the purpose being to maintain their potential benefits for a long period of time, avoiding their depreciation and their destruction.

The accumulation in a local context of a private material good, like labour or capital, today depends on its local anchorage. In modern economies, in fact, the hypermobility of production factors like capital and labour is avoided only if these factors are strongly anchored to the local environment through the presence of other less mobile factors, like knowledge and social and relational capital.[49] The latter factors, of an intangible nature, are embedded in the local society; impossible to transfer elsewhere, they become sunk costs.

As regards intangible private goods, like knowledge, these accumulate at the local level through education processes, research investments, co-operation among firms, individual and collective learning. In an indirect way, knowledge can develop through spillover effects from the research centre where it is formed because of the mobility of researchers and skilled technicians, as well as imitation and co-operation processes. The accumulation over time of knowledge at local level depends closely on the continuity and persistence of the actors that participate in its generation. Today, there are high risks of knowledge 'de-cumulation' due to the local firms' outsourcing of phases of the production processes. In the short term, this can disrupt integration and synergy processes within firms and within the local area, and impoverish the flow of technological creativity and knowledge production.

9.9.3 *Endowment and efficiency of territorial capital*

The vast and composite mosaic of local success stories and of dynamic evolutionary trajectories can be explained by diverse elements of territorial capital, endogenous and exogenous, qualitative and quantitative, traditionally functional (based on the presence of production factors and of preconditions for local efficiency) or relational in nature.

Regions do not require the presence of all the above components of territorial capital to develop and maintain over time a positive and dynamic development trajectory. The local endowment of specific assets of territorial capital results from the history of the local area, and determines its productive specializations on which a strategic growth pattern is to be built.

Most of the empirical analyses measuring the various components of territorial capital at regional level show that there is a decisive difference between the endowment of territorial capital elements and development levels; for example, the North of Italy has a much higher endowment of all territorial capital assets than the Centre and the South of Italy, this last possessing the lowest endowment of all territorial capital assets.[50] However, when the analysis is developed at a more disaggregated territorial level, a different picture emerges. An analysis conducted at provincial level (NUTS3) depicts the following situation:

- metropolitan provinces, which are endowed with above-average levels of territorial capital components;
- provinces mainly endowed with intangible elements like social and relational capital;
- provinces mainly endowed with non-material elements like transport, energy and educational infrastructure;
- finally, provinces that lack an endowment of all territorial capital elements.

Provinces that belong to one of these clusters do not have a specific geographic location; the soft elements of territorial capital, like social capital, are mostly present in the northwestern part of Italy, while the northeastern part has none. Central Italy, generally presented as a uniform socio-economic area, has provinces that belong to all four clusters. The provinces of the South of Italy, with the exception of the metropolitan provinces, mostly belong to the fourth cluster.

Even more interesting is that the same analysis shows that the simple endowment of territorial capital assets is unable to explain the rates of growth of those areas; whilst the northern part of Italy has the highest endowment of territorial capital assets, it does not record the highest growth rates. This testifies that what makes the difference in terms of growth is the efficiency with which these assets are used. The greatest efficiency in the exploitation of territorial capital resides in the integration of tangible and intangible elements, which reinforce each other.[51]

As in the case of all economic resources, the efficiency of territorial capital assets depends not on the endowment of single assets but on the presence of complementary and synergic components, and on their balanced development: an idea that recalls the balance of development theory (Chapter 4). The novelty is that today the interaction takes place among non-material resources. Econometric analyses show that the mere existence of knowledge does not explain regional growth trajectories; on the contrary, it plays an important role in those European regions with high endowments of social and relational capital.[52]

The synthesis of an area's success factors reminds us that identification of such factors takes place within conceptual approaches that are extremely different from each other. The traditional functional approach – also termed a positivist and cognitive approach – interprets the reality on the basis of deterministic, mechanical, cause–effect relationships. Another approach has recently been developed. It suggests inter-subjective relationships more complex than the deterministic ones and based on the ways in which economic actors interpret the reality, react to external stimuli, and are capable of synergic and co-operative behaviours. This new approach underlines that local competitiveness is linked more to trust and a sense of belonging than to a simple endowment of capital; more to creativity than the pure presence of skilled labour; more to relational capital than to accessibility; more to local identity than to the presence of important elements like quality of life and efficiency of the economic system.[53]

Such a rich concept is of great normative value, especially in a period when regional policies are expected to be conceptualized on the basis of differentiated strategies specific to the local context. As the 'Barca Report' of the European Union suggests, regional policy must be a place-based policy built on the basis of the specificities and elements of competitiveness of each single area through participatory and inclusive processes.[54] A conception of territorial capital that embraces and systematizes all the elements on which competitiveness can rely, and that highlights the laws of accumulation and depreciation, is therefore crucial for the appropriate design of these policies.

9.10 Conclusions

The chapter has surveyed the theories based on the idea that the role of space in economic activity does not consist solely in improvements to the static efficiency of production processes (that is, an increase in firms' revenues or a decrease in their costs). The advantages of a concentrated location of activities in space are also manifest in the innovative and creative capacity of firms, and space becomes a source of *dynamic efficiency*. In these theories, the innovation capacity of local systems is made dependent on local socio-economic conditions deeply embedded in the local area. Different concepts of proximity (physical, relational, institutional, cognitive) have been brought to the fore as elements explanatory of the innovation capacities of local firms, within a rigorous microeconomic and micro-behavioural framework.

The next chapters will consider theories developed in more recent times which include spatial aspects and the increasing returns that derive from them in macroeconomic growth models, but are only able to do so because they adopt a different conception of space: that of 'diversified-stylized' space.

Appendix: indicators of 'related variety'

In formal terms, the indicator of 'related variety' is built as an entropy indicator.[55] Applied to the concept of 'related variety', maximum entropy shows a situation of equal distribution in a region of knowledge among the different technological classes within a larger technological class to which they belong. More precisely, the indicator of 'related variety' is the sum of entropy (H_g) internal to a broad technological class (e.g. a two-digit class) (g), weighted for the share of patents in each class (S_g) on the total number of patents present in a region from *1* to *G*:

$$Related\,Variety = \sum_{g=1}^{G} S_g H_g$$

Entropy within a digit class g (H_g) is calculated as:

$$H_g = \sum_{i \in g} \frac{s_i}{S_g} \log_2 \left(\frac{1}{s_i / S_g} \right)$$

where s_i represents the share of patents in one technological class (e.g. five-digit) on the total of patents contained in a larger technological class (e.g. two-digit). When the share of patents of a larger technological class (S_g) is equal to the more detailed technological class (s_i), entropy of class g is minimum; knowledge is entirely confined within one technological class belonging to class g and knowledge complementarity is null within the technological class g. If this happened in all classes g, the indicator of 'related variety' would also be equal to zero, and it would signal the absence of related variety. By contrast, when the entropy indicator increases, the level of 'related variety' increases as well.

More recently, a related interregional variety has been suggested. This index measures the degree of knowledge complementarity within a common knowledge base of two regions; and it has been used to determine the role of interregional cognitive proximity in explaining scientific co-operations.[56] It is calculated as follows:

$$\textit{Interregional Related Variety} = \sqrt{\sum_{g=1}^{G} \frac{\left(S_{g,1} * S_{g,2}\right)}{\left|S_{g,1} - S_{g,2}\right|} \left(\sum_{i=1}^{I} \left(\left|s_{i,1} - s_{i,2}\right|\right)\right)}$$

where r_1 and r_2 represent a pair of regions, S the share of patents in the largest technological classes g (present in a number from 1_l to G), s the share of patents in classes I belonging to class g, in a number from 1 to I. The index measures the technological complementarity between regions through the differences in the patent shares in the technological classes i: in particular, the larger the share of patents in classes i between two regions, the higher the technological complementarity between regions. Moreover, the index measures the common knowledge base of the two regions through the product of the patent shares g of the two regions. When the shares of the patents within classes g of the two regions, having controlled for their relative size (measured as the difference in the share at the denominator[57]), are high, regions show a large common technological base.

Review questions

1. What is conceptualized by the knowledge spillover approach? What are the limits of this approach?
2. How would you define a '*milieu innovateur*'? What are the genetic elements of a milieu? Does the theory of the *milieu innovateur* overcome some of the limits of the knowledge spillover approach, and how?
3. What is conceptualized in the learning regions theory? How would you define a learning region? What are the main strengths and weaknesses of the theory?
4. What is the role of space in the knowledge spillover approach, in the *milieu innovateur* theory and in the learning region approach?
5. What is meant by regional innovation systems?
6. What does the evolutionary economic geography approach theorize? According to this theory, what are the determinants of industrial specialization of an area? What makes the exchange of information easier according to this approach?
7. What is meant by cognitive, relational and institutional proximities? Are they synonyms? How do they relate to the geographical proximity concept?
8. Which role does space play in the *milieu innovateur* theory, in the learning region approach and in the evolutionary economic geography approach?

9 What is the definition of a regional innovation pattern? What are the novelties of this approach?
10 What is the smart specialization strategy about? How would you distinguish the smart specialization strategy from previous innovation policies? What are the risks that still exist in relation to the implementation of this strategy?
11 How would you define a territorial capital? What are the dimensions on which a taxonomy of territorial capital elements have been produced and why? What is the usefulness of the territorial capital taxonomy?

Selected reading on empirical findings

About innovation and regional development

Aydalot P. and Keeble D. (eds) (1988), *High Technology Industry and Innovative Environment*, Routledge, London.

Breschi S. (2000), 'The Geography of Innovation: A Cross-sector Analysis', *Regional Studies*, vol. 34, no. 2, pp. 213–229.

Ciciotti E. and Wettmann R. (1981), 'The Mobilisation of Indigenous Potential', Commission of the European Community, *Internal Documentation on Regional Policy*, no. 10.

Cooke P. and Morgan K. (1994), 'The Creative Milieu: A Regional Perspective on Innovation', in M. Dodgson and R. Rothwell (eds), *The Handbook of Industrial Innovation*, Edward Elgar, Cheltenham, pp. 25–32.

Decoster E. and Tabariés M. (1986), 'L'innovation dans un Pôle Scientifique et Technologique: Le Cas de la Cité Scientifique Ile-de-France', in Ph. Aydalot (eds.), *Milieux Innovateurs en Europe*, GREMI, Paris, pp. 79–100.

Frenkel A. (2001), 'Why High-Technology Firms Choose to Locate in or near Metropolitan Areas', *Urban Studies*, vol. 38, no. 7, pp. 11083–11101.

Goddard J. and Thwaites A. (1986), 'New Technology and Regional Development Policy', in P. Nijkamp (ed.), *Technological Change, Employment and Spatial Dynamics*, Springer Verlag, Berlin, pp. 91–114.

Maggioni M. (2002), *Clustering Dynamics and the Location of High-Tech Firms*, Physica-Verlag, Berlin.

Malecki E. and Varaiya P. (1986), 'Innovation and Changes in Regional Structure', in P. Nijkamp (ed.), *Handbook of Regional and Urban Economics*, North-Holland, Amsterdam, pp. 629–645.

Oakey R., Thwaites A. and Nash P. (1980), 'The Regional Distribution of Innovative Manufacturing Establishments in Britain', *Regional Studies*, vol. 14, no. 3, pp. 235–253.

Paci R. and Usai S. (2000), 'Technological Enclaves and Industrial Districts: An Analysis of the Regional Distribution of Innovative Activity in Europe', *Regional Studies*, vol. 34, no. 2, pp. 97–114.

Schatzl E.L. and Revilla Diez J. (eds) (2002), *Technological Change and Regional Development in Europe*, Springer Verlag, Berlin.

Sternberg R. (1996), 'Reasons for the Genesis of High-tech Regions – Theoretical Explanation and Empirical Evidence', *Geoforum*, vol. 27, no. 2, pp. 205–223.

About milieux innovateurs

Amin A. and Thwaites A. (1986), 'Change in the Local Economy: The Case of the Northern Region (UK)', in Ph. Aydalot (ed.), *Milieux Innovateurs en Europe*, GREMI, Paris, pp. 129–162.

Camagni R. and Capello R. (1998), 'Innovation and Performance in SMEs in Italy: The Relevance of Spatial Aspects', *Competition and Change*, vol. 3, pp. 69–106.

Camagni R. and Rabellotti R. (1997), 'Footwear Production Systems in Italy: A Dynamic Comparative Analysis', in R. Ratti, A. Bramanti and R. Gordon (eds), *The Dynamics of Innovative Regions*, Ashgate, Aldershot, pp. 139–164.
Capello R. (1999), 'SMEs Clustering and Factor Productivity: A Milieu Production Function Model', *European Planning Studies*, vol. 7, no. 6, pp. 719–735.
Capello R. (2001), 'Urban Innovation and Collective Learning: Theory and Evidence from Five Metropolitan Cities in Europe', in M.M. Fischer and J. Froehlich (eds), *Knowledge, Complexity and Innovation Systems*, Springer, Berlin, pp. 181–208.
Gordon R. (1993), 'Structural Change, Strategic Alliances and the Spatial Reorganisation of Silicon Valley's Semiconductor Industry', in D. Maillat, M. Quévit and L. Senn (eds), *Réseaux d'innovation et Milieux Locaux*, GREMI, Neuchâtel, EDES, pp. 51–72.
Lambooy J. (1986), 'Regional Development Trajectories and Small Enterprises: the Case Study of the Amsterdam Region', in Ph. Aydalot (ed.), *Milieux Innovateurs en Europe*, GREMI, Paris, pp. 57–78.
Tabariès M. (1992), 'Nouvelles PME et cité scientifique en Formation: Ile-de-France Sud', in D. Maillat and J.-C. Perrin (eds), *Entreprises Innovatrices et Développement Territorial*, GREMI, Neuchâtel, EDES, pp. 23–40.

About knowledge spillovers and regional development

Acs Z., Audretsch D. and Feldman M. (1994), 'R&D Spillovers and Recipient Firm Size', *Review of Econmics and Statistics*, vol. 76, no. 2, pp. 336–340.
Audretsch D. and Feldman M. (1996), 'R&D Spillovers and the Geography of Innovation and Production', *American Economic Review*, vol. 86, no. 3, pp. 630–640.
de Groot H., Nijkamp P. and Acs Z. (2001), 'Knowledge Spill-overs, Innovation and Regional Development', *Papers in Regional Science*, vol. 80, no. 3, special issue.
Maier G. and Sedlacek S. (eds) (2005), *Spillovers and Innovations – Space, Environment and the Economy*, Springer Verlag, Vienna.
Ruslan L. and Plasmans J. (2003), 'Measuring Knowledge Spillovers in the New Economy Firms in Belgium Using Patent Citations', *Global Business and Economics Review*, vol. 5, no. 1, pp. 75–99.

About regional innovation systems

Bacaria J., Borràs Alomar S. and Fernàndez-Riba A. (2004), 'The Changing Institution Structure and Performance of the Catalan Innovation System', in P. Cooke, M. Heidenreich and H.-J. Braczyk (eds), *Regional Innovation Systems*, Routledge, London, pp. 63–90.
Boekhold P. and de Jager D. (2004), 'South-east Brabant: A Regional Innovation System in Transition', in P. Cooke, M. Heidenreich and H.-J. Braczyk (eds), *Regional Innovation Systems*, Routledge, London, pp. 44–62.
Cooke P. (2004), 'The Regional Innovation System in Wales: Evolution or Eclipse?', in P. Cooke, M. Heidenreich and H.-J. Braczyk (eds), *Regional Innovation Systems*, Routledge, London, pp. 214–233.
Maskell P. (2004), 'Learning the Village Economy in Denmark: The Role of Institutions and Policy in Sustaining Competitiveness', in P. Cooke, M. Heidenreich and H.-J. Braczyk (eds), *Regional Innovation Systems*, Routledge, London, pp. 154–185.

About evolutionary economic geography

Boschma R. and Iammarino S. (2009), 'Related Variety, Trade Linkages, and Regional Growth in Italy', *Economic Geography*, vol. 85, no. 3, pp. 289–311.

Boschma R., Minondo A. and Navarro M. (2012), 'Related Variety and Regional Growth in Spain', *Papers in Regional Science*, vol. 91, no. 2, pp. 241–256.
Frenken K., van Oort F.G. and Verburg T. (2007), 'Related Variety, Unrelated Variety and Regional Economic Growth', *Regional Studies*, vol. 41, no. 5, pp. 685–697.
Neffke F., Henning M. and Boschma R. (2011), 'How Do Regions Diversify Over Time? Industry Relatedness and the Development of New Growth Paths in Regions', *Economic Geography*, vol. 87, no. 3, pp. 237–265.
Tödtling F. and Trippl M. (2005), 'One Size Fits All? Towards a Differentiated Regional Innovation Policy Approach', *Research Policy*, vol. 34, no. 8, pp. 1203–1219.

About regional patterns of innovation

Camagni R. and Capello R. (2013), 'Regional Innovation Patterns and the EU Regional Policy Reform: Towards Smart Innovation Policies', *Growth and Change*, vol. 44, no. 2, pp. 355–389.
Capello R. and Lenzi C. (eds) (2013), *Territorial Patterns of Innovation: An Inquiry on the Knowledge Economy in European Regions*, Routledge, London.

About the smart specialization strategy

Boschma R. (2014), 'Constructing Regional Advantage and Smart Specialisation', *Scienze Regionali – Italian Journal of Regional Science*, vol. 13, no. 1, pp. 51–78.
Foray D. (2009), 'Understanding Smart Specialisation', in D. Pontikakis, D. Kyriakou and R. van Bavel (eds), *The Question of R&D Specialisation*, JRC, European Commission, Directoral General for Research, Brussels, pp. 19–28.
Foray D., David P. and Hall B. (2009), 'Smart Specialisation: The Concept', in *Knowledge for Growth: Prospects for Science, Technology and Innovation*, Report, EUR 24047, European Union.
Foray D., David P. and Hall B. (2011), 'Smart Specialisation: From Academic Idea to Political Instrument, the Surprising Career of a Concept and the Difficulties Involved in its Implementation', MTEI working paper, November, Lausanne.
Giannitsis T. (2009), 'Technology and Specialization: Strategies, Options, Risks', *Knowledge Economists Policy Brief*, no. 8, available online, http://ec.europa.eu/invest-in-research/pdf/download_en/kfg_policy_brief_no8.pdf?11111.
McCann P. and Ortega-Argilés R. (2014), 'The Role of the Smart Specialisation Agenda in a Reformed EU Cohesion Policy', *Scienze Regionali, Italian Journal of Regional Science*, vol. 13, no. 1, pp. 15–32.
Neffke F., Henning M. and Boschma R. (2011), 'How Do Regions Diversify Over Time? Industry Relatedness and the Development of New Growth Paths in Regions', *Economic Geography*, vol. 87, no. 3, pp. 237–265.
Pontikakis D., Chorafakis G. and Kyriakou D. (2009), 'R&D Specialization in Europe: From Stylized Observations to Evidence-based Policy', in D. Pontikakis, D. Kyriakou and R. van Bavel (eds), *The Question of R&D Specialisation*, JRC, European Commission, Directoral General for Research, Brussels, pp. 71–84.

About territorial capital

Camagni R. (2009), 'Territorial Capital and Regional Development', in R. Capello and P. Nijkamp (eds), *Handbook of Regional Growth and Development Theories*, Edward Elgar, Cheltenham, pp. 118–132.
Perucca G. (2013), 'A Redefinition of Italian Macro-areas: The Role of Territorial Capital', *Rivista di Economia e Statistica del Territorio*, vol. 2, no. 2, pp. 35–63.

Perucca G. (2014), 'The Role of Territorial Capital in Local Economic Growth: Evidence from Italy', *European Planning Studies*, vol. 22, no. 3, pp. 537–562.

Tóth B.I. (2014), 'Territorial Capital: Theory, Empirics and Critical Remarks', *European Planning Studies*, DOI: 10.1080/09654313.2014.928675.

Further reading

Camagni R. (1991), 'Local *Milieu*, Uncertainty and Innovation Networks: Towards a New Dynamic Theory of Economic Space', in R. Camagni (ed.), *Innovation Networks: Spatial Perspectives*, Belhaven-Pinter, London, pp. 121–144.

Capello R. (1999), 'Spatial Transfer of Knowledge in High-technology Milieux: Learning vs. Collective Learning Processes', *Regional Studies*, vol. 33, no. 4, pp. 353–365.

Foray D., David P. and Hall B. (2009), 'Smart Specialisation: The Concept', in *Knowledge for Growth: Prospects for Science, Technology and Innovation*, Report, EUR 24047, European Union.

Frenken K., van Oort F.G. and Verburg T. (2007), 'Related Variety, Unrelated Variety and Regional Economic Growth', *Regional Studies*, vol. 41, no. 5, pp. 685–697.

Keeble D. and Wilkinson F. (2000), *High-technology Clusters, Networking and Collective Learning in Europe*, Ashgate, Aldershot.

Lundvall B.A. (ed.) (1992), *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, Pinter, London.

Neffke F., Henning M. and Boschma R. (2011), 'How Do Regions Diversify Over Time? Industry Relatedness and the Development of New Growth Paths in Regions', *Economic Geography*, vol. 87, no. 3, pp. 237–265.

Torre A. and Wallet F. (eds) (2014), *Regional Development and Proximity Relations*, Edward Elgar, Cheltenham.

Notes

1. For a critical survey of neo-Schumpeterian theories see Moulcart and Sekia, 2003.
2. The concept of proximity gave rise to a school of thought in France that analysed the relatively greater importance for local development of organizational and cultural proximity compared with geographical proximity, Caragliu, 2015. The French school's notion of 'organizational and cultural proximity' and the one proposed here of 'relational proximity' have many features in common. On the French school of the 'economics of proximity' see Bellet et al., 1993; Rallet and Torre, 1998 and 2005; Rallet, 2002; Torre and Wallet, 2014.
3. For the French school of proximity, see among others, Bellet et al., 1993; Rallet, 2002, Rallet and Torre, 2005; Torre and Wallet, 2014. For the theory of *milieu innovateur*, see Camagni, 1991.
4. In the next sections of the chapter we shall not take into consideration organized, technological and social proximities. As regards social proximity, the concept was formulated in the industrial district theory devoted to explanation of static rather than dynamic efficiency, and it has been presented in Chapter 8. As regards organized proximity, this is a very general concept embracing different concepts of proximities at the same time, and therefore not really useful for highlighting the effects of each type of proximity. Technological proximity is interpreted by the concept of cognitive proximity, and is therefore already taken into consideration by that concept.
5. Wide-ranging empirical studies on innovative activity have been carried out in the UK by the CURDS (Centre for Urban and Regional Development Studies) of the University of Newcastle (see Oakey et al., 1980; Goddard and Thwaites, 1986) and by the SPRU (Science Policy Research Unit) of the University of Sussex (see Clark, 1971); in the USA by Malecki (see Malecki and Varaiya, 1986); and more recently also in Italy (see Breschi, 2000; Paci and Usai, 2000). Studies on the concentrated location of high-tech firms have been conducted by Keeble on the UK, Sternberg on Germany, Ciciotti on Italy, Decoster and Tabaries

on France, Malecki on the USA, Frenkel on Israel and Maggioni on a group of OECD countries. See Ciciotti, 1982; Decoster and Tabaries, 1986; Keeble, 1990; Sternberg, 1996; Frenkel, 2001; Maggioni, 2002. For detailed studies on the role of innovation in regional development see, e.g., Cappellin and Nijkamp, 1990; de Groot et al., 2004; Ewers and Allesch, 1990. For a theoretical and empirical analysis of spatial spillovers see Maier and Sedlacek, 2005.

6 Among the numerous empirical studies on knowledge spillovers, to be mentioned in particular are the seminal work by Jaffe, 1989, which was followed by other studies: Acs et al., 1994, who examined the differing abilities of small and large firms to exploit knowledge spillovers; Audretsch and Feldman, 1996 and Feldman and Audretsch, 1999, who distinguished between scientifically diversified and specialized spillovers; and Anselin et al., 2000, who defined the distance beyond which spillover effects disappear. For a critical review of the role of knowledge spillovers in regional development see de Groot at al., 2001.

7 Given expenditure on research and development as the input to innovative activity (R&D) and the number of patents (B) as the output, the knowledge production function shows the quantity of innovative input required to obtain a certain amount of innovative output:

$$B = f(R\,\&\,D) \tag{8.4n}$$

Recent econometric tests have shown the existence of a certain simultaneity between R&D and patents. This evidences that applications for patents tend to be made very early on in the innovative process and are consequently less indicative of a capacity to produce innovative output.

8 More refined methods, such as the inclusion of 'lagged' variables taking different values according to the geographical distance between the areas analysed, have recently been used. The results of the analysis do not change: they confirm the existence of technological spillovers through the greater significance of university expenditure on R&D for the innovative capacity of areas geographically closer to where the university is located, finding that 50 miles is the distance beyond which spillover effects disappear.

9 See Section 7.4.

10 The reference is to studies conducted by GREMI, Groupe de Recherche Européen sur les Milieux Innovateurs, headquartered in Paris, whose members were scholars from all the European countries. The group's research results were set out in a series of publications, most notably Aydalot, 1986; Aydalot and Keeble, 1988; Camagni, 1991; Maillat et al., 1993; Ratti et al., 1997; RERU, 1999.

11 It should be stressed that 'the concept of "innovative milieu" is necessarily abstract; the milieu must be considered an economic and territorial archetype more than an empirical reality. Its conceptualization in economic terms enables us to generalise some recent empirical findings showing the importance of relational assets in the success of some specific areas, and to find an economic rationale for the manner in which they support innovative processes. The characteristics of the innovative milieu are never fully realised in real territorial systems, however. The relationship between the presence of these characteristics and the innovative outcome has been verified in some empirical cases, and it is above all theoretically justified. But it can never be considered a precondition, either necessary or sufficient, for innovation; it is only an element which increases the probability of an innovative outcome.' See Camagni and Capello, 2002, p. 17.

12 See Bellet et al., 1993; Storper, 1995.

13 See Camagni and Capello, 2002, p. 18.

14 The concept of relational capital is similar to that of social capital developed by Putnam. See Putnam, 1993. It has been argued that the main difference between the two concepts is that social capital exists wherever a local society exists, while relational capital consists in the (rare) ability of actors to inter-relate their different skills, interact with each other, trust each other, and co-operate even at a distance with other complementary organizations. See Camagni, 2001.

15 Since the concept of collective learning was first formulated by the GREMI group (see Camagni, 1991), it has been used by numerous other authors. See Capello, 1999a; Keeble

and Wilkinson, 1999 and 2000; Lawson and Lorenz, 1999. On the concept of cognitive dimension of agglomeration economies, see Cappellin, 2003.

16 See Camagni, 1991.

17 This recalls the theory of the Marshallian industrial district and the role performed by social and cultural homogeneity in producing forms of transaction regulation which deter opportunistic behaviour. See Camagni and Rabellotti, 1997; Arrighetti et al., 2001.

18 For details on the methodology used see Capello, 1999b. After this pioneering study, subsequent analyses have also examined the effect of collective learning and the local atmosphere on the innovative activity of firms. A collection of studies is published in Camagni and Capello, 2002.

19 Romer's and Lucas's theories are set out in Chapter 11. For 'stylization' of the *milieu innovateur* theory within a neoclassical endogenous growth framework see Capello, 2002b.

20 For the main studies produced by the Danish school see Lundvall, 1992; Lundvall and Johnson, 1994; Asheim, 1996; Edquist, 1997; Maskell and Malmberg, 1999; Malberg and Maskell, 2002. For studies produced in Britain and North America see Cooke and Morgan, 1994; Morgan, 1997; Boekema et al., 2000; Cooke, 2002. The importance of institutional factors for local growth is now so widely recognized that creating institutional performance indicators for inclusion in macroeconometric growth models is considered to be essential. See Stimson et al., 2005.

21 See Amin and Thrift, 1994. The term 'institution' should of course be understood in the sense given to it by North's and Williamson's institutional economics, namely as a set of societal norms and 'rules of the game' (North, 1990, p. 3). See on this also WIlliamson, 2002.

22 'Institutions are here defined as the sets of habits, routines, norms and laws that regulate the relations between people and thus shape human interaction and learning': Lundvall and Johnson, 1994, p. 33.

23 This definition of 'learning region' has obvious links with the theories already described in Chapter 8. The concept of 'organized market' recalls that of 'community market' developed – much more convincingly and in richer form – by industrial district theory: social rules and norms regulate the market, making it more efficient and dynamic. The difference between the 'community' market and the 'organized' market resides in their outcomes: the former generates the factors that determine the co-existence and positive interaction between forms of co-operation and competition; the latter generates a dynamic process of interactive learning. Moreover, the theory of learning regions resembles that of the *milieu innovateur* when it emphasizes the importance of the 'destructive' learning that enables a region to abandon an obsolete technological trajectory. The *milieu innovateur* theory, too, stresses the importance for the local system's dynamic of avoiding 'lock-in' to knowledge that may become, like rules and norms of behaviour, 'barriers to exit' if the *milieu* must rapidly shift to a new technological trajectory. See Bianchi and Miller, 1993.

24 Lundvall and Johnson point out that 'learning economy refers not only to the importance of the scientific and technology system – universities, research organisations, in-house R&D departments and so on – but also to the learning implications of the economic structure, the organisational forms and the institutional set-up': Lundvall and Johnson, 1994, p. 26.

25 See among others, Edquist, 1997, 2005; Cooke et al., 2000; Cooke et al., 2004; Asheim and Gertler, 2005; Trippl, 2010.

26 See Tödtling and Trippl, 2005.

27 See Section 9.7.

28 On the debate concerning whether the concept of 'national innovation system' can be used to derive a 'regional' version in the form of a 'regional innovation system' see Howells, 1999; Acs et al., 2000; Fritsch, 2001.

29 In this perspective, the meaning of 'localized innovation' becomes clear: it is an innovation that takes place along specific technological trajectories within a specific paradigm. For an in-depth analysis of evolutionary theory, see Nelson and Winter, 1977; Dosi, 1982; Antonelli, 1989; Foray and Lundvall, 1996.

30 Supporters of this theory highlight in fact that in the first stages of formation of a concentrated area, space is neutral in the locational decision-making process, since it does not influence location choices. In an evolutionary perspective, once firms have by chance located

in one area, the latter becomes an attracting factor for new firms of the same industry, and space moves from being a '*neutral space*' to a '*real place*'. See Boschma and Frenkel, 2006, p. 290.

31 The first author to deal with the concept of 'related variety' was Nooteboom (2000), but it is thanks to the Dutch school of evolutionary economic geography that the concept has developed and received empirical evidence. For advanced studies on this concept, see Boschma, 2005; Frenkel et al., 2007; Boschma and Iammarino, 2009; Boschma et al., 2012.

32 See Frenkel et al., 2007. For the way in which 'related variety' indicators are built, see the Appendix in this chapter.

33 On the concept of 'absortive capacity', see Cohen and Levinthal, 1990.

34 For a detailed analyses of the negative aspects of proximity, see Boschma, 2005.

35 For the recent hermeneutic approaches to local creativity, see Cusinato and Philippopoulos-Mihalopoulos, 2015.

36 On the concept of regional patterns of innovation, see Camagni and Capello, 2013; Capello and Lenzi, 2013.

37 For the knowledge filter theory, see Acs et al., 2004.

38 This idea has opened the way to what has been later on called the 'smart specialization strategy' of the European Union. The first version of this strategy suggested innovation policies differentiated between centre and periphery. Today, the same policy is more elaborated in terms both of strategy design, and of geographical areas on which to implement different strategies. See on this issue Foray, 2009; Foray et al., 2009; McCann and Ortéga-Argiles, 2014. See Section 9.8.

39 For the absorptive capacity concept, see Cohen and Levinthal, 1990.

40 Based on indicators covering all aspects of territorial innovation patterns (from local preconditions, to intensity of knowledge and innovation creation) for all European regions, a cluster analysis has identified six (and not three) territorial patterns of innovation, witnessing the complexity of the real world. The theoretical endogenous innovation model identifies two empirical patterns, one associated with regions generating base knowledge, and another with regions generating applied knowledge. The creative application pattern distinguishes in the real world two groups of regions, applying external knowledge to internal formal knowledge, the first one, and to informal knowledge, the second one. The imitative pattern also shows two empirical groups of regions: those that imitate and those where innovation does not take place. See Capello and Lenzi, 2013, chapter 7.

41 On the 'smart specialization strategy', see, among others, Foray, 2009; Foray et al., 2009; McCann and Ortéga-Argiles, 2014.

42 See Pontikakis et al., 2009.

43 See Foray, 2009; Foray et al., 2009; Giannitsis, 2009.

44 See Foray et al., 2011

45 See Camagni and Capello, 2013; McCann and Ortega-Argilés, 2014.

46 See Frenken et al., 2007; Boschma 2011 and 2014; Neffke et al., 2011; Iacobucci, 2014.

47 See OECD, 2001 and European Commission, 2005, p. 1.

48 See Camagni, 2009.

49 To be clarified is the distinction between social and relational capital. While one can easily argue that social capital exists when a society exists, relational capital may be kept separate and be defined as that part of social capital which refers to the co-operative action of each individual, measured through the bilateral and multilateral agreements that are developed among actors, both within and outside the local area, facilitated by a friendly and trusting atmosphere based on shared behavioural rules and values. See Camagni, 2009.

50 For empirical analysis of territorial capital, see Perucca, 2013 and 2014; on Hungarian regions, see Tóth, 2014.

51 For a detailed analysis of the data and methodology adopted, and the results obtained, see Perucca, 2014.

52 For data, methodology and results, see Capello et al., 2011.

53 See Camagni, 2009.

54 See Barca, 2009.

55 The entropy principle makes it possible, under imperfect information conditions, to determine the most probable condition of a system formed of a large number of elements

(molecules) corresponding to its equilibrium condition (maximum entropy). The entropy principle, in fact, describes the process of a system (e.g. a gaseous system), oriented only in one direction, exposed to an external solicitation (e.g. a variation in temperature) towards an order system. This is the case of Prigogine's example of an iron bar exposed to a source of heat at one of its extremes, in which temperature distributes in an ordered way. When the external source of heat ceases, temperature distributes uniformly along the bar, and the system moves from a molecular order to a disorder situation, and therefore from a low to a high probability condition. See Camagni, 1992a, p. 105.

56 See Capello and Caragliu, 2012.

57 Thanks to the denominator, the greater the difference in the class between the two regions, the lower the cognitive proximity. The denominator, in fact, makes it possible to adjust high levels of the numerator due to extreme cases in which one of the two regions concentrates its patent activity in only one technological class.

Part IV

Theories of regional growth

Diversified-stylized space

10 Territorial competitiveness and cumulative demand/supply growth

10.1 Increasing returns, competitiveness and cumulative growth

The previous two chapters stressed the active role that space may play in economic development as the source of advantages for the firms located within it: static and dynamic agglomeration economies in the form of localization or urbanization economies significantly influence the productivity and innovative capacity of firms and, in aggregate terms, also of the area in which those firms are situated. Increasing returns arise from concentrated production, and they determine the efficiency of the economic system.

