Growth, Location, and Cities

A Thesis presented

by

Rafael González Val

Licenciado en Economía, 2005, Universidad de Zaragoza

DEA, 2007, Universidad de Zaragoza

directed by

Luis Fernando Lanaspa Santolaria

Profesor Titular de la Universidad de Zaragoza

and

Fernando Pueyo Baldellou

Profesor Titular de la Universidad de Zaragoza

submitted to

Departamento de Análisis Económico

Universidad de Zaragoza

February, 2010

Abstract

This doctoral thesis centres on the study of the economic concentration of activity and growth. First, we propose two theoretical models which enable us to analyse the specific influence of two factors on the distribution and growth of economic activity: the different trade policies which countries may impose, and the presence of a natural resource. Next, we carry out an empirical analysis of the evolution of growth rates in the cities of three countries: the USA, Spain and Italy. The novelty of this analysis is the use of data for all of the 20th century and all cities, with no size restriction. This thesis concludes with a study of the determinants of American urban growth, understood as the growth of either the population or the per capita income, from 1990 to 2000.

Resumen

Esta tesis doctoral se centra en el estudio de la concentración económica de la actividad y el crecimiento. En primer lugar, planteamos dos modelos teóricos que nos permiten analizar la influencia específica de dos factores tanto en la distribución de la actividad económica como en el crecimiento: las diferentes políticas comerciales que pueden llevar a cabo los países, y la presencia de algún recurso natural. A continuación, realizamos un análisis empírico de la evolución de las tasas de crecimiento de las ciudades de tres países, Estados Unidos, España e Italia. La novedad de este análisis es el empleo de datos para todo el siglo veinte de todas las ciudades sin ninguna restricción de tamaño. Por último, esta tesis concluye con un estudio de los determinantes del crecimiento urbano americano, entendido como crecimiento en población o en renta per cápita, entre los años 1990 y 2000.

Acknowledgements

This document contains a large part of my work of the last four years, and brings to a close a period of my life which has been enriching and intellectually exciting. I would like to take this opportunity to acknowledge and give thanks for the support of the people and institutions contributing, in one way or another, to this thesis.

First, none of this would have been possible without the direction of Luis Fernando Lanaspa Santolaria and Fernando Pueyo Baldellou. This was the first time they had undertaken such a task, and their work was impeccable. Fernando Sanz Gracia was also involved in this project from the beginning. I would like to thank him for his advice and the trust placed in me. I hope that all of them will be as satisfied with the final result as I am.

The members of the ADETRE research group from the Universidad de Zaragoza also contributed, to greater or lesser degrees. All of them constitute an extraordinary group of people. I would especially like to acknowledge the support and comments of the members with whom I had the opportunity to work throughout this period.

Thanks must also be given for the financial support provided by the Spanish Ministerio de Educación y Ciencia, in the form of a research grant from the

Programa Nacional de Formación de Profesorado Universitario, for this doctoral thesis.

Finally, my family can be as proud of me as I am of them. Thank you for everything. And, to my partner, the most important person in my life during the last three years, thank you for so much unconditional love and support. Thank you for always being there for me, and for going with me to the end of the world.

Contents

List of Tables

List of Figures

Introduction

At first glance, it seems obvious that the spatial distribution of economic activity is not homogenous. Thus, in developed countries, it is usual to find firms and their customers concentrated in the more populous areas. Economic Geography studies the spatial location of economic activity. However, this question is not new, and its origin can be found in the works of Von Thünen (1826), and Christaller (1933) and Lösch (1940).

Nevertheless, in both cases the models have little economic content, as they are purely geometrical characterisations in which the location of activity is an exogenous variable. According to Von Thünen's theory of concentric rings, the goods that generate the greatest profits and with the highest transport costs are produced in the places closest to the city, while according to Christaller and Lösch's theory of the central place, there is a trade-off between scale economies and transport costs, so that activity is eventually distributed in a honeycomb pattern of hexagons.

The New Economic Geography appeared in the 1990s, beginning with the pioneering work of Krugman (1991) (a recent Nobel laureate), with models with a high economic content, where location is an endogenous variable; such models acquired great importance thanks to their ability to explain regional integration processes (the paradigmatic example in the 20th century is the European Union).

In a classic framework, the location of activity would be uniform, given that markets are competitive, there is perfect information, demand is distributed uniformly in space, etc. Therefore, the assumptions of the New Economic Geography are different: mobile factors, the existence of transport costs and centrifugal or centripetal forces (centripetal forces favour the agglomeration of activity, such as increasing returns, whereas centrifugal forces favour dispersion, such as congestion costs), the presence of Marshallian external economies, and the importance of expectations and of the small initial advantages, which can eventually produce a global advantage (economics of qwerty).

The key to developing these theoretical models incorporating increasing returns to scale and transport costs, according to the founding fathers of the New Economic Geography, Fujita et al. (1999), is the use of a few modelling tricks: monopolistic competition à la Dixit and Stiglitz and iceberg transport costs as in Samuelson (1954). Although even under these specific assumptions, the models of the New Economic Geography are still complex, and often difficult to resolve analytically.

These models are usually intrinsically dynamic, as they study the evolution over time of the location of workers and firms until the equilibrium is reached. Nevertheless, until a few years ago, the question of the relationship between industrial concentration and economic growth was not dealt with specifically, despite the obviousness, intuitively, of the huge effect of the concentration of activity on growth. For this reason there are not many earlier references in the literature that include both questions.

As models integrating Economic Geography and economic growth, we can cite Martin and Ottaviano (1999), Baldwin and Forslid (2000), and Baldwin, et al. (2001). Notable models of Economic Geography parallel to those of growth are that of Palivos and Wang (1996), which present a model with a scale effect in the population size in which growth is due to an externality in total capital which makes technology linear in such capital (AK model), or that of Ioannides (1994).

From these few models, the most successful has been that of Martin and Ottaviano (1999), which has led to several extensions: Martin (1999), Martin and Ottaviano (2001), and Hirose and Yamamoto (2007). This is so because it presents a high explanatory power, despite its relative analytical simplicity. For this reason, it constitutes our theoretical framework of reference. Figure 0.1 represents schematically how the model works, although its foundations are fully explained in later chapters.

Martin and Ottaviano's model (1999) combines a framework of endogenous growth, similar to that of Romer (1990), and Grossman and Helpman (1991), with a geographical framework like that of Helpman and Krugman (1985), and Krugman (1991). This model, which only analysed the role of common international infrastructures, was extended by Martin (1999) to incorporate domestic infrastructures. Later, in Martin and Ottaviano (2001), the same authors considered the possibility that differentiated goods could be used as input in the research activity which gives rise to economic growth. Finally, Hirose and Yamamoto (2007) studied the behaviour of the model in the case of a certain degree of knowledge dissemination at the international level.

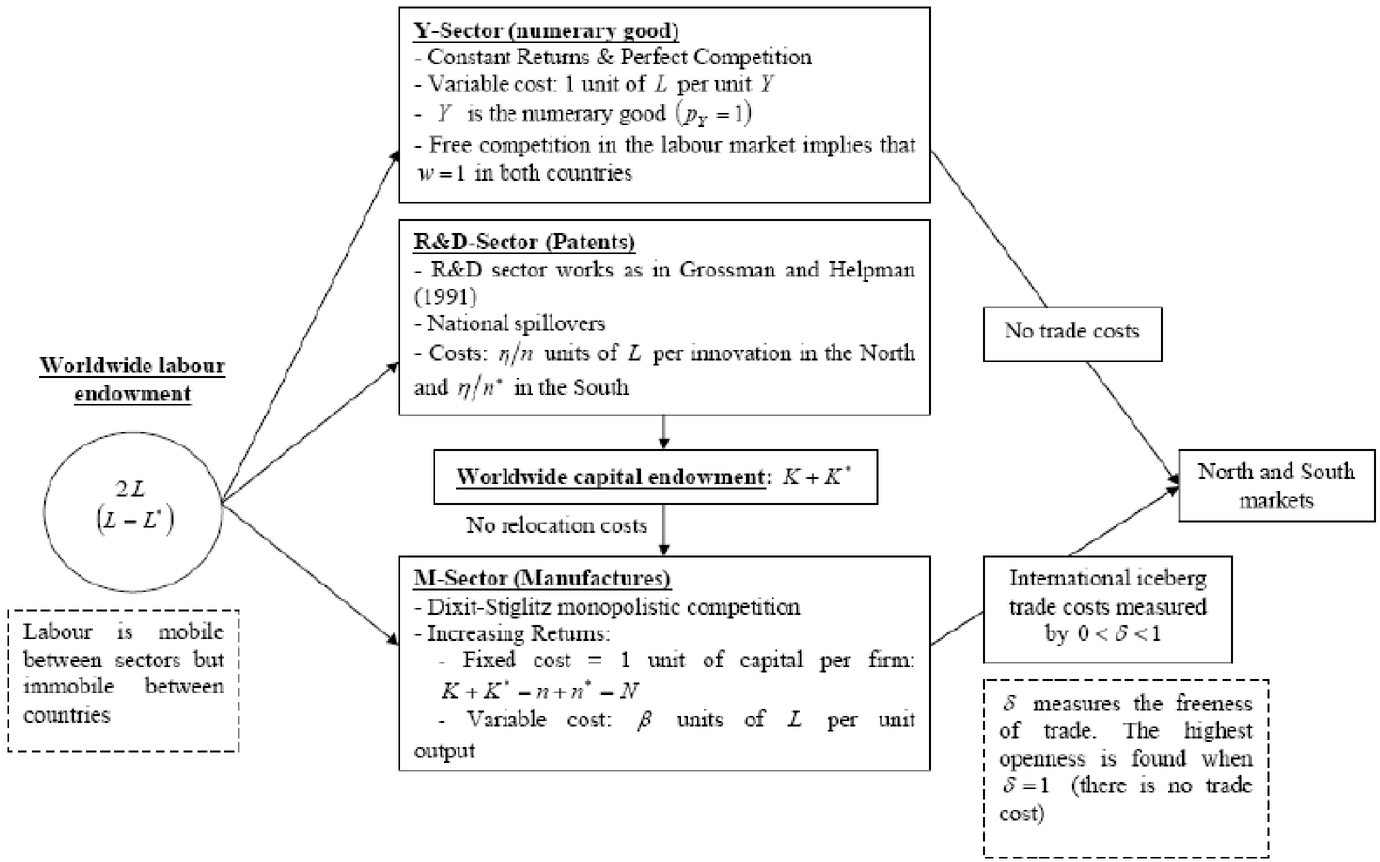


Fig. 0.1 Schematic diagram of Martin and Ottaviano's model (1999).

In the first two chapters, making up the first section of this thesis, we introduce two modifications to the initial model of Martin and Ottaviano (1999), enabling us to analyse from a theoretical point of view the specific influence of two factors in the concentration of economic activity and economic growth: the different trade policies which may be imposed by countries, and the presence of a natural resource. Thus, our spatial units for analysis are countries or regions.

While the first section is completely theoretical, the second is entirely empirical, focusing on the city as an economic unit and moving to the field of Urban Economics, but without abandoning our interest in the relationship between geography and economic growth. The evolution of urban centres presents a complex dynamic, so that their movements over time may or may not follow certain patterns of behaviour. Concretely, we are interested in researching whether urban growth depends on the initial size of the city, the empirical regularity known as Gibrat's law. Therefore, the aim of the third chapter is to test empirically whether Gibrat's law holds, using data for all of the twentieth century and all the cities, with no size restriction, of three countries: the USA, Spain and Italy.

Finally, the fourth chapter presents an analysis of the determinants of urban growth in the USA, understood as the growth of either the population or the per capita income, from 1990 to 2000. The starting point for this chapter is the idea that cities have a double nature, being on one hand population centres and on the other drivers of economic growth, and that the different externalities generated in cities can potentially have different effects on population growth and per capita income growth. To explain these possible differentiated behaviours, we examine the relationship between the urban characteristics in 1990 and the city growth (both in population and in per capita income) using a Multinomial Logit Model. The geographical situation of cities appears to play a key role in their growth.

References

Baldwin, R., and R. Forslid, (2000). The core-periphery model and endogenous growth: stabilizing and destabilizing integration. Economica 67, 307-324.

Baldwin, R., P. Martin and G.I.P. Ottaviano, (2001). Global income divergence, trade, and industrialization: the geography of growth take-offs. Journal of Economic Growth 6, 5-37.

Christaller, W., (1933). Die zentralen Orte in Suddeutschland. Gustav Fischer, Jena.

Fujita, M., P. Krugman, and A.J. Venables, (1999). The spatial economy: Cities, Regions and international Trade. MIT Press.

Grossman, G., and E. Helpman, (1991). Innovation and growth in the world economy. MIT Press, Cambridge, MA.

Helpman E., and P. Krugman, (1985). Market structure and foreign trade. MIT Press, Cambridge, MA.

Hirose, K., and K. Yamamoto, (2007). Knowledge spillovers, location of industry, and endogenous growth. The Annals of Regional Science, 41:17–30.

Ioannides, Y., (1994). Product differentiation and economic growth in a system of cities, Regional Science and Urban Economics 24, 461-484.

Krugman, P., (1991). Increasing returns and economic geography. Journal of Political Economy 99, 483-499.

Losch, A., (1940). The Economics of Location, (Jena: Fischer). English translation (New Haven, Conn: Yale University Press), 1954.

Martin, P., (1999). Public policies, regional inequalities and growth. Journal of Public Economics 73:85-105.

Martin, P., and G.I.P. Ottaviano, (1999). Growing locations: industry location in a model of endogenous growth. European Economic Review 43:281-302.

Martin, P., and G.I.P. Ottaviano, (2001). Growth and Agglomeration. International Economic Review, 42: 947-968.

Palivos, T., and P. Wang, (1996). Spatial agglomeration and endogenous growth. Regional Science and Urban Economics 26, 645-669.

Romer, P., (1990). Endogenous technical change. Journal of Political Economy 98(5):S71-S102.

Samuelson, P., (1954). The transfer problem and transport costs, II: analysis of effects of trade impediments. Economic Journal 64:264-289.

Von Thünen, J. H., (1826). Der isolierte Staat in Beziehung auf Landwirtschaft und Nationaloekonomie. F. Perthes, Hamburg.

Chapter 1
Trade Policies, Concentration, Growth and Welfare

1.1 Introduction

Since the mid-20th century, international efforts to reduce the costs of trade have intensified. The GATT (now the WTO) has been trying to reduce barriers to trade since 1947, and there are many other international agreements, such as MERCOSUR, with the same intention. The European Union deserves a special mention because, while it does not respond to trade issues alone, it is the furthest-reaching international integration process and the one in which more countries are involved.

In 1957, the founding states of the European Economic Community were in a very similar economic situation, occupied with reconstruction after WWII and with homogenous per capita incomes. Aside from political motives, the main economic aim of this union was the elimination of trade barriers. This measure meant a simultaneous reduction of import costs for member states. The favourable situation of the international economy at the time, together with the good trade results of the lowering of barriers, allowed the Customs Union to be reached two years ahead of schedule, in 1968. However, this was not the end of the story. Once the trade barriers had disappeared, member states continued to implement a series of measures which, although responding to political reasons, also had an effect on commercial relations (and on industrial location, growth and welfare). The clearest example is the harmonisation of legislation carried out through rulings and directives in fields such as competition law or consumer rights, which makes it easier for firms to export to other EU countries or even to establish themselves there.

In fact, as more countries have joined the European Union, the heterogeneity between them has been increasing (mainly, but not only, in terms of per capita income). This has motivated an increase in the number of policies and instruments devoted to homogenising and reducing trade costs.

Among the various types of policies, we can distinguish between those that affect one or several countries and those that modify the overall framework of the Union. The first include the Cohesion Funds and ERDF (European Regional Development Fund), which normally finance infrastructure projects in the member states with the lowest per capita income[1]. Rodríguez-Pose and Fratesi (2004) calculate that, for the period 1989-1999, almost half (49.6 %) of the Structural Funds for Objective 1 regions were allocated to investment in infrastructures, transport and environment. Their effects were local, as they decreased the costs of internal transactions.

We also find policies which affect only the imports or exports of one of the members. The typical political measures on trade, such as tariffs, quotas and other non-tariff barriers, are mainly related to imports. The policies affecting exports are basically export subsidies but also internal legislation in the style of the US anti-trust laws or different types of public

[1] Obviously, the main purpose of these investments is not to reduce trade costs but rather to encourage the economic and social cohesion of the poorest regions of the Union, as affirmed in article 158 of the Treaty Establishing the European Community. Paradoxically, although their effects on transaction costs are clear, it does not appear to be empirically demonstrated that the aim of reducing inequality is being achieved. Several studies (see Garcia-Milà and McGuire, 2001, or Boldrin and Canova, 2001) suggest a weak impact of EU regional funds on regional inequalities and convergence.

aid (such as direct grants, tax incentives for continued training activities in companies) for the improvement of product quality, design, packaging or marketing, which facilitate penetration in external markets. In this group we can also include any public policy devoted to the improvement of industrial technology as a way of favouring external competition (see Pollard and Storper, 1996). Sheringhaus and Rosson (1990) provide a complete set of public measures undertaken by developed countries to foster exports. Recent studies analyse the impact on exports of government export promotion assistance programs (Gencturk and Kotabe, 2001) and of export promotion agencies and their strategies (Rose, 2005; Lederman et al., 2006; and Gil Pareja, 2008). They find a positive and statistically significant effect of these public policies on exports.

Among the measures affecting the overall framework of the Union, we can highlight the introduction of the Euro in 1999, which facilitated trade by reducing uncertainties about exchange rates. Investment in trans-European networks can also be included in this group, since investment in infrastructures facilitates economic integration (Puga, 2002; Vickerman et al., 1999). These measures affect the transaction costs of all the member states at the same time.

Thus, any integration process can involve a wide set of policy measures devoted to reducing transaction costs. Martin and Rogers (1995) defined public infrastructure as any good or service provided by the state which can facilitate the connection between production and consumption. Good infrastructures mean low transaction costs; poor infrastructures represent a situation where trade is difficult because of the high costs incurred. It is evident that transport and communication media can be included among these trade infrastructures, but there are other elements, such as the legal system or the levels of public safety, which have an equally great influence on trade. We will consider the latter, besides the more common transport costs, when we analyse the effects of a reduction of transaction costs.

The aim of this chapter is to present, from a theoretical perspective, a model in which to analyse the effects of a reduction in trade costs (in a wide sense, not only transport costs). We consider iceberg costs (Samuelson, 1954): a portion of every good produced is lost in transport and, thus, not finally consumed. The portion lost can be reduced by adequate trade policies.

The interest of our analysis comes from the fact that a trade integration process not only changes internal trade and commercial relations with other countries, but also has repercussions on other key aspects of the spatial and temporal organisation of economic activity. The ample literature in the field of economic geography has discussed its implications for industrial location[2] (the basic mechanisms of transmission in models of economic geography are described, for example, in the survey of Ottaviano and Puga, 1998). But, in a dynamic context, the capacity for sustaining long-run economic growth can also be influenced. Finally, due to this diversity of influences, trade integration also has an impact on social welfare. This is an important dimension because, generally, neither the concen-

[2] Empirically, the intensity of these effects may vary according to regional characteristics. Huber (2004) analyses the case of the EU during the period 1975-2000 and considers a series of variables (wages, population growth, investment rates, productivity and wage growth) where the effect of increases has been small, although effects on regional wages and investment rates are stronger than the rest, at least in the long term. Hanson (1997, 1998) finds that the formation of the North American Free Trade Agreement (NAFTA) led to a less concentrated spatial distribution of production in Mexico because it was more beneficial for firms to be sited along the frontier with the US than in the old industrial belt of Mexico City. Sjöberg and Sjoholm (2004) conclude that the liberalisation of trade in Indonesia did not decrease spatial concentration in the manufacturing industry for the period 1980-1996.

tration of activity nor a higher growth rate are necessarily associated with greater levels of welfare (for example, Pfüger and Südekum, 2008, find, in a simple model of economic geography, that the spatial distribution of the most efficient activity depends decisively on the degree of freedom of trade).

In this chapter, we follow the model developed by Martin and Ottaviano (1999), which joins a framework of endogenous growth similar to Romer (1990) and Grossman and Helpman (1991) with a geographical framework similar to Helpman and Krugman (1985) and Krugman (1991). Martin and Ottaviano's model, which only analysed the role of common international infrastructures, was widened in Martin (1999) to include domestic infrastructures. In both cases, the role of public infrastructures was introduced following the static model of Martin and Rogers (1995), which distinguishes between domestic and common international infrastructures. We will go a step further and consider a wider casuistry, following Lanaspa and Sanz (2004), including asymmetries in what we can refer to, in a broad sense, as import and export infrastructures.

We consider two countries that are heterogeneous in two main aspects. First, in their trade policies, that is to say, the wide set of measures that can promote either imports or exports. This gives value to this chapter, since, as far as we know, no work has been devoted to analysing such policies from a theoretical point of view. Second, they are also considered different in their per capita incomes since, as we will see, the effects of any measure can depend on the characteristics of the country that introduces it.

When the integration process consists of a reduction in the costs associated with the internal trade of the rich country, its exports or the imports of the poor country, some firms move from the poor to the rich country and the growth rate accelerates. As a consequence, the welfare of both countries increases. However, when the internal trade of the poor country or its exports are promoted, the firms tend to move to the poor country and the growth rate diminishes. Although the latter effect could be associated with a loss of welfare (which would make this type of policies pointless), this is not necessarily the case for the poor country.