The previous two chapters also stressed the qualitative nature of the theories of endogenous development examined. This qualitativeness was in some cases due to an explicit methodological choice; in others to the difficulty of including increasing returns in an analytical model. In mathematical terms, the hypothesis of scale economies entails that the relations among the variables that determine development cannot be based on linear equations; instead it is necessary to have higher-level equations that inevitably require a descriptive mathematical language more complex than that of linear systems.[1] In economic terms, the existence of increasing returns (at the individual firm level) requires abandonment of the perfect competition hypothesis, and the contrary assumption of imperfect competition; a notion that was never formalized prior to the 1970s.[2]

In the 1980s, major progress was achieved in the fields of both non-linear mathematical models and of economic modelling in conditions of imperfect competition. This opened the way for new theories on local economic growth. Thanks to the advent (i) of mathematical approaches to the study of the qualitative behaviour of non-linear dynamic systems (bifurcation, catastrophe and chaos theory) and (ii) in economics, of Avinash Dixit and Joseph Stiglitz's formalized model of imperfect competition, increasing returns became the decisive factor in development, not only for qualitative theories but for analytical theories and models as well.[3]

This part of the book describes the theories of local growth that for the most part use advanced mathematical tools and draw on recent economic analytical models. They are of particular importance for local development theory because they take analysis beyond Edwin von Böventer's already mentioned distinction between 'pure and exact' regional theory without agglomeration economies, on the one hand, and 'applied regional theory', which is inexact but takes agglomeration factors into account, on the other.[4]

The first innovative feature of these more formalized theories is that they enable elegant growth models of a strictly economic nature to include agglomeration economies, in the form of increasing returns, as determinants of local development. They then demonstrate that these phenomena can be treated using the traditional tools of economic theory (optimizing choices for firms and individuals). They have thus induced orthodox economists to (re-)discover the spatial dimension of economic phenomena, and it is to this aspect that they owe a large part of their continuing success.

The second innovative feature of these approaches is their ability to escape the mechanicism of the formalized models that preceded them and to introduce elements of uncertainty into both growth trajectories and the final equilibrium towards which the development path tends. Real phenomena accompanying development trajectories – synergy and positive cumulativeness (agglomeration economies) as well as negative feedbacks (congestion or saturation in growth processes) – are incorporated into the logic of the models through the non-linearity of growth relations. This makes possible multiple equilibria associated with diverse initial conditions, with diverse values of the variables and parameters of the structural relations of development, and with convergent or divergent, explosive or implosive, stable or unstable, growth paths.

These models generate a growth path that recalls that of the theories surveyed in the previous two chapters; once again, this is a path of *cumulative, endogenous and largely selective growth*. The models now described envisage a diversified space, in fact. That is to say, they assume the existence of sharp polarities where development takes place and cumulates due to increasing returns in the form of learning processes, scale economies (at the area or firm level), and localization and urbanization economies that engender a virtuous circle of cumulative development. Moreover, because increasing returns are included in the structural relations that characterize the dynamic behaviour of the local system (or of the individual firms located in it), they are produced by the workings themselves of the local economic system, and they mark out an endogenous growth path.

These theories are all the more similar to those of the previous two chapters in that they pursue the same goal of identifying the elements that determine long-period competitiveness, and the conditions under which an area can acquire and maintain a role in the international division of labour. The increasing returns hypothesis, in fact, entails the assumption that when the market expands, either production increases with resources remaining equal, or cost decreases with production remaining equal. In other words, it entails the assumption that associated with increases in production are ever greater savings of resources, and therefore increasingly greater rises in productivity, with positive and growing effects on local competitiveness. These effects are expressed differently by each theory: in terms of a greater capacity to capture larger shares of world demand by the theory of cumulative circular development; of greater capacity to attract external capital in search of good financial and productive opportunities by the most recent models of the 'new economic geography'; and of greater capacity to (re-)create over time the conditions for constant economic growth of productive resources by the theory of endogenous growth.

However, the marked differences and discontinuities between these theories and the endogenous development theories discussed previously should also be emphasized. The first of these differences/discontinuities concerns the formalized, macroeconomic and aggregate nature of the theories that this chapter examines, which stand in sharp contrast to the micro-territorial and micro-behavioural approach taken by the models in the previous three chapters. Owing to their aggregate macroeconomic nature, the theories now presented aim to explain the growth rate of aggregate income interpreted as a synthetic indicator for the various aspects of development. Unlike those seen in the previous three chapters, these theories do not seek to provide a qualitative interpretation of all the tangible and intangible elements, economic or otherwise, which characterize the dynamic of local economic systems. Once again, therefore, the dynamic path of a local economy is interpreted by *growth theories*. But there are two major differences between these and the growth theories of the 1950s and 1960s: (i) returns are no longer constant but increasing, and (ii) the conception of growth assumed is a dynamic and long-term one. Theories seek to define the elements with which the competitiveness conditions of a local system can be maintained and re-created, rather than to highlight the mechanisms that increase long-term employment and production, or individual well-being and per capita income, as in previous theories.

A second difference with respect to the theories discussed in the previous three chapters resides in the treatment of space, which now becomes *diversified* and *stylized*. These approaches envisage the existence of polarities in space where development takes place, diversifying the level and rate of income growth even among areas of the same region. However, although diversified, space is now stylized into points devoid of any territorial dimension. Localized technological externalities do not exist in this space; nor does a set of tangible and intangible factors that may act upon firms' productivity and innovative capacity because of proximity and reduced transaction costs; nor a system of economic and social relations constituting the relational or social capital of a certain geographical space. Yet all these are elements able to differentiate spatial elements on the basis of strictly territorial aspects. These approaches thus reprise the simple – somewhat banal – view of space as the simple container of development, and they therefore necessarily abandon the more interesting and intriguing interpretation of space as an additional resource and as an independent factor in development.

These considerations introduce the third discontinuity with respect to the theories discussed in the previous two chapters: increasing returns no longer take the form of specific advantages involuntarily generated by individual firms. According to the theories now examined, increasing returns are economies of scale or of learning stylized in systems of equations which explain the structure and dynamic of a local system through non-linear relations which give rise to multiplicative effects in the aggregate growth rate.

Whilst these are the main features of the most recent theories of regional growth, this chapter also deals with models that assume the existence of increasing returns (at the firm or area level) to interpret development as resulting from a *cumulative process of demand/supply growth*. Left for treatment in the next chapter are theories that conceive growth as resulting from increasing returns on production resources, in a production function of neoclassical derivation; for these theories, growth depends exclusively on *supply elements*.

According to the logic of the models presented in this chapter, therefore, the competitiveness (exogenously assumed) of strong areas generates greater production (supply); more investments (induced by an 'acceleration' mechanism for some theories, and by the creation of greater profits in the local market for others); and higher employment, which fuels immigration. This process drives the development of a broad local market (demand) which in its turn attracts new investments and creates new employment (supply), in a circle of cumulative growth. In parallel with this circuit of increasing local production, increases come about in the level (or rate) of productivity because of technical progress embodied in capital goods, to firm- and system-level learning processes, and scale economies deriving from larger production volumes.

The modern theories now presented are rooted in a model, formulated at the end of the 1950s and then formalized in the 1970s by Nicholas Kaldor, which already conceived the existence of increasing returns intrinsic to the structural relations that characterize a local economy's aggregate growth. In this model, economies of scale are assumed to be external to firms, taking the form of learning economies – or learning-by-doing economies à la Arrow. The rich and dynamic advanced economies, with their high growth rates, also display (in these models) greater rates of productivity growth that generate a cumulative circle of growth. Reasoning on the basis of increasing returns at territorial level, the model is able to formalize these returns on the assumption of perfect competition (Section 10.3).[5]

Myrdal's and Kaldor's idea of giving increasing returns a key role in local development was taken up by a school of thought that developed in the 1990s under the guidance of the well-known economist Paul Krugman. Exploiting the formalization of the imperfect competition model, Krugman and his followers produced elegant economic growth models that incorporated the location choices of firms. These were made to depend on three economic factors – transport costs, increasing returns and migratory flows – that determine, according to the values that they assume, the existence of agglomerative phenomena (what Krugman calls 'geographic concentration') or diffusion processes. When the concentration of productive activities prevails in an area, the conditions for cumulative local growth are generated (Section 10.4).

Before the theories are introduced, it may help the reader to understand the new logics of 'equilibrium' if an outline is provided of the most recent mathematical instruments used to interpret economic growth.[6]

10.2 Equilibrium in conditions of non-linearity

10.2.1 The novel aspects of the approach

There are two reasons for the great success of non-linear dynamic models since the 1980s. The first has just been mentioned: that these models make it possible to represent in stylized form real phenomena that manifestly affect the formation, dynamic and structure of economic systems. In the case of local systems in particular, they enable the inclusion in growth models of scale and agglomeration economies (diseconomies), synergies and idiosyncrasies among the various components of a complex system, oscillatory movements in variables like price, income, and technological innovation; all of which are elements likely to affect the development path of a local economic system.

The second reason for the success of these models is that it is today possible to overcome the difficulties that often accompany solution of these models by using *numerical simulations*, on the one hand, and mathematical analyses of the qualitative behaviour of non-linear dynamic systems, on the other. Recent mathematical approaches allow study to be made of the nature of solutions, rather than of their exact value. Mathematical analyses of this kind are the 'bifurcation and catastrophe theories' that emphasize the existence of multiple equilibria in which the transition from one equilibrium to another may take place through a 'break' or 'catastrophic change' in the time pattern of the variables.[7]

The distinctive feature of these theories is their ability to describe qualitative changes in the state of a system resulting from variation in:

- variables expressing the system's dynamic (also known as 'state variables'), which describe the state of a system at each moment *t*, and the values of which change rapidly in time;
- parameters (or 'control variables') which instead change relatively slowly.

The catastrophe and bifurcation theories have several innovative features. First, they allow for the existence of *multiple system equilibria*, in contrast to the unique dynamic equilibrium (whether stable or unstable) that characterized previous theories. The prevalence of one equilibrium over another, as well as the choice among possible equilibria, depends on the values of the parameters conditioning the temporal dynamic of the unknown variable and on the initial conditions, which as we shall see significantly influence the system's development trajectories.

It is therefore possible to stylize time patterns of development in which small variations in the parameter values may trigger *sudden catastrophic changes* so that, according to the alternative that prevails, entirely different growth paths ensue. These models are thus able to simulate an endogenous series of complex phenomena that in the past could only be replicated by means of exogenous shocks introduced ad hoc.

The customary distinction between stable or unstable dynamic equilibria – which represent 'dynamic stability'[8] – has been supplemented with a further meaning of 'stability' that concerns the conditions in which the nature itself of solutions may change (for example, instead of a single solution, periodic or chaotic ones are propounded). In this case, analysis centres on the 'structural stability' of systems.[9] Indeed, the most recent approaches to non-linear dynamic systems have studied this 'structural stability' of systems, the quality and nature of solutions, and the form that the system may assume. Unlike linear dynamic models, they have not concerned themselves with analysis of the existence and stability of the system's equilibrium (its 'dynamic stability'). This helps explain why we shall often find that the models do not yield unequivocal results, but instead offer a range of possible solutions according to the initial conditions and the values assumed by the parameters.

There is a further and equally important feature of these models: when time trajectories undergo an abrupt change, they are rarely able to return to their initial state. They appear to be largely *irreversible* if the direction of time is reversed, because the system spontaneously reorganizes itself around the new state; no development trajectory can be replicated by chance, and no development trajectory can move in reverse direction.

10.2.2 An example of catastrophic growth: the export-base model in conditions of non-linearity

The exposition thus far can be made clearer if we take a model already discussed – the dynamic export-base model set out in Chapter 5 – and re-work it with the aim of determining the existence and the stability of dynamic equilibrium.

Reprising the structural relations between population (P) and employment (total, in the base sector, and in services, respectively E_T, E_b, E_s) that characterized Hoyt's model:

$$P = aE_T \tag{10.1}$$
$$E_T = E_b + E_s \tag{10.2}$$
$$E_s = bP \tag{10.3}$$
$$E_b = \overline{E}_b \tag{10.4}$$

we introduce a time lag in (10.3)

$$E_s(t) = \beta P(t-1) \tag{10.5}$$

and non-linearity in the structural relations (10.1) and (10.3):

$$P = \alpha E_T \qquad \textit{with } \alpha' > 0 \tag{10.6}$$
$$E_s = \beta P \qquad \textit{with } \beta' > 0 \tag{10.7}$$

Relation (10.6) states that there exists a certain threshold of employment above which a marginal increase produces a very large increase in the population (presumably because of the positive effects of agglomeration economies). Relation (10.7) states that there exists a critical mass of the population beyond which a population increase has an extremely marked effect on the services sector.[10] On these hypotheses, (10.1) becomes:

$$P(t) = \beta\left\{E_b + \alpha\left[P(t-1)\right]\right\} \tag{10.8}$$

and the region's growth can be depicted as in Figure 10.1. It is easy to identify the differences with respect to Figure 5.1, which shows the linear growth of income (though the same figure would apply for a model expressing the physical growth of the region, to which the dynamic hypotheses describing the trend in Figure 5.1 are applied). In the model incorporating non-linearity (Figure 10.1a):

- there exist *multiple possible equilibria*: P^*, P° and P^{**};
- there exist equilibria *of different kinds*: P^* and P^{**} are stable equilibria; P° is an unstable equilibrium;[11]
- the equilibrium which comes to prevail depends on the initial conditions of the system: if the initial population of the region is less than P°, it will tend to a stable dynamic equilibrium equal to P^*; if instead the initial population is greater than P°, the system will tend to shift to size P^{**}.

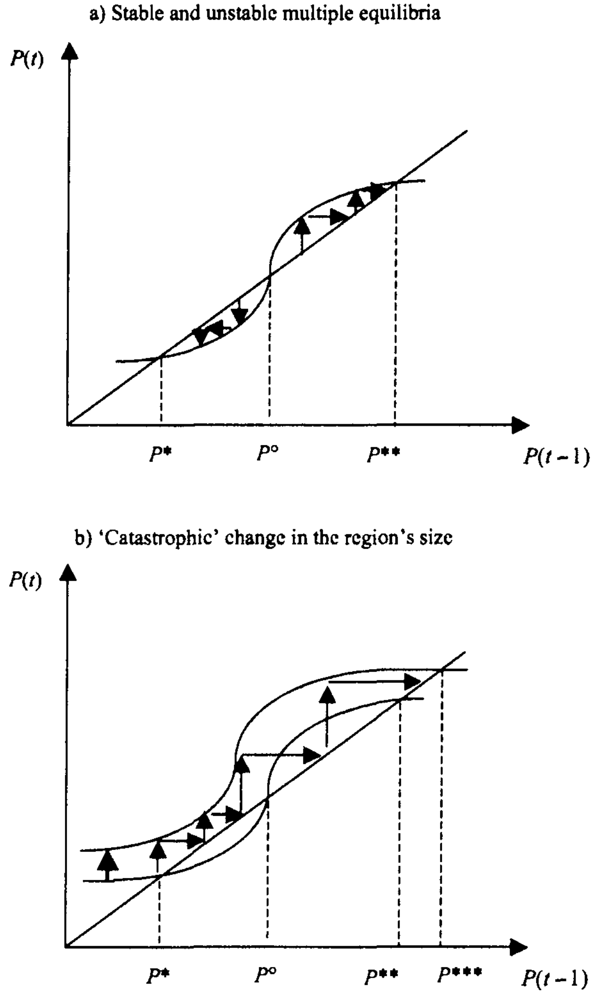


Figure 10.1 The export-base model on the hypothesis of non-linearity

Source: Miyao (1984)

Moreover, on assuming exogenous increases in base sector employment (which here performs the role of the control variable or 'parameter' mentioned earlier), the curve expressed by (10.5) shifts upwards. If the increases in base employment are small, the shift does not generate major changes in the region's size. However, there exists a value of employment increase, which may even be very small, at which a break-point, a catastrophic jump, occurs, with a shift to a very much larger size (Figure 10.1b); the size of the city may in fact grow abruptly from P^* to P^{***}. In this case, the system is rarely able to return to P^* and is more likely to reorganize itself around a new and larger size.

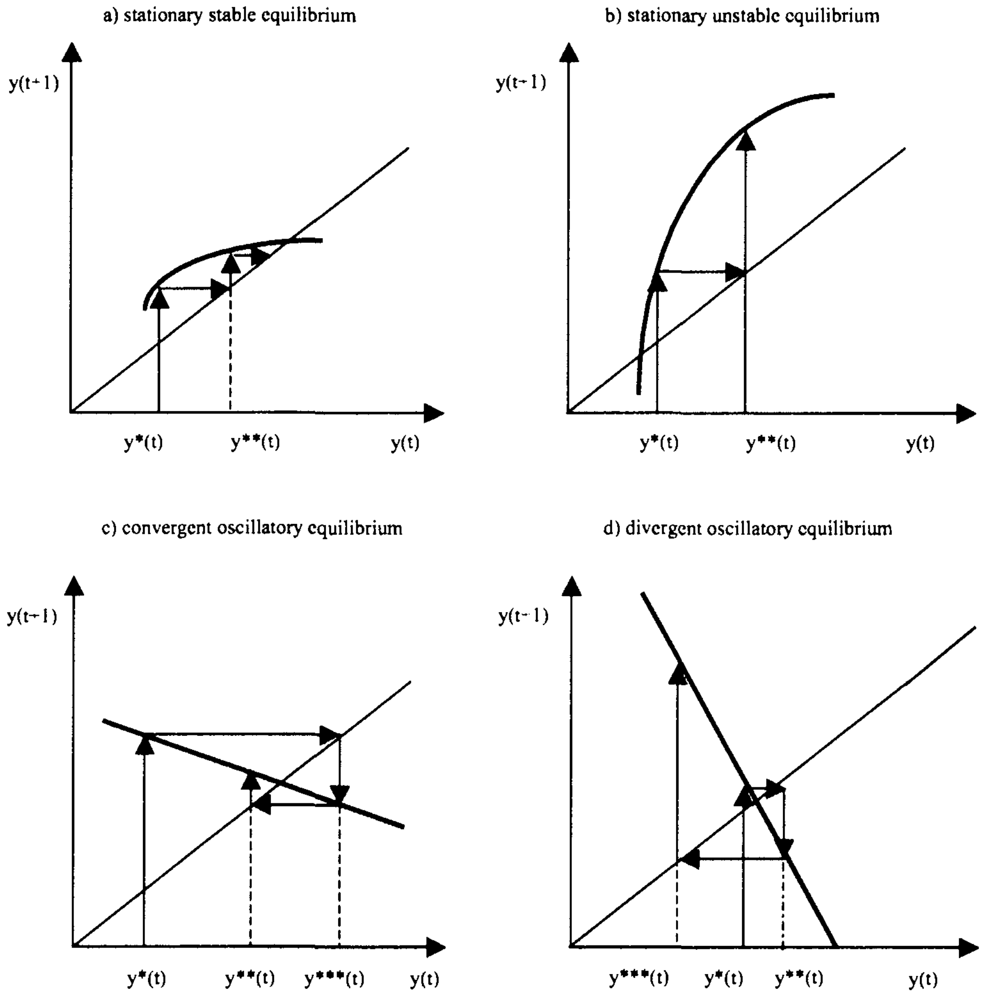


Figure 10.2 Diagram of the phases in the finite differences equations (time as the discrete variable)

10.3 Increasing returns external to the firm: the circular and cumulative causation model

At the end of the 1950s, Gunnar Myrdal formulated a model that ran counter to the neoclassical belief in the existence of spontaneous processes of re-equilibrium.[12] Myrdal's 'circular and cumulative causation model' was able to explain the persistent interregional disequilibria reported by empirical research in terms of self-fuelling virtuous/vicious circles. According to the logic of Myrdal's model, rich regions grow increasingly richer, and poor regions increasingly poorer, if spontaneous market forces alone are permitted to operate.[13]

The results yielded by Myrdal's model – which, as said, was entirely at odds with the traditional neoclassical view – are explained by the assumptions on which it was constructed:

- the existence of an investment function which depends on the real or expected level of demand (accelerator theory), rather than on the rate of return on capital, as suggested by the neoclassicals;
- the existence of increasing returns at territorial level; that is, agglomeration economies generated by the territorial concentration of productive activities and by the accumulation of knowledge embodied in capital goods – as opposed to the constant returns of the neoclassical production function. Assuming this hypothesis signified (for the first time in the history of economic thought on development) that a role must be given to *increasing returns* when the trajectories of local economic growth are defined.[14] It also means relinquishing a single production function with equal technological progress across regions, as imposed by the neoclassical logic, and instead accepting the more realistic assumption that richer regions are endowed with superior technologies that partly explain their greater productivity and competitiveness.

Under these hypotheses, two virtuous processes operate in strong regions (Figure 10.3):

- these regions attract workers because they have high levels of production (which the model assumes to be exogenous) and a consequent strong demand for labour. Unlike in the neoclassical theory, where the production factors are assumed to be homogeneous, this is a selective migratory process that involves more highly skilled human capital and consequently deprives the weak areas of better-quality labour. The migratory flows to the rich areas expand the local market, stimulate new investments and attract new capital, in a virtuous circle of development;
- at the same time, the close concentration of production activities in a particular area generates agglomeration economies that act upon the area's productivity and competitiveness, boosting development. Greater supply generates further labour demand, increased (internal and external) demand for locally produced goods, new investments, new business start-ups, closer concentration, greater advantages deriving from concentrated locations and further productivity increases, in a virtuous demand/supply circle.

Conversely, the reverse processes of emigration, capital loss, decreasing internal demand, and a decline in productivity due to diminished agglomeration economies

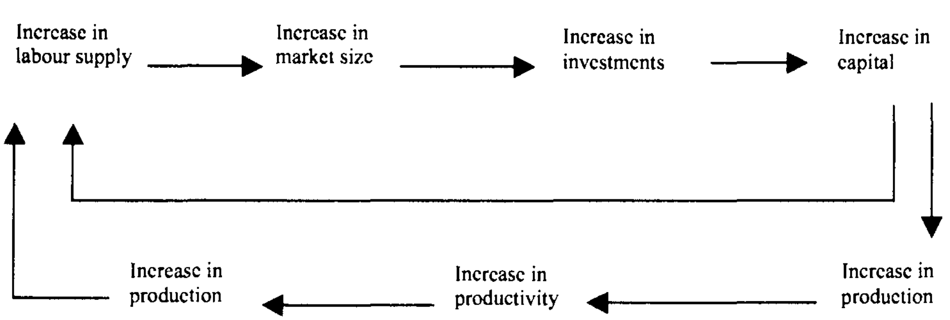


Figure 10.3 Myrdal's virtuous circle of cumulative development

characterize the poor areas; according to the logic of the model, these are bound to suffer desertification and poverty.

However, Myrdal's model sets limits on the infinite evolution of the circular cumulative process; limits that consist mainly in territorial and supply-side factors. A constant and concentrated development process generates diffusion effects ('spread effects' in Myrdal's terminology) due to physical congestion, the growing scarcity of the production factors, and their increasing costs. These diffusion processes may arise in the area because of spatial contiguity and then spread along transport and communication axes, or they may 'filter down' through the branches of the urban hierarchy.[15]

The great explanatory potential of Myrdal's model was realized by Nicholas Kaldor, who in the 1970s produced a formalized model of cumulative circular causation.[16]

The dynamic of local income (y) is made to depend on the growth of exports (e).[17] The latter exhibits a dynamic that depends partly on exogenous factors connected with the development of the world economy (b) and partly on endogenous elements connected with the trend of local competitiveness, which depends on domestic price variation (p). In its turn, domestic price variation is explained by variation in the cost of labour per unit of output, also termed 'efficiency wage', and therefore by the difference between the rates of wage growth (w) and of productivity growth (π). Finally, productivity growth is governed by the well-known 'Verdoorn's Law',[18] according to which the rate of productivity growth consists of an exogenous component (d) and a component endogenous to the system expressed by the output growth rate.[19] This last relation states that more than proportional productivity growth rates are associated with higher output growth rates; the existence of scale economies and learning effects explain this relation, and they are comprised in the positive parameter (f) of the mathematical equation:[20]

$$y = ae \qquad a > 0 \tag{10.9}$$

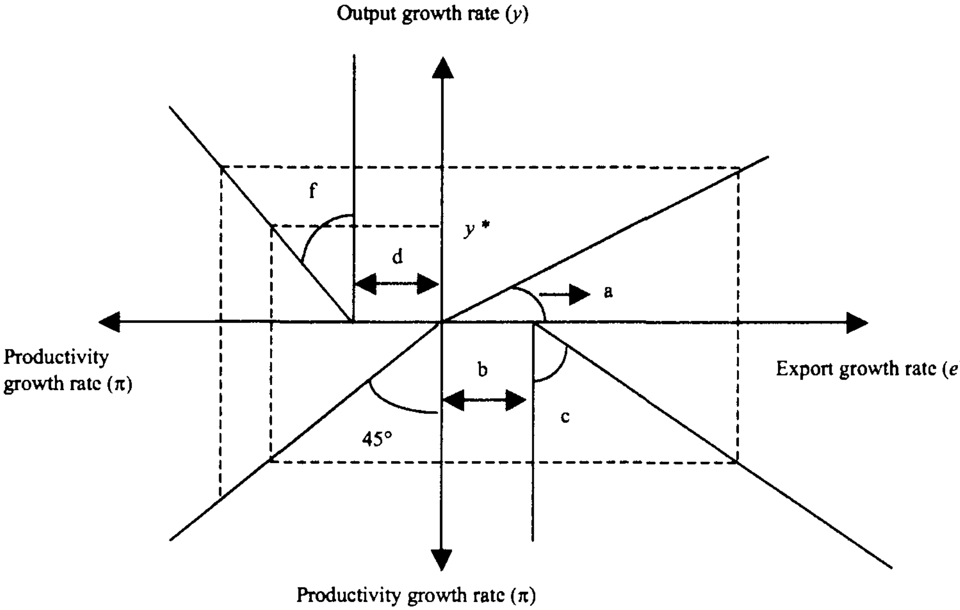
$$e = b - cp = b - c(w - \pi) \qquad b > 0, c > 0 \tag{10.10}$$

$$\pi = d + fy \qquad d > 0, f > 0 \tag{10.11}$$

These three relations are depicted by Figure 10.4, where equation (10.9) is shown in the upper-right quadrant, equation (10.11) in the upper-left quadrant and equation (10.10) in the lower-right quadrant. The lower-left quadrant transposes the variables on the axes. It is easy to see from the figure that, according to the values taken by parameters a, c and f, the system starts from an initial growth rate y^* and enters either a virtuous and cumulative circle of development (Figure 10.4a) or a vicious one of underdevelopment (Figure 10.4b). The economic conditions that determine a trajectory of growth rather than of decline are the following:

- greater elasticity of demand for exports (parameter a);[21]
- higher increasing returns which associate output growth with productivity growth (parameter f);
- greater elasticity of exports to variation in productivity (and in domestic prices) (parameter c).

Moreover, it is evident that when the economic system suffers from weak structural conditions – expressed by a low initial rate of output growth (y), limited growth of *a) Cumulative regional development (explosive growth)*



b) Cumulative regional decline (implosive growth)
(for values of a, c, f, different from those in case a)

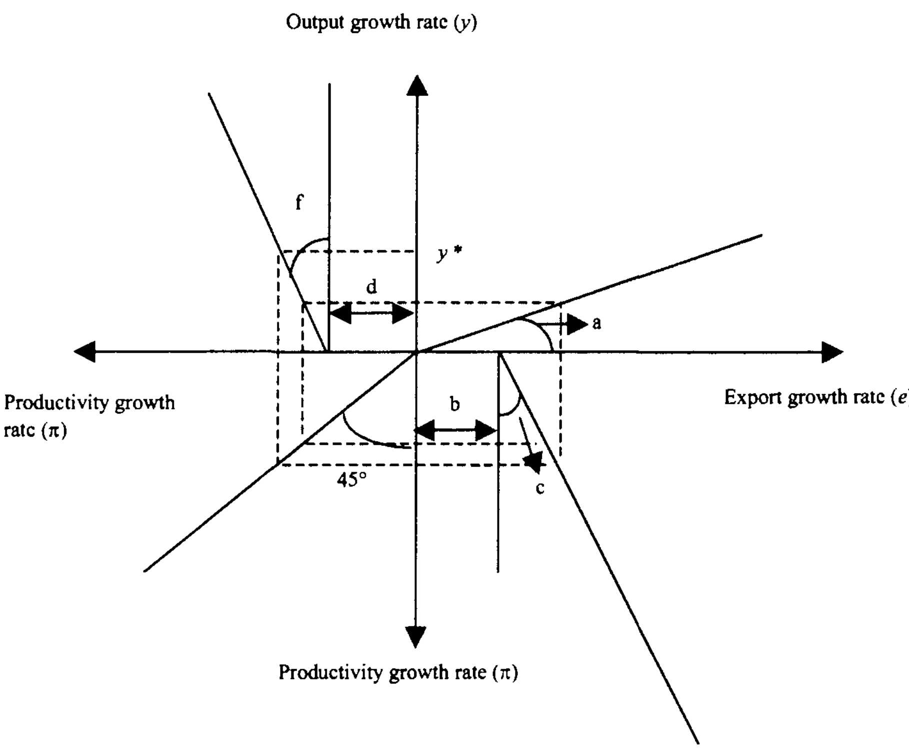


Figure 10.4 The process of cumulative circular causation

(*Continued*)

c) Cumulative regional decline (implosive growth)
(for values of b and d different from those in case a)

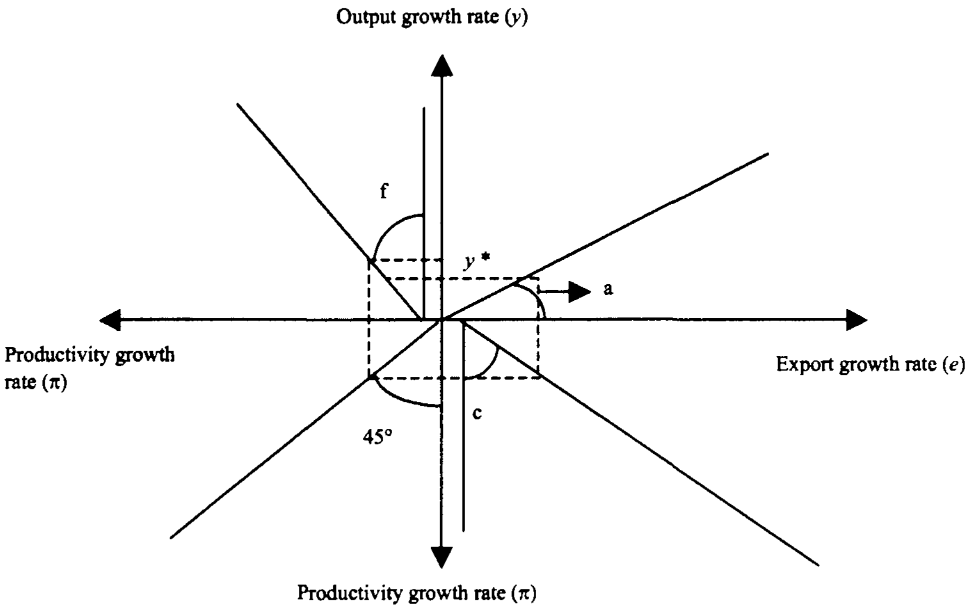


Figure 10.4 (Continued)

the exogenous component of productivity (d) and of competitiveness (b) – a vicious circle of decline ensues, even with parity of endogenous conditions represented by equal values for the parameters a, c and f (Figure 10.4c).

The same result can be obtained if the dynamic properties of the system are analysed. Solving equations (10.9), (10.10) and (10.11), and introducing a time lag into the last of them, yields:

$$y(t) = a(b - cw + cd) - acfy(t - 1) \tag{10.12}$$

(10.12) is represented graphically by Figure 10.5, where it is again evident that equilibrium depends on the parameter values and the initial conditions. If $acf > 1$ (i.e. if the endogenous components of competitiveness are favourable to development), the system is unstable and diverges from the equilibrium development rate, undergoing an explosive or implosive process according to the initial conditions (Figure 10.5a). An initially low growth rate, accompanied by slight exogenous components of productivity and competitiveness, leads to economic decline (Figures 10.4c and 10.5a for initial growth rates less than y^*). Better initial structural conditions, by contrast, engender explosive cumulative development (Figures 10.4a and 10.5a for initial growth rates greater than y^*). These conditions of explosive or implosive development bear out the theoretical expectations of the first proponents of the model, Myrdal and Kaldor.[22]

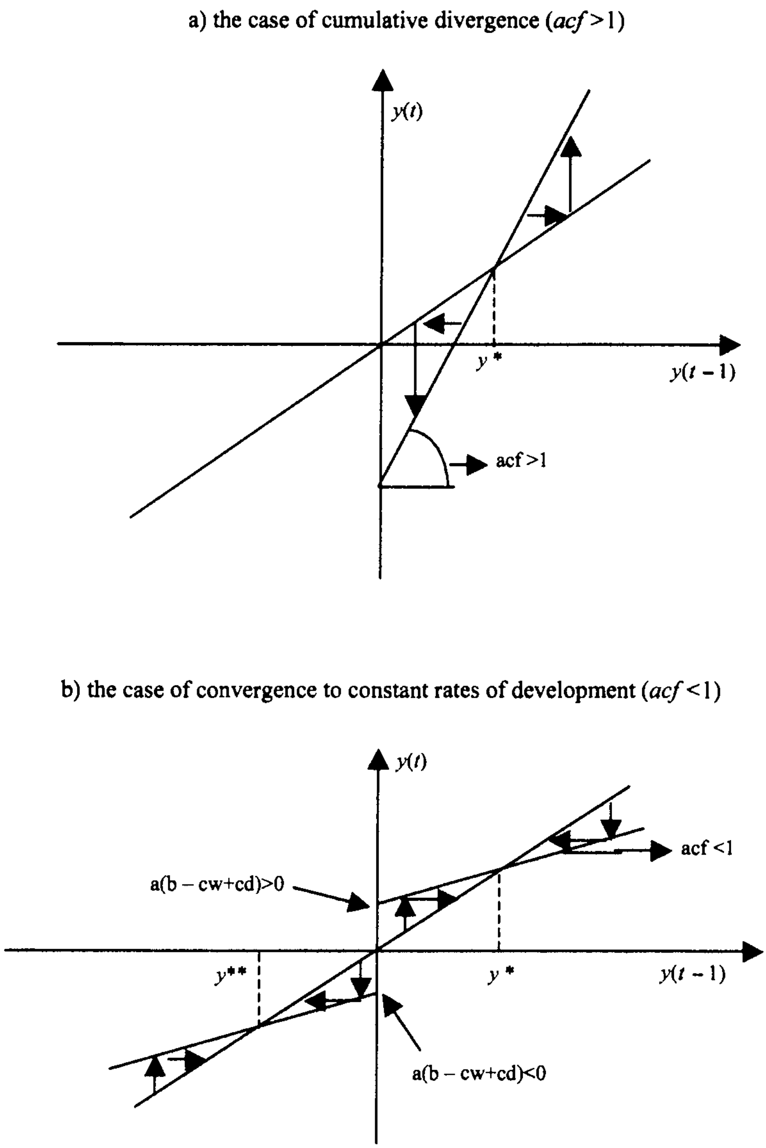


Figure 10.5 The dynamic equilibrium of the cumulative circular causation model

Source: Miyao (1984)

By contrast, if $acf < 1$, the system converges on a constant rate y^*. The growth rate is positive if the first term of the right-hand member is positive, i.e. if $w < (b - cd) / c$, that is, if wages do not grow to such an extent that they compromise the external competitiveness of the local economic system (Figure 10.5b). If wages are instead so high that they thwart the competitiveness of the local system, the latter will nevertheless tend towards a steady-state growth rate, but this growth rate will be negative (y^{**} in Figure 10.5b).

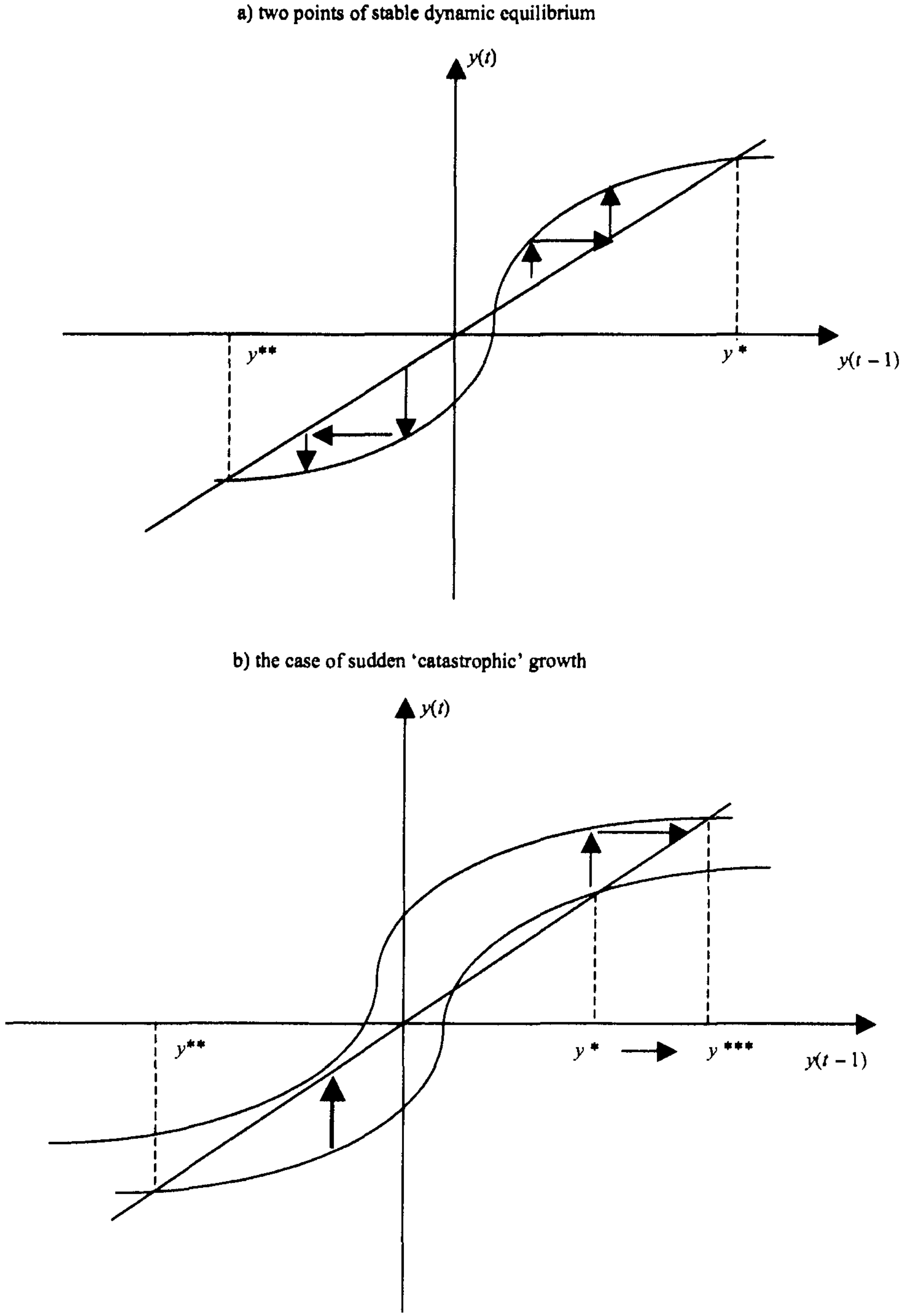


Figure 10.6 The cumulative circular causation model on the hypothesis of non-linearity in returns

Source: Miyao (1984)

In this case, too, non-linearities can be introduced into the system. One can hypothesize in particular that the productivity growth rate increases with very high growth rates, as follows:

$$\pi = d + \varphi y \qquad d > 0, \varphi' > 0 \tag{10.13}$$

(10.12) thus becomes:

$$y(t) = a(b - cw + cd) - ac\varphi y(t-1) \tag{10.14}$$

(10.14), which is expressed by Figure 10.6a, shows that there are always stable rates of growth y^* or decline, but if the local system experiences an increase in the independent component of productivity growth (d), in exports (b), or in wages beyond a certain threshold, the system will abruptly 'jump' to much higher rates of income growth (y^{***} in Figure 10.6b), even eliminating any possibility of negative growth.

10.4 Increasing returns internal to the firm: the new economic geography

10.4.1 The specific nature of the approach

Myrdal and Kaldor's model conceived increasing returns as economies of scale external to the firm: a simple assumption which enables formalization of the growth process in accordance with a market logic of perfect competition.

In the 1990s, thanks to the prolific work of its founder, Paul Krugman, a current of thought known as 'new economic geography' arose in regional studies, provoking considerable criticism by claiming its independence from regional science and from the discipline's 'founding fathers'.[23]

The most distinctive feature of this approach is that it eschews the direct assumption of economies external to the firm. Instead, it highlights local externalities as the consequences of market interactions among firms able individually to exploit internal scale economies, and it does so by making necessary reference to a market structure of imperfect competition.[24]

The aim of 'new economic geography' is to interpret the phenomena of industrial agglomeration – or of 'geographic concentration' to use Krugman's expression – on which local growth processes depend, and to do so by going beyond a simple explanation based on an unequal spatial distribution of resources and production factors.[25] This aim was in fact achieved by the first versions of the model, on the hypothesis of an initially homogeneous distribution of resources, through analysis of the location choices of firms and individuals within a neoclassical framework of the maximization of profits and individual well-being.

Like location theory, these models conceive the organization of productive activities in space as resulting from centrifugal and centripetal forces. In economic terms, the centrifugal forces are represented by the tendency of firms to cover spatially diffused demand and to avoid direct competition with other firms on small local markets. The centripetal forces enable firms to exploit increasing returns for broader outlet markets, and individuals to access markets offering a wider range of goods at more competitive prices and a higher standard of living.

Three fundamental elements affect agglomerative phenomena: increasing returns, transport costs and migratory movements. Increasing returns encourage activities to concentrate in space because they guarantee that relocation will be profitable, and also that profitability will increase because of the local market's expansion. The second element, transport costs, induces firms to locate close to broad outlet markets. And the third, migratory movements, influences both an area's labour pool and the size of the local market, both of which affect potential profits and incentivize agglomeration.

The agglomeration (dispersion) in space of firms and households generates cumulative conditions of growth (decline) in production; the process is irreversible unless contrary external forces intervene.

Krugman's base model incorporates the cumulative development model à la Myrdal and Kaldor as the increasing size of the market. Entry by new firms into a local market attracts new workers and population; these enlarge the local market, increase potential profits and offset the downturn in profits suffered by local firms because of greater local competition (competition effect). The larger size of the local market then stimulates entry by new firms, in a virtuous circle of agglomeration and development.

Operation of the cumulative mechanism is guaranteed by the externalities accruing to the firms located in the area. These externalities are generated by market interactions among firms that individually exploit internal economies of scale. In the presence of imperfect competition, in fact, the decision by a firm to enter a market unintentionally influences demand for a good produced by another firm – the entering firm's potential supplier – that is already operating in the area. This latter firm obtains pecuniary advantages from the expansion of its production because its average production costs diminish. Increasing returns thus turn into externalities which are termed 'pecuniary' in that they come about solely by virtue of trade activity, and take the form of greater potential profits for local firms.[26]

Innumerable models have been developed by 'new economic geographers' over the past decade. They can be distinguished according to the different ways in which inter-firm relations generate externalities: in some cases, these are associated with demand elements; in others with input/output relations among firms; and in others with research and development activities producing knowledge spillovers for local firms.[27] However, the logical framework within which agglomerative phenomena are studied does not change. It is set out in the next section in its original and simplest form.

10.4.2 *The centre–periphery model: the 'demand effect' and the 'cost effect'*

The base model, which goes by the name of the 'centre–periphery' model, seeks to explain the concentration of industrial activities on the hypothesis of an initially homogeneous distribution of productive resources – that is, in the absence of geographic and economic elements that might easily account for the agglomeration of manufacturing activities.[28]

The model's reasoning is based on the following assumptions:

- there are two regions, with two sectors: agriculture and manufacturing. The good in the manufacturing sector is produced at increasing returns, i.e. in conditions of imperfect competition, while the agricultural good is produced at constant returns, i.e. in conditions of perfect competition;[29]

- each manufacturing firm produces a good of different quality;
- the utility of consumers is influenced by both the quantity and the quality of the manufacturing goods produced;[30]
- transport costs exist if the manufacturing good is produced in one region and purchased in another. Transport costs are presumed to take Samuelson's 'iceberg' form; that is, they are calculated on the basis of the fraction of the good which does not reach the destination;[31]
- the labour force of the agricultural sector is immobile, while that of the manufacturing sector is free to migrate from one region to another;
- the labour force of the agricultural sector is homogeneously distributed between the two regions; and so too, at the initial stage of the process, is the labour force of the manufacturing sector.

Supposing an initially homogeneous distribution of activities (firms) in the two regions, a firm's decision (introduced exogenously into the model) to move from one region to the other alters the initial equilibrium[32] and triggers the following two effects (Figure 10.7a):

- *a competition effect.* Made possible by the hypothesis of increasing returns in the manufacturing sector, the new firm's entry into the market on the one hand squeezes the market shares of the other firms, and on the other reduces the prices of the goods produced, with the inevitable consequence that profits fall on the local market and the location becomes less profitable;
- *a demand (or market-size) effect.* The new firm's presence on the market increases the demand for labour. The wage differential between the two regions consequently widens in favour of the expanding local market, which attracts new workers, and with them new residents. The local market grows further in size, with positive effects on the profits of local firms. The increased profit makes the location more attractive.

These two effects therefore have opposite impacts on the profitability of the new location.[33] It is obvious that agglomeration of manufacturing activities will only come about when the net effect on profits generated by the entry of a new firm into the market is positive – that is, when the demand effect surpasses the competition effect. Whichever effect prevails over the other depends on the values assumed by certain parameters, which either amplify or reduce those effects. As regards the competition effect in particular, its intensity depends on:

- *the substitution elasticity among the goods produced by the manufacturing firms*: the greater the substitution elasticity among goods, the larger the fall in prices due to increased competition, and hence the larger the fall in the profits of already-existing firms;
- *transport costs*: the more transport costs increase, the more the prices of goods diminish. In this case, competition by firms situated in the other area is low, and any relocation will markedly alter competition.

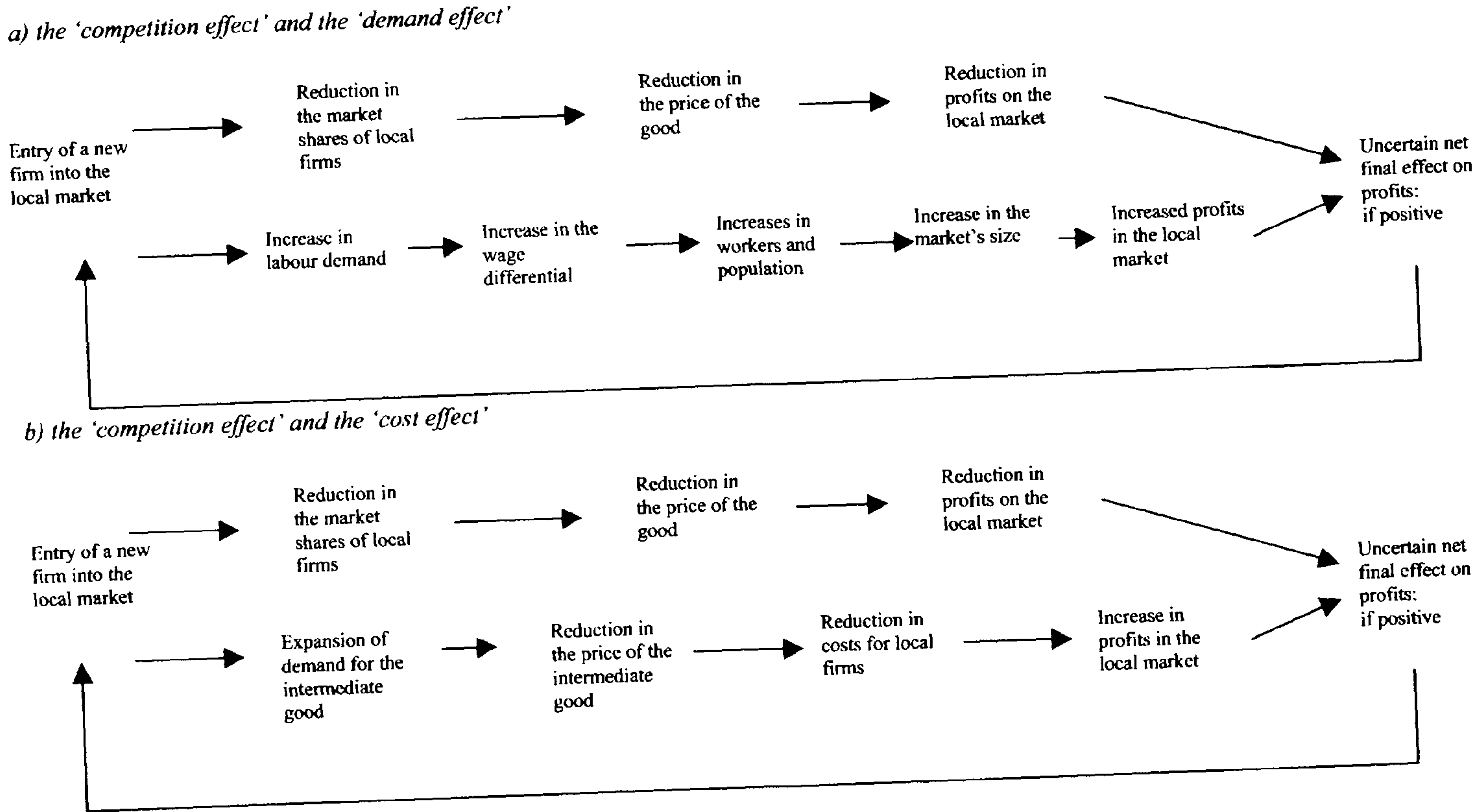


Figure 10.7 The virtuous cumulative development circle of the 'new economic geography'

As for the demand effect, its intensity depends on:

- *increasing returns*: higher increasing returns ensure wider profit margins for entrant firms, so that a larger number of firms are attracted into the local market. This determines the size of the market and the amount of profit that firms (new entrants and already-existing ones) can realistically expect;
- *the share of income spent on manufacturing goods*: the more this share of income increases, the greater the demand effect generated by a new firm's entry into the market.

If an area is to grow, the demand effect must exceed the competition effect. This happens if the varieties are difficult to substitute, returns to scale are intense, transportation costs are low and the share of income spent on agricultural goods is small. Under these conditions, the relocation of a firm produces a net increase in profit for all the firms operating in the area. Rising profits attract new firms, which further expand the size of the market, in a process of circular causation and cumulative growth.

Since empirical evidence has shown scant labour force mobility in Europe compared with the United States, and this despite wide wage differentials among the European countries and also among regions of the same country, various refinements have been made to the centre–periphery model.[34] One of the best known is the model developed by Anthony Venables, who reprises the original centre–periphery model. In his version of the model Venables stresses pecuniary externalities deriving from the presence of input–output linkages with other firms ('cost effect') instead of those due to a larger number of workers (and therefore of consumers) in the region ('demand effect').[35] In this manner, Venables is able to derive the mechanism of cumulative circular causation also on the assumption of labour force immobility.

In this version of the centre–periphery model, a new firm's location in a region reduces the costs of its goods, and it generates pecuniary advantages for firms downstream from its production. Furthermore, the presence of the new firm increases the demand of intermediate goods for its manufacturing process; and this demand increases the profits of upstream firms.

Also in this model, two effects ensue from a new firm's entry into the market (Figure 10.7b):

- a *competition effect*, as already described, which causes profits to decrease in the local market because of a reduction in the good's price resulting from greater local competition;
- a *cost effect*, which conversely generates an increase in profits through expansion of the market for the intermediate good; this expansion results from a decrease in the intermediate good's price for downstream firms, and an increase in the size of the market for upstream ones – on the assumption that all local firms use the same intermediate good.

Once again, the net effect on profit depends largely on the values of certain parameters. As before, the competition effect is strong if the varieties can easily be substituted for each other, and if transportation costs are high. The cost effect is strong if – on the assumption that the manufacturing good is at once an intermediate good for firms and a final good for consumers – final demand for the good by consumers is less than intermediate demand for it by firms.

It is evident from the foregoing discussion that the final outcome is largely indeterminate in the models of the 'new economic geography'. Concentration in a single region or the equal distribution of manufacturing activities on the market result from different values of the parameters. One way to analyse the dynamic properties of the system is to conduct numerical simulation.[36]

The base centre–periphery model demonstrates that multiple equilibria – of the concentration or equidistribution of activities – may exist according to the values of the parameters. However, if the tendency is towards concentration, whichever of the regions comes to 'host' the concentration of productive activities will depend on the regions' historical endowment of industrial activities. In this sense, history determines the economy's growth path towards one rather than another stationary equilibrium just as in Myrdal's model a region's initial economic structure determines – endogenous competitiveness conditions remaining equal – an explosive or implosive process of 'circular causation'.

10.5 A critical assessment of the model

As already pointed out, the strength of the theory developed by the 'new economic geography' is its capacity to include firm-level increasing returns in models of location choice, and at the same time express them with the elegant modelling of imperfect competition;[37] these features represent the main innovations introduced by the new economic geography.[38]

The success of the new economic geography approach resides in the formal elegance with which it accounts for spatial phenomena – such as location choices, the concentration of activities, and the economic growth deriving from agglomeration economies – within a framework of general economic equilibrium (final equilibrium in the markets of final goods and production factors). The location choices of firms and workers are matched by economic choices for profit maximization by firms and for welfare maximization by individuals, in a strictly neoclassical economic framework.

Furthermore, this neoclassical economic logic comprises positive (or negative) feedback mechanisms that render the process cumulative and have it tend towards concentration (rather than diffusion) and to growth (or decline). The process prior to equilibrium is therefore a path-dependent one in which the well-known elements of cumulativeness, learning and cross-fertilization of knowledge, and an economic system's feedbacks find appropriate systematization. This family of models conceives choices as being rendered *irreversible* by increasing local advantages or disadvantages which induce firms and workers – on the basis of entirely rational decisions – either to concentrate or diffuse. With these effects of path-dependency and irreversibility, these models closely resemble the modern economic theories of innovation and technological development,[39] and even more closely Myrdal and Kaldor's theory of cumulative circular development; hence, development is a cumulative process, and intrinsic to its workings are self-reinforcing mechanisms that push inexorably towards one or other extreme, concentration or dispersion, explosive development or decline, with no evident possible reverse trajectories.

The success and enthusiasm aroused by the 'new economic geography' in its early years of formulation tended to obscure the interpretative shortcomings of the approach.

The first of these shortcomings is that the theory identifies the sources of increasing returns in elements that are economic but not necessarily territorial. Increasing returns, in fact, arise within the firm and then become external economies taking the form of pecuniary externalities, cost advantages and profits generated by market relations, which do not necessarily require a territorial logic for their explanation.[40] The proximity advantages – physical but above all social and relational – of such importance for regional economists do not perform a central role in the generation of agglomeration economies. An implicit consequence is the necessary sacrifice of a finding now well established by local development theory: that the territory is an independent factor, an additional economic resource and an active determinant of the development process. Not surprisingly, therefore, as happens with all aspatial theories, the logic of these models does not change when they are applied to countries or to regions or to urban areas.[41] The new economic geography approach banally conceives space as punctiform, and, although it is space characterized by physical poles at which growth cumulates, as the mere container of development, with scant economic-geographic influence upon it.

A second shortcoming is the fact that a firm's location or its decision to relocate (according to whether the model hypothesizes an initially uneven or homogeneous distribution of activities), enters the model exogenously and determines its final result of concentration in one or the other region according to the values assumed by certain fundamental parameters. In a more recent version of his model, Krugman has investigated the role of history in determining the final equilibrium and introduced the possibility that in reality the development path follows the profit expectations of economic actors. But he lapses into the same type of paradox: these profit expectations are not determined by the model but are assumed to be exogenous.[42] Under some very specific conditions, their existence may indeed induce firms and workers to choose locations contrary to the historical development path, and they may give rise to equilibria completely opposed to those determined by history. Yet the model provides no definitions of the elements that determine the profit expectations of firms and workers in economic systems, and no explanations as to how those expectations are fulfilled. One may accordingly state that the models of the 'new economic geography' suffer from the same weakness as Perroux's and Boudeville's 'development poles' model, where the source of an area's growth, an *industrie motrice* for the former and export capacity for the latter, is left entirely unexplained.

Moreover, because of the underlying theoretical structure, it is impossible to introduce limits to growth and concentration into the model, so that it generates a cumulative accumulation of activities without this ever encountering physical obstacles (congestion) or economic ones (shortages of land and productive resources). Yet it would be more realistic to set limits on what is otherwise infinite growth by incorporating net disadvantages of concentration into the model. The onset of these disadvantages (even if only foreseen) would generate profit expectations in regions with lower locational advantages, thus explaining why spatial concentration may go into reverse.

To conclude, the new economic geography makes a commendable effort to include space in strictly economic models. Also to be commended is the implicit merging in its theoretical structure of the various conceptions of space put forward over the years; the merging, that is, of the physical-metric space represented by transport costs with the diversified space which assumes the hypothesis of the existence of certain territorial polarities where growth cumulates. However, the new economic geography is still unable to combine the economic laws and mechanisms that explain growth with territorial factors springing from the intrinsic relationality present at local level. An approach that did so would represent the maximum of cross-fertilization among location theory, development theory and macroeconomic growth theory and would give rise to a framework able to blend specifically local territorial features into a single macroeconomic model. Today, the frontier of knowledge in regional economics consists precisely in defining the territorial foundations of macroeconomic growth models. An endeavour in this direction is presented in Chapter 12.

10.6 Conclusions

The chapter has examined the first large group of the most recent theories on growth, whose distinctive feature is the resumption of macroeconomic models based on the increasing returns hypothesis. These new theories are rooted in Myrdal and Kaldor's model of cumulative circular causation, which they employ to interpret growth as a cumulative, endogenous and selective process. The models described hypothesize the existence of specific polarities in which development comes about as a result of increasing returns in the form of learning processes, economies of scale (at area or firm level), localization economies and urbanization economies that set off a virtuous circle of cumulative development.

The models illustrated in this chapter are non-linear dynamic systems. As in all the most recent approaches to such systems, they analyse the 'structural stability' of systems, the nature and quality of solutions and the form that the system may assume. For this reason, these models do not yield unequivocal results; rather, they offer a range of possible solutions, which vary according to the initial conditions and to the values assumed by the parameters.

These models envisage growth as a cumulative demand/supply process. But those examined in the next chapter include increasing returns to productive resources in a production function of neoclassical derivation. As we shall see, these theories regard growth as depending exclusively on *supply-side elements.*

Review questions

1. What is meant by diversified-stylized space?
2. What have been the obstacles to the analytical formulation of increasing returns up to the middle of the 1980s and what elements have recently been allowed to overcome these obstacles?
3. What is meant by an equilibrium in non-linearity conditions? What are the peculiarities of such an equilibrium?

4 What are the main aims of the circular and cumulative causation model? What is explained by the Verdoorn law? How are the increasing returns embedded in this theory?
5 Does the circular and cumulative causation model contain a concept of regional divergence? Explain your reasoning.
6 What are the main new elements contained in the 'new economic geography' theory?
7 What is meant by 'competition effect' and 'cost effect'?
8 Does the 'new economic geography' explain concentration or diffusion of activities in space?
9 What are the weaknesses and strengths of the 'new economic geography'?
10 What aspects contained in the 'new economic geography' were already contained in the circular and cumulative causation model?

Selected reading on empirical findings

About the new economic geography

Davis D.R. and Weinstein D.E. (1999), 'Economic Geography and Regional Production Structure: An Empirical Investigation', *European Economic* Review, vol. 43, no. 2, pp. 379–407.

Giarratani F., Gruver G. and Jackson R. (2005), 'Empirical Evidence Concerning the Economic Geography of Market Entry in the U.S. Steel Industry: Plant Location and the Advent of Slab Casting by U.S. Steel Minimills', Center for Industries Study Research, University of Pittsburgh.

Hanson G.H. (1998), 'Market Potential, Increasing Returns, and Geographic Concentration' Working Paper No. 6429, National Bureau of Economic Research.

Midelfart-Knarvik K. and Steen F. (1999), 'Self-reinforcing Agglomerations? An Empirical Industry Study', *Scandinavian Journal of Economics*, vol. 101, no. 4, pp. 515–532.

Mucchieli J.L. and Mayer T. (2004), *Multinational Firms' Location and the New Economic Geography*, Edward Elgar, Cheltenham.

Naudé W.A., Krugell W.F. and Gries T. (2005), 'The New Economic Geography: Empirical Evidence from South Africa', paper presented at the Regional Studies Association's International Conference on 'Regional Growth Agendas', University of Aalborg, Denmark, 28–31 May.

Overman H.G. and Puga D. (2002), 'Unemployment Clusters Across European Regions and Countries', *Economic Policy*, vol. 34, no. 2, pp. 115–147.

Redding S. and Venables A. (2000), 'Economic Geography and International Inequality', Discussion Paper No. 2568, Centre for Economic Policy Research, Reading.

Further reading

Barentsen W. and Nijkamp P. (1989), 'Modelling Non-Linear Processes in Time and Space', in Å. Andersson, D. Batten, B. Johansson and P. Nijkamp (eds), *Advances in Spatial Theory and Dynamics*, North-Holland, Amsterdam, pp. 175–192.

Dixit A. and Stiglitz J. (1977), 'Monopolistic Competition and Optimum Product Diversity', *American Economic Review*, vol. 67, no. 3, pp. 297–308.

Feenstra R.E. (2003) *Advanced International Trade: Theory and Evidence*. Princeton University Press, Princeton, NJ.

Fujita M., Krugman P. and Venables A.J. (1999), *The Spatial Economy: Cities, Regions and International Trade*, MIT Press, Cambridge, MA.
Krugman P. (1991), *Geography and Trade*, MIT Press, Cambridge, MA.
Krugman P. and Venables A.J. (1996), 'Integration, Specialisation and Adjustment', *European Economic Review*, vol. 40, no. 3–5, pp. 959–967.
Myrdal G. (1957), *Economic Theory of Under-developed Regions*, Gerald Duckworth & Co., London.
Verdoorn P.J. (1949), 'Factors That Determine the Growth of Labour Productivity', in D. Ironmonger, J.O.N. Perkins and T. Van Hoa (eds) (1988), *National Income and Economic Progress; Essays in Honour of Colin Clark*, Macmillan, Basingstoke, pp. 199–207.

Notes

1 Linear equations (or systems) are equations (or systems) in which the variables are raised to the power one, and in which there are no products of different variables. Higher-level equations (systems) are equations (systems) which comprise terms as the product of two variables (yx) or of variables with superscripts (y^2, y^3, etc.).
2 The increasing returns hypothesis entails that firms have surplus productive capacity to exploit when the market expands. In other words, as the market expands, firms are able to increase their output, moving along the decreasing cost curve and obtaining increasing returns. Perfect competition instead hypothesizes that firms produce in conditions of minimum average cost.
3 Dixit and Stiglitz produced the first formalized model of imperfect competition à la Chamberlin. All the models which introduce increasing returns into growth paths presented in this chapter are based on Dixit and Stiglitz's original formulation. See Dixit and Stiglitz, 1977. For surveys of the literature on non-linear dynamic models applied to the dynamic of territorial systems see, among others, Wilson, 1981; Lung, 1987; Nijkamp and Reggiani, 1988, 1992 and 1993; Barentsen and Nijkamp, 1989; Reggiani, 2000.
4 See von Böventer, 1975, p. 3.
5 Some seminal ideas in Myrdal's theory had already been propounded by Young. See Young, 1928.
6 Given that it has been decided here to address these new theories from within a strictly economic framework, the treatment that follows will often use graphical tools, without excessively encumbering itself with mathematics. References will be provided to the specific literature for mathematical aspects.
7 Catastrophe theory originated in a work published in 1972 by the French mathematician, René Thom. It consists in a mathematical account of morphogenesis, or the formation of a system's structure. Like bifurcation theory, it analyses non-linear dynamic systems characterized by multiple equilibria in which the passage from one equilibrium to another may be triggered by a sudden and slight variation in the parameters determining the system's dynamic. See Thom, 1972.
8 A 'stable dynamic equilibrium' exists when, although the system may have been distanced by an external force from a certain value of the unknown variable, for instance y^*, it is able to return to that value. In this case, y^* is also called the 'attractor' of the time path, or of y's trajectory. In the contrary case, it is called the 'repulsor' and the equilibrium is termed an 'unstable dynamic equilibrium'.
9 Emphasized in the case of non-linear models is the difference between a locally stable equilibrium point and a globally stable equilibrium point. A point y^* is said to be of locally stable equilibrium if it is possible to define at will a small neighbourhood of y^* such that, for initial conditions within the interval, the function $y(t)$ tends to y^* for $t \rightarrow \infty$. A point y^* is instead called globally stable if the function $y(t)$ tends to y^* for $t \rightarrow \infty$ for every initial condition of y. See Barentsen and Nijkamp, 1989.
10 See Miyao, 1987b.
11 The stability of the equilibrium can be straightforwardly verified by using a 'phase diagram'; that is, by setting the value of the variable (the population in the case of Figure 10.1)

respectively at time t and $t-1$ on the two axes. If the function is positively sloped, as it is in the present case, there is a stable dynamic equilibrium if the slope is below 1 (see also Figure 10.2.a), and an unstable dynamic equilibrium if the slope is above 1 (Figure 10.2.b). The former case is that of growth at decreasing rates in the variable over time (slope below 1); the latter case is that of growth at increasing rates (slope above 1). If the function is negatively sloped, it depicts an equilibrium reached in oscillatory and convergent manner if the slope is below 1 (Figure 10.2.c), and in oscillatory and divergent manner if the slope is above 1 (Figure. 10.2.d).

12 We saw in Chapter 6 that convergence is an inevitable result of the one-sector neoclassical model if the growth rates of regions are assumed to be initially different. This result troubled the authors of the model because it was not confirmed by empirical evidence. Using the mathematical tools of the time, Burns and Stein managed to prove divergence with the two-sector model, on the hypothesis of initially equal growth rates among regions. There matters stood until the 1980s, when the neoclassicals were able – by introducing non-linearity into the original model – to prove divergence for certain parameter values even on the assumption of different initial rates of growth. See Chapter 11.

13 See Myrdal, 1957.

14 Young had previously suggested the importance of increasing returns for the development of an economic system. See Young, 1928. In the same years, Perroux stressed, within a microeconomic framework, the importance of agglomeration economies for local development.

15 See Myrdal, 1957, chap. 3.

16 See Kaldor, 1970. A study on a convergent or divergent development of Kaldor's model is contained in Dixon and Thirlwall, 1975.

17 The influence of the export-base theory is evident here.

18 In an article published in 1949, Verdoorn demonstrated empirically, using a typically Smithian approach, the existence of a relationship between the size of the market and productivity gains, and a positive relationship between the rates of output growth and productivity growth.

19 See Verdoorn, 1949. Empirical verification of Verdoorn's Law is still a matter of much controversy. The first empirical tests conducted by Kaldor were criticized for assuming an endogenous relationship between the dependent variable (productivity) and the independent variable (employment), the reason being that the latter is by definition at the denominator of the productivity index. See Rowthorn, 1975. Kaldor rebutted these criticisms with empirical proof that the relationship between the output growth rate (in its turn correlated with the employment growth rate) and the productivity growth rate does not hold in some sectors, in particular agriculture and trade. The debate is still animated today. See Kaldor, 1975 and Thirlwall, 1983. For empirical tests of Verdoorn's Law see Leon-Ledesma, 1998, for Spain; McCombie and de Ridder, 1984, for the United States; Rid and Lau, 1998, for the UK; Soro, 2003, for Italy.

20 The presence of a constant d, which explains the exogenous growth of productivity independently of output, can resolve the dispute among regional economists on whether it is the output growth rate that determines the productivity growth rate, as Verdoorn's Law postulates, or vice-versa whether it is the productivity growth rate that determines the output growth rate. An empirical test of Verdoorn's Law for the Italian regions has shown that the independent component of productivity not explained by the output growth rate accounts for the largest share of the overall growth of productivity. See Soro, 2003.

21 The presence of relatively more export-oriented sectors and of less import-dependent sectors facilitates the onset of virtuous development circles – as was seen earlier (in strictly Keynesian terms) in the case of Thirlwall's Law. With respect to the latter, however, in Myrdal and Kaldor's model development depends on other elements, like the competitiveness of the local system expressed in the growth rates of wages and domestic prices, and increasing returns: these are entirely absent from Thirlwall's Law.

22 Although these considerations had already been put forward by Richardson, it was Miyao who introduced non-linearity into the structural relations conceived by Kaldor. See Richardson, 1978, p. 148, and Miyao, 1984.

23 In 1991 Krugman wrote: 'I could have entitled this book "Location and Trade". I was afraid, however, that this would convey too narrow an idea of what I was trying to say. Although the intellectual tradition of location theory is both wide and deep, what is thought is usually a very narrow set of geometric tricks involving triangles and hexagons (. . .). "Location" seemed too restrictive a term for this field. Location theory, however, is part of a much broader field, that of economic geography' (Krugman, 1991a, pp. x–xi). For criticism of Krugman's claimed independence from regional economics see Martin, 1999. See Gans and Shepherd, 1994, for a well-known critique of Krugman's studies: 'it's obvious, it's wrong and anyway they said it years ago'.

24 The theoretical underpinning of these models is Dixit and Stiglitz's model of monopolistic competition. See Dixit and Stiglitz, 1977.

25 Krugman's interest in analysing the location of productive activities stems from a simple observation: the dense concentration of manufacturing activity in the so-called 'manufacturing belt' of the northeastern United States, which according to an estimate by Perloff and colleagues, in 1957 already accounted for 64 per cent of manufacturing employment in the country. See Krugman, 1991a; Perloff et al., 1960.

26 Krugman regards pecuniary externalities as the great merit of his model. Defined as the externalities (or advantages) that arise from trade, pecuniary externalities can be easily quantified by variations in profit; for this reason they are more readily identifiable than technological externalities, which are generated by proximity among firms and are difficult to quantify and model. See Krugman, 1991b. I would suggest, however, that precisely because pecuniary externalities are tied to market relations, they may arise independently of geographical, social and cultural proximity among firms, impoverishing the role of territorial factors in the determination of local advantages. See Chapters 8 and 9 in this book. See Krugman, 1991a, p. 15; Fujita and Thisse, 1996 and 2002.