The chapter is structured as follows. Section 1.2 presents the basic characteristics of the theoretical model. Section 1.3 deals with geography, that is, with the equilibrium distribution of firms. Section 1.4 determines the steady state growth rate, which depends on industrial location, but also influences it through the resulting distribution of income. Section 1.5 analyses the effects of economic integration through the impact of different trade policies. Section 1.6 considers a simpler framework to analyse specific policy measures whose effects are not conclusive in a general framework. Finally, in Section 1.7, some simulations are carried out to analyse the effects of trade policies on welfare. The chapter ends with the conclusions.

1.2 The model

The diagram in Figure 1.2 describes schematically how the model works. Changes from Martin and Ottaviano's model are highlighted in red.

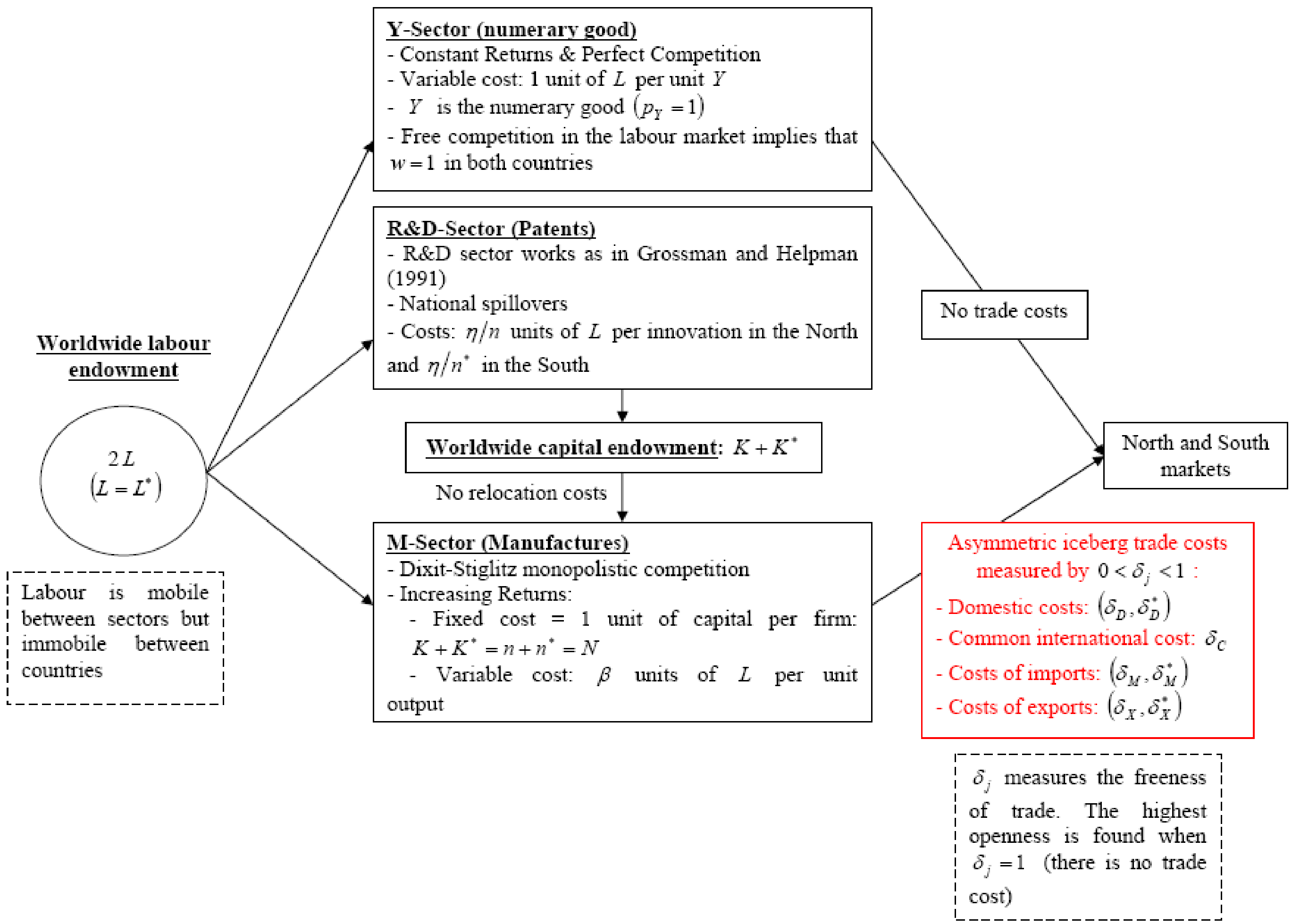


Fig. 1.2. Schematic diagram of the model.

We consider two countries, North and South, which trade with each other. They are identical except for their initial level of capital, K_0 in the North and K_0^* in the South, and their trade policies. We suppose that the initial endowment of capital is greater in the North: $K_0 > K_0^*$. Both countries are inhabited by representative households playing the part of consumers, workers and researchers. There are L families, both in the North and in the South. Labour is mobile between sectors but immobile between countries, which excludes accumulative causation and impedes a catastrophic agglomeration.

Given that the model is almost symmetrical, we will focus on the description of the economy of the North (an asterisk denotes that the variables correspond to the South). The preferences are instantaneously nested-CES and intertemporally CES, with an elasticity of intertemporal substitution equal to the unit:

$$U_0 = \int_0^\infty \log \left[D(t)^\alpha Y(t)^{1-\alpha} \right] e^{-\rho \cdot t} dt, \quad 0 < \alpha < 1, \tag{1.1}$$

where $\rho > 0$ is the intertemporal discount rate, Y is the numerary good and D is a composite good à la Dixit-Stiglitz, which consists of a number of different varieties:

$$D(t) = \left[\int_0^{N(t)} D_i(t)^{1-\frac{1}{\sigma}} di \right]^{\frac{1}{\left(1-\frac{1}{\sigma}\right)}}, \quad \sigma > 1, \tag{1.2}$$

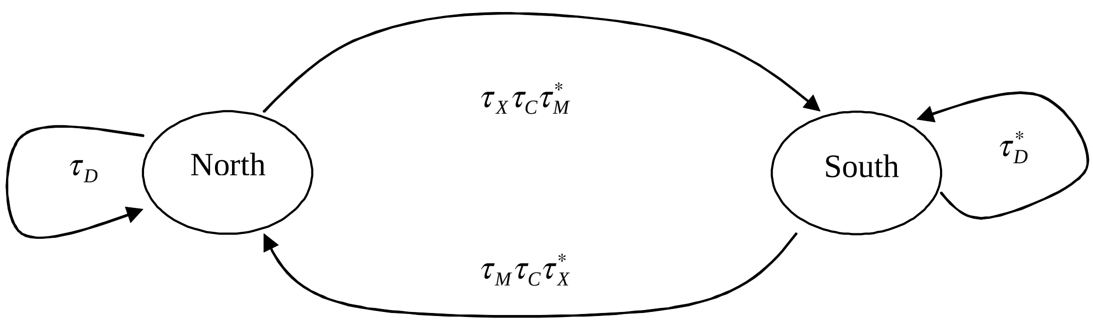
N is the total number of varieties produced between the two countries, and σ is the elasticity of substitution between varieties, assuming that N is high enough. It can be shown that σ is also the demand price elasticity of the demand for each variety. Growth is produced through an increase in the number of varieties.

The value of spending per capita E in terms of the numerary Y is:

$$\int_{i \in n} \tau_D p_i D_i di + \int_{j \in n^*} \tau_M \tau_C \tau_X^* p_j^* D_j dj + Y = E, \quad \tau > 1. \tag{1.3}$$

The number of manufactured goods produced in each country, n and n^*, is endogenous, with $N = n + n^*$.

Domestic trade costs are represented by the parameters τ_D and τ_D^*, and common international costs by τ_C. The latter affect the flow of trade between the countries, as do the costs of imports (τ_M, τ_M^*) and exports (τ_X, τ_X^*). All the $\tau > 1$ are considered to be iceberg costs, as in Samuelson (1954), which affect only the differentiated good (not the homogenous one), according to the following outline:



That is to say, for the North, for each unit of good traded, only $\tau_D^{-1} < 1$ is available for consumption and, of each unit sent to the South from the North, only $(\tau_X\tau_C\tau_M^*)^{-1} < 1$ really reaches the consumer. These transaction costs affect internal trade (τ_D for the North and τ_D^* for the South) and international trade ($\tau_X\tau_C\tau_M^*$ for a sale from the North and $\tau_M\tau_C\tau_X^*$ for a sale from the South). Thus, any τ_j captures that a portion of the good is lost in transit and, as in Martin and Rogers (1995), constitutes a measure of how easy trade is. Reductions in any τ_j denote lower transaction costs and, thus, indicate that trade is easier.

Note that the import costs of a country do not coincide with the export costs of the other because they capture the trade policy decisions taken by each country individually.

For example, the North may decide to establish export subsidies (τ_X decreases), which make the market of the other country more accessible for their products. A similar effect can be achieved by the South through a reduction in their import tariffs (τ_M^* decreases). But, although they can have the same effect, they should be considered conceptually as different because the country that introduces the policy is different.

From here on, we will assume that $\tau_D < \tau_M \tau_C \tau_X^*$ and that $\tau_D^* < \tau_X \tau_C \tau_M^*$: it is more expensive in terms of transaction costs to buy a differentiated good from abroad than to buy one made in the same country. We also assume that $\tau_D < \tau_D^*$: the domestic trade costs are lower in the rich country.

The numerary good Y is produced using only labour, subject to constant returns in a perfectly competitive sector. As labour is mobile between sectors, the constant returns in this sector tie down the wage rate w in each country at each moment. We assume throughout the chapter that the parameters of the model are such that the numerary is produced in both countries, that is, that the total demand for the numerary is big enough so as not to be satisfied with its production in a single country. In this way, wages are maintained constant and identical in both countries. A unit of labour is needed to produce a unit of Y, so free competition in the labour market implies that $w = 1$ in both countries.

The differentiated goods are produced with identical technology in an industry with monopolistic competition with increasing returns in the production of each variety. To begin to produce a variety of a good, a unit of capital is needed; this fixed cost is the source of scale economies. Also, β units of labour are used to produce a unit of differentiated good. The standard rule of monopolistic competition determines the price of any variety as a margin over the cost of labour: $p^* = p = \beta\sigma / (\sigma - 1)$. The operating profits of a producer are:

$$\pi = p_i x_i(p_i) - \beta x_i(p_i) = \frac{\beta x}{(\sigma - 1)}, \tag{1.4}$$

where x is the scale of a representative firm, equal for all varieties because of symmetry.

Investment is necessary to produce a new variety, whether in a physical asset (machinery) or an intangible one (patent). The concept of capital used corresponds to a mixture of the two types of investment. The value of the firm which produces a new variety is the value of its unit of capital. The total number of varieties and firms is determined by the stock of capital at any given time: $N = n + n^* = K + K^*$. Once the investment is made, each firm produces the new variety in a situation of monopoly and chooses where to locate its production (we assume that there are no costs of relocating the capital from one country to the other). Unlike firms, households (workers/researchers/consumers) are immobile and, therefore, their incomes are geographically fixed although firms move. In other words, if a firm owner decides to locate production in a country where he does not reside, he will repatriate the profits.

Finally, we assume there is a safe asset which pays an interest rate r whose market is characterised by freedom of financial movement between the two countries ($r = r^*$).

Resolving the first order conditions of the problem of the consumer in the North, we obtain the demands for each variety in the North (D_i) or the South (D_j), and that of the numerary good:

$$D_i = \frac{\sigma - 1}{\beta\sigma} \cdot \frac{\delta_D \alpha E}{\tau_D \left(n\delta_D + n^* \delta_M \delta_C \delta_X^*\right)}, \tag{1.5}$$

$$D_j = \frac{\sigma - 1}{\beta\sigma} \cdot \frac{\delta_M \delta_C \delta_X^* \alpha E}{\tau_M \tau_C \tau_X^* \left(n\delta_D + n^* \delta_M \delta_C \delta_X^*\right)}, \tag{1.6}$$

$$Y = (1 - \alpha)E, \tag{1.7}$$

where $\delta_j = \tau_j^{1-\sigma}$ ($j = D, M, X, C$) are parameters between 0 and 1 which measure the openness of trade. The highest openness is found when $\delta_j = 1$ (there are no trade costs). The expressions of the demand of a consumer from the South will be analogous to the above.

The intertemporal optimisation of the consumers implies that the spending growth rate is, both in the North and in the South, $\frac{\dot{E}}{E} = \frac{\dot{E}^*}{E^*} = r - \rho$, that is, the difference between the interest rate and the intertemporal discount rate (the standard Euler equation). In the steady state, E and E^* must be constant, so $r = \rho$, as we shall see below.

1.3 Geography

The geographical part of the model refers to the location of the firms, given that the population is immobile between countries[3]. The equilibrium location between firms is determined by four equilibrium conditions. The first two indicate that, when differentiated goods are produced in both countries, the total demand, coming from both North and South, of each

---

[3] The populations are tied to their country but they are very interested in the location of firms because the more firms in the region, the lower the price index they have to bear. The price indexes are:

$P = N^{\frac{1}{1-\sigma}} \left(\frac{\beta\sigma}{\sigma-1}\right) \left[S_n\delta_D + (1 - S_n)\delta_M\delta_C\delta_X^*\right]^{\frac{1}{1-\sigma}}$ in the North and

$P^* = N^{\frac{1}{1-\sigma}} \left(\frac{\beta\sigma}{\sigma-1}\right) \left[(1 - S_n)\delta_D^* + S_n\delta_X\delta_C\delta_M^*\right]^{\frac{1}{1-\sigma}}$ in the South, where $S_n = \frac{n}{N}$ is the share of manufactured goods produced in the North.

variety (including transport costs) must equal supply. Thus, starting from (1.5) and (1.6):

$$x = \frac{\alpha L(\sigma-1)}{\beta\sigma} \cdot \left(\frac{\delta_D E}{N\left[S_n \delta_D + (1-S_n)\,\delta_M \delta_C \delta_X^*\right]} + \frac{\delta_X \delta_C \delta_M^* E^*}{N\left[(1-S_n)\,\delta_D^* + S_n \delta_X \delta_C \delta_M^*\right]} \right), \tag{1.8}$$

$$x^* = \frac{\alpha L(\sigma-1)}{\beta\sigma} \cdot \left(\frac{\delta_D^* E^*}{N\left[(1-S_n)\,\delta_D^* + S_n \delta_X \delta_C \delta_M^*\right]} + \frac{\delta_M \delta_C \delta_X^* E}{N\left[S_n \delta_D + (1-S_n)\,\delta_M \delta_C \delta_X^*\right]} \right). \tag{1.9}$$

The third condition is the consequence of the free movement of capital between countries ($r = r^*$), which implies an equal retribution via profits:

$$\pi = \pi^*, \tag{1.10}$$

so, in agreement with (1.4), the same quantity is produced of all the varieties (whether in the North or the South), $x = x^*$. Finally, the fourth condition, already shown, indicates that the total number of varieties is fixed by the world supply of capital at each moment:

$$n + n^* = K + K^* = N. \tag{1.11}$$

Resolving the system formed by these four equations, the optimum size of each firm in equilibrium is obtained:

$$x = x^* = \frac{\alpha L(\sigma-1)}{\beta\sigma} \cdot \frac{E+E^*}{N}. \tag{1.12}$$

The proportion of firms in the North ($S_n = \frac{n}{N}$) is:

$$S_n = \frac{S_E \delta_D^*}{(\delta_D^* - \delta_X \delta_C \delta_M^*)} - \frac{(1-S_E)\,\delta_M \delta_C \delta_X^*}{(\delta_D - \delta_M \delta_C \delta_X^*)}, \tag{1.13}$$

where $S_E = \frac{E}{E+E^*}$ is the participation of the North in the total spending. The location equilibrium of firms depends on the national spending -greater spending implies a bigger domestic market, which attracts more firms who want to take advantage of increasing returns (home market effect)- and on all the parameters which represent trade costs.

1.4 Growth and income inequality

1.4.1 Economic growth

Firstly, we will focus on the growth rate of the economy. Starting from the solution of the problem of the intertemporal optimisation of the consumer, we know that, in equilibrium, $\frac{\dot{E}}{E} = \frac{\dot{E}^*}{E^*} = r - \rho$; as the capital flows are free, $r = r^*$ and the growth rate of spending will be the same in both countries. From (1.13), this implies that the ratio of producing firms in the North, S_n, is also constant in time and, thus, n, n^* and N grow at the same constant rate $g = \frac{\dot{N}}{N} = \frac{\dot{n}}{n} = \frac{\dot{n}^*}{n^*}$.

There are national spillovers in the innovation sector, that is to say, the more firms producing different manufactured goods in the same country, the less expensive the R&D activity[4]. This sector follows Grossman and Helpman (1991), with $\frac{\eta}{n}$ being the cost in terms of labour of an innovation in the North and $\frac{\eta}{n^*}$ in the South. The immediate conclusion of this formulation of the sector is that, for efficiency reasons, research activity will only take place in one of the two countries, namely, the one with more firms producing the manufactured goods (which will be the rich country, the North). No researcher would have an incentive to start R&D activity in the other country. This formulation makes the

[4] This type of knowledge spillovers is closer to the concept of Jacobs (1969) than to that of Marshall-Arrow-Romer (MAR). The empirical evidence for these external effects between different industries in the same geographical unit is documented in, for example, Glaeser et al. (1992) and Henderson et al. (1995).

analytical treatment of the model easier. More generally, if there is a certain degree of diffusion of knowledge at international level (Hirose and Yamamoto, 2007), R&D would be concentrated in the country with fewer innovation costs; in this case, trade policies could generate changes in the location of this activity.

The value of the firm is given by its unit of capital and, as the market is competitive, this value (v) will be the production cost of the unit of capital, $v = \frac{\eta}{n} = \frac{\eta}{NS_n}$. Thus, v decreases at the same rate at which N increases: $\frac{\dot{v}}{v} = -g$. As the number of varieties rises, the profits and value of each firm diminish, which can also be interpreted as the future flow of discounted profits: $v(t) = \int_t^\infty e^{-[\bar{r}(s)-\bar{r}(t)]} \frac{\beta x(s)}{\sigma-1} ds$, where $\bar{r}$ represents the cumulative discount factor. Taking into account the condition of arbitrage between the capital market and the safe asset, the relation between interest rates and the value of capital will be:

$$r = \frac{\dot{v}}{v} + \frac{\pi}{v}. \tag{1.14}$$

The restriction of world resources, $E + E^* = 2 + (r\eta) / (LS_n)$, ensures that spending is constant over time; so, in the steady state $r = \rho$. Finally, we must take into account the restriction of the world labour market: total labour is distributed between the production of differentiated goods, the production of the numerary good and R&D:

$$\eta \frac{g}{S_n} + \frac{\sigma - \alpha}{\sigma} L(E + E^*) = 2L. \tag{1.15}$$

In the steady state (its calculation is given in Appendix A), the variables will grow at a constant rate. By substituting in (1.14) the benefits obtained in (1.4) and the optimum size of the firms in the equilibrium given by (1.12), and taking into account (1.15) and the condition $r = \rho$, we obtain that the steady state growth rate of K and K^* (the same for both countries) is given by:

$$g = \frac{2L}{\eta} \cdot \frac{\alpha}{\sigma} S_n - \left(\frac{\sigma - \alpha}{\sigma} \right) \rho = g\left(S_n\right). \tag{1.16}$$

This rate depends upon the structural parameters of the model ($L, \eta, \alpha, \sigma, \rho$); it also has a linear dependence on the variable that represents geography (S_n).

1.4.2 World income distribution

Secondly, we want to know how this growth rate affects the inequality of income between the countries; let us remember that we have assumed that the North is initially the richer ($K_0 > K_0^*$). The per capita income of each country is the sum of labour income, which we have already seen is the unit, plus the capital income, which is the value of per capita wealth multiplied by the equilibrium interest rate. Thus, for the North, it will be $E = 1 + r\frac{Kv}{L} = 1 + \rho\frac{Kv}{L}$. If we substitute v applying the arbitrage equation (1.14), the equilibrium profits given by (1.4) and the optimum production scale obtained in (1.12), it is possible to express spending as a function of g:

$$E = 1 + \frac{2\alpha\rho S_K}{\left(\sigma - \alpha\right)\rho + \sigma g}, \tag{1.17}$$

where $S_K = \frac{K}{K+K^*}$ is the share of capital owned by the North, which is maintained constant because K, K^* and N grow at the same rate g in the steady state. Similarly, for the South:

$$E^* = 1 + \frac{2\alpha\rho\left(1 - S_K\right)}{\left(\sigma - \alpha\right)\rho + \sigma g}. \tag{1.18}$$

We previously defined the ratio $S_E = \frac{E}{E+E^*}$, which represents the participation of the North in the total income or spending. Substituting the expressions (1.17) and (1.18), we obtain:

$$S_E = \frac{1}{2} \cdot \frac{\sigma(\rho + g) + \alpha\rho(2S_K - 1)}{\sigma(\rho + g)}. \tag{1.19}$$

If, as we have assumed, the North is richer and $S_K > \frac{1}{2}$, then $S_E > \frac{1}{2}$. However, the relationship of S_E with the growth rate is negative: a greater number of varieties diminishes the value of capital and, given that the North has more capital, the distance is reduced in relative terms.