27 See, e.g., Faini, 1984; Krugman and Venables 1996; Venables, 1996; Baldwin 1998; Ottaviano and Puga, 1998; Baldwin et al., 1999; Martin and Ottaviano, 1999; Ottaviano and Thisse, 2001.

28 The term 'centre–periphery' denotes the geographic dichotomy between an area in growth (the centre) and one in decline (the periphery) generated by the model under certain conditions.

29 Increasing returns to scale are formalized by hypothesizing that the good is produced at a fixed cost:

$$L_M = \alpha + \beta x_M \tag{10.1n}$$

where L_M is the quantity of workers necessary to produce the variety M of a generic manufacturing good, α the fixed share of workers, x_M the quantity of the manufacturing good M produced, and β the share of workers proportional to the quantity produced. The consequence of this hypothesis is that each firm produces one and only one variety in equilibrium.

30 Consumer interest in the variety of the good – that is, the notion that consumers obtain greater utility, the quantity consumed remaining equal, the larger is the number of varieties of the differentiated good available – is formalized by means of the following utility function:

$$U = x_M^{\pi} x_A^{(1-\pi)} \tag{10.2n}$$

This states that an individual's well-being depends on possession of both the agricultural good (x_A) and the composite industrial good (x_M). π represents the share of spending on products from the manufacturing sector, and its complement to one *($1 - \pi$)* is the share of spending on the agricultural good. The quantity consumed of the composite industrial good x_M is a function of the consumption of individual industrial goods:

$$x_M = \left(\sum_{M=1}^{m} x_i^q \right)^{1/q} \tag{10.3n}$$

m represents the number of varieties available, of which there exist a large number, although not all of them are produced, and q the intensity of the 'preference for variety'. When the value of q is close to 1, the manufactured goods are almost perfect substitutes for each other. Conversely, if q tends to 0, the desire of individuals to consume a wide variety of goods is very high, with the consequence that the value of x_M is high as well. Setting $\sigma = 1/(1-q)$ this is substitution elasticity between two different varieties of manufactures, and equation (10.3n) becomes:

$$x_M = \left(\sum_{M=1}^{m} x_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{10.4n}$$

31 'Iceberg' transport costs were formally introduced by Samuelson in 1954. But von Thünen had hypothesized costs of this kind as early as 1826 when he stated that the transport cost of grain could be identified in the quantity of the grain consumed by the horses used to transport it. See Samuelson, 1954; von Thünen, 1826. Formalization of this hypothesis is straightforward: if for each unit of a good transported from one region to another, only a fraction $\tau < 1$ arrives at the destination, then in order to transport one unit of the good from one region to another, a quantity $1 + {}^{1}\!/_{\tau}$ must be shipped. ${}^{1}\!/_{\tau}$ denotes the transport costs incurred by the firm in order to transport the good from one region to another; τ represents the fourth parameter influencing the growth processes of the two regions.

32 Interestingly, if the model does not hypothesize increasing returns, the agglomeration process immediately meets an obstacle: the non-existence of extra profits, which renders the relocation unprofitable, so that the process is blocked at the outset.

33 The formalized version of the model defines the individual demand curve through maximization of the individual's utility (equation 10.2n), given the budget constraint and assuming that all individuals have the same utility functions. Straightforwardly obtained from the individual demand curves is the market curve of a typical variety produced in the region:

$$p = x_M^{-1/\sigma} P(n,\tau)^{1-(1/\sigma)} (A+L)^{1/\sigma} \tag{10.5n}$$

where p represents the price of the variety produced (assumed equal for all varieties), x_M the quantity of the generic variety M required by the market, P the price index, σ substitution elasticity between two different varieties of manufactures, τ the transport costs, n the number of firms present in the market, and $(A + L)$ the number of residents and therefore the size of the local market. p is therefore a function of the elasticity of demand among the varieties produced, the number of firms in the area, the transport costs, and the demand for labour, which depends on the fixed cost at which the firm produces the variety of the good.

Given the market demand curve of the variety of the good, the firm fixes the quantity and sale price of the variety of the good on the basis of its marginal costs, the purpose being to maximize its profit. The arrival of a new firm in the area alters the equilibrium reached by each firm. Market demand for each firm on the one hand diminishes because of the fall in prices due to greater competition, but on the other hand increases because of the expansion of the market. Which of the two effects will prevail depends on the values of the other parameters, which represent the elasticity of demand among the varieties produced, the demand for labour (which determines the number of firms in the area), the transport costs and the substitution elasticity among the goods produced by the manufacturing firms.

34 See Blanchard and Katz, 1992, for an empirical analysis of the United States; and Decressin and Fatàs, 1995, for an analysis of Europe.

35 See Venables, 1996.

36 See Fujita et al., 1999, for a numerical simulation.

37 As said, apart from attempts (some very successful) to model location choices in imperfect competition, for example by Hotelling and Lösch, the tendency in regional economics has been to use more easily formalized models with perfect competition, and to assume increasing returns external to the firm (localization or urbanization economies).

38 In one of his articles, Krugman himself has summarized the innovative elements of the 'new economic geography' as these: increasing returns in an imperfect competition model à la Dixit and Stiglitz, transport costs à la Samuelson, multiple equilibria and numerical simulation for empirical analysis. See Krugman, 1998.
39 See Dosi 1982; David, 1985; Arthur 1989 and 1990; Dosi et al., 1989.
40 In a subsequent refinement of their model, Krugman and Venables furnish an explanation for the tendency of local economies to specialize (an aspect hitherto not included in the formalization); on their approach, specialization results from the circular causality mechanism operating in input–output relationships among local firms. The difference with respect to Venables' model presented above is that the advantages generated by a new firm in the market are selective; they derive solely from the firms of the sector in which the new entrant operates, while the greater competition on the final goods and labour markets impoverishes all the firms in the area. The region is thus induced to specialize in the sector that enjoys agglomeration advantages. See Krugman and Venables, 1996; Venables, 1996.
41 It is important to note that the 'new economic geography' has become an important topic in the most recent handbooks on the economics of international trade. See Feenstra, 2003.
42 See Krugman, 1991c.

11 Territorial competitiveness and endogenous growth

11.1 Endogenous growth and increasing returns

This chapter examines the last group of growth models, which represent – together with those of the 'new economic geography' – the most recent of such models. They closely resemble the ones presented in the previous chapter in that they have a high level of formalization and a strictly dynamic structure. As the chapter proceeds, we shall again be dealing with models that investigate the endogenous determinants of an aggregate growth rate, as opposed to the individual microeconomic and micro-territorial elements of competitive or locational advantage typical of development theories. Once again these will be models with a high degree of mathematical formalization, which conceive increasing returns as economies of scale or learning processes, and which stylize them in equations explaining the growth rate of per capita output.

These are therefore theories and models that conceive space as diversified-stylized; a space in which growth results from increasing returns but does not have a real and proper territorial dimension. More specifically, increasing returns are included in a neoclassical production function, where they offset the effect of the marginal productivity of the individual factors, which the traditional neoclassical approach assumes to be decreasing.

The strictly neoclassical logic of these models accounts for the interpretation that they give to growth. As in the neoclassical models discussed in Chapter 6, growth is once again associated with an increase in labour productivity, with a rise in per capita income, and therefore with an increase in individual well-being.

The origin of these new neoclassical growth models is Robert Solow's well-known model developed in the 1960s. On the assumption that the only reproducible factor (capital) is characterized by decreasing marginal returns, Solow demonstrated that the economy is bound to register nil per capita output growth in the long period unless the existence of technical progress is exogenously hypothesized. By so doing, however, Solow identified the engine of economic growth as an exogenous factor linked to the progress of knowledge.

However, the assumption that increases in factor productivity stem from endogenous factors – such as innovation, scale economies, and learning processes – requires the removal of perfect competition and constant returns from the theoretical framework, and the inclusion in their stead of increasing returns or imperfect markets. This shift requires complex modelling based on the only recently developed theoretical and analytical tools outlined in the previous chapter.

Chronologically, the first of these models introduced advantages external to firms and therefore continued to assume perfect competition. Subsequent ones included aggregate increasing returns, or constant marginal productivity of a single accumulable factor, in their production functions, but still envisaged a perfectly competitive market structure. Finally, the most recent models introduce technological innovation endogenous to firms in conditions of monopoly or monopolistic competition.[1] On each of these assumptions, the possibility arises of prolonged and balanced growth at a constant and positive rate. For these models, such growth is possible because the economy is able over time to accumulate a resource that yields non-decreasing returns and is a perpetual source of development.

The aim of the models now described – called 'models or theories of endogenous growth' – is therefore to identify the conditions *endogenous* to the productive system that ensure long-term positive growth. The latter is made to depend solely on increasing returns to productive resources (individual or in aggregate), and therefore on supply-side elements. It is this feature that differentiates these models from those analysed in the previous chapter, for which increasing returns gave rise to virtuous circles of demand/supply development.

The next section will examine two initial models that identified the sources of growth in local knowledge embodied in physical capital, and learning – elements already emphasized by some of the development models examined in Chapter 9. Then an interesting application of increasing returns is presented. This adopts a more strictly territorial production function in which the physical size of the region (or the city) is the factor that generates increasing returns. The final section shows how the results of the neoclassical interregional growth model discussed in Chapter 6 change if non-linearity is introduced into economic processes.

11.2 The endogenous sources of growth: the knowledge stock and learning

11.2.1 The limitations of the traditional model

We saw in Chapter 6 that the neoclassical model of regional growth is based on technical progress on the one hand, and on growth of the production factors on the other. These are synthesized by the model into a regional economy production function of generally Cobb–Douglas type:

$$Y = AK^{\alpha}L^{1-\alpha} \qquad 0<\alpha<1 \tag{11.1}$$

where Y is income, A technical progress, K capital, L labour, and α and $1-\alpha$ are the contributions of capital and labour respectively (and consequently their distributive shares).

In logarithms, the variation income Y over time is written as:[2]

$$y = a + \alpha k + (1-\alpha)l \tag{11.2}$$

where the lower-case symbols y, a, k and l denote the growth rates of income, technical progress, capital and labour respectively. Equation (11.2) states that the possibility of growth in local output depends on the growth of technical progress, of capital and labour. Equation (11.2) can also be written as:

$$y - l = a + \alpha(k - l) \tag{11.3}$$

The steady state (i.e. long-term dynamic equilibrium) in which the rate of output growth is constant is guaranteed if and only if the rate of capital growth is equal to that of labour. This situation is equivalent to nil long-term growth of per capita income, unless one assumes an increase in technical progress as represented by parameter a. This parameter, which is also called 'the Solow residual', represents the part of an economy's growth not due to growth of the production factors, and which is therefore not 'explained' by the model.[3]

As pointed out in Chapter 6, the model of interregional growth suggests that regions grow at the same rate in the long run. This is because the distribution of the production factors is equal among regions, and so too is technical progress, given the assumption that all regions have the same production function.

This result is perplexing, for several reasons. First, the sole long-term determinant of growth, namely technical progress, is exogenous to the model; nothing explains the real capacity of a system to grow. Second, as seen earlier when the diffusion of innovation was discussed, the capacity to utilize external and available technical progress differs greatly among regions, and the assumption that parameter a is equal for all regions can only be accepted if it is hedged about with caveats. Besides these theoretical problems, the lack of systematicity in the empirical results on convergence among the growth rates of countries and regions casts further doubt on the validity of the theoretical model.[4]

All the models now described seek to determine an endogenous mechanism that explains the growth rate of per capita output.[5] They identify this mechanism in non-decreasing returns, and in externalities which may have various origins:[6]

- investment cumulated in physical capital and the consequent increase in 'technological capacities' over time (learning-by-doing);
- the aggregate impact of investment by individual firms, which generates a positive externality (and increasing returns to scale) at aggregate level;
- the constancy of the marginal return on capital if this includes all the accumulable factors, among them 'human capital';
- the investment in human capital, scientific and technical knowledge that improves the physical productivity of labour;
- investment in R&D to foster the technological innovation that improves the physical productivity of all the factors – i.e. the creation of intermediate and final goods with high value added.

Among the numerous models that have been propounded, discussed here are those developed by Paul Romer and Robert Lucas, for which the sources of growth are factors – the knowledge stock and learning – already identified by the theory of local development, and in particular by the theory of the *milieu innovateur*.[7] Moreover, as regards formalization, these models employ two methods to endogenize growth:

- in Romer's model: by introducing a source of externalities that converts decreasing returns into constant or increasing ones;

- alternatively, in Lucas's model: by introducing a production factor into the production function for which is hypothesized a law of motion such that the share of the factor used is regenerated in the same quantity (whatever that quantity may be).

11.2.2 *The knowledge stock: Romer's model*

One of the first models of endogenous growth was formulated by Paul Romer. Its central assumption concerning the source of growth was that the externalities generated by technical knowledge, and then embodied in the investments accumulated in fixed capital until a certain time t, have the nature of public goods. They are in fact available to all firms whether or not these have participated in creation of that knowledge.[8] Romer incorporates knowledge into his model as 'public capital', which is a further accumulable resource besides private fixed capital. Its existence gives rise to economies of scale in aggregate factor productivity even though the returns on the individual production factors are decreasing.

Romer's model assumes the existence of N identical firms (i) with the same production function, as follows:

$$Y_{it} = K_{it}^{\alpha} L_{it}^{1-\alpha} K_t^{\beta} \qquad where\ 0 < \alpha < 1 \quad 0 < \beta < 1 \tag{11.4}$$

As well as the usual production factors – capital (K_i) and labour (L_i) – the production function comprises a third factor (K) which represents the state of technical knowledge at time t, doing so with typical logic à la Arrow whereby knowledge is embodied in accumulated experience; or in other words, embodied in the stock of accumulated investments in capital until time t. The difference between the two types of capital is that the former (with the index) is the traditional physical capital wholly exploited by only the firm which possesses it; the latter is the capital represented by the stock of technical knowledge acquired through action by all firms; and it is a public good because it is available to all firms.

In this model, therefore, firms benefit not only from their investments but also from the knowledge acquired by other firms. Capital and labour combine with the usual decreasing returns to the factors K_i and L_i. The third factor, without the index, is also characterized by decreasing returns, but its presence – which takes the form of an externality – enables firms to offset the decreasing returns on individual factors so that aggregate factor productivity increases (the sum of the exponents is greater than 1).[9]

Using logarithms, deriving with respect to time, and denoting the growth rates of the variables with lower-case letters, we obtain:

$$y_{it} = \beta k_t + \alpha k_{it} + (1-\alpha) l_{it} \tag{11.5}$$

which can be rewritten as:

$$y_{it} - l_{it} = \beta k_t + \alpha (k_{it} - l_{it}) \tag{11.6}$$

Equation (11.6) shows that the growth rate of per capita output – i.e. the growth rate of average labour productivity (left-hand member of the equation) – increases the higher the capital/labour ratio ($k - l$) and the greater the amount of knowledge that the firm obtains from outside as an externality (k). The steady state – i.e. the condition in which the rate of growth of per capita capital is nil ($k - l = 0$) – is here averted by the presence of a public good embedded in the knowledge externalities that firms exploit.

Omitting the index i to simplify the notation, (11.6) can be rewritten as:

$$y_t - l_t = (\alpha + \beta)k_t - \alpha l_t \tag{11.7}$$

On the hypotheses that $0 < \alpha < 1$ and $0 < \beta < 1$, but $\alpha + \beta > 1$, as in the above model, the growth of per capita income will be positive and cumulative, and the economy will have a perpetual source of growth of productivity, and consequently of individual well-being.

11.2.3 Learning and human capital: the Lucas model

The model developed by Robert Lucas envisages two types of capital: physical and human. Combined in a production function of usually Cobb–Douglas type, these give rise to a certain level of output:[10]

$$Y_t = AK_t^{\alpha}(u_t H_t L_t)^{1-\alpha} H_t^{\phi} \tag{11.8}$$

where A is a proportionality factor constant in time (and which can therefore be eliminated by choosing an appropriate unit of measurement: it is *not* an indicator of technical progress as previously), K is physical capital, L the number of workers, u the fraction of their time that individuals devote to work, and H is the average amount of knowledge possessed by workers (i.e. it is an indicator of the quality of human capital).

Lucas hypothesizes that workers accumulate knowledge by taking time off work in order to acquire skills ('learning by schooling'), under the following law:[11]

$$h_t = H_t \varphi(1 - u_t) \tag{11.9}$$

where h denotes the rate of growth of human capital over time, H the stock of human capital (or the average amount of knowledge possessed by workers), ($1 - u$) the time devoted to education, which is indicated as a percentage of the total amount of time available to individuals, and where φ is learning ability, which is assumed to be positive and linear with respect to the level of knowledge attained.[12]

In steady state, u must be a value such that workers are able to produce the tangible goods that directly generate utility and well-being but also have sufficient opportunities to accumulate knowledge and to increase the labour productivity that indirectly influences utility. In this model, human capital is simultaneously the result of a productive process and the source of increased labour productivity, and therefore of greater per capita income.

On the hypothesis that u is constant over time, on switching to logarithms, the rate of output growth can be straightforwardly obtained from (11.8) and (11.9):

$$y_t = \alpha k_t + (1 - \alpha)l_t + (1 - \alpha + \phi)h_t \tag{11.10}$$

where y, k, l and h are respectively the rates of growth of output, physical capital, labour and human capital at time t.

Recalling that human capital grows according to the law represented by (11.9)[13] and rewriting (11.10) in terms of the rate of growth of per capita output, we obtain:

$$y_t - l_t = \alpha(k_t - l_t) + (1 - \alpha + \phi)\varphi(1 - u_t) \qquad (11.11)$$

In steady state, where the growth rate of capital equals that of labour, $\alpha(k_t - l_t)$ is equal to zero. In this situation, there are two endogenous elements generating growth in per capita output: the externalities of a skilled labour market, expressed by the parameter ϕ, which enable the economic system to achieve increasing returns; and learning ability φ, which determines the law of human capital accumulation.

Interestingly, contrary to Romer's model, even if there are no external effects in knowledge ($\phi = 0$), the economy's growth is endogenous, and it depends on learning ability. According to the logic of the model, in fact, the growth rate converges on a steady state and is equal to a constant rate $\varphi(1 - u)$.

11.3 A critical assessment

One of the main merits of these models is their ability to endogenize the elements responsible for growth by considering increasing returns in the form of local externalities or laws of direct resource accumulation, without these having first to be converted into output.[14] These models thus elegantly deal with the problem encountered by Solow: when estimating the growth of per capita output in the US economy, he found that the largest part of it could be explained by technical progress, and therefore by precisely the element which his model did not explicitly consider.

A further interesting aspect of the Romer's and Lucas' models is their conception of growth, which in many respects resembles that of the more traditionally territorial development theories set out in Chapter 9, and the *milieu innovateur* theory especially. The similarities between Romer's model and the *milieu innovateur* theory concern:

- *the element that determines growth*, which consists in the positive externalities resulting from a process of collective learning. In Romer's model, growth derives from the increasing marginal returns that knowledge generates in production. This process converts decreasing marginal returns to the production factors into increasing returns and thus enables the economic system to grow. In the *milieu innovateur* theory, the engine of local development is the presence of advantages external to firms that generate dynamic efficiency. Among these advantages are the 'collective' learning mechanisms that enable the spatially concentrated system of small firms to become more innovative.
- *the specific features of knowledge*. Both theories conceive knowledge as a public good. In Romer's model, knowledge is a public good because it is available to all firms; in the *milieu innovateur* theory, learning is likewise a public good because all the firms belonging to the *milieu* can access it, doing so, for example, through high mobility of the local labour force;
- *external effects of knowledge on the growth process* as a consequence of its nature as a public good. In Romer, knowledge generates positive externalities for all firms and enables them to offset the effects of decreasing marginal returns

to the production factors. This idea is exactly the same as the *milieu innovateur* theory's conception of collective learning as coming about through socialization, rather than through an explicitly co-operative process, in a dynamic local labour market with high internal turnover;

- *hypotheses on returns to the production factors – in particular returns to knowledge*. Romer maintains that knowledge, like the other production factors, yields decreasing marginal returns. Also the *milieu innovateur* theory assumes decreasing returns to local knowledge, which the local system must overcome lest it be trapped on a development path at decreasing returns due to its inability to shift to more advanced technological trajectories.

There are also similarities between Lucas's model and the *milieu innovateur* theory, as follows:

- *learning as the key to growth*. In Lucas, human capital accumulation is the source of economic growth because it stimulates labour productivity and generates productive capacity. The *milieu innovateur* theory likewise stresses learning by human capital as the source of innovative capacity, and therefore of local development; for this theory, the accumulation of knowledge is facilitated by spatial, cultural and institutional proximity among firms;
- *the external acquisition of human capital with positive effects on growth*. In both theories, the acquisition of knowledge from outside has positive effects on growth. In Lucas's model, the positive effect of human capital accumulated externally to the firm ($\phi = 0$) is amplified by a typical externality mechanism. In the *milieu innovateur* theory, knowledge accumulated externally to the *milieu* drives the long-term development of the area and prevents the system from falling irremediably victim to the decreasing returns on local knowledge.

However, it should be stressed that endogenous growth models have a serious weakness: their *aspatiality*, which makes them very different from the *milieu innovateur* theory. Like all endogenous growth models, those of Romer and Lucas suffer from the lack of any active role performed by territorial variables. Indeed, the same models apply at different territorial levels, whether national or regional (the criticism holds for the new economic geography as well). Romer's and Lucas's models consequently differ radically from the *milieu innovateur* theory, which assumes that territorial features – spatial, cultural and institutional proximity, the area's sectoral specialization, the structure and dynamic of the local labour market – are the determinants of knowledge socialization, and of local development. The Romer and Lucas models instead more closely resemble those of the new economic geography, or even the early models of regional growth, which were typically aspatial.

Attempts to remedy this detachment from the territorial context have been made by the numerous empirical analyses that go by the name of β-conditional estimation methods. These seek to identify socio-economic variables – such as human capital, schooling, the infrastructural level of the country or region – that explain why advanced regions achieve higher growth rates than do backward regions, also taking account of differences at territorial level.[15] From the theoretical point of view, this means moving beyond Solow's traditional model to consider the more recent models of endogenous development; but above all it requires the introduction into specific models of the idea that growth results from the structural and socio-economic features of the local economy.

11.4 The neoclassical interregional model with increasing returns

An interesting application of increasing returns in a strictly territorial production function is the neoclassical interregional model at increasing returns proposed by Takahiro Miyao. This model takes the physical size of the region (or city) to be the factor that generates increasing returns:

$$Y = L^{\beta} K^{\alpha} L^{1-\alpha} \tag{11.12}$$

where $0 < \alpha < 1$ and $0 < (1 - \alpha + \beta) < 1$, in which Solow's generic technical progress is replaced by the size of the region (indicated by the number of workers L), and with agglomeration economies or diseconomies depending on whether β is positive or negative.

As in the original neoclassical model, the time paths of the variables K and L depend on the accumulation rate, and on differences with respect to the other regions in factor remuneration.[16] In symbols, this means that the growth rate of capital k is given by:

$$k = \frac{sY}{K} + \mu(i_r - i_w) \tag{11.13}$$

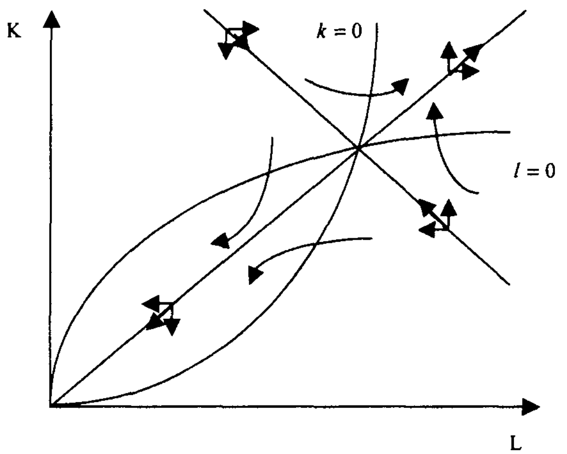
where sY represents the total savings available for the financing of investment (ΔK), i_r the rate of capital remuneration in the region, and i_w in the rest of the world. In its turn, the rate of labour growth l is given by:

$$l = n + \lambda(w_r - w_w) \tag{11.14}$$

where n is the natural growth rate of the population, $w_r - w_w$ the difference in wage remuneration between the region and the rest of the world, and μ and λ respectively the sensitivity of capital and labour to change in wage differentials.

Assuming as known the parameters (μ, λ, w_w, i_w, s, n) of equations (11.2n) and (11.3n), which represent the steady-state equilibrium curves (that is, the constant growth trajectories of the production factors), it is possible to examine the properties of the possible solutions by drawing these curves as in Figure 11.1.[17] In the presence of agglomeration economies, that is for β greater than zero, the steady-state equilibrium is unstable. As Figure 11.1a shows, when the *K/L* ratio is less than the steady-state ratio (and therefore lies below the main diagonal), labour increases at a greater rate than capital, so that the *K/L* ratio is increasingly distanced from the level that ensures dynamic equilibrium. If the *K/L* ratio is instead greater than the steady-state ratio, capital grows at a higher rate than labour, and the consequence is once again that the local economy cannot achieve constant growth. In other words, the area's economic growth explodes or implodes according to the initial factor endowment; a situation very different from the tendency to convergence expressed by the base model of the 1960s. Conversely, in the presence of agglomeration diseconomies, the region's economic growth converges on a steady-state equilibrium with constant values of K and L (z in Figure 11.1b).[18]

a) Presence of agglomeration economies (unstable equilibrium $\beta > 0$)



b) Presence of agglomeration diseconomies (stable equilibrium $\beta < 0$)

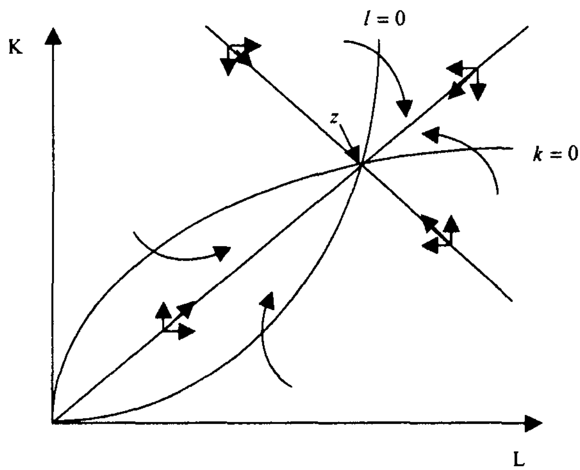


Figure 11.1 A neoclassical growth model with agglomeration economies and diseconomies
Source: Miyao (1984)

It is also possible to hypothesize the case in which agglomeration economies exist up to a size threshold L_0 beyond which these economies turn into diseconomies. In this case the production function is:

$$Y = \left(\frac{L}{L_o}\right)^{\beta} K^{\alpha} L^{1-\alpha} \tag{11.15}$$

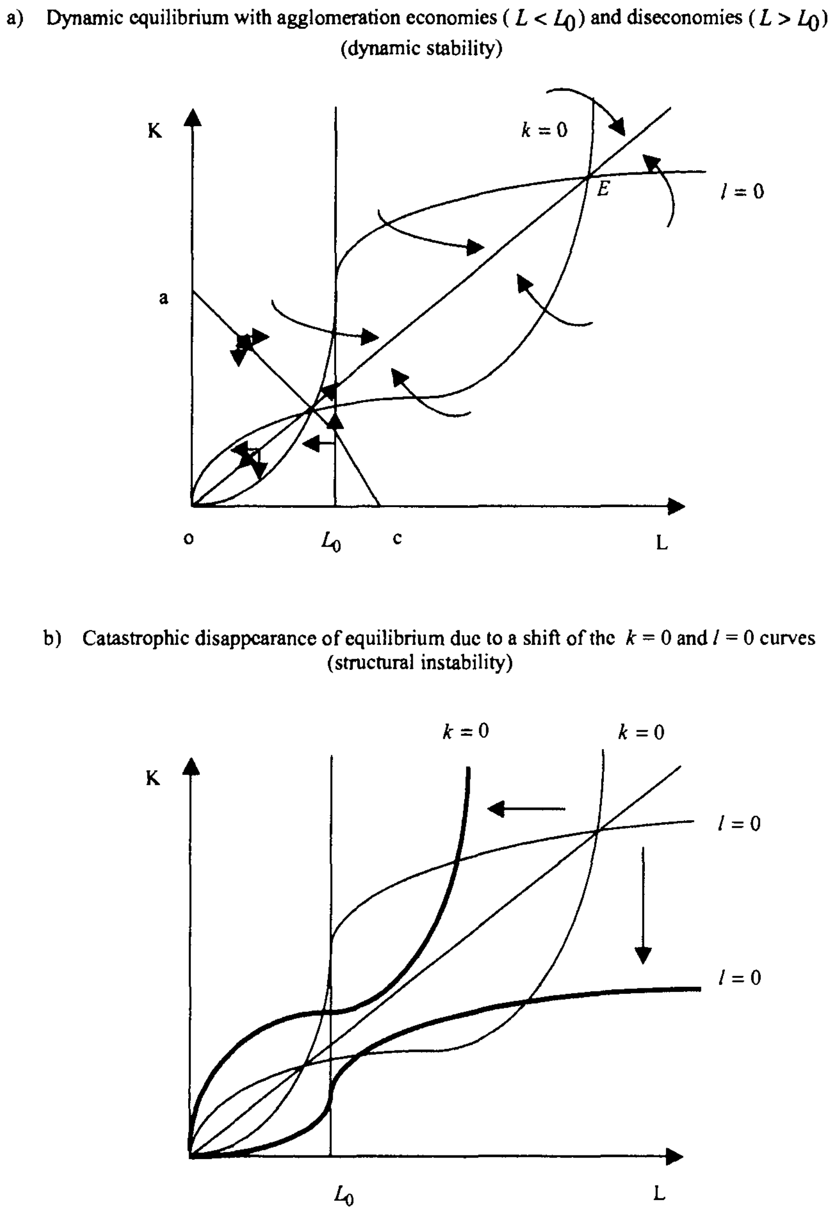


Figure 11.2 A neoclassical growth model with agglomeration economies and diseconomies: dynamic stability and structural instability

Source: Miyao (1984)

where $\beta > 0$ for $L < L_0$ and $\beta < 0$ for $L > L_0$. For values of *K/L* at which the $k = 0$ curve lies above the $l = 0$ curve,[19] the region's economy converges on a steady-state equilibrium with *K* and *L* constant (*E* in Figure 11.2a).

Finally, it is interesting to analyse what happens in the case of 'slow' exogenous shocks on the values of the parameters. In the case of a decrease in the parameters representing the forces (rate of saving or the rate of natural population growth) of the region's economy, or if those representing the economic forces (interest or wage rates) in the 'outside world' increase, the $k = 0$ and $l = 0$ curves shift (to the position shown in bold) until they reach a point at which the steady-state equilibrium suddenly disappears, and the local economy declines in a catastrophic process until disintegration (Figure 11.2b).

It is therefore evident that, by introducing agglomeration economies (diseconomies), the neoclassical model successfully simulates a series of behaviours and tendencies, both continuous and 'catastrophic', very different from the mechanicistic and univocal ones predicted by the simplified model of the 1960s.

11.5 Conclusions

The chapter has surveyed the last group of modern growth models. In these, cumulativeness is stylized in increasing returns to productive resources (individual and in aggregate) and growth consequently comes to depend solely on supply-side elements.

Like those of the previous chapter, these models are mathematically formalized as non-linear dynamic systems that enable increasing returns – in the form of scale economies or learning – to be inserted into equations that explain the growth rate of per capita output. In these models, increasing returns are included in a neoclassical production function, where they offset the effect of the marginal productivity of factors traditionally assumed to be decreasing.

As in the case of previous models, the ones examined in this chapter only achieve their objective by conceiving a diversified-stylized space in which growth is generated by increasing returns; a space, however, bereft of a genuinely territorial dimension. This is the main shortcoming of the most recent models of regional economics, and finding a remedy for it is the challenge that regional economists must address in the next decade.

Review questions

1 What is the aim of endogenous growth models?
2 What is the conception of growth behind the endogenous growth models?
3 What are the ways to make a growth model endogenous?
4 What are the similarities and differences between the theory of the '*milieu innovateur*' and Romer's and Lucas' models?
5 What is demonstrated by the interregional neoclassical growth model with increasing returns?

Selected reading on empirical findings

About human capital and regional development

Bryson P.J. (2005), 'Regional Development and Human Capital Infrastructure in China: Lessons from European Union Regional Policy Experience', Proceedings of International Conference on Regional Disparities, Economic Integration and Development, pp. 1–26, Wuhan University, Wuhan, China, December.

Garlick S. (2005), 'Regional Growth, Enterprising Human Capital and Community Engagement', working paper, Institute for Sustainability, Health and Regional Engagement (iSHaRE), University of the Sunshine Coast, Queensland, Australia, available online, https://publications.qld.gov.au/storage/f/2014-02-06T06%3A58%3A01.599Z/garlick-steve-final.pdf.
Mackay R.R. (1993), 'Local Labour Markets, Regional Development and Human Capital', *Regional Studies*, vol. 27, no. 8, pp. 783–795.
Mankin N., Romer D. and Weil D. (1992), 'A Contribution to the Empirics of Economic Growth', *Quarterly Journal of Economics*, vol. 107, no. 2, pp. 739–774.

Further reading

Aghion P. and Howitt P. (1997), *Endogenous Growth Theory*, MIT Press, Cambridge, MA.
Baldwin R.E., Braconier H. and Forslid R. (1999), 'Multinationals, Endogenous Growth and Technological Spillovers: Theory and Evidence', paper presented at the seminar of the Research Institute of Industrial Economics (IUI) on *Multinational Production, International Mergers and Welfare Effects in a Small Open Economy*, Stockholm, June.
Magrini S. (2004), 'Regional (di)convergence', in V. Henderson and F.J. Thisse (eds), *Handbook of Regional and Urban Economics*, vol. 4, North-Holland, Amsterdam.
Nijkamp P., Stough R. and Verhoff E. (eds) (1998), 'Endogenous Growth in a Regional Context', special issue, *Annals of Regional Science*, vol. 32, no. 1.
Romer P. (1986), 'Increasing Returns and Long-Run Growth', *Journal of Political Economy*, vol. 94, no. 5, pp. 1002–1037.
Sala-i-Martin X. (1996), 'Regional Cohesion: Evidence and Theories of Regional Growth and Convergence', *European Economic Review*, vol. 40, no. 6, pp. 1325–1352.

Notes

1 Numerous endogenous growth models have been propounded; among the best known of them are Romer 1986, 1987, 1990; Lucas, 1988; Barro, 1990; Grossman and Helpman, 1991; Rebelo, 1991; Aghion and Howitt, 1992.
2 For the mathematical steps see Chapter 6, note 5, p. 170.
3 As already mentioned in Chapter 7, note 30 (p. 202), when estimating equation (11.3) for the economy of the United States, Solow found that more than 40 per cent of US growth between 1900 and 1949 was due to a factor (the 'residual' as he termed it) different from factor growth. This he called 'technical progress'.
4 Romer has written that it was the statistically non-significant and non-robust (or at any rate equivocal) results of regression analyses on the initial income and growth levels of countries that stimulated his interest in formulating a new model of growth. See Romer, 1994, p. 4.
5 These models in fact pursue an even more complex goal, which stems from their profoundly neoclassical nature. They embrace the idea of rational and optimizing behaviour by economic actors who choose a temporal consumption path that enables them to optimize an intertemporal utility function while respecting a dynamic constraint of capital stock growth per unit of product constant in time. They are therefore models which explain aggregate macroeconomic growth on the basis of microeconomic behaviour. To this end they draw on Ramsey's model of intertemporal consumption as subsequently applied to growth models by Cass and Koopmans. See Ramsey, 1928; Cass, 1965; Koopmans, 1965.
6 See Aghion and Howitt, 1997 and Solow, 2000, among others, for comprehensive surveys of the theoretical and empirical issues raised by modern growth theory. On investment cumulated in physical capital and the consequent increase over time in 'technological capacities', see Romer, 1986; on the impact at aggregate level of investment by individual firms which generates a positive externality (and increasing returns to scale) at aggregate level, see Romer, 1989; on the constancy of the marginal return on capital if this includes all the accumulable factors, among them 'human capital', see Rebelo, 1991, or private and public goods, see Barro, 1990; on investment in human capital, scientific and technical knowledge

that improves the physical productivity of labour, see Lucas, 1988; finally, on investment in R&D to foster the technological innovation that enhances the physical productivity of all the factors, that is the creation of intermediate and final goods with higher value added, see Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992.

7 For the theory of the ***milieu innovateur***, see Chapter 9.

8 See Romer, 1986. The model's logic closely resembles that of the theory of knowledge spillovers, which here undergoes economic modelling. See Chapter 9.

9 In other models this role is performed by public capital (e.g. infrastructures) or by other public goods: see Barro, 1990; Rebelo, 1991. However, the mechanism of growth endogenization remains the same: the public goods enter the aggregate production function as externalities and convert the decreasing returns to the individual production factors into constant or increasing returns at aggregate level.

10 See Lucas, 1988, pp. 17–27. Lucas drew on Uzawa, 1964.

11 In a second model, Lucas hypothesizes that knowledge is accumulated through experience: 'learning by doing'. See Lucas, 1988, pp. 27–31.

12 Although this latter hypothesis is essential for determining the endogenous growth mechanism, it has major implications; it entails, in fact, that a country's initial conditions do not influence the growth of its economy. This is obviously a highly unrealistic hypothesis.

13 (11.9) can be easily rewritten as:

$$\frac{h_t}{H_t} = \varphi(1 - u_t) \tag{11.1n}$$

where the left-hand member is the rate of growth of human capital over time.

14 For a regional approach to the theory of endogenous growth see Nijkamp and Poot, 1998; Nijkamp et al., 1998 (special issue of *Annals of Regional Science*); Button and Pentecost, 1999.

15 The first time series and cross-section analyses of convergence or divergence among regions used a methodology known as 'σ-convergence' which measured the standard deviation in the distribution of income among regions or countries: if the standard deviation decreases, this indicates convergence (called 'strong' convergence) among the growth rates of regions and countries. A subsequent widely used methodology, which has examined 'β-convergence' or 'weak' or 'absolute' convergence, does not measure the standard deviation in the distribution of income among regions but rather the slope of a linear regression line connecting the rates of income growth in the high-income regions in the initial period: a negative slope indicates a higher growth rate in countries or regions with lower levels of income, and vice versa. In other words, this result confirms the hypotheses of Solow's model, which predicts convergence among the growth rates of countries because the decreasing productivity of the factors entails that, in advanced countries with higher levels of per capita capital, the productivity of capital is less than it is in the backward countries. The latter therefore inevitably record higher levels of accumulation and development. See Solow, 1957. For empirical studies on convergence and divergence see, among many others, Baumol, 1986; Mankin et al., 1992; Chatterji, 1994; Romer, 1994, Barro and Sala-i-Martin, 1995; and among regions, Sala-i-Martin, 1996; Magrini, 1997; Cuadrado-Roura and Parellada, 2002; Pompili, 1992; Terrasi, 2002; Rodriguez-Pose and Fratesi 2004. For a critical survey of methods used to measure regional disparities see Magrini, 2004; for a survey of empirical results see Abreu et al., 2005.

16 See Chapter 6.

17 The steady-state equilibrium curves (i.e. the constant growth trajectories of the production factors), are obtained by imposing nil variation in the growth rates. They are expressed by the following equations, the first obtained by setting $k = 0$ and the second by setting $l = 0$:

$$K = \left(\frac{s + \mu L}{i_w \mu}\right)^{\frac{1}{1-\alpha}} L^{\frac{1-\alpha+\beta}{1-\alpha}} \tag{11.2n}$$

$$K = \left(\frac{w_w\lambda - n}{\lambda(1-\alpha)}\right)^{\frac{1}{\alpha}} L^{\frac{\alpha-\beta}{\alpha}} \tag{11.3n}$$

For the mathematical solution see Chapter 6, note 6, p. 170.

18 See Rabenau, 1979; Miyao, 1984.

19 This happens for high initial values of *K/L*, external to the *aoc* area in Figure 11.2a, and on the condition that:

$$\left(\frac{s + \mu L}{i_w \mu}\right) > \left(\frac{w_w\lambda - n}{\lambda(1-\alpha)}\right) \tag{11.4n}$$

See Rabenau, 1979; Miyao, 1984.

12 Towards a theoretical convergence

The territorial foundations of macroeconomic regional growth models

12.1 The critical elements in local development today

It seems appropriate to conclude this book on the economic theory of space, and in particular on the economics of local growth/development, by emphasizing a number of implications that have emerged from the analysis.

In recent years, regional growth theories have evolved considerably in their interpretation of the concept itself of growth. Demand-oriented theories on short-term processes of employment creation, in conditions of given but largely under-utilized productive resources, have given way to supply-oriented approaches, developed first in regard to the achievement of individual well-being, and then to determination of the real productive capacities of local systems.

Without wishing to deny the interpretative capacity and interest of the former theories, it is today evident that the problem of growth should be associated with the endogenous local elements that generate local competitiveness. The latter is the keystone of development, and upon it depends the survival itself of a local economy in the current circumstances of fierce worldwide competition.

It is also necessary to develop theories able to explain an absolute, not relative, competitiveness; a competitiveness that springs from real productive and innovative capacities, and by virtue of which regions or territories can acquire a specific role in the international division of labour, and maintain it over time.

It is by now well-established that regions do not compete according to the principle of 'comparative advantage' – the principle that assigns to each area a role within the international division of labour regardless of its productive efficiency. The macroeconomic mechanisms (exchange rates, wage and price flexibility) that protect countries against competition do not operate at regional level, and thus theoretically impose a principle of absolute advantage. On this view, local development is substantially a problem of identifying the elements on which to construct and maintain this 'absolute' or 'competitive' advantage.

History teaches first that these elements should be sought within the area itself, and that they arise from the area's specific nature. This supports the idea of 'generative regional growth', or an endogenous development based on the efficient and creative use of local resources. Second, a region's capacity for autonomous development is driven above all by the increasing returns and agglomeration economies generated at local level. The territory should be viewed as an active element in the development process, the generator of advantages for firms and for local actors. It thus becomes the source of agglomeration economies – and consequently of locational advantages – when internal synergies and the local production system reduce static and dynamic uncertainty, production and transaction costs; and today more than ever, when they foster processes of collective learning, technological innovation, and new organizational and managerial methods.

History also teaches that regional and local development is a complex process. It results from the balanced presence of tangible and intangible elements originating in the economic and social spheres; from which derives the importance of theoretical approaches to development that take these elements into account. The endowment of production factors, infrastructures and human capital should be linked with the presence of social capital, a propensity for co-operation among actors, and the ability of people, firms and local institutions to learn. Of equal importance in this 'balanced development' are the endowment of production factors, the ability of suppliers and customers to establish synergies and co-operation, the infrastructural endowment, the tacit diffusion of information, the availability of real or financial capital, processes of collective learning in workforce training, the evolution of managerial styles, and the use of new technologies.

The success of a territorial system therefore does not depend solely on the quantity and quality of the material resources with which it is endowed. It also, and crucially, depends on the richness of economic and social relations, and on local 'social' or 'relational' capital. Because capital is hyper-mobile, the competitiveness of regions depends on intangible resources and their development. Intangible elements connected with culture and innovative capacity accumulate through slow processes of individual and collective learning fuelled by information, interaction and investments in research and training. They are therefore intrinsically localized and cumulative, embodied in human capital and local relational networks, in the labour market and in the local context – and they are consequently highly selective in spatial terms.

All these factors have major repercussions on regional disparities. It is likely that the territorial embeddedness of intangible resources boosts the centripetal and cumulative forces of development (economies of scale and scope, increasing returns of various kinds) and the centrifugal forces of territorial exclusion and decline. The divergence between strong and weak areas, between areas with greater or lesser endowments of intangible resources such as knowledge and the ability to learn, is therefore more likely than their convergence – at least in the short-to-medium term. This tendency will strengthen because the traditional elements giving absolute advantage to weak areas, such as low labour costs, tend to disappear in the medium-to-long period owing to social factors (migration), institutional factors (the imposition of national-level collective bargaining), and cultural ones (social expectations within a monetary union).

Necessary as a consequence is a modern conception of local development that concerns itself with the creative and innovative use of local tangible and intangible resources, and with the creation of the knowledge and models of co-operation and decision-making on which the innovativeness of firms is based.

However, a modern conception of local development cannot deny that regions are part of larger economic systems, with which they share growth and decline. This remark is even more apposite in a period of crisis like the present one, during which the macroeconomic conditions of countries and the limitations imposed by participation in supranational monetary unions have had a huge influence on the growth trajectories of single countries, and of their regions, as we shall see below.

12.2 The role of macroeconomic elements in regional growth

Modern theories of local development have stressed with evidence that regional development trajectories today depend on endogenous elements, and it is on these elements that regional competitiveness relies. However, the experience of the crisis period demonstrates that one cannot ignore macroeconomic, aggregate demand elements, since these decisively influence the destinies of the single national economies, and their regions. Conditions of public debt and deficit, deflation, country reliability, trends in public and private consumption and investments determine the growth trajectories of a country and of all regions belonging to that country.

Whilst it is true that the destiny of a region cannot be very different from that of its nation, it is also true that the impacts of macroeconomic trends differ among regions, as previously mentioned several times in this book. Already highlighted in Chapter 8 was the important role played by the Italian lira devaluation of the 1970s and of 1992 in the success of the local district areas; in fact, their export-oriented production gained in terms of price competitiveness.

The regional asymmetries of the effects of macroeconomic policies do not limit themselves to devaluation policies. Other examples can be provided in this regard. The widening of the spread – the risk premium required with respect to riskless bonds – characterized the past years of the crisis because international markets associated a higher probability of default with certain government debts. The increase in the spread in those countries with specific economic problems (Italy, Spain, Greece, Portugal, Ireland) generated three main macroeconomic effects, and interestingly, each of them involved different actors and different regions:

- strong control on public expenditure and its forced reduction was imposed by the EU, especially in the above-mentioned countries. The effects of this reduction were expected to be stronger in those regions with a higher share of public demand than those with a higher share of private demand, the former being generally the poorer and less productive regions;
- private investments decreased as a consequence of the increase in interest rates on private loans and bonds, which penalized private actors, and particularly industrial regions with large shares of SMEs;
- a credit crunch came as a consequence of the decision of financial intermediaries to prefer investing in public bonds rather than the private sector, when sufficient guarantees existed on possible sovereign default; the manufacturing sector and the most productive regions hosting it were once again penalized more than others.

Supply-side elements, related to the structural characteristics of local areas and to a differentiated availability of territorial capital, are an immediate and logical explanation for the differentiated spatial impacts of the crisis. The same sources of crisis generated different effects at regional level. In the first phase (2007–2009), when the crisis was triggered by the crash of the real-estate mortgage market, regions specializing in financial activities were those that were hit hardest. This was the case of London, or of New York, areas in which the highest unemployment growth rates were recorded. In the second phase (2009–2011), the crisis moved to the production industry due to the decrease in world consumption; the regions hardest hit in this phase were the export-oriented industrial ones. Finally, in the third period (2011–2013), the crisis again hit the public sector and its financiers, as a consequence of large speculations on sovereign debts. The ensuing private credit crunch negatively impacted on the construction sector, commerce and production industries, and consequently on consumption and investments. Industrial regions were once again those most exposed to these phenomena, registering high unemployment rates.

The capacity of a region to react to external negative shocks resides in its endowment of private and public territorial capital, mentioned in different theories of this textbook and synthesized in the matrix of territorial capital. As in the previous examples, productive specialization, the settlement structure, and openness to foreign markets can all provide a partial explanation of the capacity to react to external shocks. To these elements should be added the capacity to draw on untapped resources through processes of industrial reconversion, functional modernization, innovation in local governance and the exploitation of all territorial externalities that stem from equilibrated urban systems. All these processes are highly dependent on the quality of territorial capital assets and their efficient use.

Despite the obvious importance of an integrated role of macroeconomic and territorial elements in regional growth, it is evident from the overview of theories presented thus far in this textbook that there is no model that explicitly embraces both approaches. The macroeconomic regional growth models of Chapters 5 and 6, but also the most recent ones set out in Chapters 10 and 11, present aggregate macroeconomic growth patterns without leaving space for real territorial elements built within an aspatial logic. The bottom-up models, like those presented in Chapters 7, 8 and 9, instead lack an aggregate macroeconomic perspective, in that they consider only genuine local aspects. It is therefore necessary to create a macroeconomic 'model' able to encompass real territorial, behavioural and intangible aspects.

12.3 The territorial foundations of a macroeconomic growth model: the MASST model

A first step towards a growth model able to integrate macroeconomic elements with endogenous, territorial ones is represented by the MASST model, whose acronym contains the different dimensions – macroeconomic, sectoral, social and territorial – that must be considered in order to interpret regional growth patterns.[1]

The model's important innovative step forward is that it does not present a new theory; rather, it integrates different theories within a logical and organic framework where all crucial macroeconomic aspects and endogenous territorial assets find a role. The internal logic of the model, in fact, is an elegant merger of two different approaches: macroeconomic Keynesian growth theory as regards national growth, and the theory of endogenous development as regards the regional growth differential. The model consists of a series of equations, whose logic is reported in the flow chart synthesizing the cause/effect chain and all feedbacks conceived in the model (Figure 12.1).

The first theoretical feature of the model is the dependence of the national growth rate on demand elements, each of which finds its determinants in the Keynesian theory. Thanks to these components, the model is able to consider all the macroeconomic trends and policies mentioned above (devaluation, increase in the spread, decrease of public expenditure, etc.). Added to this theoretical element is a second one, which

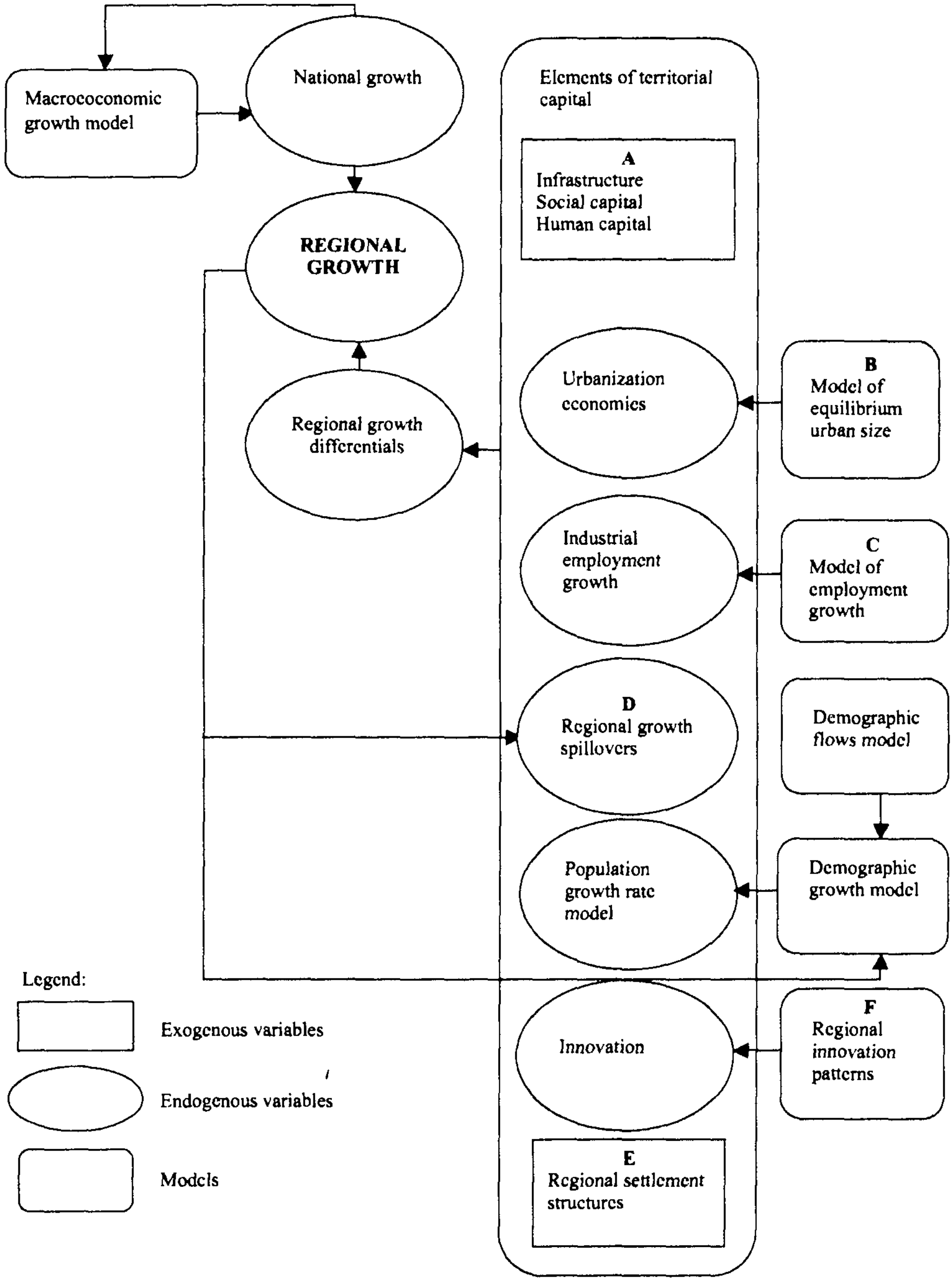


Figure 12.1 The intrinsic logic of the MASST model: interaction between macroeconomic and regional elements

interprets the regional growth differential through supply elements that generate differentiated effects at the regional level; through these elements, the above-mentioned policies and macroeconomic processes impact in different ways on the individual regions. The sum of the national growth rate and of the differential growth rate gives rise to the regional growth rate.

The territorial structures present in MASST represent both the propulsive forces of regional growth and the factors that explain local responses to exogenous aggregate trends. This is done in different ways.

First, the model includes traditional and innovative elements of territorial capital in the explanation of local dynamics. Moreover, the model measures the interactions and synergies among single elements, that are important for regional growth as we learnt throughout the whole textbook, by merging traditional elements like physical accessibility, new intangible elements like human capital à la Lucas and Romer, and relational capital as explained in the '*milieu innovateur*' theory (elements A in Figure 12.1).

Moreover, the model attributes a distinctive role to advantages stemming from an urban environment; advantages which, in their turn, depend on the specificities of single cities, and of the urban system as a whole, as explained in Chapter 8. This source of externality is inserted into the model through a sub-model that defines an equilibrium size for each city obtained when marginal costs equal marginal benefits. Both benefits and costs depend, in their turn, on the specificities of single cities: amenities, industrial diversity and high-level functions explain the benefits, while urban land, social conflicts and sprawl explain the costs (element B of Figure 12.1).

Moreover, the model includes the advantages stemming from an industrial specialization – the source of localization economies, or district economies, as explained in Chapter 8. Industrial specialization, in its turn, defines the industrial employment growth patterns, as rightly explained by the export-base theory, and the regional differential growth (element C in Figure 12.1).

The model also considers the advantages that stem from integration among regions; that is, from the possibility of taking advantage of proximity to tangible and intangible factors present in a neighbouring region in the form of technological externalities or growth spillovers, as explained in the theories of Chapter 9 (element D in Figure 12.1).

Last, but not least, the model takes into consideration differentiated regional effects of exogenous aggregate trends, at both European and national level. This differentiation is captured through specificities in the settlement structure of a region (urban, rural or agglomerated region), which become explanatory variables in industrial dynamics, regional differential growth, migration flows and indirectly population growth (element E in Figure 12.1). As regards the innovation model, the effects on regional growth are captured through the different innovation patterns of each region, as suggested by the regional innovation pattern theory presented in Chapter 9 (element F in Figure 12.1).

Through this logical structure and its theoretical bases, the model is able to overcome the distinction between a regional growth dependent on the pure top-down distribution of an aggregate national growth rate, and a regional growth obtained bottom-up through the real local competitive capacities. Given the manner in which it has been conceived, the MASST model is distributive and generative at the same time thanks to horizontal feedbacks (among regions, in the form of growth spillovers) and vertical ones (between nations and their regions, and vice versa). A macroeconomic effect propagates in a different manner among the regions according to the different structural characteristics of the local economy, and it retro-acts both on the other regions and on the nation in an aggregate manner. Likewise, changes at the local level (generated for example by regional policies) influence the regional growth trend, and through it the national growth and the growth of all other regions belonging to the same nation, in a cumulative circular process.

The model can be used to generate regional growth scenarios that have to be interpreted as a coherent set of assumptions on future values of exogenous variables. In the simulation phase, the variables enter the model both through the exogenous values of the economic and territorial variables, and through a differentiation of the effects of the external shocks linked to some specific characteristics of regional economic and settlement structures.

A last innovative aspect of the MASST model is the simultaneous presence of co-operation and competition among regions. Competition is guaranteed by the generative part of the model, which defines regional growth on the basis of relative competition. Co-operation among regions is instead guaranteed by growth spillovers, which transfer the success of a region to other regions through the well-known accessibility and proximity effects.

12.4 Conclusions

The theories described in this handbook have highlighted the increasingly complex and intriguing ways in which models of economic growth treat space. The simple (and in certain respects trivial) interpretation of space as uniform-abstract and straightforwardly relatable to administrative units – a space conceived as internally homogeneous and uniform, and which can therefore be synthesized into a vector of aggregate socio-economic-demographic features – has in recent years been replaced by a notion of diversified-relational space that restores to theories of regional development some of the founding principles of location theory: agglomeration economies and spatial interaction.

It is this more complex interpretation of space that has enabled regional economics to take decisive steps forward in analysis of local dynamics by conceiving space as the source of increasing returns and positive externalities. The development process also depends on the efficiency of the territorial organization of production, rather than solely on the quantity of economic resources available. Not only are the tangible elements of development (for example, the quantity of existing productive resources) important, so too are the intangible ones: the learning processes, local relational networks and governance mechanisms that have increasing weight in defining an area's development path.

Finally, most recent years have seen an endeavour to escape from the impasse that caught regional economics between, on the one hand, growth theories of pure macroeconomic origin formalized into elegant models, and on the other, theories that abandon the rigour of formal treatment to consider new qualitative and territorial elements synthesizable – with due caution – into the concept of agglomeration economies. The most recent theories on local growth are able to incorporate increasing returns into the economic and formal logic of macroeconomics, and they are viewed (sometimes all too enthusiastically) as a new way to conceive space – as a means to merge previous conceptions together. Space is conceived as diversified, while territorial development is conceived as selective, cumulative and at increasing returns, and it is interpreted on the basis of a macroeconomic growth model.

It has been emphasized that this merger is in fact only an initially positive result. More detailed analysis shows that space is indeed conceived as diversified, but it receives no territorial explanation apart from one taking the form of the agglomeration/non-agglomeration dichotomy. The territorial features (and the above-mentioned intangible elements) that play an important role in diversified-relational space theories by explaining and interpreting the level of competitiveness achieved entirely disappear in the macroeconomic models. The MASST model is an attempt to merge territorial elements with macroeconomic ones for the explanation of regional growth, and it is in this sense an innovative step forward.

Further reflection is necessary to refine some territorial elements that in the present version of MASST are considered exogenous. Finding a way to interpret them endogenously would enrich the model with important theoretical elements and give it even greater interpretative capacity.

Despite the efforts made with the MASST model, however, there is still space for future conceptualizations and traditional reflections. Whilst the MASST model has made it possible to insert territorial foundations into an aggregate macroeconomic regional growth model, further analysis is required of the territorial micro-foundations of regional growth models through the study of the behaviours of single firms and individuals in space.

The remedying of these shortcomings, and especially the strengthening of the territorial micro-foundations of macroeconomic regional growth models, are the challenges that face regional economists in the years to come.

Review questions

1 How would you define local development today?
2 What are the main elements of local development today?
3 Why should one take into consideration macroeconomic aspects in a regional growth model?
4 What are the features of the MASST model? What is new in its structure?
5 What are the theoretical aspects that require further attention today?

Further reading

Capello R. (2007), 'A Forecasting Territorial Model of Regional Growth: The MASST Model', *The Annals of Regional Science*, vol. 41, no. 4, pp. 753–787.

Capello R. and Fratesi U. (2012), 'Modelling Regional Growth: An Advanced MASST Model', *Spatial Economic Analysis*, vol. 7, no. 3, pp. 293–318.

Capello R., Caragliu A. and Fratesi U. (2014), 'Modelling Regional Growth between Competitiveness and Austerity Measures: The MASST3 Model', *International Regional Science Review*, DOI: 10.1177/0160017614543850.

Note

1 For in-depth explanations of the MASST model, see Capello, 2007; Capello and Fratesi, 2012; Capello et al., 2014.

References

Abernathy W. and Utterback J. (1978), 'Patterns of Industrial Innovation', *Technology Review*, Vol. 80, no. 7, pp. 121–133.

Abreu M., de Groot H.L.F. and Florax R. (2005), 'A Meta-analysis of β-Convergence: The Legendary 2%', *Journal of Economic Surveys*, vol. 19, no. 3, pp. 389–420.

Ackley G. and Spaventa L. (1962), 'Emigrazione e Industrializzazione nel Mezzogiorno', *Moneta e Credito*, vol. 15, no. 58, pp. 144–159.

Acs Z.J., Audretsch D.B., Braunerhjelm P. and Carlsson B. (2004), 'The Missing Link: The Knowledge Filter and Entrepreneurship in Endogenous Growth', CEPR Discussion Paper 4783.

Acs Z., Audretsch D. and Feldman M. (1994), 'R&D Spillovers and Recipient Firm Size', *Review of Econmics and Statistics*, vol. 76, no. 2, pp. 336–340.

Acs Z., Brooksbank D., O'Gorman J., Colm A., Terjesen S. and Pickernell D. (2012), 'The Knowledge Spillover Theory of Entrepreneurship: An Application to Foreign Direct Investment', *International Journal of Entrepreneurship and Small Business*, vol. 15, no. 2, pp. 237–261.

Acs Z., de la Mothe J. and Paquet G. (2000), 'Regional Innovation: In Search of an Enabling Strategy', in Z. Acs (ed.), *Regional Innovation, Knowledge and Global Change*, Pinter, London, pp. 37–52.

Aghion P. and Howitt P. (1992), 'A Model of Growth through Creative Destruction', *Econometrica*, vol. 60, no. 2, pp. 323–351.

Aghion P. and Howitt P. (1997), *Endogenous Growth Theory*, MIT Press, Cambridge, MA.

Allen K.J. (1969) 'The Regional Multiplier: Some Problems in Estimation', in J.B. Cullingworth and S.C. Orr (eds), *Regional and Urban Studies: a Social Science Approach*, Allen & Unwin, London, pp. 80–96.

Alonso, W. (1960), 'A Theory of the Urban Land Market', *Papers and Proceedings of the Regional Science Association*, vol. 6, no. 1, pp. 149–157.

Alonso W. (1964a), 'Location Theory', in J. Friedmann and W. Alonso (eds), *Regional Development and Planning: A Reader*, MIT Press, Cambridge, MA, pp. 78–106.

Alonso W. (1964b), *Location and Land Use: Towards a General Theory of Land Rent*, Harvard University Press, Cambridge, MA.

Alonso W. (1971), 'The Economics of Urban Size', *Papers and Proceedings of the Regional Science Association*, vol. 26, pp. 67–83.

Alonso W. (1973), 'Urban Zero Population Growth', *Daedalus*, vol. 102, no. 4, pp. 191–206.

Amin A. and Thrift N. (eds) (1994), *Globalisation, Institutions, and Regional Development in Europe*, Oxford University Press, Oxford.

Amin A. and Wilkinson F. (eds) (1999), 'Learning, Proximity and Industrial Performance', *Cambridge Journal of Economics*, vol. 23, no. 2, special issue.

Anas A. and Dendrinos D. (1976), 'The New Urban Economics: a Brief Survey', in Y.Y. Papageorgiou (ed.), *Mathematical Land Use Theory*, Lexington Books, Lexington, MA, pp. 23–51.

Andrews R. (1953), 'Mechanics of the Urban Economic Base', *Land Economics*, vol. 29, no. 2, pp. 161–167.
Andrews R. (1954), 'Measuring the Urban Economic Base', *Land Economics*, vol. 30, no. 3, pp. 52–60.
Anselin L., Varga A. and Acs Z. (2000), 'Geographic and Sectoral Characteristics of Academic Knowledge Externalities', *Papers in Regional Science*, vol. 79, no. 4, pp. 435–443.
Antonelli C. (1989), 'The Diffusion of Information Technology and the Demand for Telecommunication Services', *Telecommunications Policy*, vol. 13, no. 3, pp. 255–264.
Archibald G. (1967), 'Regional Multiplier Effects in the United Kingdom', *Oxford Economic Papers*, vol. 19, no. 1, pp. 22–45.
Armstrong H. and Taylor J. (2000), *Regional Economics and Policy*, Blackwell, Oxford.
Arrighetti, A., Serravalli, G. and Wolleb, G. (2001), 'Social Capital, Institutions and Collective Action between Firms', paper presented at the EURESCO Conference, Social Capital: Interdisciplinary Perspectives, Exeter, 15–20 September.
Arthur W.B. (1989), 'Competing Technologies, Increasing Returns and Lock–in by Historical Events', *The Economic Journal*, vol. 99, no. 394, pp. 116–133.
Arthur W.B. (1990), 'Silicon Valley's Location Clusters: When Do Increasing Returns Imply Monopoly?', *Mathematical Social Sciences*, vol. 19, no. 3, pp. 235–251.
Aschauer D.A. (1989), 'Is Public Expenditure Productive?', *Journal of Monetary Economics*, vol. 23, no. 2, pp. 177–200.
Aschauer D. A. (1990), 'Why is Infrastructure Important?', in A.H. Munnell (ed.), *Is there a Shortfall in Public Capital Investment?* Proceedings of a Conference Sponsored by the Federal Bank of Boston, Federal Reserve Bank of Boston, Boston.
Asheim B. (1996), 'Industrial Districts as "Learning Regions": A Condition for Prosperity?', *European Planning Studies*, vol. 4, no. 4, pp. 379–400.
Asheim B. and Gertler M. (2005), 'The Geography of Innovation', in J. Fagerberg, D. Mowery and R. Nelson (eds), *The Oxford Handbook of Innovation*, Oxford University Press, Oxford, pp. 291–317.
Audretsch D. and Feldman M. (1996), 'R&D Spillovers and the Geography of Innovation and Production', *American Economic Review*, vol. 86, no. 3, pp. 630–640.
Aydalot Ph. (1985), *Economie Régionale et Urbaine*, Economica, Paris.
Aydalot Ph. (ed.) (1986), *Milieux Innovateurs en Europe*, GREMI, Paris.
Aydalot Ph. and Keeble D. (eds) (1988), *High Technology Industry and Innovative Environment*, Routledge, London.
Bachtler J. and Downes R. (1999), 'Regional Policy in the Transition Countries: A Comparative Assessment', *European Planning Studies*, vol. 7, no. 6, pp. 793–808.
Bagnasco A. (1977), *Tre Italie: La Problematica Territoriale dello Sviluppo Italiano*, Il Mulino, Bologna.
Bagnasco A. (1983), 'Il Contesto Sociale', in G. Fuà and C. Zacchia (eds), *Industrializzazione senza Fratture*, Il Mulino, Bologna, pp. 149–165
Bagnasco A. (1985), 'La Costruzione Sociale del Mercato: Strategie d'Impresa e Esperimenti di Scala in Italia', *Stato e Mercato*, no. 13, pp. 9–45.
Bagnasco A. and Trigilia C. (1984), *Società e Politica nelle Aree di Piccola Impresa: Il Caso di Bassano*, Arsenale, Venice.
Balassa B. (1975), 'Trade Creation and Diversion in the European Common Market: An Appraisal of the Evidence', in B. Balassa (ed.), *European Economic Integration*, North-Holland, Amsterdam, pp. 79–118.
Baldwin R.E. (1998), 'Agglomeration and Endogenous Capital', Discussion Paper, CEPR, London.
Baldwin R.E., Braconier H. and Forslid R. (1999), 'Multinationals, Endogenous Growth and Technological Spillovers: Theory and Evidence', paper presented at the seminar of the Research Institute of Industrial Economics (IUI), Multinational Production, International Mergers and Welfare Effects in a Small Open Economy, Stockolm, June.

Banister D. (1998), 'Barriers to Implementation of Urban Sustanability', *International Journal of Environment and Pollution*, vol. 10, no. 1, pp. 65–83.

Barba Navaretti G. and Venables A. (2004), *Multinational Firms in the World Economy*, Princeton University Press, Princeton, NJ.

Barca F. (2009), 'An Agenda for the Reformed Cohesion Policy', Report to the Commissioner for Regional Policy, Brussels, April.

Barca F. (2011), 'Alternative Approaches to Development Policy: Intersections and Divergences', in OECD, *Regional Outlook 2011*, Paris, pp. 215–225.

Barentsen W. and Nijkamp P. (1989), 'Modelling Non–Linear Processes in Time and Space', in Å. Andersson, D. Batten, B. Johansson and P. Nijkamp (eds), *Advances in Spatial Theory and Dynamics*, North-Holland, Amsterdam, pp. 175–192.

Barker T. (1977), 'International Trade and Economic Growth: An Alternative to the Neoclassical Approach', *Cambridge Journal of Economics*, vol. 1, no. 2, pp. 153–172.

Barro R.J. (1990), 'Government Spending in a Simple Model of Endogenous Growth', *Journal of Political Economy*, vol. 98, no. 5, pp. S103–S125.

Barro R.J. and Sala-i-Martin X. (1995), *Economic Growth*, McGraw-Hill, New York.

Bartik T. (1985), 'Business Location Decisions in the United States: Estimates of the Effects of Unionization, Taxes, and Other Characteristics of States', *Journal of Business and Economic Statistics*, vol. 3, no. 2, pp. 14–22.

Basile R., Castellani D. and Zanfei A. (2008), 'Location Choices of Multinational Firms in Europe: The Role of National Boundaries and EU Policy', *Journal of International Economics*, vol. 74, no. 2, pp. 328–340.