Finally, to carry out the analysis in the following section, we want to relate geography (S_n) with the growth rate g. To do so, we substitute (1.19) in (1.13), obtaining:

$$S_n = \frac{1}{2} \cdot \left\{ \begin{array}{c} \left[\frac{\delta_D^*}{\left(\delta_D^* - \delta_X \delta_C \delta_M^*\right)} - \frac{\delta_M \delta_C \delta_X^*}{\left(\delta_D - \delta_M \delta_C \delta_X^*\right)} \right] + \\ + \left[\frac{\delta_D^*}{\left(\delta_D^* - \delta_X \delta_C \delta_M^*\right)} + \frac{\delta_M \delta_C \delta_X^*}{\left(\delta_D - \delta_M \delta_C \delta_X^*\right)} \right] \cdot \frac{\alpha\rho(2S_K - 1)}{\sigma(\rho+g)} \end{array} \right\} = S_n\left[S_E(g)\right]. \tag{1.20}$$

1.5 Effects of trade integration

As we explained in the introduction, the purpose of this chapter is to analyse the consequences of a trade integration process, which is represented through a decrease in trade costs. We distinguish between domestic and international costs and, within the latter, common, import and export costs. Remember that, while domestic and common international costs can be easily associated with physical trade infrastructures (although they are not necessarily so), import and export costs also include the whole range of trade policies that affect the sales to, or the purchases from, abroad. We carry out our analysis from the perspective of the effects of these policies on industrial location, growth rate and welfare.

We adopt the simplifying assumption that the reduction of any of these costs requires no financing from either of the countries, either because the measures involve no cost or because the financing comes from, for example, an international organisation. There are two reasons for this. Firstly, we are using a broad concept of costs that encompasses very different elements. Dismantling administrative obstacles (like tariff barriers) is essentially free of cost, but physical infrastructures do require a strong outlay. Given this heterogeneity of trade policies, it is difficult to choose a single formula of financing that applies to all of them. Secondly, the explicit consideration of means of financing would make the treatment of the model cumbersome without adding qualitatively different results.

In the previous section, where we considered that the two countries are different (one rich and one poor), we obtained two equations, (1.16) and (1.20), which relate the growth rate with the distribution of firms, and vice versa. The function $g = g(S_n)$ is linear and increasing: given that technological spillovers are local, the higher the industrial concentration, the lower the innovation costs and the greater the growth rate. The function $S_n = S_n(g)$ is convex and decreasing. Remember that this equation includes the inequality of incomes, $S_n = S_n[S_E(g)]$, and that this decreases as g increases via the reduction of the monopolistic profits of the firms. At the same time, as the differences in income vanish, industrial concentration and the rich country's market size decrease as a consequence of the "home market effect". These functions are represented in Figure 1.3.

Below, we look at the different effects of trade integration according to the specific type of trade costs that are reduced.

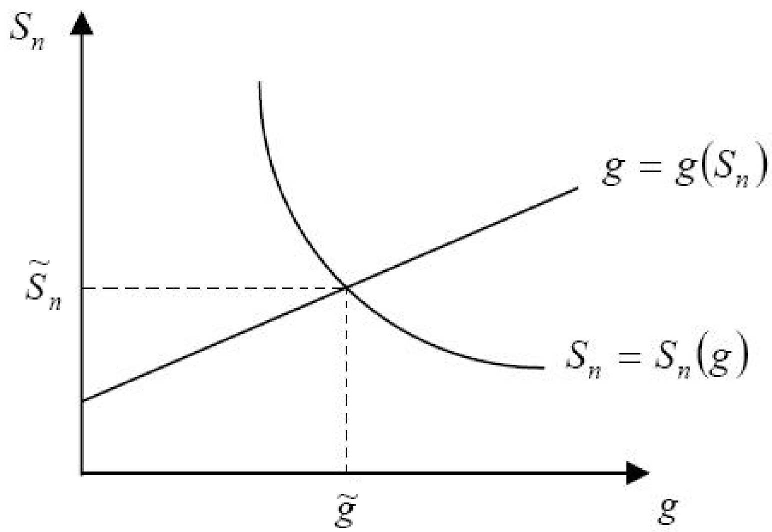


Fig. 1.3. Growth rate and firms location.

1.5.1 Domestic trade costs

When commercial integration takes the form of a reduction in domestic costs, it increases the effective internal demand (domestic consumers bear fewer transport costs) and thus attracts firms to the country in which the policy has been implemented. If it is the North which is the richer country and thus has a bigger market, firms will decide to move to the North to take more advantage of increasing returns and the ratio S_n will increase (remember that there are no relocation costs).

Concentration improves the growth rate (Figure 1.4); the more manufacturers located in the North, the lower the innovation costs in the R&D sector:

$$\frac{\partial S_n}{\partial \delta_D} > 0, \frac{\partial g}{\partial \delta_D} > 0.$$

The same reasoning is applied if the poor country reduces its domestic trade costs: more firms are attracted, wanting to better exploit its growing returns and its increasing

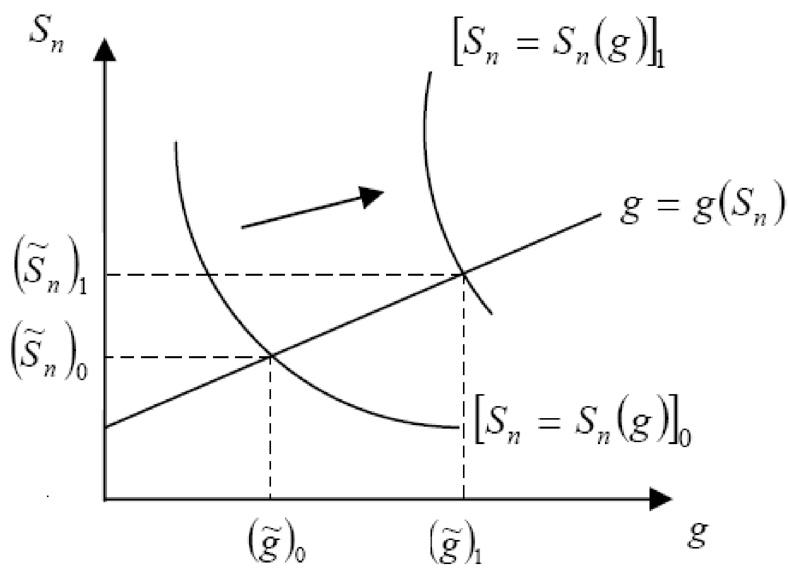


Fig. 1.4. Reduction in domestic trade costs in the rich country.

effective internal demand. But the effect on the growth rate is different: as firms move to the South, the externalities derived from the variety of products in the North are reduced, the costs of the research sector are increased, and the economy's growth rate (Figure 1.5) decreases:

$$\frac{\partial S_n}{\partial \delta_D^*} < 0, \frac{\partial g}{\partial \delta_D^*} < 0.$$

However, this benefits the inhabitants of the North, because the rate at which their capital loses value (g) is reduced; this implies from (1.19) that income inequality increases.

1.5.2 Import-enhancing policies

When one of the countries reduces its import costs, it becomes easier to reach its market from the other country (think, for example, of a lowering of tariffs). If the North carries out such measures, the effective demand of the consumers of the North for the goods produced

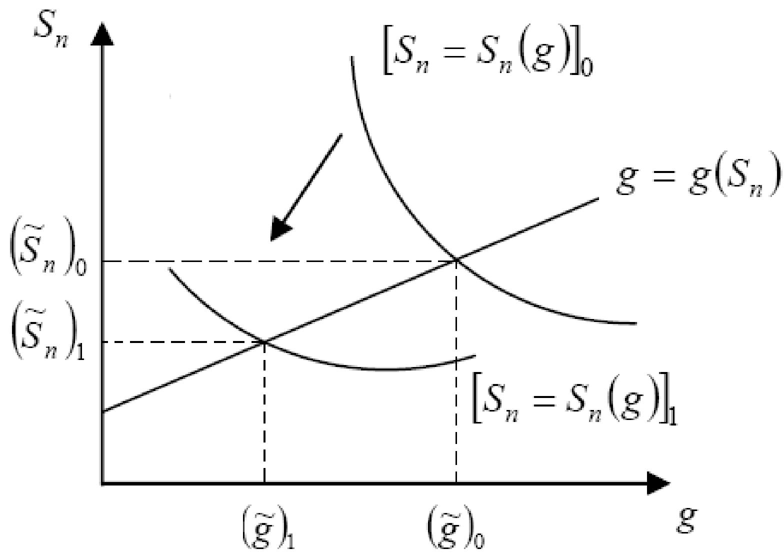


Fig. 1.5. Reduction in domestic trade costs in the poor country.

in the South will increase, so some firms located in the North will decide to move to the South. For these firms, this movement means two advantages. On the one hand, they can better satisfy the demand of consumers in the South when the costs of sending a unit from the North ($\tau_X \tau_C \tau_M^*$) are replaced by just the internal transport costs (τ_D^*) of the South (remember that $\tau_D^* < \tau_X \tau_C \tau_M^*$). On the other hand, firms moving to the South benefit from the increase in effective demand due to the reduction in import costs (from which they would not benefit if they stayed in the North). The relocation of manufacturing firms also has a negative effect on R&D as it provokes an increase in research costs through the reduction of externalities in the North and, thus, also negatively affects the growth rate of the economy (a similar situation to that of Figure 1.5):

$$\frac{\partial S_n}{\partial \delta_M} < 0, \frac{\partial g}{\partial \delta_M} < 0.$$

When the South is the country that introduces policies that reduce its import costs, firms will move to the North, benefiting from the increase in effective demand of the consumers in the South for the goods produced in the North and from the reduction in transport costs when substituting $\tau_M \tau_C \tau_X^*$ by τ_D ($\tau_D < \tau_M \tau_C \tau_X^*$). But the effects on growth will be positive: with the increase in the number of firms located in the North, the cost of R&D decreases and the economy's growth rate increases (a similar situation to Figure 1.4):

$$\frac{\partial S_n}{\partial \delta_M^*} > 0, \frac{\partial g}{\partial \delta_M^*} > 0.$$

Moreover, from (1.19), the inequality of income between the two countries is reduced.

To sum up, the country that decides to dismantle trade barriers (thus reducing import costs) will lose firms, so neither country has incentives to carry out this measure unilaterally. If the two countries decide to carry out an equal and simultaneous change, that is, a joint lowering of trade barriers, the effect on the location of industries and growth is indeterminate. Thus, a joint lowering of trade barriers can lead to a rise in the growth rate only in some specific cases (Appendix B).

1.5.3 Export-enhancing policies

The policies that reduce export costs make it easier to penetrate the market of the other country, i.e., the market potential is increased for the firms of the country implementing these policies. If the North carries out such measures, more firms will come from the South because it is now less expensive to send goods from the North to the South, and in the richer country increasing returns can be exploited better. With the increase in the number of firms, research costs will be reduced, thanks to the national spillovers, and the growth rate will rise (a similar situation to Figure 1.4):

$$\frac{\partial S_n}{\partial \delta_X} > 0, \frac{\partial g}{\partial \delta_X} > 0.$$

But if the poor country reduces its export costs, it will also be capable of attracting firms. The lowering of transport costs means an increase in the effective demand of consumers in the North for the goods produced in the South; the firms which move to the South can benefit from this increase, as well as from the reduction in trade costs associated to sales in the South (from $\tau_X \tau_C \tau_M^*$ to τ_D^*). The effect on the growth rate will be negative (a situation identical to Figure 1.5), as the number of firms in the country where the most efficient R&D sector is located will decrease:

$$\frac{\partial S_n}{\partial \delta_X^*} < 0, \frac{\partial g}{\partial \delta_X^*} < 0.$$

In this case, the trade policy will achieve its aim of attracting more firms and increasing economic activity; the country which improves its export costs receives new firms, while the other country loses them. However, what happens if the two countries agree to stop subsidising exports? As in the case of the joint lowering of tariffs, a co-ordinated action produces an indeterminate result. Again, depending on the initial conditions, trade integration based on a joint action on export infrastructures might increase growth (Appendix B).

1.5.4 International common trade costs

Determining the effect of a reduction in common international trade costs is more difficult. Differentiating (1.16) and (1.20) and grouping terms, we obtain:

$$dS_n = \left[1 + \left(\frac{\delta_D^*}{(\delta_D^* - \delta_X\delta_C\delta_M^*)} + \frac{\delta_M\delta_C\delta_X^*}{(\delta_D - \delta_M\delta_C\delta_X^*)}\right)\frac{\rho\,(2S_K - 1)\,\frac{L}{\eta}\alpha^2}{[\sigma\,(\rho + g)]^2}\right]^{-1} \cdot$$
$$\cdot \left[\left(\frac{\delta_D^*\delta_X\delta_M^*}{(\delta_D^* - \delta_X\delta_C\delta_M^*)^2} + \frac{\delta_M\delta_X^*\delta_D}{(\delta_D - \delta_M\delta_C\delta_X^*)^2}\right)S_E - \frac{\delta_M\delta_X^*\delta_D}{(\delta_D - \delta_M\delta_C\delta_X^*)^2}\right]d\delta_C \gtrless 0$$

The effect remains indeterminate. The first bracket has a positive sign as long as the North is richer ($S_K > 1/2$) but, in the second bracket, we can identify two opposite effects:

1. A positive effect corresponding to the first term, whose magnitude depends on S_E and which we identify with the home market effect; the change in δ_C means that it is less costly to send goods to the South, so it is more attractive for firms to be located in the North, where the market is bigger and they can take better advantage of increasing returns. Evidently, if S_E is greater, concentration will probably be accentuated due to a greater home market effect.

2. But the reduction in common trade costs also means a decrease in the cost of sending goods from the South to the North. And the lower the costs of importing in the North and exporting in the South, the bigger the decrease in transaction costs from South to North in relative terms, and the more firms will decide to relocate in the South. This negative effect is captured in the second term. The competition effect is also operative. We must not forget that the monopolistic profits decrease with the number of firms so, if S_n decreases, so does the growth rate of new firms.

1.6 Homogenous countries

In the previous section, we have analysed the effects of integration in the most general framework possible, that is, considering that countries can differ both in wealth and in their trade policies. However, when analysing changes in common trade costs or simultaneous variations in export or import costs, we could not determine their effects on the growth rate or on industrial location. In order to carry out this analysis, we have to adopt a simplifying assumption that the countries are homogenous and differ only in their international trade policies[5].

Thus, we assume now that the two countries have the same wealth ($S_E = 1/2$)[6], that is, equal supply of capital ($K_0 = K_0^*$) which implies that $S_K = 1/2$, and the same domestic trade costs ($\delta_D = \delta_D^*$). With this simplification we eliminate one of the two effects (the home market effect) that cause the indetermination. Under these conditions ($S_K = 1/2$ and $\delta_D = \delta_D^*$), if the two countries have the same market size ($S_E = 1/2$), the distribution of firms and the growth rate will depend only on the different trade costs and (1.20) changes to:

$$S_n = \frac{1}{2}\left[\frac{\delta_D}{(\delta_D - \delta_X \delta_C \delta_M^*)} - \frac{\delta_M \delta_C \delta_X^*}{(\delta_D - \delta_M \delta_C \delta_X^*)}\right],$$

while condition (1.16) remains intact:

$$g = g\left(S_n\right).$$

[5] The conclusions obtained in the previous section on changes carried out in only one country in import or export costs are maintained under this simplification.

[6] This assumption implies that equality of income is maintained constant over time and is independent of the growth rate.

Remember that R&D will take place only in the country with more firms. In the earlier sections, it was the North, due to its superior initial supply of capital and the home market effect, which allowed us to affirm that most firms would be set up in the country with a bigger market and, thus, more demand. Now the location of firms will depend solely on the different trade costs and firms will locate in the country with the easiest access to the market of the other.

Let us suppose that the North has more firms ($S_n > 1/2$); this is the case when the condition $\delta_M \delta_C \delta_X^* < \delta_X \delta_C \delta_M^*$ holds, that is, it is easier to send goods to the South from the North than vice versa. When asymmetries in per capita income and domestic trade costs are eliminated, this makes the North more attractive to firms than the South, and most will be located there. If this condition is fulfilled, S_n will be a horizontal straight line with positive order in the origin.

We will analyse the effects of integration when the countries are obliged to implement some kind of policy simultaneously.

1.6.1 Import-enhancing policies

Let us consider an equal and simultaneous lowering of import costs in the two countries ($d\delta_M = d\delta_M^*$), similar to lowering trade barriers:

$$\frac{dS_n}{d\delta_M} = \frac{1}{2}\delta_D \delta_C \left[\frac{\delta_X}{(\delta_D - \delta_X \delta_C \delta_M^*)^2} - \frac{\delta_X^*}{(\delta_D - \delta_M \delta_C \delta_X^*)^2} \right] \gtrless 0.$$

As can be seen, even maintaining the condition $\delta_M \delta_C \delta_X^* < \delta_X \delta_C \delta_M^*$, the effects on industrial concentration and on the growth rate remain indeterminate and depend on the export costs in both countries. If the North has lower export costs ($\delta_X \geq \delta_X^*$), the effect is positive. Firms take advantage of the lowering of trade barriers to locate in the country with lower export costs, that is, the country where it is easier to have access to the market of the other, and so industrial concentration will increase. And the increase in the number of firms in the North will provoke an increase in the growth rate of the economy. But, if $\delta_X < \delta_X^*$, the effect remains indeterminate, depending on the values of the different trade costs.

1.6.2 Export-enhancing policies

Now we will analyse a simultaneous and identical reduction in export costs in both countries ($d\delta_X = d\delta_X^*$). The expression obtained is the following:

$$\frac{dS_n}{d\delta_X} = \frac{1}{2}\delta_D\delta_C\left[\frac{\delta_M^*}{(\delta_D - \delta_X\delta_C\delta_M^*)^2} - \frac{\delta_M}{(\delta_D - \delta_M\delta_C\delta_X^*)^2}\right] \gtrless 0.$$

Maintaining the condition $\delta_M\delta_C\delta_X^* < \delta_X\delta_C\delta_M^*$, if the South has lower import costs ($\delta_M^* \geq \delta_M$), this measure constitutes an incentive for firms to move to the country which is more difficult to access from outside, the North (remember that we can identify import costs with tariffs and trade barriers in general). Thus, the effect on industrial concentration is positive, which benefits the R&D sector and increases the growth rate of the economy. However, if $\delta_M^* < \delta_M$, the sign is indeterminate, depending again on the relation between the trade costs of the two countries.

1.6.3 Common international trade costs

In the simplified framework of this section, the reduction of common international trade costs has a clear and positive sign: $\frac{dS_n}{d\delta_C} > 0$, given that a reduction in δ_C makes it even easier to reach the market of the South from the North. The greater concentration of firms in the North has the known positive effect on the growth rate of the economy, as research costs are decreased. This change is represented in Figure 1.6.

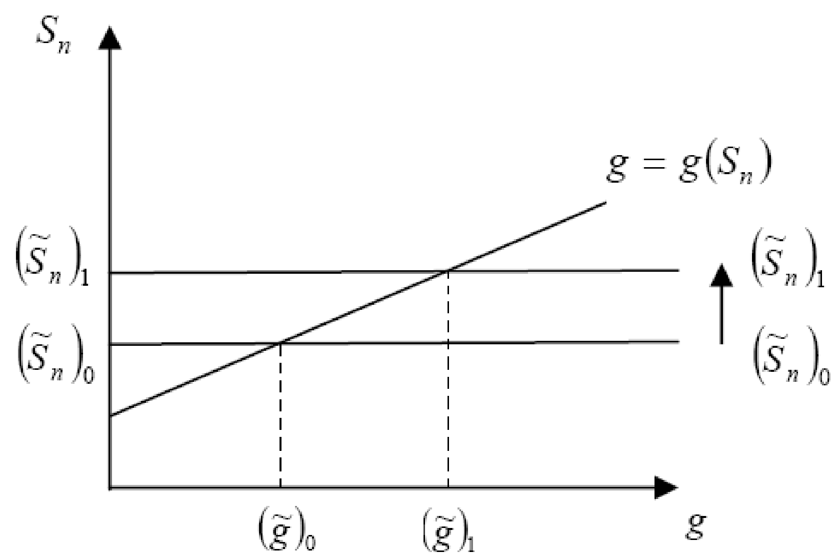


Fig. 1.6. Improvement of common international infrastructures.

1.7 Welfare

In previous sections, we have seen how trade integration could, depending on certain conditions, carry firms to the rich country, or not, and improve the growth rate, or not. However, nothing has been affirmed about the desirability of one or the other situation. In this section, we will address this gap.

The indirect utility function of a household in the North is:

$$V = \frac{1}{\rho}\ln\left\{\begin{array}{c}\alpha^{\alpha}\left(1-\alpha\right)^{1-\alpha}\left(\frac{\sigma-1}{\beta\sigma}\right)^{\alpha}\left(1+\frac{2\alpha\rho S_K}{(\sigma-\alpha)\rho+\sigma g}\right)N_0^{\frac{\alpha}{\sigma-1}}\cdot \\ \cdot\left(S_n(\delta_D-\delta_M\delta_C\delta_X^*)+\delta_M\delta_C\delta_X^*\right)^{\frac{\alpha}{\sigma-1}}e^{\frac{\alpha g}{\rho(\sigma-1)}}\end{array}\right\}. \tag{1.21}$$

Similarly, for the South:

$$V^* = \frac{1}{\rho}\ln\left\{\begin{array}{c}\alpha^{\alpha}\left(1-\alpha\right)^{1-\alpha}\left(\frac{\sigma-1}{\beta\sigma}\right)^{\alpha}\left(1+\frac{2\alpha\rho(1-S_K)}{(\sigma-\alpha)\rho+\sigma g}\right)N_0^{\frac{\alpha}{\sigma-1}}\cdot \\ \cdot\left(\delta_D^*-S_n(\delta_D^*-\delta_X\delta_C\delta_M^*)\right)^{\frac{\alpha}{\sigma-1}}e^{\frac{\alpha g}{\rho(\sigma-1)}}\end{array}\right\}. \tag{1.22}$$

Differentiating the indirect utility for the North with respect to S_n, and taking into account that $\frac{\partial g}{\partial S_n}=\frac{2L}{\eta}\cdot\frac{\alpha}{\sigma}$, the impact of an increase in industrial concentration is obtained as:

$$\frac{\partial V}{\partial S_n} = -\frac{4\alpha^2 L S_K}{\eta\left[(\sigma-\alpha)\rho+\sigma g+2\eta\alpha\rho S_K\right]\left[(\sigma-\alpha)\rho+\sigma g\right]}+\frac{2\alpha^2 L}{\sigma\eta\rho^2\left(\sigma-1\right)}+$$
$$+\frac{\alpha}{\rho\left(\sigma-1\right)}\cdot\frac{\left(\delta_D-\delta_M\delta_C\delta_X^*\right)}{\left(S_n\delta_D+(1-S_n)\delta_M\delta_C\delta_X^*\right)}\gtrless 0.$$

The effect on welfare remains indeterminate. As in Martin and Ottaviano (1999), there are three effects:

a) The first is the negative impact of an increase of g on the wealth of the North, captured in the first addend of the above expression. The increase of industrial concentration in the North diminishes the cost of R&D and raises the growth rate. This provokes a reduction of monopolistic profits and, thus, further lowers per capita income in the North.

b) The second addend captures the positive impact on the growth rate, which increases the utility of individuals due to the love-of-variety effect implied by their preferences.

c) The last addend captures the increase in welfare due to the decrease in transport costs for consumers in the North when S_n rises.

A similar expression is obtained for the South:

$$\frac{\partial V^*}{\partial S_n} = -\frac{4\alpha^2 L\,(1-S_K)}{\eta\left[(\sigma-\alpha)\,\rho+\sigma g+2\eta\alpha\rho\,(1-S_K)\right]\left[(\sigma-\alpha)\,\rho+\sigma g\right]} + \frac{2\alpha^2 L}{\sigma\eta\rho^2\,(\sigma-1)} - \\ -\frac{\alpha}{\rho\,(\sigma-1)}\cdot\frac{(\delta_D^* - \delta_X\delta_C\delta_M^*)}{(S_n\delta_X\delta_C\delta_M^* + (1-S_n)\delta_D^*)} \gtrless 0,$$

with the difference that the sign of the third effect will be the opposite as an increase in industrial concentration in the North provokes an increase in the transport costs which will have to be borne by the consumers of the South, reducing their welfare.

Both the indirect utility functions and their derivatives are too complex to evaluate analytically, so we will carry out the analysis via simulations. We want to know how welfare will vary in the two regions after changes in trade costs which provoke variations in the distribution of firms. To simplify, we analyse discrete variations, focusing only on the sign of change in the utility of the two countries, which is obtained in (1.21) and (1.22) by evaluating the direct impact of the changes in the trade policies as well as the indirect effects due to the changes in S_n and g.

Table 1.1 shows the effects of reductions in trade costs, both domestic and international, of the rich and poor countries[7].

Table 1.1. EFFECTS OF TRADE POLICIES ON CONCENTRATION, GROWTH AND WELFARE

δ_D	δ_D^*	$\delta_X\delta_C\delta_M^*$	$\delta_X^*\delta_C\delta_M$	ΔS_n	Δg	ΔV	ΔV^*
0.95	0.8	0.7	0.6				
0.96	-	-	-	(+)	(+)	(+)	(+)
-	0.81	-	-	(-)	(-)	(-)	(+)
-	-	0.71	-	(+)	(+)	(+)	(+)
-	-	-	0.61	(-)	(-)	(-)	(+)

7 The remaining parameters are N_o = 10, $\sigma = 4$, $\alpha = 0.6$, $\beta = 1$, η = 12, L = 3, ρ = 0.06, S_k = 0.6, similar values to those habitually used in the literature.

In general, the results suggest that, if trade integration through the reduction of any kind of trade costs leads to industrial concentration in the rich country, the overall growth rate of the economy will rise, generating improvements in the welfare of both the rich and the poor country. If industrial concentration decreases in the rich country and industry moves to the poor country, the growth rate falls although, in spite of this, the poor country can increase its welfare.

The key parameters for an increase of welfare in the poor country are the distribution of capital, which directly influences the equilibrium location of firms and, thus, influences the growth rate, and the relation between domestic and international transport costs. From (1.20), we can see how the nearer S_K is to $1/2$, the less impact there will be from a change in any trade policy on the proportion of firms in the rich country and, thus, on the growth rate. Therefore, the nearer S_K is to $1/2$, the more likely it is that the welfare of the poor country will improve.

These results highlight the advantages that can be derived for the countries from co-ordinated policy measures, but they also open the way to the possibility of individual strategic behaviours which could lead to trade wars to try to claim a greater share of world demand.

1.8 Conclusions

This chapter analyses the consequences of trade integration between countries that may be different in their wealth and in their trade policies, or similar. The aim is to study how industrial concentration, the growth rate of the economy and welfare are affected.

Productivity growth is endogenous and based on spillovers in innovation. Since these spillovers are assumed to be local, R&D activities are only carried out in one of the countries; thus, the greater the industrial concentration in that country, the greater the growth rate of the economy. The theoretical framework follows Martin and Ottaviano (1999), but include asymmetries which allow countries to present different trade policies: our model distinguishes between domestic and international trade costs and, within the latter, common, import and export costs.

The differences in factor endowments (and, thus, in income), on the one hand, and in trade policies, on the other, are two elements which interact when determining both the spatial distribution of economic activity, growth and welfare and their evolution.

The results we find with this classification coincide with Martin (1999) on the effects of domestic trade costs but differ on the effects of international costs. In Martin (1999), an unambiguous positive relationship between lower international transaction costs and growth was found. In our framework, the trade integration process can lead to a rise in industrial concentration, the growth rate and welfare, or not, depending on which country adopts the measures needed for integration and which specific costs are reduced. If the rich country reduces its domestic or export costs, or the poor country reduces its import costs, industrial concentration rises in the rich country, while the economic growth rate and welfare rise in both countries. However, when the poor country reduces its domestic costs or introduces export-enhancing policies (or the rich country reduces its import costs), integration leads to a lowering of industrial concentration and a drop in the growth rate. Although less growth could be associated with lower welfare, we find that this is not necessarily the case, since the poor country can improve its welfare, which would justify the adoption of these policies.

In order to make the analysis clearer for some specific policies, such as the reduction of common international trade costs, a joint lowering of tariffs or a simultaneous reduction of export subsidies, we have had to simplify the model by considering more homogeneous countries. Any of these measures could lead to an increase in industrial concentration, the growth rate and welfare in both countries, as long as firms concentrate in the country whose market potential has been widened by the policies.

To sum up, we can say that a trade integration process does not have a monotonic effect on industrial concentration, the economic growth rate or the welfare of the countries involved in it, and the final result depends on which country introduces the political measures and what type of trade cost is affected by that decision.

1.A Appendix A: Steady state equilibrium

Starting from (1.13), (1.16) and (1.19), the value of S_n in the steady state is the solution of the second degree equation:

$$\begin{gathered}
2S_n^2 L\left(\delta_D^* - \delta_X \delta_C \delta_M^*\right)\left(\delta_D - \delta_M \delta_C \delta_X^*\right) - \\
-S_n \left[\begin{array}{c} L\delta_D^*\left(\delta_D - \delta_M \delta_C \delta_X^*\right) - L\delta_M \delta_C \delta_X^*\left(\delta_D^* - \delta_X \delta_C \delta_M^*\right) \\ -\rho\eta\left(\delta_D^* - \delta_X \delta_C \delta_M^*\right)\left(\delta_D - \delta_M \delta_C \delta_X^*\right) \end{array}\right] - \\
-\rho\eta\left[S_K \delta_D^*\left(\delta_D - \delta_M \delta_C \delta_X^*\right) - \left(1 - S_K\right)\delta_M \delta_C \delta_X^*\left(\delta_D^* - \delta_X \delta_C \delta_M^*\right)\right] = 0.
\end{gathered}$$

The valid solution is:

$$S_n = \frac{\left[\begin{array}{c} L\delta_D^*\left(\delta_D-\delta_M\delta_C\delta_X^*\right)-L\delta_M\delta_C\delta_X^*\left(\delta_D^*-\delta_X\delta_C\delta_M^*\right) \\ -\rho\eta\left(\delta_D^*-\delta_X\delta_C\delta_M^*\right)\left(\delta_D-\delta_M\delta_C\delta_X^*\right)\end{array}\right]}{4L\left(\delta_D^*-\delta_X\delta_C\delta_M^*\right)\left(\delta_D-\delta_M\delta_C\delta_X^*\right)} + \\ + \frac{\sqrt{\Delta}}{4L\left(\delta_D^*-\delta_X\delta_C\delta_M^*\right)\left(\delta_D-\delta_M\delta_C\delta_X^*\right)},$$

where

$$\Delta = \left[\begin{array}{c} L\delta_D^*\left(\delta_D-\delta_M\delta_C\delta_X^*\right)-L\delta_M\delta_C\delta_X^*\left(\delta_D^*-\delta_X\delta_C\delta_M^*\right) \\ -\rho\eta\left(\delta_D^*-\delta_X\delta_C\delta_M^*\right)\left(\delta_D-\delta_M\delta_C\delta_X^*\right)\end{array}\right]^2 + \\ +8L\left(\delta_D^*-\delta_X\delta_C\delta_M^*\right)\left(\delta_D-\delta_M\delta_C\delta_X^*\right)\cdot \\ \cdot\rho\eta\left[S_K\delta_D^*\left(\delta_D-\delta_M\delta_C\delta_X^*\right)-\left(1-S_K\right)\delta_M\delta_C\delta_X^*\left(\delta_D^*-\delta_X\delta_C\delta_M^*\right)\right].$$

The other root is greater than the unit. From this equilibrium value of S_n, we can obtain that of g starting from (1.16) and that of S_E from (1.19).

1.B Appendix B: Simultaneous variations in import and export costs

We first suppose a simultaneous and identical variation in import costs ($d\delta_M = d\delta_M^*$). Differentiating (1.16) and (1.20) and grouping terms, we obtain:

$$dS_n = \left[1+\left(\frac{\delta_D^*}{\left(\delta_D^*-\delta_X\delta_C\delta_M^*\right)}+\frac{\delta_M\delta_C\delta_X^*}{\left(\delta_D-\delta_M\delta_C\delta_X^*\right)}\right)\frac{\rho\left(2S_K-1\right)\frac{L}{\eta}\alpha^2}{\left[\sigma\left(\rho+g\right)\right]^2}\right]^{-1}\cdot \\ \cdot\left[\left(\frac{\delta_D^*\delta_X\delta_C}{\left(\delta_D^*-\delta_X\delta_C\delta_M^*\right)^2}+\frac{\delta_C\delta_X^*\delta_D}{\left(\delta_D-\delta_M\delta_C\delta_X^*\right)^2}\right)S_E-\frac{\delta_C\delta_X^*\delta_D}{\left(\delta_D-\delta_M\delta_C\delta_X^*\right)^2}\right]d\delta_M \gtrless 0.$$

For an identical and simultaneous improvement in export costs ($d\delta_X = d\delta_X^*$), we obtain a similar expression:

$$dS_n = \left[1 + \left(\frac{\delta_D^*}{(\delta_D^* - \delta_X\delta_C\delta_M^*)} + \frac{\delta_M\delta_C\delta_X^*}{(\delta_D - \delta_M\delta_C\delta_X^*)}\right)\frac{\rho\,(2S_K - 1)\,\frac{L}{\eta}\alpha^2}{[\sigma\,(\rho + g)]^2}\right]^{-1} \cdot$$
$$\cdot \left[\left(\frac{\delta_D^*\delta_M^*\delta_C}{(\delta_D^* - \delta_X\delta_C\delta_M^*)^2} + \frac{\delta_C\delta_M\delta_D}{(\delta_D - \delta_M\delta_C\delta_X^*)^2}\right)S_E - \frac{\delta_C\delta_M\delta_D}{(\delta_D - \delta_M\delta_C\delta_X^*)^2}\right]d\delta_X \gtrless 0.$$

As in the case of common trade costs, described at the end of Section 1.5, the sign of the variation in industrial location (and, thus, of the variation in the growth rate) remains indeterminate.

References

Boldrin, M., and F. Canova, (2001). Inequality and convergence: Reconsidering European Regional Policies. Economic Policy 16 (32), 205-253.

Garcia-Milá, T., and T. McGuire, (2001). Do interregional transfers improve the economic performance of poor regions? The case of Spain. International Tax and Public Finance 8 (5), 281-95.

Gençtürk, E. F., and M. Kotabe, (2001). The Effect of Export Assistance Program Usage on Export Performance: A Contingency Explanation. Journal of International Marketing 9, 51-72.

Gil Pareja, S., R. Llorca Vivero and J. A. Martínez Serrano, (2008). Measuring the impact of regional export promotion: The Spanish case. Papers in Regional Science 87 (1), 139-146.

Glaeser, E. L., H. Kallal, J. Scheinkman and A. Schleifer, (1992). Growth in cities. Journal of Political Economy 100, 1126-1152.

Grossman, G., and E. Helpman, (1991). Innovation and growth in the world economy. MIT Press, Cambridge, MA.

Hanson, G. H., (1997). Increasing returns, trade and the regional structure of wages. Economic Journal 107, 113-33.

Hanson, G. H., (1998). North American economic integration and industry location. Oxford Review of Economic Policy 14, 30-44.

Helpman, E., and P. Krugman, (1985). Market structure and foreign trade. MIT Press, Cambridge, MA.

Henderson, V., A. Kuncoro and M. Turner, (1995). Industrial development in cities. Journal of Political Economy 103 (5), 1067-1090.

Hirose, K., and K. Yamamoto, (2007). Knowledge spillovers, location of industry, and endogenous growth. Annals of Regional Science 41, 17-30.

Huber, P., (2004). An empirical analysis of the regional effects of European integration. Manuscript, WIFO, Vienna.

Jacobs, J., (1969). Economy of Cities. Vintage, New York.

Krugman, P., (1991). Increasing returns and economic geography. Journal of Political Economy 99, 483-499.

Lanaspa, L., and F. Sanz, (2004). Regional policy and industrial location decisions. Investigaciones Económicas XXVIII (1), 67-87.

Lederman, D., M. Olarreaga and L. Payton, (2006). Export promotion agencies: What works and what doesn't. World Bank Policy Research Working Paper 4044.

Martin, P., (1999). Public policies, regional inequalities and growth. Journal of Public Economics 73, 85-105.

Martin, P., and G. I. P. Ottaviano, (1999). Growing locations: industry location in a model of endogenous growth. European Economic Review 43, 281-302.

Martin, P., and C. A. Rogers, (1995). Industrial location and public infrastructure. Journal of International Economics 39, 335-351.

Ottaviano, G. I. P., and D. Puga, (1998). Agglomeration in the global economy: A survey of the 'New Economic Geography'. World Economy 21 (6), 707-731.

Pflüger, M., and J. Südekum, (2008). Integration, agglomeration and welfare. Journal of Urban Economics 63, 544-66.

Pollard, J., and M. Storper, (1996). Metropolitan employment change in dynamic industries in the 1980s. Economic Geography 72, 1-22.

Puga, D., (2002). European regional policies in light of recent location theories. Journal of Economic Geography 2, 373-406.

Rodríguez-Pose, A., and U. Fratesi, (2004). Between Development and Social Policies: The Impact of European Structural Funds in Objective 1 Regions. Regional Studies 38 (1), 97-113.

Romer, P., (1990). Endogenous technical change. Journal of Political Economy 98: 338-54.

Rose, A., (2005). The foreign service and foreign trade: Embassies as export promotion. NBER Working Paper 11111.

Samuelson, P., (1954). The transfer problem and transport costs, II: analysis of effects of trade impediments. Economic Journal 64, 264-289.

Sheringhaus, F. H., and P. Rosson, (1990). Government Export Promotion. Routledge.

Sjöberg, Ö., and F. Sjoholm, (2004). Trade liberalization and the geography of production: agglomeration, concentration, and dispersal in Indonesia's manufacturing industry. Economic Geography 80 (3), 287-310.

Vickerman, R, K. Spiekemann, and M. Wegener, (1999). Accesibility and economic development in Europe. Regional Studies 33 (1), 1-15.

Chapter 2
First Nature vs. Second Nature Causes: Industry Location and Growth in the Presence of an Open-Access Renewable Resource

2.1 Introduction

There are many factors influencing the distribution of economic activity. It is traditional to distinguish between characteristics linked to the physical landscape, such as temperature, rainfall, access to the sea, the presence of natural resources or the availability of arable land, and factors relating to human actions and economic incentives (for example, scale economies or knowledge spillovers). The first group of factors, related to natural geographical circumstances, are called “first nature causes”, and the second group are called “second nature causes”.

A great deal of effort has been dedicated to researching the influence of second nature causes, especially after the pioneering work of Krugman (1991), who demonstrated how economic forces (increasing returns and transport costs) determine the distribution of activity. However, the models of New Economic Geography are usually based on the assumption that the space is homogenous, thus controlling the first nature causes. This means that less work has been invested in the theoretical study of the effect of first nature causes, even though many empirical studies demonstrate their important influence on economic growth and the concentration of economic activity.

For the case of the United States, Ellison and Glaeser (1999) state that natural advantages, such as the presence of a natural harbour or a particular climate, can explain at least half of the observed geographic concentration. Glaeser and Shapiro (2003) find that in the 1990s people moved to warmer, dryer places. Black and Henderson (1998) conclude that the extent of city growth and mobility are related to natural advantages, or geography. Beeson et al. (2001) show that access to transport networks, either natural (oceans) or produced (railroads) was an important source of growth during the period 1840-1990, and that climate is one of the factors promoting population growth. And Mitchener et al. (2003) find that some geographical characteristics account for a high proportion of the differences in productivity levels between American states.

The aim of this chapter is to provide a theoretical model which enables us to analyse the influence of one of the first nature causes, the presence of natural resources, on the concentration of economic activity and growth. To do this we will build a model in which firms can choose to locate in one of two countries which trade with each other, which we will call North and South. This model integrates characteristics of the New Economic Geography, the theory of endogenous growth and the economy of natural resources.

We will follow the model developed by Martin and Ottaviano (1999), which combines a model of endogenous growth similar to that of Romer (1990), and Grossman and Helpman (1991), with a geographical framework like that of Helpman and Krugman (1985), and Krugman (1991). Economic growth is supported by an endogenous framework with national spillovers in innovation, causing research activities to take place in a single country, and thus, the greater the industrial concentration in that country, the higher the economic growth rate.

To this model, we add an open access renewable natural resource, used by firms as a productive input. This introduces an additional element that conditions firms' decisions about whether to locate in the North or in the South, besides the traditional home market effect and the existence of trade costs. The relative importance of these three forces determines a non-symmetrical location of firms. The industrial geography here relates to the natural resource in two ways. First, the natural resource is located in only one of the two countries, namely, the South. And, second, the international trade of the natural resource is subject to a transport cost.

There are other theoretical models which study how the presence of natural resources affects international trade, focusing on factors such as comparative advantages and relative prices (Brander and Taylor, 1997a, 1997b, 1998a, 1998b), or differences in property rights of the resources (Chichilnisky, 1994). This chapter proposes a different approach, as the natural resource has an influence not only on international trade, but also on the distribution of firms among countries, which is endogenously determined. In turn, the distribution of economic activity also affects the equilibrium stock of the natural resource.

The following results are obtained. After a decrease in any of the transport costs, firms decide to move to the country with the greatest domestic demand and market size. Despite the cost advantage of locating in the South due to the presence of the resource, firms prefer to move to the North, the rich country, where they can take more advantage of increasing returns. In turn, concentration improves the economic growth rate, given the national nature of the spillovers. The concentration of firms in the North also has a positive effect on the stock of the natural resource, which increases. This means that in the framework of our model, second nature causes (the home market effect), acting centripetally, have greater weight in firm decisions than the advantages of natural geographic circumstances (first nature causes) which act centrifugally.

However, the South can increase the importance of the first nature cause by introducing public policies to reinforce the cost advantage of the resource's presence for firms located in the South. We will consider two public policies: imposing restrictions on the international trade of the resource and promoting a technological change to a technology which uses the resource more intensively. In both cases, after such policies the South attracts firms from the North, producing decreases in the growth rate and in the stock of the natural resource in equilibrium. The effect on welfare remains undetermined.