Basile R., Castellani D. and Zanfei A. (2009), 'National boundaries and the Location of Multinational Firms in Europe', *Papers in Regional Science*, vol. 88, no. 4, pp. 733–748.

Batten D. (1995), 'Network Cities: Creative Urban Agglomerations for the 21st Century', *Urban Studies*, vol. 32, no. 2, pp. 313–237.

Baumol W.J. (1986), 'Productivity Growth, Convergence and Welfare: What the Long Run Data Show', *American Economic Review*, vol. 76, no. 5, pp. 1072–1085.

Beaudry C. and Schiffauerova A. (2009), 'Who's Right, Marshall or Jacobs? The Localization Versus Urbanization Debate', *Research Policy*, vol. 38, no. 2, pp. 318–337.

Becattini G. (1979), 'Dal Settore Industriale al Distretto Industriale. Alcune Considerazioni sull'Unità di Indagine dell'Economia Industriale', *Rivista di Economia e Politica Industriale*, no. 1, pp. 35–48; English edn (1989), 'Sectors and/or Districts: Some Remarks on the Conceptual Foundations of Industrial Economics', in E. Goodman and J. Bamford (eds), *Small Firms and Industrial Districts in Italy*, Routledge, London, pp. 123–135.

Becattini G. (1990), 'The Marshallian Industrial District as a Socio-economic Notion', in F. Pyke, G. Becattini and W. Sengenberger (eds), *Industrial Districts and Interfirm Cooperation in Italy*, International Institute of Labour Studies, Geneva, pp. 37–51.

Becattini G. (ed.) (1975), *Lo Sviluppo Economico della Toscana con Particolare Riguardo all'Industrializzazione Leggera*, Guaraldi, Florence.

Becattini G. (ed.) (1987), *Mercato e Forze Locali: il Distretto Industriale*, Il Mulino, Bologna.

Becattini G. (ed.) (2004), *Industrial Districts: A New Approach to Industrial Change*, Edward Elgar, Cheltenham.

Beckmann M.J. (1958), 'City Hierarchies and the Distribution of City Sizes', *Economic Development and Cultural Change*, vol. 6, no. 1, pp. 243–248.

Beckmann M.J. (1968), *Location Theory*, Random House, New York.

Beckmann M.J. (1969), 'On the Distribution of Urban Rent and Residential Density', *Journal of Economic Theory*, vol. 18, no. 1, pp. 60–68.

Beckmann M.J. and Mc Pherson J. (1970), 'City Size Distribution in a Central Place Hierarchy: An Alternative Approach', *Journal of Regional Science*, vol. 10, no. 1, pp. 25–33.

Beguin H. (1984), 'The Shape of City-Size Distribution in a Central Place System', *Environment and Planning A*, vol. 16, no. 6, pp. 749–758.

Beguin H. (1988), 'La Région et les Lieux Centraux', in C. Ponsard (ed.), *Analyse Economique Spatiale*, Presses Universitaires de France, Paris, pp. 231–275.

Bellandi M. (1989), 'The Industrial District in Marshall', in E. Goodmand and J. Bamford (eds), *Small Firms and Industrial Districts in Italy*, Routledge, London, pp. 136–152; orig. edn (1982), 'Il Distretto Industriale in Alfred Marshall', *L'Industria*, no. 3, July–September, pp. 355–375.

Bellet M., Colletis G. and Lung Y. (1993), 'Introduction au Numéro Special sur l'Economie de Proximité', *Révue d'Economie Régionale et Urbaine*, no. 3, pp. 357–364.

Benedetti E. and Camagni R. (1983), 'Riflessioni sulla Periferia', *Economia e Politica Industriale*, no. 39, pp. 19–37.

Berry B. and Garrison W. (1958), 'Recent Development of Central Place Theory', *Papers and Proceeding of the Regional Science Association*, vol. 13, no. 4, pp. 107–120.

Bianchi G. (1994), 'Requiem per la Terza Italia? Sistemi territoriali di piccola impresa e transizione post-industriale', in G. Garofoli and R. Mazzoni (eds), *Sistemi produttivi locali: struttura e trasformazione*, FrancoAngeli, Milan, pp. 59–90.

Bianchi P. and Miller L. (1993), 'Collective Action, Strategic Behaviour and Endogenous Growth', Working Papers 160, Economic Department, University of Bologna.

Biehl D. (1986), *The Contribution of Infrastructure to Regional Development*, Regional Policy Division, European Community, Brussels.

Blanchard O.J. and Katz L.F. (1992), 'Regional Evolutions', *Brookings Papers on Economic Activity*, vol. 23, no. 1, pp. 1–75.

Blomstrøm M. and Kokko A. (1998), 'Multinational Corporations and Spillovers', *Journal of Economic Surveys*, vol. 12, no. 3, pp. 247–277.

Blonigen B. (2005), 'A Review of the Empirical Literature on FDI Determinant', *Atlantic Economic Journal*, vol. 33, no. 4, pp. 383–403.

Boekema F., Morgan K., Bakkers S. and Rutten R. (1997), 'Introduction to Learning Regions: A New Issue for Analysis?', in F. Boekema, K. Morgan, S. Bakkers and R. Rutten (eds), *Knowledge, Innovation and Economic Growth*, Edward Elgar, Cheltenham, pp. 3–16.

Boix R. (2004), 'Redes de Ciudades y Externalidades', *Investigaciones Regionales*, no. 4, pp. 5–27.

Borensztein, E., De Gregorio, J. and Lee, J.W. (1998), 'How Does Foreign Direct Investment Affect Economic Growth?', *Journal of International Economics*, vol. 45, no. 1, pp. 115–135.

Borts G.H. (1960), 'The Equalisation of Returns and Regional Economic Growth', *The American Economic Review*, vol. 50, no. 1, pp. 319–347; reprinted in McKee D., Dean R. and Leahy W. (eds) (1970), *Regional Economics: Theory and Practice*, The Free Press, New York, pp. 147–176.

Borts G.H. and Stein J.L. (1964), *Economic Growth in a Free Market*, New York, Columbia University Press.

Borts G.H. and Stein J.L. (1968), 'Regional Growth and Maturity in the Unites States: A Study of Regional Structural Change', in L. Needleman (ed.), *Regional Analysis*, Penguin Books, Harmondsworth, pp. 159–197; orig. edn (1962), in *Schweizerische Zeitschrift für Volkswirtscahft und Statistik*, vol. 98, pp. 290–321.

Boschma R. (2005), 'Proximity and Innovation: A Critical Assessment', *Regional Studies*, vol. 39, no. 1, pp. 61–74.

Boschma R. (2011), 'Regional Branching and Regional Innovation Policy', in K. Kourtit, P. Nijkamp and R.R. Stough (eds), *Drivers of Innovation, Entrepreneurship and Regional Dynamics*, Springer Verlag, Berlin/Heidelberg, pp. 359–368.

Boschma R. (2014), 'Constructing Regional Advantage and Smart Specialisation', *Scienze Regionali – Italian Journal of Regional Science*, vol. 13, no. 1, pp. 51–78.

Boschma R. and Frenken K. (2006), 'Why Is Economic Geography Not an Evolutionary Science? Towards an Evolutionary Economic Geography', *Journal of Economic Geography*, vol. 6, no. 3, pp. 273–302.

Boschma R. and Iammarino S. (2009), 'Related Variety, Trade Linkages, and Regional Growth in Italy', *Economic Geography*, vol. 85, no. 3, pp. 289–311.
Boschma R., Minondo A. and Navarro M. (2012), 'Related Variety and Regional Growth in Spain', *Papers in Regional Science*, vol. 91, no. 2, pp. 241–256.
Boudeville J.-R. (1964), *Les Espaces Economiques*, Presses Unversitaires de France, Paris, English edition (1966) *Problems of Regional Economic Planning*, Edinburgh University Press, Edinburgh.
Breschi S. (2000), 'The Geography of Innovation: A Cross–sector Analysis', *Regional Studies*, vol. 34, no. 2, pp. 213–229.
Brown L. (1981), *Innovation Diffusion: A New Perspective*, Methuen, London.
Bruinsma F., Nijkamp P. and Rietveld P. (1990) 'Regional Economic Transformation and Social Overhead Investments', *Serie Research Memoranda*, no. 55, Free University of Amsterdam.
Brusco S. (1982), 'The Emilian Model: Productive Decentralisation and Social Integration', *Cambridge Journal of Economics*, vol. 6, no. 2, pp. 167–184.
Brusco S. (1990), 'The Idea of the Industrial District: Its Genesis', in F. Pyke, G. Becattini and W. Sengenberger (eds), *Industrial Districts and Interfirm Cooperation in Italy*, International Institute of Labour Studies, Geneva, pp. 10–19.
Burger, M.J., Meijers, E.J., Hoogerbrugge, M.M. and Masip Tresserra, J. (2014), 'Borrowed Size, Agglomeration Shadows and Cultural Amenities in North-West Europe', *European Planning Studies*, online first. DOI: 10.1080/09654313.2014.905002.
Butler J., Gaspar J.M.B. and Jeppesen E. (1986) 'Telecommunications and Regional Development in Portugal', *Arbejdsrapport*, vol. 16, no. 3, Aarhus University.
Button K. and Pentecost E. (1999), *Regional Economic Performance within the European Union*, Edward Elgar, Cheltenham.
Camagni R. (1982), 'L'Impatto sull'Economia Sarda della Spesa e dell'Investimento Turistico in Costa Smeralda', *Quaderni Sardi di Economia*, no. 4, pp. 371–413.
Camagni R. (1985), 'Spatial Diffusion of Pervasive Process Innovation', *Papers of the Regional Science Association*, vol. 58, no. 8, pp. 83–95.
Camagni R. (1988), 'Functional Integration and Locational Shifts in the New Technology Industry', in Ph. Aydalot and D. Keeble (eds), *High Technology Industry and Innovative Environment*, Routledge, London, pp. 48–64.
Camagni R. (1991), 'Local *Milieu*, Uncertainty and Innovation Networks: Towards a New Dynamic Theory of Economic Space', in R. Camagni (ed.), *Innovation Networks: Spatial Perspectives*, Belhaven-Pinter, London, pp. 121–144.
Camagni R. (1992a), *Economia Urbana: Principi e Modelli Teorici*, La Nuova Italia, Rome; French ver. (1996) *Economie Urbaine*, Economica, Paris.
Camagni R. (1992b), 'Development Scenarios and Policy Guidelines for the Lagging Regions in the 1990s', *Regional Studies*, vol. 26, no. 4, pp. 361–74.
Camagni R. (1994), 'From City Hierarchy to City Networks: Reflections about an Emerging Paradigm', in T.R. Lakshmanan and P. Nijkamp (eds), *Structure and Change in the Space Economy: Festschrift in Honor of Martin Beckmann*, Springer Verlag, Berlin, pp. 66–87.
Camagni R. (1998), 'Sustainable Urban Development: Definition and Reasons for a Research Programmme', *International Journal of Environment and Pollution*, vol. 10, no.1, pp. 6–26.
Camagni R. (1999a), *La Teoria dello Sviluppo Regionale*, Cusl Nuova Vita, Padua.
Camagni R. (1999b), The City as a Milieu: Applying the GREMI Approach to Urban Evolution', *Révue d'Economie Régionale et Urbaine*, no. 3, pp. 591–606.
Camagni R. (2001), 'Policies for Spatial Development', in OECD (ed.), *Territorial Outlook*, Paris, pp. 147–169.
Camagni R. (2002), 'On the Concept of Territorial Competitiveness: Sound or Misleading?', *Urban Studies*, vol. 39, no. 13, pp. 2395–2411.
Camagni R. (2009), 'Territorial Capital and Regional Development', in R. Capello and P. Nijkamp (eds), *Handbook of Regional Growth and Development Theories*, Edward Elgar, Cheltenham, pp. 118–132.

Camagni R. and Capello R. (1990), 'Towards a Definition of the Manoeuvring Space of Local Development Initiatives: Italian Success Stories of Local Development – Theoretical Conditions and Practical Experiences', in W. Stöhr (ed.), *Global Challenge and Local Responses*, Mansell, London, pp. 328–353.
Camagni R. and Capello R. (1998), 'Indivisibilità e Irreversibilità nei Processi di Adozione delle Tecnologie Rinnovabili in Ambiente Urbano', in F. Boscacci and L. Senn (eds), *I Luoghi della Trasformazione e dell'Innovazione*, SEAT, Turin, pp. 223–246.
Camagni R. and Capello R. (2002), 'Milieux Innovateurs and Collective Learning: From Concepts to Measurement', in Z. Acs, H. de Groot and P. Nijkamp (eds), *The Emergence of the Knowledge Economy: A Regional Perspective*, Springer Verlag, Berlin, pp. 15–45.
Camagni R. and Capello R. (eds) (2002), *Apprendimento Collettivo e Competitività Territoriale*, FrancoAngeli, Milan, pp. 57–84.
Camagni R. and Capello R. (2004), 'The City Network Paradigm: Theory and Empirical Evidence', in R. Capello and P. Nijkamp (eds), *Urban Dynamics and Growth: Advances in Urban Economics*, Elsevier, Amsterdam, pp. 495–529.
Camagni R. and Capello R. (2013), 'Regional Innovation Patterns and the EU Regional Policy Reform: Towards Smart Innovation Policies', *Growth and Change*, vol. 44, no. 2, pp. 355–389.
Camagni R. and Capello R. (2015), 'Rationale and Design of EU Cohesion Policies in a Period of Crisis', *Regional Science Policy and Practice*, vol. 7, no. 1, pp. 25–49.
Camagni R., Capello R. and Caragliu A. (2013), 'One or Infinite Optimal City Sizes? In Search of an Equilibrium Size for Cities', *The Annals of Regional Science*, vol. 51, no. 2, pp. 309–341.
Camagni R., Capello R. and Caragliu A. (2014), 'The Rise of Second Rank Cities: What Role for Agglomeration Economies?', *European Planning Studies*, DOI: 10.1080/09654313.2014.904999.
Camagni R., Capello R. and Caragliu A. (2015), 'Agglomeration Economies and Urban Growth: Spatial Context and Structural Evolution behind Urban Growth', in *Papers in Regional Science*, DOI:10.1111/pirs.12182.
Camagni R., Diappi L. and Leonardi G. (1986), 'Urban Growth and Decline in a Hierarchical System: A Supply-oriented Dynamic Approach', *Regional Science and Urban Economics*, vol. 16, no. 1, pp. 145–160.
Camagni R. and Pompili T. (1990), 'Competence, Power and Waves of Urban Development in the Italian City System', in P. Nijkamp (ed.), *Sustainability of Urban Systems: A Cross-national Evolutionary Analysis of Urban Innovation*, Avebury, Aldershot, pp. 37–86.
Camagni R. and Rabellotti R. (1997), 'Footwear Production Systems in Italy: A Dynamic Comparative Analysis', in R. Ratti, A. Bramanti and R. Gordon (eds), *The Dynamics of Innovative Regions*, Ashgate, Aldershot, pp. 139–164.
Cantwell J. (1989), *Technological Innovation and Multinational Corporations*, Blackwell, Oxford.
Cantwell J. and Iammarino S. (1998), 'MNCs, Technological Innovation and Regional Systems in the EU: Some Evidence in the Italian Case', *International Journal of the Economics of Business*, vol. 5, no. 3, pp. 383–408.
Cantwell J. and Piscitello L. (2002), 'The Location of Tecnological Activities of MNCs in European Regions: The Role of Spillovers and Local Competencies', *Journal of International Management*, vol. 8, no. 1, pp. 69–96.
Capello R. (1988), 'La Diffusione Spaziale dell'Innovazione: Il Caso del Servizio Telefonico', *Economia e Politica Industriale*, no. 58, pp. 141–175.
Capello R. (1994), *Spatial Economic Analysis of Telecommunications Network Externalities*, Avebury, Aldershot.
Capello R. (1999a), 'Spatial Transfer of Knowledge in High-technology Milieux: Learning vs. Collective Learning Processes', *Regional Studies*, vol. 33, no. 4, pp. 353–365.

Capello R. (1999b), 'A Measurement of Collective Learning Effects in Italian High-tech Milieux', *Révue d'Economie Régionale et Urbaine*, no. 3, pp. 449–468.

Capello R. (2000), 'The City-Network Paradigm: Measuring Urban Network Externalities', *Urban Studies*, vol. 37, no. 11, October, pp. 1925–1945.

Capello R. (2001), 'Urban Innovation and Collective Learning: Theory and Evidence from Five Metropolitan Cities in Europe', in M.M. Fischer and J. Froehlich (eds), *Knowledge, Complexity and Innovation Systems*, Springer, Berlin, pp. 181–208.

Capello R. (2002a), 'Urban Rent and Urban Dynamics: the Determinants of Urban Development in Italy', *The Annals of Regional Science*, vol. 36, no. 4, pp. 593–611.

Capello R. (2002b), 'Apprendimento Collettivo e Teorie della Crescita Endogena: una Prospettiva Territoriale', in R. Camagni and R. Capello (eds), *Apprendimento Collettivo e Competitività Territoriale*, FrancoAngeli, Milan, pp. 57–84.

Capello R. (2007), 'A Forecasting Territorial Model of Regional Growth: The MASST Model', *The Annals of Regional Science*, vol. 41, no. 4, pp. 753–787.

Capello R. (2012), 'Territorial Patterns of Innovation', in R. Cappellin, F. Ferlaino and P. Rizzi (eds), *La città nell'economia della conoscenza*, FrancoAngeli, Milan, pp. 51–79.

Capello R. and Caragliu A. (2012), 'Proximities and the Intensity of Scientific Relations: Synergies and Non–linearities', paper presented at the Tinbergen Institute Seminar, Amsterdam, 7–8 May.

Capello R., Caragliu A. and Fratesi U. (2014), 'Modelling Regional Growth between Competitiveness and Austerity Measures: The MASST3 Model', *International Regional Science Review*, DOI: 10.1177/0160017614543850.

Capello R., Caragliu A. and Nijkamp P. (2011), 'Territorial Capital and Regional Growth: Increasing Returns in Knowledge Use', *Tijdschrift voor Economische en Sociale Geographie (TESG)*, vol. 102, no. 4, pp. 385–405.

Capello R. and Fratesi U. (2012), 'Modelling Regional Growth: An Advanced MASST Model', *Spatial Economic Analysis*, vol. 7, no. 3, pp. 293–318.

Capello R. and Lenzi C. (eds) (2013), *Territorial Patterns of Innovation: An Inquiry on the Knowledge Economy in European Regions*, Routledge, London.

Capello R., Nijkamp P. and Pepping G. (1999), *Sustainable Cities and Energy Policies*, Springer Verlag, Berlin.

Capello R. and Spairani A. (2004), 'The Role of Collective Learning in ICT Adoption and Use', in H. de Groot, P. Nijkamp and R. Stough (eds), *Entrepreneurship and Regional Economic Development: A Spatial Perspective*, Edward Elgar, Cheltenham, pp. 198–224.

Cappellin R. (1980), 'Teorie e Modelli dello Sviluppo Spaziale delle Attività di Servizio', *Giornale degli Economisti e Annali di Economia*, no. 3–4, pp. 205–231.

Cappellin R. (1983), 'Osservazioni sulla Distribuzione Inter e Intraregionale delle Attività Produttive', in G. Fuà and C. Zacchia (eds), *Industrializzazione senza Fratture*, Il Mulino, Bologna, pp. 241–271.

Cappellin, R. (2003), 'Territorial Knowledge Management: Towards a Metrics of the Cognitive Dimension of Agglomeration Economies', *International Journal of Technology Management*, vol. 26, no. 2/3/4, pp. 303–325.

Cappellin R. and Grillenzoni C. (1983), 'Diffusion and Specialisation in the Location of Service Activities in Italy', *Sistemi Urbani*, vol. 5, no. 2, pp. 249–282.

Cappellin R. and Nijkamp P. (eds) (1990), *The Spatial Context of Technological Development*, Avebury, Aldershot.

Caragliu A. (2015), 'The Economics of Proximity', PhD dissertation, Free University of Amsterdam, Amsterdam.

Carlino G. (1980), 'Contrast in Agglomeration: New York and Pittsburgh Reconsidered', *Urban Studies*, vol. 17, no. 3, pp. 343–351.

Carney J. (1980), 'Regions in Crisis: Accumulation, Regional Problems and Crisis Formation', in J. Carney, R. Hudson and J. Lewis (eds), *Regions in Crisis*, Croom Helm, London, pp. 28–59.

Carney J., Hudson R. and Lewis J. (eds) (1980), *Regions in Crisis*, Croom Helm, London.

Casi L. and Resmini L. (2014a), 'Spatial Complexity and Interactions in the FDI Attractiveness of Regions', *Papers in Regional Science*, 93(S1), S51–S78.

Casi L. and Resmini L. (2014b), 'FDI and Growth: Can Different Regional Identities Shape the Returns to Foreign Capital Investments?', paper presented at the ERSA Conference, St Petersburg, 26–31 August.

Cass D. (1965), 'Optimal Growth in an Aggregative Model of Capital Accumulation', *Review of Economic Studies*, vol. 32, no. 3, pp. 233–240.

Cecchini P. (1988), *The European Challenge: 1992*, Wildwood House, Aldershot.

Chamberlin E.H. (1936), *The Theory of Monopolistic Competition*, Harvard University Press, Cambridge, MA.

Chatterji M. (1994), 'Convergence Clubs and Endogenous Growth', *Oxford Review of Economic Policy*, vol. 8, no. 4, pp. 57–69.

Chenery H. (1962), 'Development Policies for Southern Italy', *Quarterly Journal of Economics*, vol. 76, no. 4, pp. 515–547.

Chenery H., Clark P. and Pinna C. (1953), *The Structure and Growth of the Italian Economy*, U.S. Mutual Security Agency, Special Mission to Italy for Economic Co-operation.

Chinitz B. (1961), 'Constrast in Agglomeration: New York and Pittsburgh', *American Economic Review*, vol. 51, no. 2, pp. 279–289.

Christaller W. (1933), *Die Zentralen Orte in Süddeutschland*, Wissenschaftlische Buchgesellschaft, Darmstadt; English edition (1966), *The Central Places in Southern Germany*, Prentice-Hall, Englewood Cliffs, NJ.

Ciccone A. (2002), 'Agglomeration Effects in Europe', *European Economic Review*, vol. 46, no. 2, pp. 213–227.

Ciccone A. and Hall R.E. (1996), 'Productivity and the Density of Economic Activity', *American Economic Review*, vol. 86, no. 1, pp. 54–70.

Ciciotti E. (1982), 'Differenze Regionali nella Capacità Innovativa', *Politica ed Economia*, no. 3, pp. 42–48.

Ciciotti E. and Wettmann R. (1981), 'The Mobilisation of Indigenous Potential', Commission of the European Community, *Internal Documentation on Regional Policy*, no. 10, Brussels.

Clark C. (1940), *The Conditions of Economic Progress*, Macmillan, London.

Clark C. (1945), 'The Economic Functions of a City in Relation to Its Size', *Econometrica*, vol. 13, no. 2, pp. 97–113.

Clark N.G. (1971), 'Science, Technology and Regional Economic Development', *Research Policy*, vol. 8, no. 1, pp. 296–319.

Cobb C.W. and Douglas P.H. (1928), 'A Theory of Production', *American Economic Review*, vol. 18, no. 1, March, pp. 139–165.

Coehlo P. and Ghali M.A. (1971), 'The End of the North–South Wage Differential', *The American Economic Review*, vol. 61, no. 5, pp. 932–937.

Coehlo P. and Shepherd J. (1979), 'The Impact of Regional Differences in Prices and Wages in Economic Growth: the United States in 1890', *Journal of Economic History*, vol. 39, no. 1, pp. 69–85.

Cohen W.M. and Levinthal D.A. (1990), 'Absorptive Capacity: A New Perspective on Learning and Innovation', *Administrative Science Quarterly*, vol. 35, no. 1, pp. 128–152.

Combes P.P., Duranton G. and Gobillon L. (2008), 'Spatial Wage Disparities: Sorting Matters!', *Journal of Urban Economics*, vol. 63, no. 2, pp. 723–742.

Cooke Ph. (2002), *Knowledge Economies: Clusters, Learning and Cooperative Advantage*, Routledge, London.

Cooke Ph., Boekholt P. and Tödling F. (2000), *The Governance of Innovation in Europe*, Pinter, London.

Cooke Ph., Heidenreich M. and Braczyk M. (2004), *Regional Innovation Systems*, 2nd edn, Routledge, London.

Cooke Ph. and Morgan K. (1994), 'The Creative Milieu: A Regional Perspective on Innovation', in M. Dodgson and R. Rothwell (eds), *The Handbook of Industrial Innovation*, Edward Elgar, Cheltenham, pp. 25–32.

Costa J. Da Silva, Ellson R.W. and Martin R.C. (1987), 'Public Capital, Regional Output and Development: Some Empirical Evidence', *Journal of Regional Science*, vol. 27, no. 3, pp. 419–435.

Courlet C. and Pecqueur B. (1992), 'Les Systèmes Industriels Localisés en France: Un Nouveau Modèle de Développement', in G. Benko and A. Lipietz (eds), *Les Régions qui Gagnent. Districts et Réseaux: Les Nouveaux Paradigmes de la Géograqhie Economique*, Presses Universitaires de France, Paris, pp. 81–102.

Crescenzi R., Pietrobelli C. and Rabellotti R. (2012), 'Innovation Drivers, Value Chains and the Geography of Multinational Firms in European Regions', LSE 'Europe in Question' Discussion Paper Series No. 3/2012.

Cuadrado-Roura J.R. and Parellada M. (eds) (2002), *Regional Convergence in the European Union*, Springer Verlag, Berlin.

Cusinato A. and Philippopoulos-Mihalopoulos A. (eds) (2015), *Knowledge-creating Milieus in Europe: Firms, Cities, Territories*, Berlin, Springer Verlag.

Dacey M.F. (1964), 'A Note on Some Number Properties of a Hexagonal Hierarchical Plane Lattice', *Journal of Regional Science*, vol. 5, no. 2, pp. 63–67.

Damette F. (1980), 'The Regional Framework of Monopoly Exploitation: New Problems and Trends', in J. Carney, R. Hudson and J. Lewis (eds), *Regions in Crisis*, Croom Helm, London, pp. 76–92.

Davelaar E.J. (1991), *Regional Economic Analysis of Innovation and Incubation*, Avebury, Aldershot.

Davelaar E.J. and Nijkamp P. (1990), 'Industrial Innovation and Spatial Systems: The Impact of Producer Services', in H. Ewers and J. Allesch (eds), *Innovation and Regional Development*, de Gruyter, Berlin, pp. 83–122.

David P. (1985), 'Clio and the Economics of Qwerty', *AEA Papers and Proceedings*, vol. 75, no. 2, pp. 332–337.

Davies D. (1979), *The Diffusion of Process Innovations*, Cambridge University Press, Cambridge.

de Groot H., Nijkamp P. and Acs Z. (eds) (2001), 'Knowledge Spill-overs, Innovation and Regional Development', *Papers in Regional Science*, vol. 80, no. 3, special issue.

de Groot H., Nijkamp P. and Stough R. (eds) (2004), *Entrepreneurship and Regional Economic Development: A Spatial Perspective*, Edward Elgar, Cheltenham.

Decoster E. and Tabariés M. (1986), 'L'innovation dans un Pôle Scientifique et Technologique: Le Cas de la Cité Scientifique Ile–de–France', in Ph. Aydalot (ed.), *Milieux Innovateurs en Europe*, GREMI, Paris, pp. 79–100.

Decressin J. and Fatàs A. (1995), 'Regional Labour Market Dynamics in Europe', *European Economic Review*, vol. 39, no. 9, pp. 1627–1655.

Dei Ottati G. (1995), *Tra Mercato e Comunità: Aspetti Concettuali e Ricerche Empiriche sul Distretto Industriale*, FrancoAngeli, Milan.

Dei Ottati G. (2003), 'The Governance of Transactions in the Industrial District: The "Community Market"', in G. Becattini, M. Bellandi, G. Dei Ottati and F. Sforzi (eds), *From Industrial Districts to Local Development*, Edward Elgar, Cheltenham, chap. 4; orig. edn (1987), 'Il Mercato Comunitario', in G. Becattini (ed.), *Mercato e Forze Locali: il Distretto Industriale*, Il Mulino, Bologna, pp. 117–141.

Dematteis G. (1994), 'Global Networks, Local Cities', *Flux*, no. 15, pp. 17–13.

Deno K.T. (1988), 'The Effects of Public Capital on U.S. Manufacturing Activity: 1970 to 1978', *Southern Economic Journal*, vol. 55, no. 1, pp. 400–411.

Derycke P.-H. (1996), 'Equilibre Spatial Urbain', in P.-H. Derycke , J.-M. Huriot and D. Pumain (eds), *Penser la Ville*, Anthropos, Paris, pp. 53–90.

Devereux M.P., Griffith R. and Simpson H. (2004), 'The Geographic Distribution of Production Activity in the UK', *Regional Science and Urban Economics*, vol. 34, no. 5, 533–564.

Ding L. and Haynes K.E. (2004), 'The Role of Telecommunications Infrastructure in Regional Economic Growth of China', paper presented at the Telecommunications Policy Research Conference, Washington, DC, 1–3 October.

Dixit A. and Stiglitz J. (1977), 'Monopolistic Competition and Optimum Product Diversity', *American Economic Review*, vol. 67, no. 3, pp. 297–308.

Dixon R. and Thirlwall A.P. (1975), 'A Model of Regional Growth Rate Differences on Kaldorian Lines', *Oxford Economic Papers*, vol. 27, no. 2, pp. 201–214.

Dobkins L.H. and Ioannides Y.M. (2001), 'Spatial Interaction Among U.S. Cities: 1900–1990', *Regional Science and Urban Economics*, vol. 31, no. 6, pp. 701–731.

Domar E.D. (1957), *Essays in the Theory of Economic Growth*, Oxford University Press, London.

Dosi G. (1982), 'Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change', *Research Policy*, vol. 11, no. 3, pp. 147–162.

Dosi G., Freeman C., Nelson R., Silverberg G. and Soete L. (eds) (1988), *Technical Change and Economic Theory*, Pinter, London.

Duffy-Deno K.T. and Eberts R.W. (1991), 'Public Infrastrutcure and Regional Economic Development: A Simultaneous Equation Approach', *Journal of Urban Economics*, vol. 30, no. 3, pp. 329–343.

Eaton C. and Lipsey R. (1982), 'An Economic Theory of Central Places', *The Economic Journal*, vol. 92, no. 365, pp. 56–72.

Eberts R.W. (1986), *Estimating the Contribution of Urban Public Capital Stock to Regional Growth*, working paper no. 8610, Federal Reserve Bank of Cleveland, Cleveland.

Edquist C. (ed.) (1997), *Systems of Innovation*, Pinter, London.

Edquist C. (2005), 'Systems of Innovation – Perspectives and Challenges', in J. Fagerberg, D. Mowery and R. Nelson (eds), *The Oxford Handbook of Innovation*, Oxford University Press, Oxford, pp. 181–208.

Elhance A. and Lakshmanan T.R. (1988), 'Infrastructure-Production System Dynamics in National and Regional Systems', *Regional Science and Urban Economics*, vol. 18, no. 4, pp. 511–531.

Ellison G. and Glaeser E.L. (1997), 'Geographic Concentration in US Manufacturing Industries: A Dartboard Approach', *Journal of Political Economy*, vol. 105, no. 5, 889–927.

Emanuel C. (1988), 'Recenti Trasformazioni nell'Organizzazione Urbana e Regionale: il Caso Padano', Papers and Proceedings of the IX Italian Regional Science Conference, Turin.

Emanuel C. and Dematteis G. (1990), 'Reti Urbane Minori e Deconcentrazione Metropolitana nella Padania Centro–Occidentale', in D. Martellato and F. Sforzi (eds), *Studi sui Sistemi Urbani*, FrancoAngeli, Milan, pp. 233–261,

ESDP – *European Spatial Development Perspective* (1998), Final Version, Glasgow, 8 June.

European Commission (2004), *A New Partnership for Cohesion*, Third Report on Economic and Social Cohesion, Office for Offical Publications of the European Communities, Luxembourg.

European Commission (2005), *Territorial State and Perspectives of the European Union*, scoping document and summary of political messages, May OOPEC, Luxembourg.

European Commission (2011), *The Urban and Regional Dimension of Europe 2020 Seventh Progress Report on Economic, Social and Territorial*, OOPEC Luxembourg.

Evans A. (1985), *Urban Economics: An Introduction*, Basil Blackwell, Oxford.

Ewers H.-J. and Allesch J. (ed.) (1990), *Innovation and Regional Development*, De Gruyter, Berlin.

Fabiani S. and Pellegrini G. (1997), 'Education, Infrastructure, Geography and Growth: An Empirical Analysis of the Development of Italian Provinces', Banca d'Italia, *Temi di Discussione*, no. 323.

Faggian A. and Biagi B. (2003), 'Measuring Regional Multipliers: A Comparison Between Different Methodologies in the Case of the Italian Regions', *Scienze Regionali*, vol. 1, no. 1, pp. 33–58.

Faggio G., Silva O. and Strange W.C. (2013), 'Heterogeneous Industries', paper presented at the 60th North American Regional Science Conference, Atlanta, 13–16 November.

Faggio G., Silva O. and Strange, W.C. (2014), 'Heterogeneous Agglomeration', SERC Discussion Papers, SERCDP0152. Spatial Economics Research Centre (SERC), London School of Economics and Political Science, London.

Faini R. (1984), 'Increasing Returns, Non Traded Inputs and Regional Development', *Economic Journal*, vol. 94, no. 374, pp. 308–323.

Farole T., Rodriguez-Pose A. and Storper M. (2011), 'Cohesion Policy in the European Union: Growth, Geography, Institutions', *Journal of Common Market Studies*, vol. 49, no. 5, pp. 1089–1111.

Feenstra R.E. (2003) *Advanced International Trade: Theory and Evidence*. Princeton University Press, Princeton, NJ.

Feldman M. and Audretsch D. (1999), 'Innovation in Cities: Science-Based Diversity, Specialisation and Localized Competition', *European Economic Review*, vol. 43, no. 2, pp. 409–429.

Ferri G. and Mattesini F. (1997), 'Finance, Human Capital and Infrastructure: An Empirical Investigation of Post-War Italian Growth', Banca d'Italia, *Temi di Discussione*, no. 321.

Fetter F. (1924), 'The Economic Law of Market Areas', *Quarterly Journal of Economics*, vol. 38, no. 3, pp. 520–529.

Fisher A. (1933), 'Capital and the Growth of Knowledge', *The Economic Journal*, vol. 43, no. 171, pp. 379–389.

Foray D. (2009), 'Understanding Smart Specialisation', in D. Pontikakis, D. Kyriakou and R. van Bavel (eds), *The Question of R&D Specialisation*, JRC, European Commission, Directoral General for Research, Brussels, pp. 19–28.

Foray D., David P. and Hall B. (2009), 'Smart Specialisation: The Concept', in *Knowledge for Growth: Prospects for Science, Technology and Innovation*, Report, EUR 24047, European Union.

Foray D., David P. and Hall B. (2011) 'Smart Specialisation: From Academic Idea to Political Instrument, the Surprising Career of a Concept and the Difficulties Involved in Its Implementation', MTEI working paper, November, Lausanne.