The next section presents the basic characteristics of the theoretical model. Section 2.3 describes the market equilibrium of differentiated goods, with special attention given to the distribution of firms in the equilibrium. Section 2.4 describes the natural resource market and solves the corresponding equilibrium. Section 2.5 determines the steady state growth rate, which depends on geography, and also shows how economic growth in turn influences geography through income inequality. Once the general equilibrium is described, section 2.6 analyses the effect of changes in differentiated goods' and resource's transport costs. These transport costs can also be interpreted in terms of public policies, as seen in section 2.7. Finally, the chapter ends with the main conclusions.

2.2 The model

The diagram in Figure 2.7 describes schematically how the model works. We will consider two countries, North and South, which trade with each other. Both are identical except for their initial level of capital, K_0 in the North and K_0^* in the South, and the presence of a natural resource only in the South. Let us suppose that the North has a higher initial income level, such that $K_0 > K_0^*$. Both countries are inhabited by representative households playing the part of consumers, workers and researchers. There are L households, both in the North and in the South. Labour is mobile between sectors but immobile between countries.

Given that the model is nearly symmetrical, we will focus on describing the economy of the North (an asterisk denotes the variables corresponding to the South). The preferences are instantaneously nested CES, and intertemporally CES, with an elasticity of intertemporal substitution equal to the unit:

$$U_0 = \int_0^\infty \log \left[D(t)^\alpha Y(t)^{1-\alpha}\right] e^{-\rho \cdot t} dt, \quad 0 < \alpha < 1, \tag{2.23}$$

where ρ is the intertemporal discount rate, Y is the numerary good and D is a composite good which, in the style of Dixit and Stiglitz, consists of a number of different varieties:

$$D(t) = \left[\int_0^{N(t)} D_i(t)^{1-\frac{1}{\sigma}} di\right]^{\frac{1}{\left(1-\frac{1}{\sigma}\right)}}, \quad \sigma > 1, \tag{2.24}$$

N is the total number of varieties available, both in the North and the South. σ is the elasticity of substitution between varieties, and is also the demand price elasticity of the demand for each variety (assuming that N is high enough). Growth comes from an increase in the number of varieties.

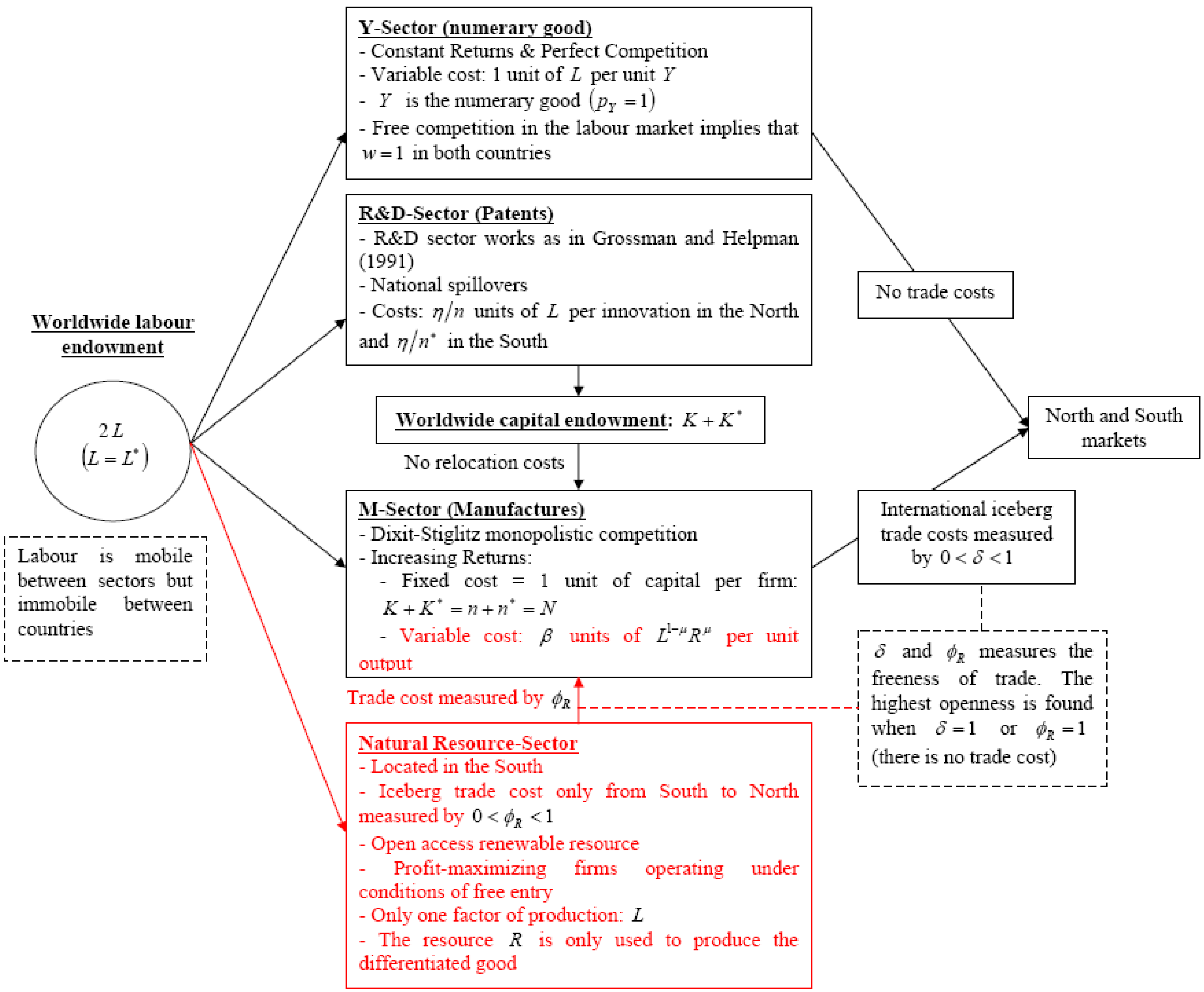


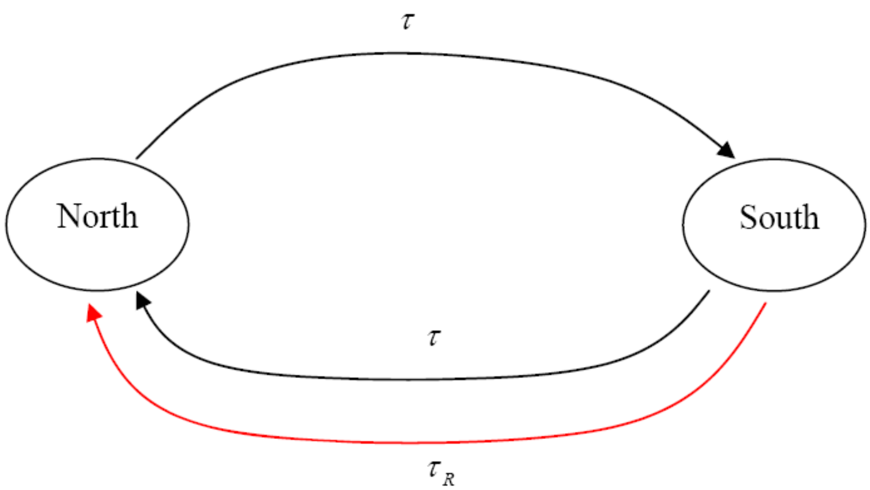
Fig. 2.7. Schematic diagram of the model.

Note that the natural resource does not appear explicitly in the structure of individual preferences, meaning that it lacks value for them (but it might have social value, as a planner might exist who decides to maintain a minimum level). This assumption is restrictive, but has a double justification. First, the indirect utility function is very difficult to analyse even without including the natural resource (see section 2.7). Including it would give rise to more indeterminacy (although, as we will see below, the resource does appear in the indirect utility function indirectly through its price). And second, individuals cannot move between countries. This means that they cannot react in any way to changes in the stock of the resource, and so introducing it into its utility function makes no sense.

The value of per capita expenditure E in terms of the numerary Y is:

$$\int_{i\in n} p_i D_i di + \int_{j\in n^*} \tau p_j^* D_j dj + Y = E. \qquad (2.25)$$

The number of manufactured goods produced in each country, n and n^*, is endogenous, with $N = n + n^*$. There is a transport cost $(\tau > 1)$ that affects international trade between the two countries. Also, international trading of the natural resource from South to North is also subject to a transport cost τ_R. τ and τ_R represent iceberg-type costs, as in Samuelson (1954), and reflect the part of the good which is lost in transit. The transport costs operate according to the following schema:



Thus, only $\tau^{-1} < 1$ of each unit of differentiated variety sent from the other country is available for consumption. Similarly, the North incurs an additional transport cost deriving from the natural resource (only $\tau_R^{-1} < 1$ of each unit of the natural resource sent from the South can be used) which the South does not bear. Decreases in τ or τ_R facilitate trade. From here we will assume that $\tau_R \leq \tau$; in other words, it is less costly or, at best, the same in terms of transaction costs to send units of the natural resource than the differentiated good[8]. Meanwhile, the numerary good is not subject to any transaction cost.

The numerary good is produced using only labour, subject to constant returns in a perfectly competitive sector. As labour is mobile between sectors, the constant returns in this sector tie down the wage rate w in each country at each moment. We assume throughout the chapter that the parameters of the model are such that the numerary is produced in both countries, that is, that the total demand for the numerary is big enough so as not to be

[8] The results are maintained even when transport cost for the resource is higher than that of the differentiated good, as long as the difference is not too great (it is a sufficient condition).

satisfied with its production in a single country. In this way, wages are maintained constant and identical in both countries. A unit of labour is needed to produce a unit of Y, so free competition in the labour market implies that $w = 1$ in both countries.

The differentiated goods are produced with identical technologies, in an industry with monopolistic competition with increasing scale returns in the production of each variety. To begin to produce a variety of a good, a unit of capital is needed; this fixed cost (FC) is the source of the scale economies. Labour (L) and natural resource (R) combine through a Cobb-Douglas type technology, $x_i = L_i^{1-\mu} R_i^{\mu}$, with a proportion $\mu \in (0,1)$ for the natural resource that represents how intensive the technology is in the use of the resource. This makes firm costs different if they are located in the North or South. If β represents the variable cost, the costs function of a representative firm in the North is as follows: $c_i = FC + \beta x_i q$, while that of a firm in the South, which does not have to bear transport costs for the natural resource (τ_R), is: $c_i = FC + \beta x_i q^*$, where q and q^* are the price indexes of the producers: $q = w^{1-\mu} (\tau_R p_R)^{\mu}$, and $q^* = w^{1-\mu} p_R^{\mu}$, and p_R is the market price of the natural resource. Therefore, firms in the South enjoy a competitive advantage in costs derived from the presence of the natural resource in its territory.

The standard rule of monopolistic competition determines the price of any variety produced either in the North or the South. The difference in costs implies that these prices are different: $p = \beta \left(\frac{\sigma}{\sigma-1}\right) (\tau_R p_R)^{\mu}$ in the North and $p^* = \beta \left(\frac{\sigma}{\sigma-1}\right) p_R^{\mu}$ in the South, where we have taken into account that $w = 1$ in both countries. Specifically, the price for any variety fixed by firms in the North is higher than the price fixed by firms in the South due to the additional transport costs for the natural resource that they bear ($p > p^*$ as $\tau_R > 1$).

The operating profits of the firms are also different depending on the country where they are located:

$$\pi = p_i x_i(p_i) - \beta x_i(p_i) q = \left(\frac{\beta x}{\sigma - 1}\right)(\tau_R p_R)^{\mu} \tag{2.26}$$

in the North, and

$$\pi^* = p_i^* x_i^*(p_i^*) - \beta x_i^*(p_i^*) q^* = \left(\frac{\beta x^*}{\sigma - 1}\right) p_R^{\mu} \tag{2.27}$$

in the South, where x and x^* are the production scale of a representative firm in the North and in the South, respectively.

In order to produce a new variety a previous investment is required, either in a physical asset (machinery) or an intangible one (patent). The concept of capital used in this chapter corresponds to a mixture of both types of investment. We assume that each new variety requires one unit of capital. Thus, the value of any firm is the value of its unit of capital. The total number of varieties and firms is determined by the aggregate stock of capital at any given time: $N = n + n^* = K + K^*$. Once the investment is made, each firm produces the new variety in a situation of monopoly and chooses where to locate its production, as there are no costs of relocating the capital from one country to the other. Unlike firms, households (workers/researchers/consumers) are immobile, so their income is geographically fixed, although the firms can move. In other words, if a firm owner decides to locate production in the country where he does not reside, he repatriates the profits.

Finally, we assume there is a safe asset which pays an interest rate r on units of the numerary, whose market is characterized by freedom of international movements $(r = r^*)$.

Solving the first order conditions of the problem of the consumer in the North we obtain the demands for each variety produced in the North (D_i), in the South (D_j), and for the numerary good:

$$D_i = \frac{\sigma - 1}{\beta\sigma} \cdot \frac{\left((\tau_R p_R)^{\mu}\right)^{-\sigma} \alpha E}{\left(n\left((\tau_R p_R)^{\mu}\right)^{1-\sigma} + n^* \delta \left(p_R^{\mu}\right)^{1-\sigma}\right)}, \tag{2.28}$$

$$D_j = \frac{\sigma - 1}{\beta\sigma} \cdot \frac{\tau^{-\sigma} \left(p_R^{\mu}\right)^{-\sigma} \alpha E}{\left(n\left((\tau_R p_R)^{\mu}\right)^{1-\sigma} + n^* \delta \left(p_R^{\mu}\right)^{1-\sigma}\right)}, \tag{2.29}$$

$$Y = (1 - \alpha)E, \tag{2.30}$$

where $\delta = \tau^{1-\sigma}$ is a parameter between 0 and 1 that measures the openness of trade: $\delta = 1$ represents a situation in which transport costs do not exist, while if $\delta = 0$ trade would be impossible due to the high transaction costs.

The intertemporal optimization of consumers implies that the growth rate of expenditure is given by the difference between the interest rate and the intertemporal discount rate: $\frac{\dot{E}}{E} = \frac{\dot{E}^*}{E^*} = r - \rho$. As we will show below, in the steady state, E and E^* will be constant, so $r = \rho$.

2.3 Equilibrium in the market of differentiated goods

The equilibrium in the differentiated goods market involves two issues. First, we have to determine x and x^*, the production scales in the equilibrium of a representative firm located in the North or in the South, respectively. Second, the distribution of firms between both countries is determined endogenously, depending directly on geography and transport costs. The geographical part of the model refers to the location of firms, as the population does not move between countries[9]

The location of firms in equilibrium is determined by four conditions. The first two refer to the fact that when differentiated goods are produced in both countries, total demand, from both North and South, for each variety (including transport costs) must equal supply. Thus, from (2.28) and (2.29):

$$x = \frac{\alpha L(\sigma-1)}{\beta\sigma} \cdot ((\tau_R p_R)^{\mu})^{-\sigma} \cdot \left(\begin{array}{c} \frac{E}{N\left(S_n((\tau_R p_R)^{\mu})^{1-\sigma} + (1-S_n)\delta\left(p_R^{\mu}\right)^{1-\sigma}\right)} + \\ + \frac{\delta E^*}{N\left(S_n\delta((\tau_R p_R)^{\mu})^{1-\sigma} + (1-S_n)\left(p_R^{\mu}\right)^{1-\sigma}\right)} \end{array} \right), \tag{2.31}$$

$$x^* = \frac{\alpha L(\sigma-1)}{\beta\sigma} \cdot (p_R^{\mu})^{-\sigma} \cdot \left(\begin{array}{c} \frac{E^*}{N\left(S_n\delta((\tau_R p_R)^{\mu})^{1-\sigma} + (1-S_n)\left(p_R^{\mu}\right)^{1-\sigma}\right)} + \\ + \frac{\delta E}{N\left(S_n((\tau_R p_R)^{\mu})^{1-\sigma} + (1-S_n)\delta\left(p_R^{\mu}\right)^{1-\sigma}\right)} \end{array} \right), \tag{2.32}$$

where $S_n = \frac{n}{N}$ is the share of varieties of the manufactured good produced in the North.

The third condition is the consequence of the free movements of capital between countries ($r = r^*$), which implies an equal retribution via profits:

$$\pi = \pi^*, \tag{2.33}$$

and, therefore, according to (2.26) and (2.27), $x = \frac{x^*}{\tau_R^{\mu}}$. Finally, the fourth condition, already mentioned, indicates that the total number of varieties is fixed by the worldwide supply of capital at each moment:

$$n + n^* = K + K^* = N. \tag{2.34}$$

---

[9] Population is tied to their native country, but individuals are affected by the location of firms, because the more firms in the country, the lower the price index they have to bear. The price indexes are:
$P = \left(\frac{\beta\sigma}{\sigma-1}\right) p_R^{\mu} \left[n\tau_R^{\mu(1-\sigma)} + n^*\delta\right]^{\frac{1}{1-\sigma}}$ in the North, and $P^* = \left(\frac{\beta\sigma}{\sigma-1}\right) p_R^{\mu} \left[n\delta\tau_R^{\mu(1-\sigma)} + n^*\right]^{\frac{1}{1-\sigma}}$ in the South.

Solving the system formed by these four equations, we obtain the optimum size of each firm in equilibrium in the North and in the South:

$$x = \frac{\alpha L(\sigma - 1)}{\beta\sigma} \cdot \frac{(E + E^*)}{N} \cdot (\tau_R p_R)^{-\mu}, \tag{2.35}$$

$$x^* = \frac{\alpha L(\sigma - 1)}{\beta\sigma} \cdot \frac{(E + E^*)}{N} \cdot p_R^{-\mu}. \tag{2.36}$$

The equilibrium production scales are different in each country. Locating in the North implies an additional cost due to the transport of the natural resource, and the firms react by producing fewer units of the differentiated good that they sell at a higher price. In turn, this different behaviour is what enables profits obtained in equilibrium to be the same in both countries.

The proportion of firms (or varieties) in the North $(S_n = \frac{n}{N})$ is given by:

$$S_n = \frac{S_E}{(1 - \delta \cdot \phi_R)} - \frac{\delta\,(1 - S_E)}{(\phi_R - \delta)}, \tag{2.37}$$

where, in turn, $S_E = \frac{E}{E+E^*}$ is the participation of the North in total expenditure and $\phi_R = \tau_R^{\mu(1-\sigma)}$ is a parameter between 0 and 1 of similar interpretation to δ, measuring the freedom of trade of the natural resource. It is also possible to demonstrate that, as long as the North has a larger domestic market[10] $\left(S_E > \frac{1}{2}\right)$, most firms are located in the North $\left(S_n > \frac{1}{2}\right)$.

The location of equilibrium of the firms depends on national expenditure – higher local expenditure or income means a larger domestic market, which attracts more firms wanting to take advantage of increasing returns (home market effect) – and the relationship between the level of openness of trade of differentiated goods (δ) and of the natural resource

[10] Below it is shown that this condition is always borne out as long as $K_0 > K_0^*$, as we have supposed.

(ϕ_R). The natural resource influences the distribution of firms in equilibrium via ϕ_R: the lower the transport cost of the natural resource, the smaller the advantage for firms located in the South. It is easy to see that $(\phi_R - \delta) > 0$ as long as $\tau_R \leq \tau$. Given that most firms are concentrated in the North, the home market effect, which we may identify as a second nature cause, acts centripetally, favouring the agglomeration of economic activity, while the cost advantage offered by the natural resource to firms located in the South, the first nature cause, acts centrifugally.

2.4 Natural resource growth

The South is endowed with a stock of natural resource (S), characterized as in Eliasson and Turnovsky (2004) or in Brander and Taylor (1997a, 1997b, 1998a, 1998b). This natural resource has some specific characteristics. It is (i) renewable, (ii) open access, (iii) used only as an input in the production of manufactured goods, and (iv) its exploitation requires only labour. These four conditions can be considered as restrictive, but are necessary to keep the model tractable. A natural resource with such characteristics is, for example, the wood from the forests of the South.

At any point of time, the net change in the stock of the resource is given by $\overset{\bullet}{S} = G(S) - R$, where $G(S)$ describes the natural growth of the resource and R is the harvested amount. We assume that the reproduction function G is a concave function depending on the current stock of the resource, and positive in the interval between $\underline{S}$ and $\overline{S}$, where $\underline{S}$ is the minimum viable stock size and $\overline{S}$ is the maximum amount which the stock can reach, given physical and natural limitations (for example, available space). $G(S)$ is analogous to a production function, with the difference that the rate of accumulation of the stock is limited. See Brown (2000) for a wider discussion of $G(S)$ and its properties.

For simplicity, we fix $\underline{S} = 0$ and assume that the growth of the resource, $G(S)$, corresponds to a logistic function:

$$G(S) = \gamma S \left(1 - \frac{S}{\overline{S}}\right), \quad \gamma > 0, \tag{2.38}$$

where γ is the intrinsic growth rate of the resource (the natural growth rate). In the absence of harvesting $(R = 0)$, S converges to its maximum sustainable stock level, $\overline{S}$. This function has been widely used in the analysis of renewable resources, and may be the simplest and most empirically plausible functional form of describing biological growth in a restricted environment.

The harvest of the natural resource requires economic resources; for the sake of simplicity, we will assume it requires only labour. We assume that harvesting is carried out according to the Schaefer harvesting production function:

$$R^S = BSL_R, \tag{2.39}$$

where L_R is the amount of labour used in the renewable resource sector (workers in the South, where the resource is located), R^S is the harvested quantity offered by the producers and B is a positive constant. If $a_{L_R}(S)$ represents the unit labour requirement in the resource sector, (2.39) implies that $a_{L_R}(S) = \frac{L_R}{R^S} = \frac{1}{BS}$. It verifies $a'_{L_R}(S) < 0$: labour requirement increases as the stock of the resource decreases.

Production is carried out by profit-maximizing firms operating under conditions of free entry (perfect competition). Therefore, the price of the resource good must equal its unit production cost:

$$p_R = wa_{L_R} = \frac{w}{BS} = \frac{1}{BS}. \tag{2.40}$$

Both B and w are in terms of the numerary good, so p_R is too. This price incorporates the assumption of open access to the resource, because the only explicit production cost is labour. There are no other explicit costs of using the resource[11].

The firms in the sector of the differentiated goods demand the natural resource as an input in the production of their varieties. Applying Shephard's lemma to the cost functions we obtain the demand for the natural resource: $\beta x \cdot \mu \left(\tau_R p_R\right)^{\mu-1}$ for a representative firm of the North and $\beta x^* \cdot \mu p_R^{\mu-1}$ for a representative firm of the South. Substituting the equilibrium production levels given by (2.35) and (2.36), and aggregating for the firms in the North (taking into account the transport cost they bear) and in the South, we obtain the worldwide demand for the resource $\left(R^D\right)$:

$$R^D = \mu p_R^{-1} \cdot \frac{\alpha(\sigma-1)}{\sigma} \cdot L\left(E+E^*\right). \tag{2.41}$$

This demand depends on some structural parameters, the price of the resource and world aggregate income, $L\left(E+E^*\right)$.

Replacing in (2.41) the price set by the producers, given by (2.40), we obtain the *resource market equilibrium condition*, which gives us the equilibrium harvest level R:

$$R\left(S\right) = \mu BS \cdot \frac{\alpha(\sigma-1)}{\sigma} \cdot L\left(E+E^*\right). \tag{2.42}$$

[11] If there were no free access to the resource, another cost would exist deriving from the reduction of the capacity for reproduction of the resource, which relates to Hotelling's rule. The resource would be exploited only by firms with property rights in a situation which would then not be perfect competition, making the final price greater than the unit cost, and generating additional income.

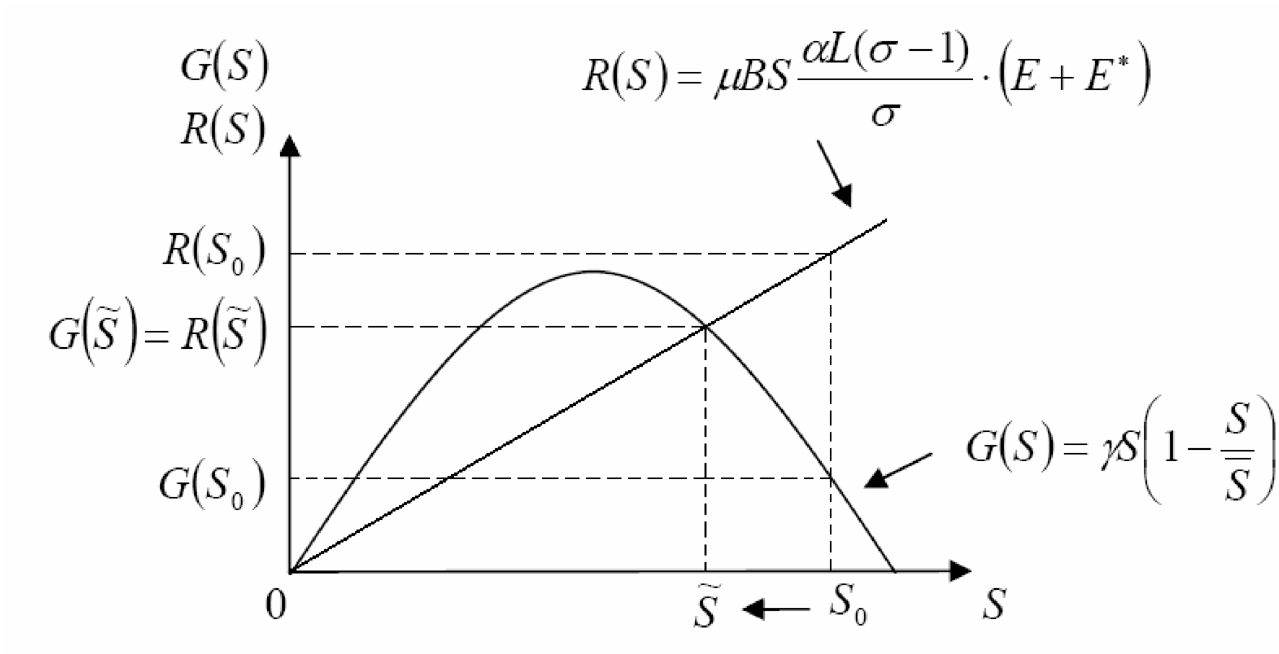


Fig. 2.8. Dynamics of the resource.

Note that this harvest level R is a function $R(S)$ that grows with the size of the stock S. Steady state is reached when the stock evolves to a level in which the harvest of the natural resource, $R(S)$, is equal to its capacity for reproduction, $G(S)$, given by equation 2.38, meaning that $\dot{S} = G(S) - R(S) = 0$. A trivial solution is reached when $S = R = 0$. The other solution is given by:

$$S^* = \overline{S}\left[1 - \mu B \cdot \frac{\alpha(\sigma - 1)}{\gamma\sigma} \cdot L(E + E^*)\right]. \tag{2.43}$$

Figure 2.8 shows how convergence is produced to the steady state level. The figure illustrates a situation in which at the initial stock S_0 the amount harvested, $R(S_0)$, exceeds natural growth, $G(S_0)$. The stock then decreases until it reaches the steady state level $\tilde{S}$. This indicates that, in steady state, the quantity of the resource used by firms is constant.

The steady state harvest level is obtained by replacing $\tilde{S}$ in $R(S)$:

$$R(S^*) = \mu B \cdot \frac{\alpha(\sigma-1)}{\sigma} \cdot L(E+E^*) \cdot \overline{S} \left[1 - \mu B \cdot \frac{\alpha(\sigma-1)}{\gamma\sigma} \cdot L(E+E^*)\right]. \quad (2.44)$$

As shown by Brander and Taylor (1997a), a positive steady state solution exists if and only if the term between brackets is positive, that is to say, if the condition $\mu B \cdot \frac{\alpha(\sigma-1)}{\sigma} \cdot (E+E^*) < \frac{\gamma}{L}$ holds. In this case the solution is globally stable (for any $S_0 > 0$). If such condition is not satisfied the resource would disappear and the unique possible steady state is $S = R = 0$. Graphically, this condition means that, in the origin, the slope of the function $R(S)$ is less than the slope of $G(S)$, thus ensuring that they cut off at some point for positive values of S.

2.5 Economic growth and income inequality

2.5.1 Economic growth

We will first examine the growth rate of the economy. Starting from the solution of the problem of the intertemporal optimization of the consumer, we know that, in equilibrium, $\frac{\dot{E}}{E} = \frac{\dot{E}^*}{E^*} = r - \rho$. As the capital flows are free, $r = r^*$, and the expenditure growth rate will be the same in both countries. From (2.37), this implies that the ratio of firms producing in the North, S_n, is also constant in time, and, therefore, n, n^* and N grow at the same constant rate $g = \frac{\dot{N}}{N} = \frac{\dot{n}}{n} = \frac{\dot{n}^*}{n^*}$.

National spillovers exist in the innovation sector, so that the more firms producing different manufactured goods are located in the same country, the less costly is R&D[12]. This

[12] This type of knowledge spillovers is closer to the concept of Jacobs (1969) than to that of Marshall-

sector follows Grossman and Helpman (1991), with $\frac{\eta}{n}$ being the cost in terms of labour of an innovation in the North and $\frac{\eta}{n^*}$ in the South. The immediate conclusion of this formulation of the sector is that, for reasons of efficiency, research activity will take place in only one of the two countries: the one with the most firms producing manufactured goods (which will be the rich country, the North, given that $S_n > \frac{1}{2}$). No researcher will have any incentive to begin R&D in the other country. This formulation makes the analytical treatment of the model easier, although the results are maintained even if a certain degree of diffusion of the knowledge exists at the international level (Hirose and Yamamoto, 2007).

The value of the firm is given by the value of its unit of capital. As the capital market is competitive, this value (v) will be given by the marginal cost of innovation, $v = \frac{\eta}{n} = \frac{\eta}{NS_n}$, which is therefore decreasing at the rate g, the rate of innovation $(\frac{\dot{v}}{v} = -g)$. As the number of varieties increases, the profits of each firm decrease, and also does its value, which can also be interpreted as the future flow of discounted profits $\left(v(t) = \int_t^{\infty} e^{-[\bar{r}(s)-\bar{r}(t)]} \frac{\beta x(s)}{\sigma-1} ds\right)$, where $\bar{r}$ represents the cumulative discount factor. Taking into account the *arbitrage condition between the capital market and the safe asset market*, the relation between the interest rate and the value of the capital is given by [13]:

$$r = \frac{\dot{v}}{v} + \frac{\pi}{v}. \tag{2.45}$$

On the other hand, the constraint of world resources, $E + E^* = 2 + (r\eta) / (LS_n)$, where the right-hand includes the sum of labour income ($w = 1$ in the two countries) and

Arrow-Romer (MAR). The empirical evidence for these external effects between different industries in the same geographical unit is documented; see, for example, Glaeser et al. (1992) and Henderson et al. (1995).

13 This condition is formulated in terms of the profits of the firms in the North (π), but applies in the same way to the South because, although the expressions of π and π^* differ (equations 2.26 and 2.27), one of the conditions of equilibrium (equation 2.33) requires that $\pi = \pi^*$.

capital returns, implies that worldwide expenditure is constant over time, so that in steady state $r = \rho$, as pointed above. Note that this restriction includes only labour and capital returns; the harvest of the natural resource does not generate additional income for either of the two countries, as it is an open access resource exploited in a competitive industry.

Finally, we must take into account the labour market. The world's labour is devoted to R&D activities (using only workers from the North), and to the production of goods. From the latter, a proportion $(1-\alpha)$ is dedicated to the production of the numerary good, and a proportion α to the production of differentiated goods. In turn, given the Cobb-Douglas technology properties, from the labor used, either directly or indirectly, in the production of manufactured goods, a proportion μ is used in the exploitation of the resource (using only workers in the South), and a proportion is used directly as an input in the production of varieties. Thus, the *world labour market equilibrium condition* is given by:

$$\eta \frac{g}{S_n} + \left(\frac{\sigma - \alpha}{\sigma}\right) L(E + E^*) = 2L. \tag{2.46}$$

In steady state (see details in Appendix A), all the variables will grow at a constant rate. Replacing in (2.45) the profits obtained in (2.26), the optimum size of firms in the equilibrium (2.35), and considering (2.46) and that in steady state $r = \rho$, we obtain the *labour and capital markets equilibrium condition*:

$$g = \frac{2L}{\eta} \cdot \frac{\alpha}{\sigma} S_n - \left(\frac{\sigma - \alpha}{\sigma}\right) \rho = g\left(S_n\right), \tag{2.47}$$

where g is the growth rate of K and K^* (the same for the two countries) in steady state[14]. This rate depends on structural parameters of the model ($L, \eta, \alpha, \sigma, \rho$), but also on S_n (the geography), lineally.

2.5.2 World income distribution

Secondly, we are interested in how this economic growth rate affects income inequality between the countries. Remember that we assumed the North to be richer initially ($K_0 > K_0^*$). The per capita income of each country is the sum of labour income (which, as we have already seen, is the unit), plus the capital income, which is r times the value of per capita wealth. Thus, it will be $E = 1 + r\frac{Kv}{L} = 1 + \rho\frac{Kv}{L}$ for any individual in the North. If we replace v from the *arbitrage condition between the capital market and the safe asset market* (2.45), the equilibrium profits (2.26), and the optimum production scale (2.35), it is possible to express Northern expenditure as a function of g:

$$E = 1 + \frac{2\alpha\rho S_K}{(\sigma - \alpha)\rho + \sigma g}, \tag{2.48}$$

where $S_K = \frac{K}{K+K^*}$ is the share of capital owned by the individuals in the North, that remains constant because K and K^* grow at the same rate g in the steady state.

Similarly, for the South:

$$E^* = 1 + \frac{2\alpha\rho(1 - S_K)}{(\sigma - \alpha)\rho + \sigma g}. \tag{2.49}$$

We have previously defined the ratio $S_E = \frac{E}{E+E^*}$, which represents the participation of the North in total income or expenditure. Replacing the expressions (2.48) and (2.49)

[14] Again the results are presented in terms of the variables of the North (π and x). Using π^* and x^* the same result is obtained (the steady state economic growth rate is the same for the two countries), taking into account that in equilibrium $\pi = \pi^*$, meaning that $x = \frac{x^*}{\tau_R^\mu}$.

we obtain:

$$S_E = \frac{1}{2} \cdot \frac{\sigma(\rho+g) + \alpha\rho(2S_K - 1)}{\sigma(\rho+g)}. \tag{2.50}$$

If, as we have supposed, the North is richer and $S_K > \frac{1}{2}$, then $S_E > \frac{1}{2}$. However, the relationship of S_E with the economic growth rate is negative: as the number of varieties increases, the value of the capital is reduced, and, as the North individuals own more capital, the income difference is reduced in relative terms.

Finally, to carry out the analysis of the next section, we need to relate the geography (S_n) with the growth rate g. To do this, we replace (2.50) in (2.37), obtaining the *differentiated goods market equilibrium condition*, indicating the distribution of firms for each value of g:

$$S_n(g) = \frac{\frac{1}{2}}{(1-\delta\cdot\phi_R)(\phi_R-\delta)} \cdot \left[\begin{array}{c} (1+\delta^2)\,\phi_R - 2\delta + \\ +\,(1-\delta^2)\,\phi_R \cdot \frac{\alpha\rho(2S_k-1)}{\sigma(\rho+g)} \end{array} \right] = S_n\left[S_E(g)\right]. \tag{2.51}$$

2.5.3 Equilibrium

We have obtained two equations, (2.47) and (2.51), representing, respectively, the *labour and capital markets equilibrium condition* and the *differentiated goods market equilibrium condition*. These functions relate the growth rate with the spatial distribution of firms, and define the equilibrium values of these variables. Since the algebraic solution is not easy, we follow a graphical approach.

The function $g = g(S_n)$ is linear and increasing: given the nature of the technological spillovers (national), the greater the concentration of firms, the lower the costs of innovation and the higher the growth rate. The function $S_n = S_n(g)$ is convex and decreasing[15].

[15] $S_n = S_n(g)$ is convex and decreasing as long as $(\phi_R - \delta) > 0$. This condition is verified if $\tau_R \leq \tau$, as

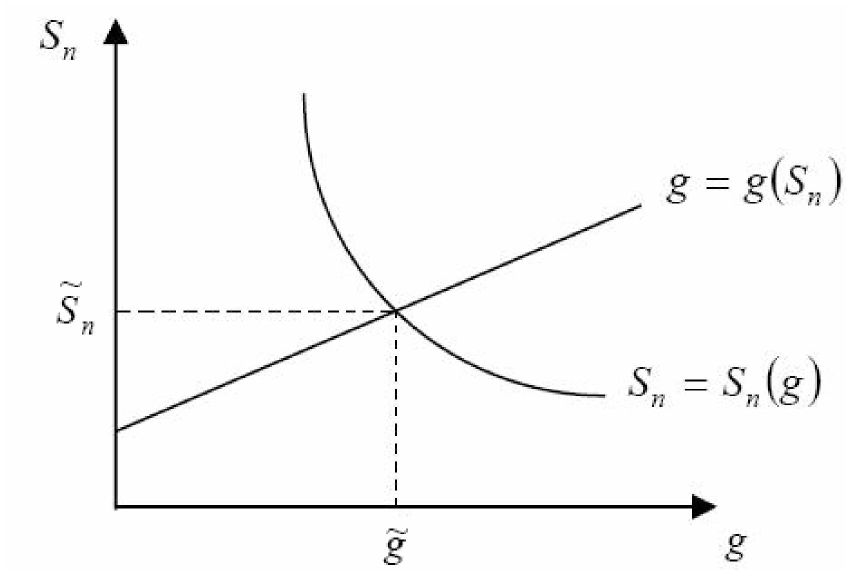


Fig. 2.9. *Labour and capital markets equilibrium condition* ($g = g(S_n)$), and the *differentiated goods market equilibrium condition* ($S_n = S_n(g)$).

Remember that this equation incorporates the inequality of income, $S_n = S_n[S_E(g)]$, and that this decreases as g increases via the reduction of monopolistic profits of firms. At the same time, as the differences in income vanish, industrial concentration and the market size of the rich country decrease due to the home market effect. These functions are represented in Figure 2.9. The intersection point determines the steady state location of firms as well as the growth rate of the economy.

2.6 Effects of reducing trade costs

As we explained in the introduction, the purpose of this chapter is explicitly to study the effect of first nature causes on the concentration of economic activity, analyzing one of the

---

we have been assuming from the begining. Additionally, $(\phi_R - \delta)$ is greater than zero even when transport cost for the resource is higher than that of the differentiated good, as long as the difference is not too great.

possible natural geographical characteristics, the role which may be played by a natural resource.

Starting from the equilibrium situation, a change in differentiated goods' or natural resource's transport cost will lead to changes in the distribution of firms. Firms move according to two types of incentives: the North attracts firms thanks to its larger domestic market, $S_E > \frac{1}{2}$, which we can identify as one of the second nature causes of concentration of firms, while the first nature causes in our model refer to the advantage in costs enjoyed by firms in the South thanks to the geographical presence of the natural resource in its territory.

Variations in any type of transaction cost do not affect the function $g = g(S_n)$, which depends only on the structural parameters of the model. It is the curve $S_n = S_n(g)$ which will reflect the changes in transport costs, moving and changing its slope. We carry out our analysis, first, from the perspective of the effects that decreasing transport costs have on the industrial localization and the growth rate. Then, the effect on the equilibrium stock of the resource is analyzed.

2.6.1 Effects on industrial concentration and economic growth

Decrease in the transport cost of differentiated goods

Let us consider first a decrease in the differentiated goods trading cost: $d\tau < 0$. After differentiating the equations (2.47) and (2.51), we obtain that $\frac{dS_n}{d\tau} < 0$, $\frac{dg}{d\tau} < 0$ and, thus, both the proportion of firms located in the North and the economic growth rate increase. This situation is represented in Figure 2.10.

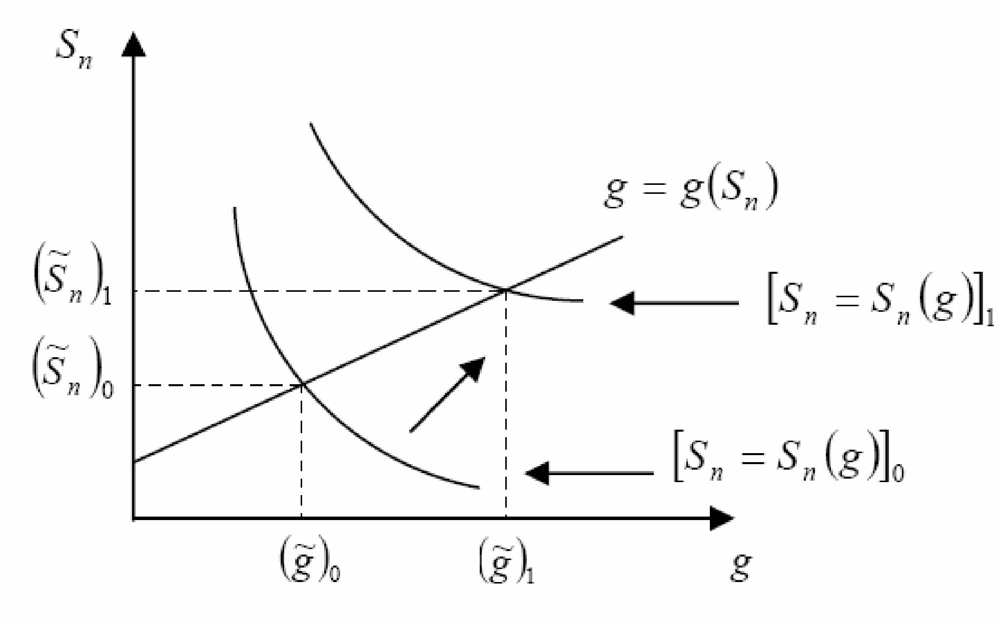


Fig. 2.10. Effect of a reduction in the transport cost of differentiated goods.

The decrease in transaction costs enables an easier access to the market of the other country, so some firms prefer to move to the North (remember that there are no relocation costs). Despite the cost advantage of locating in the South due to the presence of the natural resource, firms prefer to move to the North, the rich country and thus the bigger market, where they can take more advantage of increasing returns. This means that, in the framework of our model, the home market effect (second nature causes), acting centripetally, have a greater weight in firm decisions than the advantages of natural geographic circumstances (first nature causes), which act centrifugally.

In turn, concentration speeds up the economic growth rate, because the more manufacturing firms are located in the North, the lower the cost of innovation given the national nature of the spillovers.