Foray D. and Lundvall B.A. (eds) (1996), *Employment and Growth in the Knowledge-Based Economy*, OECD, Paris.

Fratesi U. (2008), 'Issues in the Measurement of Localization', *Environment and Planning A*, vol. 40, no. 3, pp. 733–758.

Frenkel A. (2001), 'Why High-Technology Firms Choose to Locate in or Near Metropolitan Areas', *Urban Studies*, vol. 38, no. 7, pp. 1083–1101.

Frenken K., van Oort F.G. and Verburg T. (2007), 'Related Variety, Unrelated Variety and Regional Economic Growth', *Regional Studies*, vol. 41, no. 5, pp. 685–697.

Friedmann J. (1966), *Regional Development Policy: A Case Study of Venezuela*, MIT Press, Cambridge, MA.

Fritsch M. (2001), 'Co-operation in Regional Innovation Systems', *Regional Studies*, vol. 35, no. 4, pp. 297–307.

Fuchs V. (1967), 'Differentials in Hourly Earnings by Regions and City Size, 1959', *NBER Occasional Papers*, no. 101, New York.

Fujita M. (1985), 'Existence and Uniqueness of Equilibrium and Optimal Land Use', *Regional Science and Urban Economics*, vol. 15, no. 2, pp. 295–324.

Fujita M. (1989), *Urban Economic Theory: Land Use and City Size*, Cambridge University Press, Cambridge.

Fujita M. and Mori T. (1996), 'The Role of Ports in the Making of Major Cities: Self-Organization and Hub Effects', *Journal of Development Economics*, vol. 49, no. 1, pp. 93–120.

Fujita M. and Mori T. (1997), 'Structural Stability and Evolution of Urban Systems', *Regional Science and Urban Economics*, vol. 27, no. 4, pp. 399–442.

Fujita M. and Thisse J.-F. (1996), 'Economics of Agglomeration', *Journal of the Japanese and International Economies*, vol. 10, no. 4, pp. 339–378.

Fujita M. and Thisse J.-F. (2002), *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*, Cambridge University Press, Cambridge.

Fujita M., Krugman P. and Mori T. (1999), 'On the Evolution of Hierarchical Urban Systems', *European Economic Review*, vol. 43, no. 2, pp. 209–251.

Fujita M., Krugman P. and Venables A.J. (1999), *The Spatial Economy: Cities, Regions and International Trade*, MIT Press, Cambridge, Mass.

Ganne B. (1992), 'Place et Evolution des Systèmes Industriels Locaux en France: Economie Politique d'une Transformation', in G. Benko and A. Lipietz (eds), *Les Régions qui Gagnent. Districts et Réseaux: Les Nouveaux Paradigmes de la Géograqhie Economique*, Presses Universitaires de France, Paris, pp. 315–345.

Gans J. and Shepherd G. (1994), 'How Are the Mighty Fallen: Rejected Classic Articles by Leading Economists', *Journal of Economic Perspectives*, vol. 8, no. 1, pp. 165–179.

Garofoli G. (1981), 'Lo Sviluppo delle Aree Periferiche nell'Economia Italiana degli Anni Settanta', *L'Industria*, vol. 5, no. 3, pp. 391–404.

Giannitsis T. (2009), 'Technology and Specialization: Strategies, Options, Risks', *Knowledge Economists Policy Brief*, no. 8.

Giersch H. (1949), 'Economic Union between Nations and the Location of Industries', *Review of Economic Studies*, vol. 17, no. 2, pp. 87–97.

Gillespie A., Goddard J., Hepworth M. and Williams H. (1989), 'Information and Communications Technology and Regional Development: An Information Economy Perspective', *Science, Technology and Industry Review*, no. 5, April, pp. 86–111.

Gillespie A. and Williams H. (1988), 'Telecommunications and the Reconstruction of Regional Comparative Advantage', *Environment and Planning A*, vol. 20, no. 10, pp. 1311–1321.

Glaeser E.L., Kallal H.D., Scheinkman J.A. and Shleifer A. (1992), 'Growth in Cities', *The Journal of Political Economy*, vol. 100, no. 6, pp. 1126–1152.

Goddard J. and Pye R. (1977), 'Telecommunications and Office Location', *Regional Studies*, vol. 11, no. 1, pp. 19–30.

Goddard J. and Thwaites A. (1986), 'New Technology and Regional Development Policy', in P. Nijkamp (ed.), *Technological Change, Employment and Spatial Dynamics*, Springer Verlag, Berlin, pp. 91–114.

Gottman J. (1991), 'The Dynamics of City Networks in an Expanding World', *Ekisticks*, no. 350–351, pp. 227–281.

Graziani A. (1983), 'La Teoria Economica di Vera Lutz', *Moneta e Credito*, vol. 36, no. 141, pp. 3–29.

Greenhut M. (1959a), 'Size of Markets versus Transport Costs in Industrial Location Surveys and Theory', *Journal of Industrial Economics*, vol. 8, pp. 172–184.

Greenhut M. (1959b), 'Comments on Economic Base Theory', *Land Economics*, vol. 35, no. 1, pp. 71–74.

Greenhut M. (1964), 'When Is the Demand Factor of Location Important?', *Land Economics*, vol. 40, no. 2, pp. 175–184.

Greenhut M. (1966), 'Needed – A Return to the Classics in Regional Economic Development Theory', *Kyklos*, vol. XIX, pp. 461–478; reprinted in D. Mckee, R. Dean and W. Leahy (eds) (1970), *Regional Economics: Theory and Practice*, The Free Press, New York, pp. 65–80.

Griliches Z. (1957), 'Hybrid Corn: An Exploration in the Economics of Technological Change', *Econometrica*, vol. 25, no. 4, pp. 501–525.

Grossmann G. and Helpman E. (1991), *Innovation and Growth in the Global Economy*, MIT Press, Cambridge, Mass.

Guild R.L. (1998), *Infrastructure Investment and Regional Development: Theory and Evidence*, Working Paper no. 3, Working Paper Series of the Department of Planning, University of Auckland, New Zealand.

Hägerstrand T. (1952), 'The Propagation of Innovation Waves', *Lund Studies in Geography, Human Geography*, no. 4, pp. 3–19.

Hägerstrand T. (1966), 'Aspects of the Spatial Structure of Social Communication and the Diffusion of Innovation', *Papers of the Regional Science Association*, vol. 16, no. 1, pp. 27–42.

Hägerstrand T. (1967), *Innovation Diffusion as a Spatial Process*, University of Chicago Press, Chicago.

Haggett P. (1965), *Locational Analysis in Human Geography*, Edward Arnold, London.

Haggett P., Cliff A. and Frey A. (1977), *Locational Analysis in Human Geography*, vol. I, 2nd edn, Edward Arnold, London.

Hansen E. (1987), 'Industrial Location Choice in Saõ Paulo, Brazil: A Nested Logic Model', *Regional Science and Urban Economics*, vol. 17, no. 1, pp. 89–108.

Hansen N.M. (1965a), 'Unbalanced Growth and Regional Development', *Western Economic Journal*, no. 4, pp. 3–14.

Hansen N.M. (1965b), 'The Structure and the Determinants of Local Public Investment Expenditures', *The Review of Economics and Statistics*, vol. 45, no. 2, pp. 150–162.

Hansen N.M. (1967), 'Development Pole Theory in a Regional Context', *Kyklos*, vol. 20, pp. 709–725; reprinted in D. Mckee, R. Dean and W. Leahy (eds) (1970), *Regional Economics: Theory and Practice*, The Free University Press, New York, pp. 229–242.

Harrod R.F. (1939), 'An Essay in Dynamic Theory', *The Economic Journal*, vol. 49, no. 193, pp. 14–33.

Hawtrey R. (1939), 'Mr. Harrod's Essay in Dynamic Theory', *The Economic Journal*, vol. 49, no. 195, pp. 468–475.

Heckscher E.F. (1919), 'The Effects of Foreign Trade on the Distribution of Income', *Economik Tidskrift*, pp. 497–512; English version in H.S Ellis and L.S. Metzler (eds) (1950), *Readings on the Theory of International Trade*, Allen & Unwin, London, pp. 270–300.

Henderson J.V. (1974), 'The Sizes and Types of Cities', *The American Economic Review*, vol. 64, no. 4, pp. 640–656.

Henderson J.V. (1985), *Economic Theory and the Cities*, Academic Press, Orlando, FL.

Henderson J.V. (1996), 'Ways to Think about Urban Concentration: Neoclassical Urban Systems vs. the New Economic Geography', *International Regional Science Review*, vol. 19, no. 1–2, pp. 31–36.

Henderson J.V. (2003), 'Marshall's Scale Economies', *Journal of Urban Economics*, vol. 53, no. 1, pp. 1–28.

Hepworth M. and Waterson M. (1988), 'Information Technology and the Spatial Dynamics of Capital', *Information Economics and Policy*, vol. 3, no. 2, pp. 148–163.

Herfindahl O.C. (1950), 'Concentration of the US Steel Industry', PhD dissertation, Columbia University, NY.

Hewings G.J.D. (1977), *Regional Industrial Analysis and Development*, Methuen, London.

Hewings G.J.D., Okuyama Y. and Sonis M. (2001), 'Economic Interdependence within the Chicago Metropolitan Region: A Miyazawa Analysis,' *Journal of Regional Science*, vol. 41, no. 2, pp. 195–217.

Higgins B. (1977) 'Development Poles: Do They Exist?', *Economie Appliquée*, vol. 30, no. 2, pp. 241–258.

Hildebrand G. and Mace A. (1950), 'The Employment Multiplier in an Expanding Industrial Market: Los Angeles County, 1940–47', *Review of Economics and Statistics*, vol. 32, no. 3, pp. 241–249.

Hirsch S. (1967), *Location of Industry and International Competitiveness*, Clarendon Press, London.

Hirshmann A.O. (1945), *National Power and the Structure of Foreigner Trade*, University of California Press, Berkeley, CA.
Hirschman A.O. (1957), 'Investment Policies and "Dualism" in Underdeveloped Countries', *The American Economic Review*, vol. 47, no. 5, pp. 550–570.
Hirschman A.O. (1958), *The Strategy of Economic Development*, Yale University Press, New Haven, CT.
Hirschmann A.O. (1964), 'The Paternity of an Index', *American Economic Review*, vol. 54, no. 5, p. 761.
Hirschman A.O. and Sirkin G. (1958), 'Investment Criteria and Capital Intensity Once Again', *Quarterly Journal of Economics*, vol. 72, no. 3, pp. 469–471.
Hoch I. (1972), 'Income and City Size', *Urban Studies*, vol. 9, no. 3, pp. 299–328.
Hogenbirk A. and Narula R. (2004), 'Location and Agglomeration of FDI in the Netherlands: Implications for policy'. Research Memoranda 006, MERIT, Maastricht Economic Research Institute on Innovation and Technology, Maastricht.
Holland S. (1971), 'Regional Underdevelopment in a Developed Economy: The Italian Case', *Regional Studies*, vol. 5, no. 2, pp. 71–90.
Holland S. (1977), 'Capital, Labour and the Regions: Aspects of Economic, Social and Political Inequality in Regional Theory and Policy', in H. Folmer and J. Oosterhaven (eds), *Spatial Inequalities and Regional Development*, Martinus Nijhoff, The Hague, pp. 185–218.
Hoover E.M. (1933), 'The Location of Shoe Industry in the United States', *Quarterly Journal of Economics*, vol. 47, no. 2, pp. 254–276.
Hoover E.M. (1936), 'The Measurement of Industrial Localisation', *The Review of Economics and Statistics*, vol. 18, no. 4, pp. 162–171.
Hoover E.M. (1937a), 'Spatial Price Discrimination', *The Review of Economics and Statistics*, vol. 4, no. 3, pp. 182–191.
Hoover E.M. (1937b), *Location Theory and the Shoe and Leather Industry*, Harvard University Press, Cambridge, Mass.
Hoover E.M. (1948), *The Location of Economic Activity*, McGraw-Hill, New York.
Hoover E.M. and Fisher A. (1949), 'Research in Regional Economic Growth', *Problems in the Study of Economic Growth*, N.B.E.R., New York.
Hoover E.M. and Vernon R. (1962), *Anatomy of a Metropolis*, Harvard University Press, Cambridge, MA.
Hotelling H. (1929), 'Stability in Competition', *The Economic Journal*, vol. 39, no. 153, pp. 41–57.
Howells J. (1999), 'Regional Systems of Innovation?', in D. Archibugi, J. Howells and J. Michie (eds), *Innovation Policy in a Global Economy*, Cambridge University Press, Cambridge, pp. 67–93.
Hoyt H. (1954), 'Homer Hoyt on the Development of Economic Base Concept', *Land Economics*, vol. 30, no. 2, pp. 182–187.
Hulten C.R. and Schwab R.M. (1995), 'Infrastructure and the Economy', in J.M. Pogodzinski (ed.), *Readings in Public Policy*, MIT Press, Cambridge, MA, pp. 213–234.
Huriot J.-M. (1988), *Von Thünen: Economie et Espace*, Economica, Paris.
Huriot J.-M. (1994), 'L'Utilisation du Sol', in J.-P. Auray, A. Bailly, P.-H. Derycke and J.-M. Huriot (eds), *Encyclopédie d'Economie Spatiale*, Economica, Paris, pp. 297–306.
Iacobucci D. (2013), 'Designing and Implementing a Smart Specialisation Strategy at Regional Level: Some Open Questions', *Scienze Regionali – Italian Journal of Regional Science*, vol. 13, no. 1, pp. 107–126.
Iammarino S. and McCann Ph. (2013), *Multinationals and Economic Geography*, Edward Elgar, Cheltenham.
Ioannides Y. and Overman H. (2003), 'Zipf's Law for Cities: An Empirical Examination', *Regional Science and Urban Economics*, vol. 33, no. 2, pp. 127–137.

Isard W. (1949), 'The General Theory of Location and Space', *Quarterly Journal of Economics*, vol. 63, no. 4, pp. 476–506.

Isard W. (1954), 'Location Theory and Trade Theory: Short Run Analysis', *Quarterly Journal of Economics*, vol. 68, no. 2, pp. 305–320.

Isard W. (1956), *Location and Space-Economy*, MIT Press, Cambridge, MA.

Isard W. (1970), 'Game Theory, Location Theory and Industrial Agglomeration', in H. Richardson (ed.), *Regional Economics: A Reader*, Macmillan, London.

Isard W. and Peck M. (1954), 'Location Theory and International and Interregional Trade Theory', *Quarterly Journal of Economics*, vol. 68, no. 1, pp. 97–114.

Jacobs, J. (1969), *The Economy of Cities*, Vintage, New York.

Jaffe A.B. (1989), 'Real Effects of Academic Research', *American Economic Review*, vol. 79, no. 5, pp. 957–970.

Johannisson B. and Spilling O. (1983), *Strategies for Local and Regional Self-Development*, NordREFO, Oslo.

Johnson A. (2006), 'FDI and Exports: The Case of High Performing East Asian Economies', CESIS Electronic working paper series no. 57.

Kaldor N. (1970), 'The Case of Regional Policies', *Scottish Journal of Political Economy*, vol. 17, no. 3, pp. 337–348.

Kaldor N. (1975), 'Economic Growth and the Verdoorn Law – A Comment on Mr. Rowthorn's Article', *The Economic Journal*, vol. 85, no. 340, pp. 891–896.

Kanemoto Y. (1987), 'Externalities in Space', in T. Miyao and Y. Kanemoto (eds), *Urban Dynamics and Urban Externalities*, Harwood Academic Publishers, Chur, Switzerland, pp. 43–103.

Keeble D. (1990), 'Small Firms, Innovation and Regional Development in Britain in the 1990s', *Regional Studies*, vol. 31, no. 3, pp. 281–293.

Keeble D., Offord J. and Walker S. (1988), *Peripheral Regions in a Community of Twelve Member States*, Office for Official Publications of the European Communities, Luxembourg.

Keeble D., Owens P.L. and Thompson C. (1982), 'Regional Accessibility and Economic Potential in the European Community', *Regional Studies*, vol. 16, no. 6, pp. 419–432.

Keeble D. and Wilkinson F. (1999), 'Collective Learning and Knowledge Development in the Evolution of Regional Clusters of High-techology SMEs in Europe', *Regional Studies*, vol. 33, no. 4, pp. 295–303.

Keeble D. and Wilkinson F. (eds) (2000), *High-Technology Clusters, Networking and Collective Learning in Europe*, Ashgate, Aldershot.

Keesing D. (1966), 'Labour Skills and Comparative Advantage', *The Review of Economics and Statistics*, vol. 56, no. 1/2, pp. 249–258.

Kim J. (2011), 'Non-market Effects on Agglomeration and Their Policy Responses: Can We Overcome the Mismatch?', in OECD *Regional Outlook 2011*, Paris, 195–201.

Koopmans T. (1965), 'On the Concept of Optimal Economic Growth', *Econometric Approach to Development Planning*, North-Holland, Amsterdam, pp. 225–287.

Krauss M.B. and Johnson H.G. (1974), *General Equilibrium Analysis*, Allen & Unwin, London.

Krugman P. (1991a), *Geography and Trade*, MIT Press, Cambridge, MA.

Krugman P. (1991b), 'Increasing Returns and Economic Geography', *Journal of Political Economy*, vol. 99, no. 3, pp. 484–499.

Krugman P. (1991c), 'History vs. Expectations', *The Quarterly Journal of Economics*, vol. 106, no. 2, pp. 651–667.

Krugman P. (1993), 'First Nature, Second Nature and Metropolitan Location', *Journal of Regional Science*, vol. 33, no. 2, pp. 129–144.

Krugman P. (1996a), 'Making Sense of the Competitiveness Debate', *Oxford Review of Economic Policy*, vol. 12, no. 3, pp. 17–25.

Krugman P. (1996b), *Pop Internationalism*, MIT Press, Cambridge MA.

Krugman P. (1998), 'What's New about the New Economic Geography', *Oxford Review of Economic Policy*, vol. 14, no. 2, pp. 7–17.
Krugman P. and Venables A.J. (1996), 'Integration, Specialisation and Adjustment', *European Economic Review*, vol. 40, no. 3/5, pp. 959–967.
Kugler M. (2006), 'Spillover from Foreign Direct Investment: Within or Between Industries?', *Journal of Development Economics*, vol. 80, no. 2, pp. 444–474.
Lancaster K. (1975), 'Socially Optimal Product Differentiation', *American Economic Review*, vol. 65, no. 4, pp. 567–585.
Lancaster K. (1980), 'Intra-Industry Trade Under Perfect Monopolistic Competition', *Journal of International Economics*, vol. 10, no. 2, pp. 151–175.
Launhardt W. (1882), 'Die Bestimmung des Zweckmässigsten Standorts einer Gewerblischen Anlage', *Zeitschrift des Vereins Deutscher Ingenieure*, vol. 26, no. 3, pp. 106–115.
Launhardt W. (1885), *Mathematische Begründung der Volkwirtschaftslehre*, Leipzig.
Lawson C. and Lorenz E. (1999), 'Collective Learning, Tacit Knowledge and Regional Innovation Capacity', *Regional Studies*, vol. 33, no. 4, pp. 305–317.
Leon-Ledesma M.A. (1998), 'Economic Growth and Verdoorn's Laws in the Spanish Regions, 1962–1991', Discussion Paper no. 1, Department of Economics, University of Kent.
Leontief W. (1953), 'Domestic Production and Foreigner Trade: the American Capital Position Re-examined', *Proceedings of the American Philosophical Society*, vol. 97, no. 4, pp. 332–349.
Leontief W. (1956), 'Factor Proportions and the Structure of American Trade: Further Theoretical and Empirical Analysis', *The Review of Economics and Statistics*, vol. 38, no. 4, pp. 386–407.
Linder S.B. (1961), *An Essay on Trade and Transformation*, John Wiley & Sons, New York.
Lipietz A. (1980), 'The Structuration of Space, the Problem of Land, and Spatial Policy', in J. Carney, R. Hudson and J. Lewis (eds), *Regions in Crisis*, Croom Helm, London, pp. 60–75.
Long W. (1971), 'Demand in Space: Some Neglected Aspects', *Papers and Proceedings of the Regional Science Association*, vol. 27, no. 1, pp. 45–62.
Lösch A. (1954), *The Economics of Location*, Yale University Press, New Haven, Conn., orig. edn (1940), *Die Räumlische Ordnung der Wirtschaft*, Gustav Fischer, Jena.
Lucas R. (1988), 'On the Mechanics of Economic Development', *Journal of Monetary Economics*, vol. 22, no. 1, pp. 3–42.
Lundvall B.-A. (1992), 'Introduction', in B.-A. Lundvall (ed.), *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, Pinter, London, pp. 1–19.
Lundvall B.-A. and Johnson B. (1994), 'The Learning Economy', *Journal of Industry Studies*, vol. 1, no. 2, pp. 23–42.
Lung P. (1987), 'Complexity and Spatial Dynamic Modelling, from Catastrophe Theory to Self-organising Processes: A Review of the Literature', Institut d'Economie Régionale du Sud-Ouest, mimeo.
Lutz V. (1962), *Italy – A Study in Economic Development*, Oxford University Press, London.
Maggioni M. (2002), *Clustering Dynamics and the Location of High-Tech Firms*, Physica-Verlag, Berlin.
Magrini S. (1997), 'Spatial Concentration in Research and Regional Income Disparities in a Decentralised Model of Endogenous Growth', *Research Papers in Environmental and Spatial Analysis*, no. 43, London School of Economics.
Magrini S. (2004) 'Regional (Di)convergence', in V. Henderson and F.J. Thisse (eds), *Handbook of Regional and Urban Economics*, vol. 4, North-Holland, Amsterdam, pp. 2741–2796.
Maier G. and Sedlacek S. (eds) (2005), *Spillovers and Innovations – Space, Environment and the Economy*, Springer Verlag, Vienna.
Maillat D., Quévit M. and Senn L. (eds) (1993), *Réseaux d'Innovation et Milieux Innovateurs: Un Pari pour le Développement Régional*, EDES, Neuchâtel.
Malberg A. and Maskell P. (2002), 'The Elusive Concept of Localization Economies: Towards a Knowledge-based Theory of Spatial Clustering', *Environment and Planning A*, vol. 34, no. 3, pp. 429–449.

Malecki E. and Varaiya P. (1986), 'Innovation and Changes in Regional Structure', in P. Nijkamp (ed.), *Handbook of Regional and Urban Economics*, North-Holland, Amsterdam, pp. 629–645.

Mankin N., Romer D. and Weil D. (1992), 'A Contribution to the Empirics of Economic Growth', *Quarterly Journal of Economics*, vol. 107, no. 2, pp. 739–774.

Mansfield E. (1961), 'Technological Change and the Rate of Imitation', *Econometrica*, vol. 29, no. 4, pp. 741–766.

Mansfield E. (1968), *The Economics of Technological Change*, Norton, New York.

Marelli E. (1981), 'Optimal City Size, the Productivity of Cities and Urban Production Functions', *Sistemi Urbani*, vol. 3, no. 1/2, pp. 149–163.

Markusen J. and Venables A. (1999), 'Foreign Direct Investment as a Catalyst for Industrial Development', *European Economic Review*, vol. 43, no. 2, pp. 335–356.

Marshall A. (1920), *Principles of Economics*, 8th edn, Macmillan, London.

Martellato D. (1982), 'The Tuscany Interregional Input Output Model', International Institute for Applied Systems Analysis, IIASA paper series no. 82–30, Vienna.

Martin P. and Ottaviano G. (1999), 'Growing Locations: Industry Location in a Model of Endogenous Growth', *European Economic Review*, vol. 43, no. 2, pp. 281–302.

Martin R. (1999), 'The New "Geographical Turn" in Economics: Some Critical Reflections', *Cambridge Journal of Economcs*, vol. 23, no. 1, pp. 65–91.

Maskell P. and Malmberg A. (1999), 'Localised Learning and Industrial Competitiveness', *Cambridge Journal of Economics*, vol. 23, no. 2, pp. 167–185.

Massey D. and Meegan R. (1978), 'Industrial Restructuring versus the City', *Urban Studies*, vol. 15, no. 3, pp. 273–288.

McCann P. and Ortega-Argilés R. (2014), 'The Role of the Smart Specialisation Agenda in a Reformed EU Cohesion Policy', *Scienze Regionali, Italian Journal of Regional Science*, vol. 13, no. 1, pp. 15–32.

McCann Ph. (1998), *The Economics of Industrial Location: A Logistic-Cost Approach*, Springer, Heidelberg.

McCann Ph. and Rodríguez-Pose A. (2011), 'Why and When Development Policy Should Be Place-based', in *OECD Regional Outlook 2011*, Paris, pp. 203–213.

McCann Ph. and Sheppard S. (2003), 'The Rise, Fall and Rise Again of Industrial Location Theory', *Regional Studies*, vol. 37, no. 6/7, pp. 649–663.

McCombie J.S. (1982), 'How Important Is the Spatial Diffusion of Innovation in Explaining Regional Growth Disparities?', *Urban Studies*, vol. 19, no. 2, pp. 377–382.

McCombie J.S. (1983), 'Kaldor's Law in Retrospect', *Journal of Post Keynesian Economics*, vol. 5, no. 3, pp. 414–429.

McCombie J.S. (1988), 'A Synoptic View of Regional Growth and Unemployment: I – The Neoclassical Theory', *Urban Studies*, vol. 25, no. 4, pp. 267–281.

McCombie J.S. (1992), 'Thirwall's Law and Balance of Payments Constrained Growth: More on the Debate', *Applied Economics*, vol. 24, no. 5, pp. 493–512.

McCombie J.S. and De Ridder J.R. (1984), 'The Verdoorn Law Controversy: Some New Empirical Evidence Using U.S. State Data', *Oxford Economic Papers*, vol. 36, no. 2, pp. 208–283.

McGregor P. and Swales J.K. (1985), 'Professor Thirlwall and Balance of Payments Constrained Growth', *Applied Economics*, vol. 17, no. 1, pp. 17–32.

McGuire A. (1983), 'The Regional Income and Employment Impacts of Nuclear Power Stations', *Scottish Journal of Political Economy*, vol. 30, no. 3, pp. 264–274.

Meade J.E. (1952) 'External Economies and Diseconomies in a Competitive Situation', *Economic Journal*, vol. 62, no. 245, pp. 54–67.

Meijers E. (2013), 'Cities Borrowing Size: An Exploration of the Spread of Metropolitan Amenities across European Cities', paper presented at the Association of American Geographers annual meeting, Los Angeles, 9–13 April.

Krugman P. (1998), 'What's New about the New Economic Geography', *Oxford Review of Economic Policy*, vol. 14, no. 2, pp. 7–17.

Krugman P. and Venables A.J. (1996), 'Integration, Specialisation and Adjustment', *European Economic Review*, vol. 40, no. 3/5, pp. 959–967.

Kugler M. (2006), 'Spillover from Foreign Direct Investment: Within or Between Industries?', *Journal of Development Economics*, vol. 80, no. 2, pp. 444–474.

Lancaster K. (1975), 'Socially Optimal Product Differentiation', *American Economic Review*, vol. 65, no. 4, pp. 567–585.

Lancaster K. (1980), 'Intra-Industry Trade Under Perfect Monopolistic Competition', *Journal of International Economics*, vol. 10, no. 2, pp. 151–175.

Launhardt W. (1882), 'Die Bestimmung des Zweckmässigsten Standorts einer Gewerblischen Anlage', *Zeitschrift des Vereins Deutscher Ingenieure*, vol. 26, no. 3, pp. 106–115.

Launhardt W. (1885), *Mathematische Begründung der Volkwirtschaftslehre*, Leipzig.

Lawson C. and Lorenz E. (1999), 'Collective Learning, Tacit Knowledge and Regional Innovation Capacity', *Regional Studies*, vol. 33, no. 4, pp. 305–317.

Leon-Ledesma M.A. (1998), 'Economic Growth and Verdoorn's Laws in the Spanish Regions, 1962–1991', Discussion Paper no. 1, Department of Economics, University of Kent.

Leontief W. (1953), 'Domestic Production and Foreigner Trade: the American Capital Position Re-examined', *Proceedings of the American Philosophical Society*, vol. 97, no. 4, pp. 332–349.

Leontief W. (1956), 'Factor Proportions and the Structure of American Trade: Further Theoretical and Empirical Analysis', *The Review of Economics and Statistics*, vol. 38, no. 4, pp. 386–407.

Linder S.B. (1961), *An Essay on Trade and Transformation*, John Wiley & Sons, New York.

Lipietz A. (1980), 'The Structuration of Space, the Problem of Land, and Spatial Policy', in J. Carney, R. Hudson and J. Lewis (eds), *Regions in Crisis*, Croom Helm, London, pp. 60–75.

Long W. (1971), 'Demand in Space: Some Neglected Aspects', *Papers and Proceedings of the Regional Science Association*, vol. 27, no. 1, pp. 45–62.

Lösch A. (1954), *The Economics of Location*, Yale University Press, New Haven, Conn., orig. edn (1940), *Die Räumlische Ordnung der Wirtschaft*, Gustav Fischer, Jena.

Lucas R. (1988), 'On the Mechanics of Economic Development', *Journal of Monetary Economics*, vol. 22, no. 1, pp. 3–42.

Lundvall B.-A. (1992), 'Introduction', in B.-A. Lundvall (ed.), *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, Pinter, London, pp. 1–19.

Lundvall B.-A. and Johnson B. (1994), 'The Learning Economy', *Journal of Industry Studies*, vol. 1, no. 2, pp. 23–42.

Lung P. (1987), 'Complexity and Spatial Dynamic Modelling, from Catastrophe Theory to Self-organising Processes: A Review of the Literature', Institut d'Economie Régionale du Sud-Ouest, mimeo.

Lutz V. (1962), *Italy – A Study in Economic Development*, Oxford University Press, London.

Maggioni M. (2002), *Clustering Dynamics and the Location of High-Tech Firms*, Physica-Verlag, Berlin.

Magrini S. (1997), 'Spatial Concentration in Research and Regional Income Disparities in a Decentralised Model of Endogenous Growth', *Research Papers in Environmental and Spatial Analysis*, no. 43, London School of Economics.

Magrini S. (2004) 'Regional (Di)convergence', in V. Henderson and F.J. Thisse (eds), *Handbook of Regional and Urban Economics*, vol. 4, North-Holland, Amsterdam, pp. 2741–2796.

Maier G. and Sedlacek S. (eds) (2005), *Spillovers and Innovations – Space, Environment and the Economy*, Springer Verlag, Vienna.

Maillat D., Quévit M. and Senn L. (eds) (1993), *Réseaux d'Innovation et Milieux Innovateurs: Un Pari pour le Développement Régional*, EDES, Neuchâtel.

Malberg A. and Maskell P. (2002), 'The Elusive Concept of Localization Economies: Towards a Knowledge-based Theory of Spatial Clustering', *Environment and Planning A*, vol. 34, no. 3, pp. 429–449.

Malecki E. and Varaiya P. (1986), 'Innovation and Changes in Regional Structure', in P. Nijkamp (ed.), *Handbook of Regional and Urban Economics*, North-Holland, Amsterdam, pp. 629–645.

Mankin N., Romer D. and Weil D. (1992), 'A Contribution to the Empirics of Economic Growth', *Quarterly Journal of Economics*, vol. 107, no. 2, pp. 739–774.

Mansfield E. (1961), 'Technological Change and the Rate of Imitation', *Econometrica*, vol. 29, no. 4, pp. 741–766.

Mansfield E. (1968), *The Economics of Technological Change*, Norton, New York.

Marelli E. (1981), 'Optimal City Size, the Productivity of Cities and Urban Production Functions', *Sistemi Urbani*, vol. 3, no. 1/2, pp. 149–163.

Markusen J. and Venables A. (1999), 'Foreign Direct Investment as a Catalyst for Industrial Development', *European Economic Review*, vol. 43, no. 2, pp. 335–356.

Marshall A. (1920), *Principles of Economics*, 8th edn, Macmillan, London.

Martellato D. (1982), 'The Tuscany Interregional Input Output Model', International Institute for Applied Systems Analysis, IIASA paper series no. 82–30, Vienna.

Martin P. and Ottaviano G. (1999), 'Growing Locations: Industry Location in a Model of Endogenous Growth', *European Economic Review*, vol. 43, no. 2, pp. 281–302.

Martin R. (1999), 'The New "Geographical Turn" in Economics: Some Critical Reflections', *Cambridge Journal of Economcs*, vol. 23, no. 1, pp. 65–91.

Maskell P. and Malmberg A. (1999), 'Localised Learning and Industrial Competitiveness', *Cambridge Journal of Economics*, vol. 23, no. 2, pp. 167–185.

Massey D. and Meegan R. (1978), 'Industrial Restructuring versus the City', *Urban Studies*, vol. 15, no. 3, pp. 273–288.

McCann P. and Ortega-Argilés R. (2014), 'The Role of the Smart Specialisation Agenda in a Reformed EU Cohesion Policy', *Scienze Regionali, Italian Journal of Regional Science*, vol. 13, no. 1, pp. 15–32.

McCann Ph. (1998), *The Economics of Industrial Location: A Logistic-Cost Approach*, Springer, Heidelberg.

McCann Ph. and Rodríguez-Pose A. (2011), 'Why and When Development Policy Should Be Place-based', in *OECD Regional Outlook 2011*, Paris, pp. 203–213.

McCann Ph. and Sheppard S. (2003), 'The Rise, Fall and Rise Again of Industrial Location Theory', *Regional Studies*, vol. 37, no. 6/7, pp. 649–663.

McCombie J.S. (1982), 'How Important Is the Spatial Diffusion of Innovation in Explaining Regional Growth Disparities?', *Urban Studies*, vol. 19, no. 2, pp. 377–382.

McCombie J.S. (1983), 'Kaldor's Law in Retrospect', *Journal of Post Keynesian Economics*, vol. 5, no. 3, pp. 414–429.

McCombie J.S. (1988), 'A Synoptic View of Regional Growth and Unemployment: I – The Neoclassical Theory', *Urban Studies*, vol. 25, no. 4, pp. 267–281.

McCombie J.S. (1992), 'Thirwall's Law and Balance of Payments Constrained Growth: More on the Debate', *Applied Economics*, vol. 24, no. 5, pp. 493–512.

McCombie J.S. and De Ridder J.R. (1984), 'The Verdoorn Law Controversy: Some New Empirical Evidence Using U.S. State Data', *Oxford Economic Papers*, vol. 36, no. 2, pp. 208–283.

McGregor P. and Swales J.K. (1985), 'Professor Thirlwall and Balance of Payments Constrained Growth', *Applied Economics*, vol. 17, no. 1, pp. 17–32.

McGuire A. (1983), 'The Regional Income and Employment Impacts of Nuclear Power Stations', *Scottish Journal of Political Economy*, vol. 30, no. 3, pp. 264–274.

Meade J.E. (1952) 'External Economies and Diseconomies in a Competitive Situation', *Economic Journal*, vol. 62, no. 245, pp. 54–67.

Meijers E. (2013), 'Cities Borrowing Size: An Exploration of the Spread of Metropolitan Amenities across European Cities', paper presented at the Association of American Geographers annual meeting, Los Angeles, 9–13 April.

Meyer J.R. (1963), 'Regional Economics: A Survey', *American Economic Review*, vol. 53, no. 1, pp. 19–54.