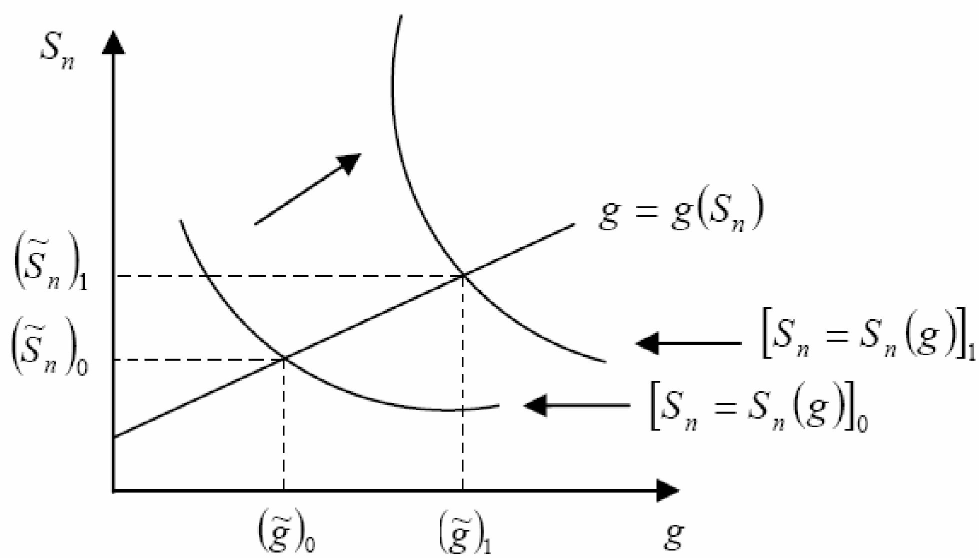


Fig. 2.11. Effect of a reduction in the transport cost of the resource.

Decrease in the transport cost of the resource

If the transport cost of the natural resource decreases, $d\tau_R < 0$, we obtain that $\frac{dS_n}{d\tau_R} < 0$, $\frac{dg}{d\tau_R} < 0$. Thus, both the proportion of firms located in the North and the economic growth rate rise: Figure 2.11 shows this situation. The difference from Figure 2.10 is that, in this case, the slope of the curve $S_n = S_n(g)$ moves upwards rather than downwards.

The lower transport cost of the natural resource means a loss in the cost advantage of the firms located in the South, close to the natural resource, over those located in the North. At the limit, if this transport cost did not exist $(\tau_R = 1)$ the firms could not extract any advantage from its location close to the resource and there would be no relationship between the distribution of natural resource and the economic geography. In other words, as the transport cost of natural resources decreases, the importance of the first nature cause (in our model, the natural resource) vanishes.

As a consequence of this decrease in relative costs in the North, firms move from the South to the North, which has a bigger domestic market and greater demand. Moreover, as the number of firms in the North increases, the cost of research decreases due to national spillovers, and the economic growth rate increases.

2.6.2 Effects on the stock of the natural resource

Any variation in the distribution of firms or in the economic growth rate, whether due to a change in the transport cost of differentiated goods or of the resource, will have an effect on the stock level of the resource in steady state. That is, changes in the geographical distribution of firms affect the market of the natural resource.

Let us remember that both the harvest level, given by the *resource market equilibrium condition* (equation 2.42), and the stock of the resource in equilibrium (equation 2.43), depend on aggregate world income $L\left(E+E^{*}\right)$. In turn, world income can be related to S_n and g, replacing in (2.45) the profits obtained in (2.26) and the optimum size of firms in the equilibrium (2.35):

$$L(E+E^{*})=\frac{\eta\sigma(\rho+g)}{\alpha S_n}. \tag{2.52}$$

If we replace this expression of world income in (2.42) and (2.43) we obtain:

$$S=\overline{S}\left[1-\mu B\cdot\frac{(\sigma-1)}{\gamma}\cdot\frac{\eta(\rho+g)}{S_n}\right], \tag{2.53}$$

$$R\left(S\right)=\mu BS\cdot(\sigma-1)\cdot\frac{\eta(\rho+g)}{S_n}. \tag{2.54}$$

From these expressions we can analyse the effects on the natural resource of the changes in the distribution of firms. Let us consider changes in the transport costs that lead to a higher proportion of firms located in the North $(dS_n > 0)$, that is, reductions in the transport cost of either the intermediate goods or the natural resource. In turn, given the national nature of the R&D spillovers, the higher concentration of firms in the North reduces the cost of innovation and raises economic growth: $dg > 0$. So, by differentiating (2.53), we obtain the effect of the reduction in transport costs on the stock of the natural resource in steady state:

$$dS = -\overline{S}\mu B \cdot \frac{(\sigma - 1)}{\gamma} \cdot \eta \frac{1}{S_n} \left[dg - \frac{1}{S_n} (\rho + g)\, dS_n \right].$$

This expression enables us to identify two opposite effects:

a) Industry localization effect: As the number of firms located in the North increases, the amount of the resource which is harvested decreases, because the firms in the North produce less units of differentiated good $(x < x^*)$ and thus require less natural resource.

b) Growth effect: As the number of firms in the North increases, the growth rate of the number of varieties also increases, so that the number of firms grows faster. More firms require a higher aggregate amount of the natural resource.

However, applying that, from (2.47), $dg = \frac{2L}{\eta} \cdot \frac{\alpha}{\sigma} dS_n$, it is possible to obtain a clear sign:

$$dS = -\overline{S}\mu B \cdot \frac{(\sigma - 1)}{\gamma} \cdot \eta \frac{1}{S_n^2} \left[\frac{-\alpha}{\sigma} \rho \right] dS_n > 0,$$

indicating that the firms localization effect dominates: more firms in the North means that less resource is consumed on average, enabling the level of stock to increase in steady state.

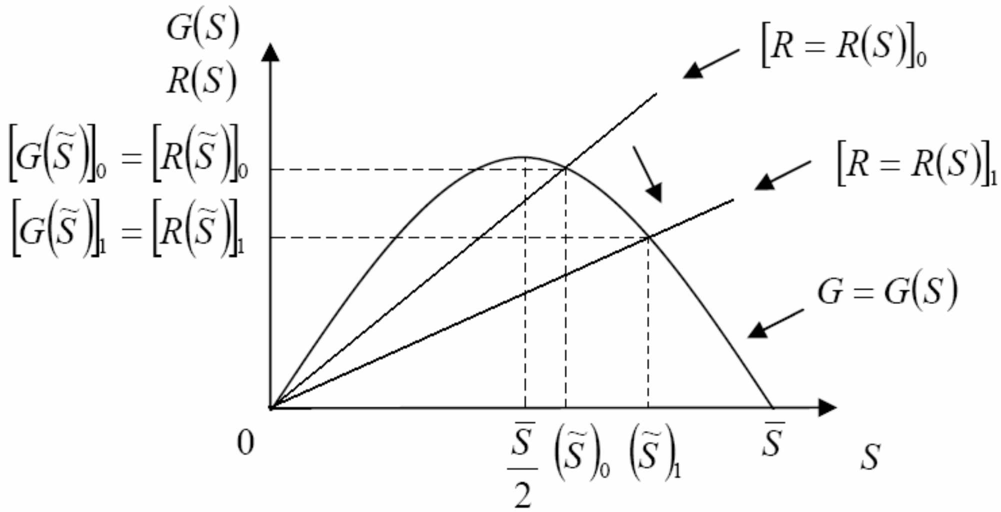


Fig. 2.12. Evolution of the stock of resource when concentration of firms in the North increases $(dS_n > 0)$: Case in which $\tilde{S} > \frac{\overline{S}}{2}$.

On the other hand, the effect on the harvested amount is not clearly determined. If we differentiate (2.54), and replace dg and dS with the expressions obtained earlier, we have:

$$dR = \mu B(\sigma - 1)\eta \frac{2}{S_n^2}\left(\frac{-\alpha}{\sigma}\rho\right)\left[S - \frac{\overline{S}}{2}\right] dS_n \gtrless 0.$$

The sign of the above expression depends on $S - \frac{\overline{S}}{2}$, that is, on whether the initial steady state stock exceeds or not $\frac{\overline{S}}{2}$. The same conclusion can be obtained if we differentiate the function $G(S)$ (equation 2.38). Graphically, it depends on whether $\tilde{S}$ is on the increasing or decreasing part of $G(S)$. Figures 2.12 and 2.13 illustrate the two possibilities.

In Figure 2.12 we consider the case $\tilde{S} > \frac{\overline{S}}{2}$, meaning that $dR < 0$ after the reduction in transport costs. In this situation, the increasing number of firms in the North is accompanied by a decrease in the amount harvested. This will be the most common solution, as it corresponds to situations where the slope of the function $R(S)$ is low. From (2.54), this

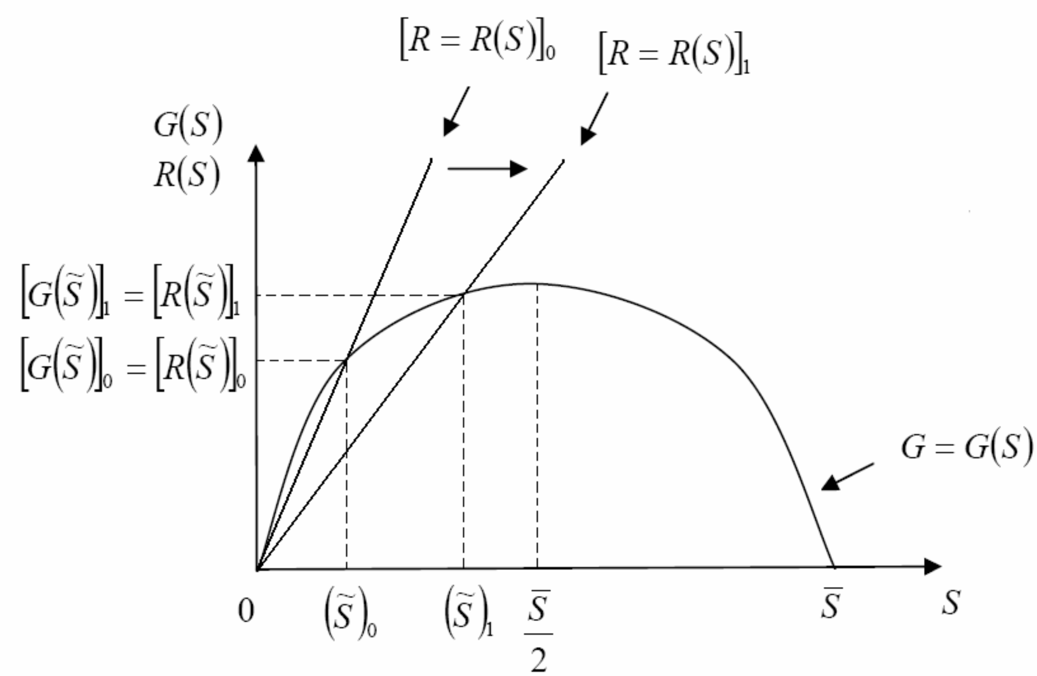


Fig. 2.13. Evolution of the stock of resource when concentration of firms in the North increases $(dS_n > 0)$: Case in which $\tilde{S} < \frac{\overline{S}}{2}$.

is the more probable case when the industry is highly concentrated in the North and/or the technology of the intermediate good firms is not very intensive in the use of the natural resource. In contrast, if the function $R(S)$ is very steep and $\tilde{S} < \frac{\overline{S}}{2}$, the amount harvested increases ($dR > 0$). This case is represented in Figure 2.13, and corresponds to situations where, despite consuming more resource, the equilibrium stock increases due to the high capacity of regeneration of the natural resource on this side of the curve $G(S)$.

2.7 Public policies: How to protect the South's natural advantage?

In the previous section we analyzed the effects of decreases in transport costs, obtaining as a result an increase in industrial concentration in the North, the rich country, and an increase in the growth rate and the stock level of the resource in steady state. Such lower transport cost of the natural resource meant that the firms of the South lost some of the cost advantage due to the closer location of the natural resource. That is, as the transport cost of the resource decreases, the less important this first nature cause becomes, configured as a centrifugal force, and the more firms concentrate in the North.

From this point of view, there is not much the South can do faced with a rich North with the home market effect in its favour, in a context of international transport costs trending downwards over time, so that sooner or later the cost advantage will disappear. However, the South can consider some public policies in order to protect the cost advantage.

2.7.1 Restrictions on international trading of the resource

A first route, the most direct, would be to influence τ_R, since higher transport costs for the resource increase the cost advantage for firms in the South. By modifying slightly the interpretation of the parameter τ_R, we can consider some ways the South could protect and even increase the cost advantage given by nature.

Martin and Rogers (1995) posited that the transport costs used in the models of Economic Geography can alternatively be interpreted as a measure of the quantity and quality of transport infrastructures, and, thus, can be modified by public policies. From this point of view, they defined public transport infrastructures as any good or service provided by the state which can facilitate the connection between production and consumption. It is evident that transport and communication media can be included among these trade infrastructures, but there are other non-physical elements, such as the legal system or the levels of public safety, which have an equally great influence on trade. Good infrastructures mean low transaction costs; poor infrastructures represent a situation where trade is difficult because of the high costs incurred. From this wide sense of the term, the parameter ϕ_R becomes an index between 0 and 1 which measures the level of infrastructures and/or legal restrictions related to the natural resource trade. The best (worst) quality in trade infrastructures is found when $\phi_R = 1$ (0). Such is also the case when there are no legal restrictions for trade of the natural resource.

In this way, the South could act through public policies and reinforce the cost advantage of Southern firms by restricting the international trade of the natural resource. The easiest way can be the introduction of exportation tariffs. The more difficult it is to access the natural resource from outside, the more firms will decide to locate in the South. This will enable to attract firms from the North, which would in turn cause a reduction in the growth rate and in the stock of the natural resource in equilibrium (because the firms in the South use more quantity of the natural resource than those in the North).

2.7.2 Technological change

There is another parameter that can influence the importance of the cost advantage which the natural resource gives to firms in the South. This is μ, which measures the degree in which the technology of the differentiated goods sector is intensive in the use of the natural resource.

Specifically, the more dependent the technology is on the natural resource, the greater the cost advantage of locating production in the South. If the South could use some kind of public policy, such as subsidising firms, to promote a change to a technology that used the resource more intensively, this would reinforce the cost advantage of its firms.

This policy can be represented as an increase in the parameter μ $(d\mu > 0)$. After differentiating the equations (2.47) and (2.51), we find that this leads to a decrease in the proportion of firms located in the North, S_n, as well as in the economic growth rate g, due to the national nature of the spillovers: $\frac{dS_n}{d\mu} < 0$, $\frac{dg}{d\mu} < 0$.

The equilibrium stock of the resource also decreases. Differentiating (2.53), and taking into account that, from (2.47), $dg = \frac{2L}{\eta} \cdot \frac{\alpha}{\sigma} dS_n$, we have:

$$dS = -\overline{S}B \cdot \frac{(\sigma - 1)}{\gamma} \cdot \eta \frac{1}{S_n} \left[(\rho + g) d\mu - \frac{\mu}{S_n} \cdot \frac{\alpha}{\sigma} \rho dS_n \right] < 0.$$

The effect on the harvested amount in equilibrium is again not clearly determined, depending on whether $\tilde{S}$ is in the increasing or decreasing side of the function $G(S)$.

Meanwhile, the effect on the variables would be the opposite if the North were to try to reduce firms' technological dependence $(d\mu < 0)$ on the natural resource not present in its territory. In this case the concentration of firms in the North and the economic growth rate would increase.

It is not difficult to find examples of this kind of policies, carried out by countries either to protect their advantages associated to the presence of natural resources, either to reduce the dependence in the case of importers. The case of oil, although it is not a renewable open access natural resource, is possibly the more representative. On one hand, the producers try to protect the profits derived from its exploitation by controlling (even reducing) the international availability of the input. On the other hand, the countries which have to import the resource promote changes in the technology and research in substitute inputs in order to reduce its dependence.

2.7.3 What about utility?

The two types of policies proposed above strengthen the influence of the first nature cause, leading firms to move from the North to the South. A question that arises at this point is whether such change would be desirable.

In order to try to answer this question, we analyse the indirect utility functions. Although it is difficult to carry out a rigorous analysis of welfare, given that any variation in the distribution of firms (the ratio S_n) has several different effects on the indirect utility function, with the global sign remaining undetermined, we can identify the different effects that consumers would experience in utility. The indirect utility function of a household in the North is given by:

$$V = \frac{1}{\rho} \ln \left\{ \begin{array}{c} \alpha^{\alpha} (1-\alpha)^{1-\alpha} \left(\frac{\sigma-1}{\beta\sigma}\right)^{\alpha} \left(\frac{1}{p_R^{\mu}}\right)^{\alpha} \left(1+\frac{\rho\eta S_k}{S_n L}\right) \cdot \\ \cdot N_0^{\frac{\alpha}{\sigma-1}} \left(S_n \left(\phi_R-\delta\right)+\delta\right)^{\frac{\alpha}{\sigma-1}} e^{\frac{\alpha g}{\rho(\sigma-1)}} \end{array} \right\}. \tag{2.55}$$

As we remarked above, although the natural resource does not appear explicitly in consumer preferences (equation 2.23), it influences the indirect utility function indirectly through its price p_R. If we replace p_R from (2.40), the utility function becomes:

$$V = \frac{1}{\rho} \ln \left\{ \begin{array}{c} \alpha^{\alpha} (1-\alpha)^{1-\alpha} \left(\frac{\sigma-1}{\beta\sigma}\right)^{\alpha} (BS)^{\alpha\mu} \left(1+\frac{\rho\eta S_k}{S_n L}\right) \cdot \\ \cdot N_0^{\frac{\alpha}{\sigma-1}} \left(S_n \left(\phi_R-\delta\right)+\delta\right)^{\frac{\alpha}{\sigma-1}} e^{\frac{\alpha g}{\rho(\sigma-1)}} \end{array} \right\}. \tag{2.56}$$

The impact of a change in the concentration of firms[16] can be obtained by differentiating the above function with respect to S_n, taking into account that, from (2.47), $dg = \frac{2L}{\eta} \cdot \frac{\alpha}{\sigma} dS_n$, and considering the expression obtained earlier for the change in the natural resource stock $dS = \overline{S}\mu B \cdot \frac{(\sigma-1)}{\gamma} \cdot \eta \frac{1}{S_n^2} \left[\frac{\alpha}{\sigma}\rho\right] dS_n$:

$$\partial V = \begin{bmatrix} -\frac{\eta S_k}{S_n^2 L + \rho\eta S_n S_k} + \frac{2L\alpha^2}{\rho^2\eta\sigma(\sigma-1)} + \\ +\frac{\alpha}{\rho(\sigma-1)} \cdot \frac{(\phi_R-\delta)}{(S_n(\phi_R-\delta)+\delta)} + \frac{\alpha^2\mu^2(\sigma-1)\bar{S}B\eta}{S_n^2\sigma\gamma S} \end{bmatrix} \partial S_n \gtrless 0.$$

The effect on a Northern household welfare is undetermined. Besides the three effects obtained by Martin and Ottaviano (1999), in our model a fourth effect deriving from the price of the natural resource arises. Thus, if the South manages, using public policies, to attract firms from the North ($dS_n < 0$), not only the economic growth rate and the level of equilibrium stock of the resource will decrease. Consumers in the North also experience four effects on utility:

a) The first element of the above derivative captures the positive impact of a decrease in the growth rate on the wealth of Northern households. Since the concentration of firms in the North is reduced, the cost of R&D rises and the economic growth rate decreases. This leads to a rise in intermediate firms' monopolistic profits and, thus, per capita income increases in the North.

b) The second element represents the negative impact on the reduction of the growth rate, which implies a slower rate of introduction of new varieties of the intermediate good, on the utility of individuals due to their structure of preferences and the love-of-variety effect.

[16] This analysis of utility is partial, as we consider that the change in S_n is exogenous. In the concrete case that the cause of the variation in the concentration of firms were a change in τ_R or in μ, additional effects would exist that would increase indeterminacy. See Appendix B.

c) The third term captures the decrease in welfare due to rising trade costs for consumers in the North when S_n decreases, since a higher range of varieties have to be imported. This effect depends on the differential $(\phi_R - \delta)$. It is easy to see that $(\phi_R - \delta) > 0$ as long as $\tau_R \leq \tau$, as we supposed. Thus, a lower proportion of firms located in the North, imply that Northern consumers will bear higher transport costs.

d) The last element represents the negative effect of a lower concentration of firms in the North on the price of the natural resource. As the proportion of firms in the North decreases, so does the stock of the natural resource in equilibrium, $\frac{dS}{dS_n} > 0$, and this leads to an increase in its price (equation 2.40). In turn, this increase in the price of the input translates to the price of the differentiated goods, with consumers losing utility.