Mills E. (1972), *Urban Economics*, Scott Foresman and Co., Glenview, IL.

Mills E. and Lav M. (1964), 'A Model of Market Areas with Free Entry', *The Journal of Political Economy*, vol. 72, no. 3, pp. 278–288.

Mirrlees J.A. (1972), 'The Optimum Town', *Swedish Journal of Economics*, vol. 74, no. 1, pp. 114–135.

Mishan E.J. (1971), 'The Postwar Literature on Externalities: An Interpretative Essay', *Journal of Economic Literature*, vol. 9, no. 1, pp. 1–28.

Miyao T. (1981), *Dynamic Analysis of the Urban Economy*, Academic Press, New York.

Miyao T. (1984), 'Dynamic Models of Urban Growth and Decay: A Survey and Extensions', paper presented at the Second World Conferences of Arts and Sciences, Rotterdam, 4–15 June.

Miyao T. (1987a), 'Dynamic Urban Models', in E. Mills (ed.), *Urban Economics; Handbook of Regional and Urban Economics*, vol. 2, North-Holland, Amsterdam, pp. 877–925.

Miyao T. (1987b), 'Urban Growth and Dynamics', in T. Miyao and Y. Kanemoto (eds), *Urban Dynamics and Urban Externalities*, Harwood Academic Publishers, Chur, Switzerland, pp. 1–41.

Miyao T. and Kanemoto Y. (eds), *Urban Dynamics and Urban Externalities*, Harwood Academic Publishers, Chur, Switzerland.

Montesano A. (1972), 'A Restatement of Beckmann's Model on the Distribution of Urban Rent and Residential Density', *Journal of Economic Theory*, vol. 4, no. 2, pp. 329–354.

Moomaw R.L., Mullen J.K. and Martin W. (1995), 'The Interregional Impact of Infrastructure Capital', *Southern Economic Journal*, vol. 61, no.3, pp. 830–845.

Moomaw R.L. and Williams M. (1991), 'Total Factor Productivity in Manufacturing: Further Evidence from the States', *Journal of Regional Science*, vol. 31, no. 1, pp. 17–34.

Morgan K. (1997), 'The Learning Region: Institutions, Innovation and Regional Renewal', *Regional Studies*, vol. 31, no. 5, pp. 491–503.

Moroney J.R. and Walker J.M. (1966), 'A Regional Test of the Heckscher–Ohlin Hypothesis', *Journal of Political Economy*, vol. 74, no. 6, pp. 573–586.

Mouleart F. and Sekia F. (2003), 'Territorial Innovation Models: A Critical Survey', *Regional Studies*, vol. 37, no. 3, pp. 289–302.

Mulligan G. (1979), 'Additional Properties of a Hierarchical City-Size Model', *Journal of Regional Science*, vol. 10, no. 1, pp. 1–42.

Mulligan G. (1984), 'Agglomeration and Central Place Theory: A Review of the Literature', *International Regional Science Review*, vol. 9, no. 1, pp. 1–42.

Munnell A.H. (1990), 'How Does Public Infrastructure Affect Regional Economic Performance?', in A.H. Munnell (ed.), *Is there a Shortfall in Public Capital Investment?* Proceedings of a Conference Sponsored by the Federal Bank of Boston, Boston.

Muth R. (1961), 'Economic Change and Rural–Urban Land Conversions', *Econometrica*, vol. 29, no. 1, pp. 1–23.

Muth R. (1968), 'Urban Residential Land and Housing Market', in H. Perloff and L. Wingo (eds), *Issues in Urban Economics*, Johns Hopkins University Press, London, pp. 285–333.

Muth R. (1969), *Cities and Housing*, University of Chicago Press, Chicago.

Myrdal G. (1957), *Economic Theory of Under-developed Regions*, Gerald Duckworth & Co., London.

Nauwelaers C., Reid A. and Desterbecq H. (1988), *Firmes Multinationales et Développement Régional*, Editore RIDER, Université Catholique de Louvain, Louvain-la-Neuve.

Neffke F., Henning M. and Boschma R. (2011), 'How Do Regions Diversify Over Time? Industry Relatedness and the Development of New Growth Paths in Regions', *Economic Geography*, vol. 87, no. 3, pp. 237–265.

Nelson R. and Winter S. (1977), 'In Search of a Useful Theory of Innovation', *Research Policy*, vol. 6, no. 1, pp. 36–76.

Nelson R.R. and Norman V.D. (1977), 'Technological Change and Factor Mix Over the Product Cycle: A Model of Dynamic Comparative Advantage', *Journal of Development Economics*, vol. 4, no. 1, pp. 3–24.

Niedercorn J. H. and Bechdolt B.V. (1969), 'An Economic Derivation of the Law of Spatial Interaction', *Journal of Regional Science*, vol. 9, no. 2, pp. 273–282.

Niedercorn J. H. and Bechdolt B.V. (1972), 'An Economic Derivation of the Law of Spatial Interaction: Further Reply and a Reformulation', *Journal of Regional Science*, vol. 12, no. 1, pp. 127–136.

Nijkamp, P. (1999) 'Environment and Regional Economics', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Edward Elgar, Cheltenham, pp. 525–538.

Nijkamp P. and Perrels A. (1994), *Sustainable Cities in Europe*, Earthscan, London.

Nijkamp P. and Poot J. (1998), 'Spatial Perspectives on New Theories of Economic Growth', *Annals of Regional Science*, vol. 32, no. 1, pp. 7–38.

Nijkamp P. and Reggiani A. (1988), 'Entropy, Spatial Interaction Models and Discrete Choice Analysis: Static and Dynamic Analogies', *European Journal of Operational Research*, vol. 36, no. 2, pp. 186–196.

Nijkamp P. and Reggiani A. (1992), *Interaction, Evolution and Chaos in Space*, Springer Verlag, Berlin.

Nijkamp P. and Reggiani A. (1993), *Non-Linear Evolution of Spatial Economic Systems*, Springer Verlag, Berlin.

Nijkamp P., Stough R. and Verhoef E. (eds) (1998), 'Endogenous Growth in a Regional Context', *Annals of Regional Science*, vol. 32, no. 1, special issue.

Nooteboom B. (2000), *Learning and Innovation in Oragnizations and Economies*, Oxford University Press, Oxford.

North D. (1955), 'Location Theory and Regional Economic Growth', *Journal of Political Economy*, vol. 63, no. 3, pp. 243–258.

North D. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge University Press, Cambridge.

Norton D. and Rees J. (1979), 'The Product Cycle and the Spatial Decentralisation of American Manufacturing', *Regional Studies*, vol. 13, no. 2, pp. 141–151.

Nurkse R. (1952), 'Some International Aspects of the Problem of Economic Development', *The American Economic Review*, vol. 42, no. 2, pp. 571–583.

Oakey R., Nash P. and Thwaites A. (1980), 'The Regional Distribution of Innovative Manufacturing Establishments in Britain', *Regional Studies*, vol. 13, no. 3, pp. 141–151.

OECD (2001), *OECD Territorial Outlook*, OECD, Paris.

Ohlin B. (1933), *Interregional and International Trade*, Harvard University Press, Cambridge, MA.

Okun B. and Richardson R. (eds) (1961), *Studies in Economic Development*, Holt, Rinehart & Winston, New York.

Onida F. (1984), *Economia degli Scambi Internazionali*, Il Mulino, Bologna.

Ottaviano G. and Puga D. (1998), 'Agglomeration in the Global Economy: A Survey of the New Economic Geography', *The World Economy*, vol. 21, no. 6, pp. 707–731.

Ottaviano G. and Thisse J.-F. (2001), 'On Economic Geography in Economic Theory: Increasing Returns and Pecuniary Externalities', *Journal of Economic Geography*, vol. 1, no. 2, pp. 153–179.

Paci M. (1973), *Mercato del Lavoro e Classi Sociali in Italia*, Il Mulino, Bologna.

Paci R. and Saddi S. (2002), 'Capitale Pubblico e Produttività nelle Regioni Italiane', *Scienze Regionali*, vol. 2, no. 3, pp. 5–26.

Paci R. and Usai S. (2000), 'Technological Enclaves and Industrial Districts: An Analysis of the Regional Distribution of Innovative Activity in Europe', *Regional Studies*, vol. 34, no. 2, pp. 97–114.

Paelink J. (1965), 'La Théorie du Développement Régional Polarisé', Cahiers de l'ISEA, Série Economie Régionale, no. 3, pp. 5–47.

Paelink J. and Nijkamp P. (1976), *Operational Theory and Method in Regional Economics*, Saxon House, Westmead.

Palander T. (1935), *Beitrage zur Standortstheorie*, Almqvist & Wiksells Boktryckeri, Uppsala.

Parr J. (1978), 'Models of the Central Place System: A More General Approach', *Urban Studies*, vol. 15, no. 1, pp. 35–49.

Parr J. (1981), 'Temporal Change in a Central-Place System', *Environment and Planning A*, vol. 13, no. 1, pp. 97–118.

Parr J. (1985), 'A Note on the Size Distribution of Cities over Time', *Journal of Urban Economics*, vol. 18, no. 2, pp. 199–212.

Parr J. (1999a), 'Growth Pole Strategies in Regional Economic Planning: A Retrospective View. Part 1. Origins and Advocacy', *Urban Studies*, vol. 36, no. 7, pp. 1195–1216.

Parr J. (1999b), 'Growth Pole Strategies in Regional Economic Planning: A Retrospective View. Part 2. Implementation and Outcome', *Urban Studies*, vol. 36, no. 8, pp. 1247–1268.

Parr J.B. (2002), 'Agglomeration Economies: Ambiguities and Confusions', *Environment and Planning A*, vol. 34, no. 4, pp. 717–731.

Perloff H. (1957), 'Interrelations of State Income and Industrial Structure', *The Review of Economics and Statistics*, vol. 39, no. 2, pp. 162–171.

Perloff H., Dunn E., Lampard E. and Muth R. (1960), *Region, Resources and Economic Growth*, John Hopkins University Press, Baltimore, MD.

Perroux F. (1950), 'Economic Space: Theory and Applications', *The Quarterly Journal of Economics*, vol. 64, no. 1, pp. 89–104.

Perroux F. (1955), 'Note sur la Notion de Pôle de Croissance', *Economie Appliquée*, vol. 7, no. 1/2, pp. 307–320.

Perucca G. (2013), 'A Redefinition of Italian Macro-areas: The Role of Territorial Capital', *Rivista di Economia e Statistica del Territorio*, no. 2, pp. 35–63.

Perucca G. (2014), 'The Role of Territorial Capital in Local Economic Growth: Evidence from Italy', *European Planning Studies*, vol. 22, no. 3, pp. 537–562.

Petrakos G. (2000), 'The Spatial Impact of East–West Integration in Europe', in G. Petrakos, G. Maier and G. Gorzelak (eds), *Integration and Transition in Europe*, Routledge, London, pp. 38–68.

Pietrobelli C. (1998), 'The Socio-economic Foundations of Competitiveness: An Econometric Analysis of Italian Industrial Districts', *Industry and Innovation*, vol. 5, no. 2, pp. 139–155.

Piore M. and Sabel C.F. (1984), *The Second Industrial Divide: Possibilities for Prosperities*, Basic Books, New York.

Pompili T. (1992), 'The Role of Human Capital in Urban System Structure and Development: The Case of Italy', *Urban Studies*, vol. 29, no. 6, pp. 905–934.

Pontikakis D., Chorafakis G. and Kyriakou D. (2009), 'R&D Specialization in Europe: From Stylized Observations to Evidence–Based Policy', in D. Pontikakis, D. Kyriakou and R. van Bavel (eds), *The Question of R&D Specialisation*, JRC, European Commission, Directoral General for Research, Brussels, pp. 71–84.

Poratt M. (1977), *The Information Economy: Definition and Measurement*, special publication 77–12(1), Office of Telecommunications, US Department of Commerce, Washington, DC.

Posner M.V. (1961), 'International Trade and Technical Change', *Oxford Economic Papers*, vol. 13, no. 3, pp. 323–359.

Pratt R. (1968), 'An Appraisal of Minimum Requirement Technique', *Economic Geography*, vol. 44, no. 2, pp. 117–124.

Puga, D. (2010), 'The Magnitude and Causes of Agglomeration Economies', *Journal of Regional Science*, vol. 50, no. 1, pp. 1203–1219.

Pumain D. and Saint-Julien T. (eds) (1996), *Urban Networks in Europe*. John Libbey Eurotext, Paris.

Pusterla F. and Resmini L. (2007), 'Where Do Foreign Firms Locate in Transition Countries? An Empirical Investigation', *Annals of Regional Science*, vol. 41, no. 4, pp. 835–856.

Putnam R.D. (1993), *Making Democracy Work*, Princeton University Press, Princeton, NJ.

Quévit M. (1992), 'The Regional Impact of the Internal Market: A Comparative Analysis of Traditional Industrial Regions and Lagging Regions', *Regional Studies*, vol. 26, no. 4, pp. 349–360.

Rabellotti R. (1997), *External Economies and Cooperation in Industrial Districts. A Comparison of Italy and Mexico*, Macmillan, London.

Rabenau B.V. (1979), 'Urban Growth with Agglomeration Economies and Diseconomies', *Geographia Polonica*, vol. 42, pp. 77–90.

Rallet A. (2002), 'L'économie de Proximités: Propos d'étape', *Etudes et Recherches*, no. 33, 11–26.

Rallet A. and Torre A. (1998), 'On Geography and Technology: Proximity Relations in Localised Innovation Networks', in M. Steiner (ed.), *From Agglomeration Economies to Innovative Clusters*, Pion Editor, London, pp. 41–56.

Ramsey F. (1928), 'A Mathematical Theory of Saving', *Economic Journal*, vol. 38, no. 152, pp. 543–559.

Ratti R., Bramanti A. and Gordon R. (eds) (1997), *The Dynamics of Innovative Regions*, Ashgate, Aldershot.

Rebelo S. (1991), 'Long-run Policy Analysis and Long-run Growth', *Journal of Political Economy*, vol. 99, no. 3, pp. 500–521.

Reggiani A. (2000), 'Introduction: New Frontiers in Modelling Spatial and Economic Systems', in A. Reggiani (ed.), *Spatial Economic Science*, Springer Verlag, Berlin, pp. 1–11.

Reilly W.J. (1931), *The Law of Retail Gravitation*, Knickerbocker Press, New York.

RERU (1999), 'Le Paradigme du Milieu Innovateur dans l'Economie Contemporaine', *Revue d'Economie Régionale et Urbaine*, no. 3, special issue.

Resmini L. (2007), 'Regional Patterns of Industry Location in Transition Countries: Does the Economic Integration with the EU Matter?', *Regional Studies*, vol. 41, no. 6, pp. 747–764.

Ricardo D. (1971), *Principles of Political Taxonomy and Taxation*, Penguin Books, Hardmondsworth; orig. edn 1817.

Richardson H.W. (1969), *Regional Economics*, World University, Redwood Press, Trowbridge.

Richardson H.W. (1972), 'Optimality in City Size, Systems of Cities and Urban Policy: A Sceptic's View', *Urban Studies*, vol. 9, no. 1, pp. 29–47.

Richardson H.W. (1973a), *Regional Growth Theory*, Macmillan, London.

Richardson H.W. (1973b), *The Economics of Urban Size*, Saxon House, Farnborough.

Richardson H.W. (1977), *The New Urban Economcs: and Alternatives*, Pion, London.

Richardson H.W. (1978), *Regional and Urban Economics*, Penguin Books, Harmondsworth.

Rid H. and Lau E. (1998), 'Verdoorn's Law and Increasing Returns to Scale in the UK Regions, 1968–91: Some New Estimates Based on the Cointegration Approach', *Oxford Economic Papers*, vol. 50, no. 2, pp. 201–219.

Rietveld P. (1989), 'Infrastructure and Regional Development – A Survey of Multiregional Economic Models', *The Annals of Regional Science*, vol. 23, no. 6, pp. 255–274.

Robinson J. (1934), *The Economics of Imperfect Competition*, Macmillan, London.

Rodriguez-Clare A. (1996), 'Multinationals, Linkages, and Economic Development', *The American Economic Review*, vol. 86, no. 4, pp. 852–873.

Rodriguez-Pose A. and Fratesi U. (2004), 'Between Development and Social Policies: The Impact of European Structural Funds in Objective 1 Regions', *Regional Studies*, vol. 38, no. 1, pp. 97–113.

Romer P. (1986), 'Increasing Returns and Long-Run Growth', *Journal of Political Economy*, vol. 94, no. 5, pp. 1002–1037.
Romer P. (1987), 'Growth Based on Increasing Returns due to Specialisation', *American Economic Review, Papers and Proceedings*, vol. 77, no. 2, pp. 56–67.
Romer P. (1989), 'Capital Accumulation in the Theory of Long-Run Growth', in R. Barro (ed.), *Modern Business Cycle Theory*, Oxford, Basil Blackwell, pp. 51–127.
Romer P. (1990), 'Endogenous Technological Change', *Journal of Political Economy*, vol. 98, no. 5, pp. S71–S102.
Romer P. (1994), 'The Origins of Endogenous Growth', *Journal of Economic Perspective*, vol. 8, no. 1, pp. 3–22.
Rosenthal S.S. and Strange W.C. (2003), 'Geography, Industrial Organization, and Agglomeration', *Review of Economics and Statistics*, vol. 85, no. 2, pp. 377–393.
Rosenthal S.S. and Strange W.C. (2001), 'The Determinants of Agglomeration', *Journal of Urban Economics*, vol. 50, no. 3, pp. 191–229.
Rosestein-Rodan P.N. (1943), 'Problems of Industrialisation of Eastern and South-Eastern Europe', *The Economic Journal*, vol. 53, no. 210/211, pp. 202–211.
Rosestein-Rodan P.N. (1959), 'Due Lezioni sui Problemi di Sviluppo', *L'Industria*, no. 4, pp. 422–442.
Rostow W.W. (1960), *The Stages of Economic Growth*, Cambridge University Press, Cambridge.
Rowthorn R.E. (1975), 'What Remains of Kaldor's Law?', *The Economic Journal*, vol. 85, no. 340, pp. 10–19.
Sala-i-Martin X. (1996), 'Regional Cohesion: Evidence and Theories of Regional Growth and Convergence', *European Economic Review*, vol. 40, no. 6, pp. 1325–1352.
Samuelson P. (1954), 'The Transfer Problem and Transport Costs. Part II: Analysis of the Effects of Trade Impediments', *Economic Journal*, vol. 64, no. 254, pp. 264–289.
Samuelson P. (1983), 'Thünen at Two Hundred', *Journal of Economic Literature*, vol. 21, no. 4, pp. 1468–1488.
Sapir A. (2003), *An Agenda for a Growing Europe*, The Sapir Report to the EU, Brussels, July.
Sassen S. (ed.) (2002), *Global Networks, Linked Cities*, Routledge, New York.
Schmitz H. (1995), 'Collective Efficiency: Growth Path for for Small-scale Industry', *Journal of Development Studies*, vol. 31, no. 4, pp. 529–566.
Schmitz H. (1998), 'Collective Efficiency and Returns to Scale', *Cambridge Journal of Economics*, vol. 23, no. 4, pp. 465–483.
Schmitz H. and Musyck B. (1994), 'Industrial District in Europe: Policy Lessons for Developing Countries?', *World Development*, vol. 22, no. 6, pp. 889–910.
Scitovski T. (1954), 'Two Concepts of External Economies', *Journal of Political Economy*, vol. 62, no. 2, pp. 143–151.
Scitovski T. (1958), *Economic Theory and Western Europe Integration*, Allen & Unwin, London.
Secchi B. (1974), *Squilibri Regionali e Sviluppo Economico*, Marsilio, Padua.
Segal D. (1976), 'Are there Returns to Scale in City Size?', *Review of Economics and Statistics*, vol. 58, no. 3, pp. 339–350.
Segal D. (1977), *Urban Economics*, Richard Irwin, Homewood, Ill.
Shefer D. (1973), 'Localization Economies in SMSA'S: A Production Function Analysis', *Journal of Regional Science*, vol. 13, no. 1, pp. 55–64.
Signorini L.F. (2000), *Lo Sviluppo Locale: Un'Indagine della Banca d'Italia sui Distretti Industriali*. Meridiana Libri Publisher, Corigliano Calabro.
Sirkin G. (1959), 'The Theory of Regional Economic Base', *The Review of Economics and Statistics*, vol. 41, no. 4, pp. 426–429.
Smith D.M. (1966), 'A Theoretical Framework for Geographical Studies of Industrial Location', *Economic Geography*, vol. 42, no. 2, pp. 95–113.

Smith D.M. (1971), *Industrial Location: An Economic Geographical Analysis*, Wiley & Sons, London.

Smith D.M. (1974), 'Regional Growth, Interstate and Intersectoral Factor Reallocation', *Review of Economics and Statistics*, vol. 61, pp. 353–359.

Smith D.M. (1975), 'Neoclassical Growth Models and Regional Growth in the U.S.', *Journal of Regional Science*, vol. 15, no. 2, pp. 165–181.

Solow R. (1957), 'Technical Change and the Aggregate Production Function', *Review of Economics and Statistics*, vol. 39, no. 3, pp. 312–320.

Solow R. (1972), 'Congestion, Density and the Use of Land in Transportation', *Swedish Journal of Economics*, vol. 74, no. 1, pp. 161–173.

Solow R. (2000), *Growth Theory: An Exposition*, 2nd edn, Oxford University Press, Oxford.

Soo K.T. (2005), 'Zipf's Law for Cities: A Cross-Country Investigation', *Regional Science and Urban Economics*, vol. 35, no. 3, pp. 239–263.

Soro B. (2003), 'Fattori che Regolano lo Sviluppo della Produttività del Lavoro. Fifty Years On', in J. McCombie, M. Pugno and B. Soro (eds), *Productivity Growth and Economic Performance: Essays on Verdoorn's Law*, Palgrave Macmillan, Basingstoke, chap. 3, pp. 37–63.

Spaventa L. (1959), 'Dualism in Economic Growth', *Banca Nazionale del Lavoro Quarterly Review*, no. 51, pp. 386–434.

Stabler J.C. (1968), 'Exports and Evolution: The Process of Regional Change', *Land Economics*, vol. 44, no. 1, pp. 11–23; reprinted in D. McKee, R. Dean and W. Leahy (eds) (1970), *Regional Economics: Theory and Practice,* The Free Press, New York, pp. 49–64.

Steele D.B. (1969), 'Regional Multipliers in Great Britain', *Oxford Economic Papers,* vol. 21, no. 2, pp. 268–292.

Sternberg R. (1996), 'Reasons for the Genesis of High-tech Regions – Theoretical Explanation and Empirical Evidence', *Geoforum*, vol. 27, no. 2, pp. 205–223.

Stimson R.J., Stough R.R. and Salazar M. (2005), 'Leadership and Institutional Factors in Endogenous Regional Economic Development', *Investigationes Regionales*, no. 7, pp. 23–52.

Stöhr W. (1990), 'On the Theory and Practice of Local Development in Europe', in W. Stöhr (ed.), *Global Challenge and Local Responses*, Mansell, London, pp. 35–54.

Stöhr W. and Tödtling F. (1977), 'Spatial Equity. Some Anti-theses to Current Regional Development Doctrine', *Papers of the Regional Science Association*, vol. 38, no. 1, pp. 33–53.

Stoneman P. (1986), 'Technological Diffusion: The View Point of Economic Theory', *Ricerche Economiche*, vol. 40, no. 4, pp. 585–606.

Storper M. (1995), 'La Géographie des Conventions: Proximité Territoriale, Intérdependance Non-marchandes et Développement Economique', in A. Rallet and A. Torre (eds), *Economie Industrielle et Economie Spatiale*, Economica, Paris, pp. 111–128.

Stouffer S.A. (1940), 'Intervening Opportunities: A Theory Relating Mobility and Distance', *American Sociological Review*, vol. 5, no. 6, pp. 845–867.

Stouffer S.A. (1960), 'Intervening Opportunities and Competing Migrants', *Journal of Regional Science*, vol. 12, no. 2, pp. 26–35.

Subirats J. (ed.) (2002), *Redes, Territorios y Gobierno*, UIMP, Barcelona.

Sweikauskas L. (1975), 'The Productivity of City Size', *Quarterly Journal of Economics*, vol. 89, no. 3, pp. 393–413.

Taylor P.S. (2001), 'Specification of the World City Network', *Geographical Analysis*, vol. 33, no. 2, pp. 181–194.

Terrasi M. (2002), 'National and Spatial Factors in EU Regional Convergence', in J.R. Cuadrado-Roura and M. Parellada (eds), *Regional Convergence in the European Union*, Springer Verlag, Berlin, pp. 185–209.

Thirlwall A.P. (1974), 'Regional Economic Disparities and Regional Policy in the Common Market', *Urban Studies*, vol. 11, no. 1, pp. 1–12.

Thirlwall A.P. (1980), 'Regional Problems are Balance of Payments Problems', *Regional Studies*, vol. 14, no. 5, pp. 419–425.

Weber A. (1929), *Alfred Weber's Theory of the Location of Industries*, University of Chicago Press, Chicago; orig. edn (1909) *Über der Standort der Industrien*, Verlag Mohr, Tübingen.

Weimer A. and Hoyt H. (1939), *Principles of Urban Real Estate*, Ronald Press, New York.

Weiss S. and Gooding E. (1968), 'Estimation of Differential Employment Multipliers in a Small Region', *Land Economics*, vol. 44, no. 2, pp. 235–244.

Wheaton W. (1979), 'Monocentric Models of Urban Land Use; Contributions and Criticisms', in P. Mieszkowski and M. Straszheim (eds), *Current Issues in Urban Economics*, Johns Hopkins University Press, London, pp. 105–129.

Williamson J.G. (1965), 'Regional Inequality and the Process of National Development: A Description of the Patterns', *Economic Development and Cultural Change*, vol. 13, no. 4, pp. 3–45.

Williamson O. (1975), *Markets and Hierarchies: Analysis and Antitrust Implications*, The Free Press, New York.

Williamson O. (2002), 'The Lens of Contract: Private Ordering', *American Economic Review, Papers and Proceedings*, vol. 92, no. 2, pp. 438–453.

Wilson A. (1970), *Entropy in Urban and Regional Modelling*, Pion, London.

Wilson A. (1971), 'A Family of Spatial Interaction Models and Associated Developments', *Environment and Planning*, vol. 14, no. 3, pp. 1–32.

Wilson A. (1981), *Catastrophe Theory and Bifurcation*, Croom Helm, London.

Wilson T. (1968), 'The Regional Multiplier – A Critique', *Oxford Economic Papers*, vol. 20, no. 3, pp. 374–393.

Wingo L. (1961), *Transportation and Urban Land*, Resources for the Future, Washington, DC.

World Bank (2009), *World Development Report*, World Bank, Washington, DC.

Yilmaz S. and Dinc M. (2002), 'Telecommunications and Regional Development: Evidence from the U.S. States', *Economic Development Quarterly*, vol. 16, no. 3, pp. 211–228.

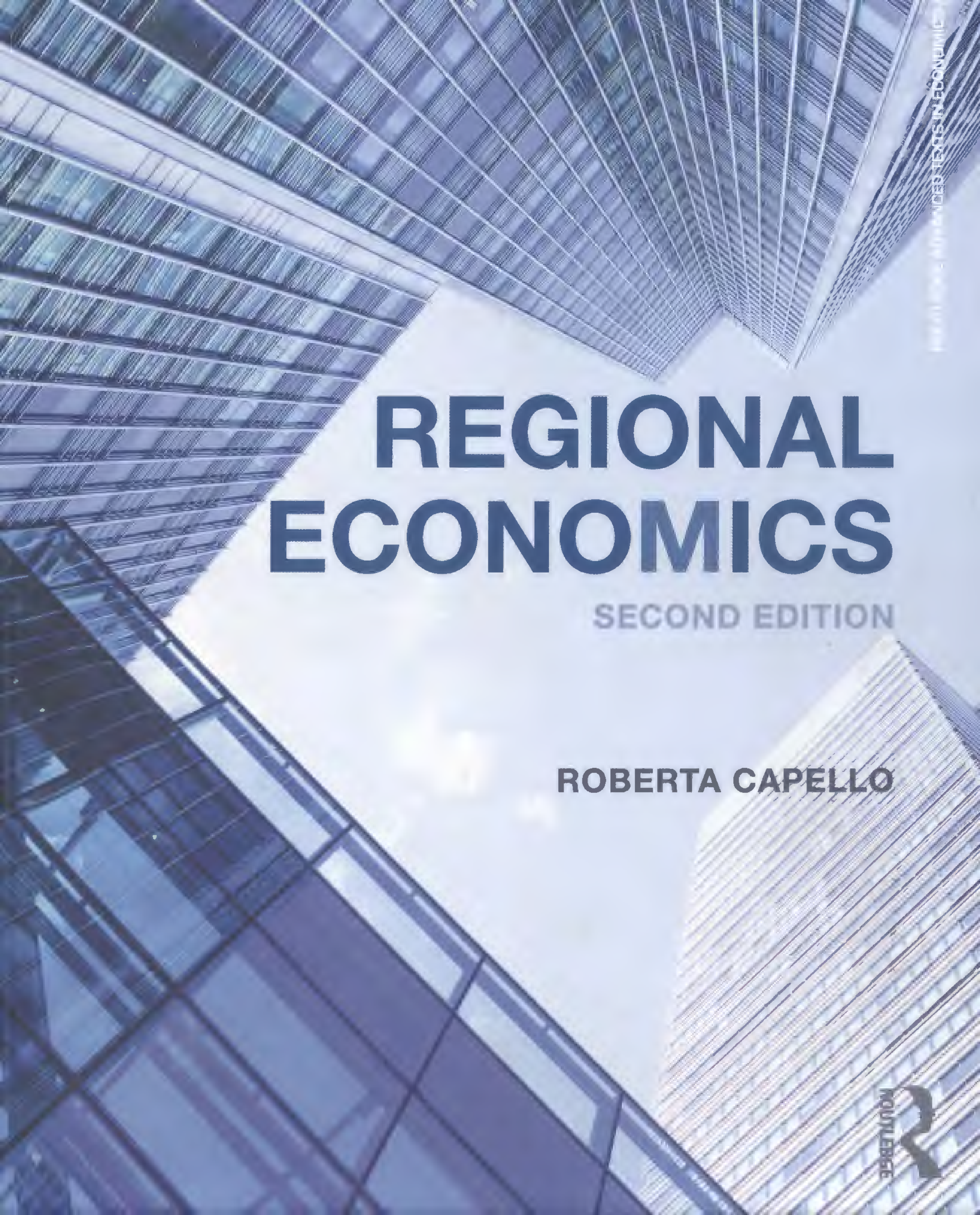
Young A. (1928), 'Increasing Returns and Economic Progress', *The Economic Journal*, vol. 38, no. 152, pp. 527–542.

Young S., Hood N. and Dunlop S. (1988), 'Global Strategies, Multinational Subsidiary, Roles and Economic Impact in Scotland', *Regional Studies*, vol. 22, no. 6, pp. 487–497.

Zipf G.K. (1949), *Human Behaviour and the Principle of Least Effort*, Addison Wesley, Cambridge, MA.

Thirlwall A.P. (1983), 'A Plain Man's Guide to Kaldor's Law', *Journal of Post Keynesian Economics*, vol. 5, no. 3, pp. 345–358.
Thom R. (1972), *Stabilité Structurelle et Morphogenèse: Essai d'une Théorie Générale des Modèles*, Inter Editions, Paris.
Tiebout C. (1956), 'The Urban Economic Base Reconsidered', *Land Economics*, vol. 32, no. 1, pp. 95–99.
Tiebout C. (1957), 'Regional and Interregional Input–Output Models: An Appraisal', *Southern Economic Journal*, vol. 24, no. 4, pp. 140–147.
Tiebout C. (1960), 'The Community Income Multiplier: A Case Study', in R. Pfouts (ed.), *The Techniques of Urban Economic Analysis*, Chandler-Davis, London.
Tödtling F. and Trippl M. (2005), 'One Size Fits All? Towards a Differentiated Regional Innovation Policy Approach', *Research Policy*, vol. 34, no. 8, pp. 1203–1219.
Torre A. and Rallet A. (2005), 'Proximity and Localization', *Regional Studies*, vol. 39, no. 1, pp. 47–59.
Torre A. and Wallet F. (eds) (2014), *Regional Development and Proximity Relations*, Edward Elgar, Cheltenham.
Torrens R. (1815), *An Essay on the External Corn Trade*, J. Hatchard, London.
Tóth B.I. (2014), 'Territorial Capital: Theory, Empirics and Critical Remarks', *European Planning Studies*, DOI: 10.1080/09654313.2014.928675.
Traistaru I., Nijkamp P. and Resmini L. (eds) (2003), *The Emerging Economic Geography in EU Accession Countries*, Ashgate, Aldershot.
Trigilia C. (1985), 'La Regolazione Localistica: Economia e Politica nelle Aree di Piccola Impresa', *Stato e Mercato*, no. 14, pp. 181–228.
Trippl M. (2010), 'Developing Cross-Border Regional Innovation Systems: Key Factors and Challenges', *Tijdschrift voor Economische en Sociale Geografie*, vol. 101, no. 2, pp. 150–160.
Ullman E. (1941), 'A Theory of Location for Cities', *The American Journal of Sociology*, vol. 46, no. 6, pp. 853–864.
Ullman E. and Dacey M. (1960), 'The Minimum Requirements Approach to the Urban Economic Base', *Papers and Proceedings of the Regional Science Association*, vol. 6, pp. 175–194.
Uzawa H. (1964), 'Optimal Growth in a Two Sector Model of Capital Accumulation', *Review of Economic Studies*, vol. 31, no. 1, pp. 1–24.
Valavanis S. (1955), 'Lösch on Location', *American Economic Review*, vol. 45, no. 4, pp. 637–644.
Vásquez-Barquero A. (2002), *Endogenous Development*, Routledge, London.
Venables A.J. (1996), 'Equilibrium Location of Vertically Linked Industries,' *International Economic Review*, vol. 37, no. 2, pp. 341–359.
Verdoorn P.J. (1949), 'Factors that Determine the Growth of Labour Productivity', in D. Ironmonger, J.O.N. Perkins and T. Van Hoa (eds) (1988), *National Income and Economic Progress; Essays in Honour of Colin Clark*, Macmillan, Basingstoke, pp. 199–207.
Vernon R. (1957), 'Production and Distribution in the Large Metropolis', *The Annals of the American Academy of Political and Social Sciences*, vol. 314, no. 1, pp. 15–29.
Vernon R. (1960), *Metropolis 1985*, Harvard University Press, Cambridge, MA.
Vernon R. (1966), 'International Investment and International Trade in the Product Cycle', *Quarterly Journal of Economics*, vol. 80, no. 2, pp. 190–207.
Vickerman R. (1991), 'Infrastructure and Regional Development: Introduction', in R. Vickerman (ed.), *Infrastructure and Regional Development*, Pion, London, no. 1, pp. 1–8.
von Böventer E. (1975), 'Regional Growth Theory', *Urban Studies*, vol. 12, pp. 1–29.
von Thünen J.H. (1826), *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Puthes, Hamburg.

Index



REGIONAL ECONOMICS

SECOND EDITION

ROBERTA CAPELLO