Similarly, the indirect utility function of a household in the South is:

$$V^* = \frac{1}{\rho} \ln \left\{ \begin{array}{c} \alpha^\alpha (1-\alpha)^{1-\alpha} \left(\frac{\sigma-1}{\beta\sigma}\right)^\alpha (BS)^{\alpha\mu} \left(1 + \frac{\rho\eta(1-S_K)}{S_n L}\right) \cdot \\ \cdot N_0^{\frac{\alpha}{\sigma-1}} \left(1 - S_n \left(1 - \phi_R \delta\right)\right)^{\frac{\alpha}{\sigma-1}} e^{\frac{\alpha g}{\rho(\sigma-1)}} \end{array} \right\}. \tag{2.57}$$

And, by differentiating this function with respect to S_n, we obtain an analogous expression to that above:

$$\partial V^* = \left[\begin{array}{c} -\frac{\eta(1-S_K)}{S_n^2 L + \rho\eta S_n (1-S_K)} + \frac{2L\alpha^2}{\rho^2 \eta\sigma(\sigma-1)} - \\ -\frac{\alpha}{\rho(\sigma-1)} \cdot \frac{(1-\phi_R\delta)}{(1-S_n(1-\phi_R\delta))} + \frac{\alpha^2\mu^2(\sigma-1)\bar{S}B\eta}{S_n^2 \sigma\gamma S} \end{array} \right] \partial S_n \gtrless 0,$$

with the difference that the sign of the third effect is the opposite, since a lower concentration of firms in the North causes a decrease in the transport costs borne by consumers in the South, so that their welfare increases via prices.

In this situation, in which both the concentration of firms in the North and the economic growth rate decrease, two negative effects on welfare are shared by the individuals of both countries: the love-of-variety effect (negative as the consequence of a slower growth rate of the number of varieties), and the negative effect of the increased price of the natural resource on the price of the differentiated goods. In the opposite, the reduction in the growth rate causes monopolistic profits of intermediate good producers to rise, and thus increase per capita income in both countries.

Only the trading cost effect has an opposite impact on each country. While Northern consumers lose utility because they have to import more varieties and bear higher transport costs, the opposite holds for Southern individuals, which gain utility. This enables us to affirm that, when the South succeeds in attracting firms from the North, either consumers in the South lose utility, although less than the consumers in the North (in which case the public policy would be pointless), or they would gain utility, depending on the concrete values of the parameters. Therefore, in some situations (for a certain range of parameters), the South will be interested in applying such public policies that enable it to increase the cost advantage of the presence of the natural resource in its territory, the first nature cause, thus attracting firms from the other country.

2.8 Conclusions and future lines of research

In this chapter, we present a model integrating characteristics of the New Economic Geography, the theory of endogenous growth, and the economy of natural resources. This theoretical framework enables us to study explicitly the effect of first nature causes in the concentration of economic activity, analyzing one of the possible natural geographical characteristics, the presence of a natural resource in the territory.

Geography enters the model via transport costs, which condition the distribution of firms which attempt to take advantage of increasing returns in a market of monopolistic competition. Economic growth is supported by an endogenous framework with national spillovers in innovation, causing research activities to take place in a single country (the North), and thus, the greater the industrial concentration in that country, the higher the economic growth rate. And the natural resource appears as a localized input in one of the two countries (the South), giving firms located in that country a cost advantage.

After a decrease in any of the transport costs, firms decide to move to the country with the greatest domestic demand and market size. Despite the cost advantage of locating in the South, due to the presence of the natural resource, firms prefer to move to the North, where they can take more advantage of increasing returns. In turn, concentration improves the economic growth rate, given the national nature of the spillovers.

Finally, the concentration of firms in the North would also have a positive effect on the stock of the natural resource in steady state, which would increase. Despite identifying two opposite effects, an industry localization effect and a growth effect, the industry localization effect dominates. Firms located in the North use a lower amount of natural resource, enabling the stock in steady state to increase. This is so because the firms in the North react to the cost advantage of firms in the South by producing a lower quantity of the differentiated good (and thus using less natural resource) and selling them at a higher price.

This means that, in the framework of our model, the home market effect (second nature causes), acting centripetally, have greater weight in firm decisions than the advantages of natural geographic circumstances (first nature causes), which act centrifugally.

However, the South can increase the importance of the first nature cause by introducing public policies to reinforce the cost advantage due to the natural resource presence in its territory. We have considered two different public policies: imposing restrictions on the international trade of the natural resource and promoting a technological change towards a technology which uses the resource more intensively. In both cases, the South attracts firms from the North, causing both the economic growth rate and the stock level of the natural resource in equilibrium to decrease. The effect of such policies on welfare, both for Northern and Southern households, is undetermined.

However, our results depend on the particular characteristics of the natural resource considered in our model: (i) it is renewable, (ii) with open access, (iii) used as an input only in the production of manufactured goods, and (iv) it is exploited using only labour. These assumptions have enabled us to build the simplest possible model in analytical terms, which we can call the basic model. Variations in any of these characteristics can produce extensions of the model.

In particular, there are two possible extensions which could add to our knowledge of the relationship between natural resources and the distribution of economic activity. Firstly, since at present most natural resources used in the production of manufactured goods are derived from oil or mining, it would be interesting to analyse how our model changes when the natural resource is not renewable. Secondly, another very interesting aspect would be to consider alternative mechanisms for the property rights of the natural resource. If the resource were not open access, the sector would generate additional income which, if most property rights were owned by Southern households, could have a positive impact on the size of the South market. This income effect, added to the advantage in costs which already appears in our model, would increase the weight of the first nature causes in the decisions made by firms.

2.A Appendix A: Steady state equilibrium

The value of S_n in the steady state equilibrium is the solution of the second degree equation:

$$(1-\delta\cdot\phi_R)(\phi_R-\delta)\,2L\cdot S_n^2+$$
$$+S_n\left[(1-\delta\cdot\phi_R)(\phi_R-\delta)\,\rho\eta-\left[(\phi_R-\delta)+\delta(1-\delta\cdot\phi_R)\right]L+2\delta(1-\delta\cdot\phi_R)\,L\right]-$$
$$-\rho\eta\left(\left[(\phi_R-\delta)+\delta(1-\delta\cdot\phi_R)\right]S_k-\delta(1-\delta\cdot\phi_R)\right)=0.$$

The valid solution is given by:

$$S_n=\frac{\left[\left[(\phi_R-\delta)+\delta(1-\delta\cdot\phi_R)\right]L-(1-\delta\cdot\phi_R)(\phi_R-\delta)\,\rho\eta-2\delta(1-\delta\cdot\phi_R)\,L\right]+\sqrt{\Delta}}{4L(1-\delta\cdot\phi_R)(\phi_R-\delta)},$$

where

$$\Delta=\left[(1-\delta\cdot\phi_R)(\phi_R-\delta)\,\rho\eta-\left[(\phi_R-\delta)+\delta(1-\delta\cdot\phi_R)\right]L+2\delta(1-\delta\cdot\phi_R)\,L\right]^2+$$
$$+8L(1-\delta\cdot\phi_R)(\phi_R-\delta)\cdot\rho\eta\left(\left[(\phi_R-\delta)+\delta(1-\delta\cdot\phi_R)\right]S_k-\delta(1-\delta\cdot\phi_R)\right).$$

The other root is greater than the unit and thus has no economic meaning. From this equilibrium value of S_n, which indicates the location of firms, we can obtain the steady state growth rate g in (2.47), and the North share in aggregate expenditure S_E in (2.50).

2.B Appendix B: Public policies and changes in utility

Section 2.7 gives an overall analysis of utility, in which we considered directly a change in S_n without paying attention to its causes. But if such variation in the concentration of firms is the consequence of any of the public policies suggested (a change in τ_R or in μ), additional effects on welfare appear which increase the aggregate indeterminacy, as both parameters appear in the indirect utility function.

In the case of $d\tau_R > 0$, after differentiating the indirect utility for the North in (2.56) we obtain:

$$dV = \begin{bmatrix} -\frac{\eta S_k}{S_n^2 L+\rho\eta S_n S_k} + \frac{2L\alpha^2}{\rho^2\eta\sigma(\sigma-1)} + \\ +\frac{\alpha}{\rho(\sigma-1)}\cdot\frac{(\phi_R-\delta)}{(S_n(\phi_R-\delta)+\delta)} + \frac{\alpha^2\mu^2(\sigma-1)\bar{S}B\eta}{S_n^2\sigma\gamma S} \end{bmatrix} dS_n - \frac{\alpha\mu S_n}{\rho\phi_R\left(S_n(\phi_R-\delta)+\delta\right)}\cdot d\tau_R \gtrless 0.$$

And, similarly for the South, after differentiating (2.57):

$$dV^* = \begin{bmatrix} -\frac{\eta(1-S_K)}{S_n^2 L+\rho\eta S_n(1-S_K)} + \frac{2L\alpha^2}{\rho^2\eta\sigma(\sigma-1)} - \\ -\frac{\alpha}{\rho(\sigma-1)}\cdot\frac{(1-\phi_R\delta)}{(1-S_n(1-\phi_R\delta))} + \frac{\alpha^2\mu^2(\sigma-1)\bar{S}B\eta}{S_n^2\sigma\gamma S} \end{bmatrix} dS_n - \frac{\alpha\mu\delta S_n}{\rho\phi_R\left(1-S_n\left(1-\phi_R\delta\right)\right)}\cdot d\tau_R \gtrless 0.$$

A new term appears which affects the utility of consumers in the North and in the South. This last term, with a negative sign, represents the loss of utility experienced by consumers in both countries when the transport cost of the natural resource is increased.

In the case of $d\mu > 0$, after differentiating the indirect utility for the North in (2.56), we obtain:

$$dV = \begin{bmatrix} -\frac{\eta S_k}{S_n^2 L+\rho\eta S_n S_k} + \frac{2L\alpha^2}{\rho^2\eta\sigma(\sigma-1)} + \\ +\frac{\alpha}{\rho(\sigma-1)}\cdot\frac{(\phi_R-\delta)}{(S_n(\phi_R-\delta)+\delta)} + \frac{\alpha^2\mu^2(\sigma-1)\bar{S}B\eta}{S_n^2\sigma\gamma S} \end{bmatrix} dS_n +$$
$$+\left[\frac{\alpha}{\rho\left(\sigma-1\right)}\cdot\frac{S_n\left(1-\sigma\right)\phi_R\ln\left(\tau_R\right)}{\left(S_n(\phi_R-\delta)+\delta\right)} + \frac{\alpha}{\rho}\cdot\ln\left(BS\right)\right] d\mu \gtrless 0.$$

And, similarly for the South, after differentiating (2.57):

$$dV^* = \left[\begin{array}{c} -\frac{\eta(1-S_K)}{S_n^2 L+\rho\eta S_n(1-S_K)} + \frac{2L\alpha^2}{\rho^2\eta\sigma(\sigma-1)} - \\ -\frac{\alpha}{\rho(\sigma-1)} \cdot \frac{(1-\phi_R\delta)}{(1-S_n(1-\phi_R\delta))} + \frac{\alpha^2\mu^2(\sigma-1)\bar{S}B\eta}{S_n^2\sigma\gamma S} \end{array} \right] dS_n +$$
$$+ \left[\frac{\alpha}{\rho(\sigma-1)} \cdot \frac{\delta S_n (1-\sigma)\phi_R \ln(\tau_R)}{(1-S_n(1-\phi_R\delta))} + \frac{\alpha}{\rho} \cdot \ln(BS) \right] d\mu \gtrless 0.$$

To the effects noted above a new term affecting the utility of consumers in the North and the South appears. It represents the direct impact on utility that would be caused by changing to a technology which uses the resource more intensively, and has an indeterminate sign.

References

Beeson, P.E., D. N. DeJong, and W. Troesken, (2001). Population Growth in US Counties, 1840-1990. Regional Science and Urban Economics, 31: 669-699.

Black, D., and V. Henderson, (1998). Urban evolution in the USA. Brown University Working Paper No. 98-21.

Brander, J. A., and M. S. Taylor, (1997a). International Trade and Open-Access Renewable Resources: The Small Open Economy Case. The Canadian Journal of Economics / Revue canadienne d'Economique, 30(3): 526-552.

Brander, J. A., and M. S. Taylor, (1997b). International trade between consumer and conservationist countries. Resource and Energy Economics 19, 267–298.

Brander, J. A., and M. S. Taylor, (1998a). Open access renewable resources: Trade and trade policy in a two-country model. Journal of International Economics, 44: 181–209.

Brander, J. A., and M. S. Taylor, (1998b). The simple economics of Easter Island: a Ricardo-Malthus model of renewable resource use. American Economic Review, 88: 119– 138.

Brown, G. M., (2000). Renewable natural resource management and use without markets. Journal of Economic Literature 38, 875-914.

Chichilnisky, G., (1994). North-south trade and the global environment. American Economic Review, 84: 851 - 874.

Eliasson, L., and S. J. Turnovsky, (2004). Renewable Resources in an Endogenously Growing Economy: Balanced Growth and Transitional Dynamics. Journal of Environmental Economics and Management, 48, 1018-1049.

Ellison, G., and E. L. Glaeser, (1999). The geographic concentration of industry: Does natural advantage explain agglomeration? American Economic Review Papers and Proceedings 89(2), 311–316.

Glaeser, E. L., H. Kallal, J. Scheinkman, and A. Schleifer, (1992). Growth in cities. Journal of Political Economy 100, 1126-1152.

Glaeser, E. L., and J. Shapiro, (2003). Urban Growth in the 1990s: Is city living back? Journal of Regional Science, 43(1): 139-165.

Grossman, G., and E. Helpman, (1991). Innovation and growth in the world economy. MIT Press, Cambridge, MA.

Helpman, E., and P. Krugman, (1985). Market structure and foreign trade. MIT Press, Cambridge, MA.

Henderson, V., A. Kuncoro, and M. Turner, (1995). Industrial development in cities. Journal of Political Economy 103 (5), 1067-1090.

Hirose, K., and K. Yamamoto, (2007). Knowledge spillovers, location of industry, and endogenous growth. Annals of Regional Science 41, 17-30.

Jacobs, J., (1969). Economy of Cities. Vintage, New York.

Krugman, P., (1991). Increasing returns and economic geography. Journal of Political Economy 99, 483-499.

Martin, P., and G. I. P. Ottaviano, (1999). Growing locations: industry location in a model of endogenous growth. European Economic Review 43:281-302.

Martin, P., and C. A. Rogers, (1995). Industrial location and public infrastructure. Journal of International Economics 39:335-351.

Mitchener, K. J., and I. W. McLean, (2003). The Productivity of US States since 1880. Journal of Economic Growth, 8: 73-114.

Romer, P., (1990). Endogenous technical change. Journal of Political Economy 98(5): S71-S102.

Samuelson, P., (1954). The transfer problem and transport costs, II: analysis of effects of trade impediments. Economic Journal 64:264-289.

Chapter 3
Gibrat's Law for Cities Revisited

3.1 Introduction

The relationship between the growth rate of a quantifiable phenomenon and its initial size is a question with a long history in statistics: do larger entities grow more quickly, or more slowly? On the other hand, perhaps no relationship exists and the rate is independent of size. A fundamental contribution to this debate is that of Gibrat (1931), who observed that the distribution of size (measured by sales or the number of employees) of firms could be approximated well with a lognormal, and that the explanation lay in the growth process of firms tending to be multiplicative and independent of their size. This proposition became known as Gibrat's Law and prompted a deluge of work exploring the validity of this law for the distribution of firms (see the surveys of Sutton (1997) and Santarelli et al. (2006)). Gibrat's Law establishes that no regular behaviour of any kind can be deduced between growth rate and initial size.

The fulfilment of this empirical proposition also has consequences for the distribution which follows the variable; in the words of Gibrat (1931) himself "*the law of proportionate effect will therefore imply that the logarithms of the variable will be distributed following the (normal distribution)*". Some years later, Kalecki (1945), in a classical article, tested this statistical relationship between lognormality and proportionate growth under certain conditions, consolidating the conceptual binomial Gibrat's Law – lognormal distribution.

In the field of urban economics, Gibrat's Law, especially since the 1990s, has given rise to numerous empirical studies contrasting its validity for city size distributions, arriving at a majority consensus, though not absolute, that it holds in the long term. Gibrat's Law presents the added advantage that, as well as explaining relatively well the growth of cities, it can be related to another empirical regularity well known in urban economics, Zipf's Law, which appears when the so-called Pareto distribution exponent is equal to the unit[17]. The term was coined after a work by Zipf (1949), which observed that the frequency of the words of any language is clearly defined in statistical terms by constant values. This has given rise to theoretical works explaining the fulfilment of Gibrat's Law in the context of external urban local effects and productive shocks, relating them with Zipf's Law and associating them directly to an equilibrium situation. These theoretical works include Gabaix (1999), Duranton (2006, 2007), and Córdoba (2008).

Returning to the empirical side, there is an apparent contradiction in these studies, as they normally accept the fulfilment of Gibrat's Law but at the same time affirm that the distribution followed by city size is a Pareto distribution, very different to the lognormal. Recently, Eeckhout (2004) was able to reconcile both results, by demonstrating (as Parr and Suzuki (1973) affirmed in a pioneering work) that, if size restrictions are imposed on the cities, taking only the upper tail, this skews the analysis. Thus, if all cities are taken, it can be found that the true distribution is lognormal, and that the growth of these cities is independent of size. However, to date, Eeckhout (2004) is the only study to consider the

[17] If city size distribution follows a Pareto distribution, the following expression can be deduced: $\ln R = a - b \cdot \ln S$, where R is rank (1 for the biggest city, 2 for the second biggest and so on), S is the size or population and a and b are parameters, this latter being known as the Pareto exponent. Zipf's Law is fulfilled when b equals the unit.

entire city size distribution. But this is a short term analysis[18], when the phenomenon under study (Gibrat's Law) is, by definition, a long term result.

The aim of this chapter is to test empirically the validity of Gibrat's Law in the growth of cities, using data for all the twentieth century of the complete distribution of cities (without any size restrictions or with no truncation point) in three countries: the US, Spain and Italy. The following section offers a brief overview of the literature on Gibrat's Law and cities and the results obtained. Section 3.3 presents the databases, with special attention to the US census.

From the results we deduce that, when we consider the complete distribution of cities in the short term (section 3.4), a tendency to divergence is seen. However, the empirical evidence (section 3.5) shows that this does not impede city size distribution being adequately approximated as a lognormal distribution. Finally, in section 3.6 a long term viewpoint is taken. Panel data unit root tests confirm the validity of Gibrat's Law in the upper tail distribution (section 3.6.1), and we find evidence in favour of a weak Gibrat's Law (size affects the variance of the growth process but not its mean) when using non-parametric methods which relate growth rate with city size (section 3.6.2). The chapter ends with our conclusions.

---

[18] Eeckhout (2004) takes data from the United States census of 1990 and 2000, possibly because they are the only ones to be available online. Levy (2009), in a comment to Eeckhout (2004), and Eeckhout (2009) in the reply, also consider no truncation point, but only for the 2000 US Census data.

3.2 Gibrat's Law for cities. An overview of the literature

In the 1990s numerous studies began to appear which empirically tested the validity of Gibrat's Law. Table 3.2 shows the classification of all the studies on urban economics that we know of. While the countries considered, the statistical and econometric techniques used and the sample sizes are heterogeneous, the predominating result is the acceptance of Gibrat's Law.

Thus, both Eaton and Eckstein (1997) and Davis and Weinstein (2002) accept its fulfilment for Japanese cities, although they use different sample sections (40 and 303 cities, respectively), and time horizons. Davis and Weinstein (2002) affirm that long-run city size is robust even to large temporary shocks and, in studying the effect of Allied bombing in the Second World War, deduce that the effect of these temporary shocks disappears completely in less than 20 years.

Brakman et al. (2004) come to the same conclusion when analysing the impact of the bombardment on Germany during the Second World War, concluding that, for the sample of 103 cities examined, bombing had a significant but temporary impact on post-war city growth. Nevertheless, nearly the same authors in Bosker et al. (2008) obtain a mixed result with a sample of 62 cities in West Germany: correcting for the impact of WWII, Gibrat's Law is found to hold only for about 25% of the sample.

Meanwhile, both Clark and Stabler (1991) and Resende (2004) also accept the hypothesis of proportionate urban growth for Canada and Brazil respectively. The sample size used by Clark and Stabler (1991) is tiny (the 7 most populous Canadian cities), although the main contribution of their work is to propose the use of data panel methodology

Table 3.2. EMPIRICAL STUDIES ON CITY GROWTH

Study	Country	Period	Truncation point	Sample size	GL	EcIss
Eaton and Eckstein (1997)	France and Japan	1876-1990 (F)	Cities > 50,000 inhabitants (F)	39 (F), 40 (J)	A	non par (tr mat, lz)
		1925-1985 (J)	Cities > 250,000 inhabitants (J)			
Davis and Weinstein (2002)	Japan	1925-1965	Cities > 30,000 inhabitants	303	A	par (purt)
Brakman et al. (2004)	Germany	1946-1963	Cities > 50,000 inhabitants	103	A	par (purt)
Clark and Stabler (1991)	Canada	1975-1984	7 most populous cities	7	A	par (purt)
Resende (2004)	Brazil	1980-2000	Cities > 1,000 inhabitants	497	A	par (purt)
Eeckhout (2004)	US	1990-2000	All cities	19361	A	par (gr reg); non par (ker)
Ioannides and Overman (2003)	US	1900-1990	All MSAs	112 (1900) to 334 (1990)	A	non par (ker)
Gabaix and Ioannides (2004)	US	1900-1990	All MSAs	112 (1900) to 334 (1990)	A	non par (ker)
Black and Henderson (2003)	US	1900-1990	All MSAs	194 (1900) to 282 (1990)	R	par (purt)
Guérin-Pace (1995)	France	1836-1990	Cities > 2,000 inhabitants	675 (1836) to 1782 (1990)	R	par (corr)
Soo (2007)	Malaysia	1957-2000	Urban areas > 10,000 inhabitants	44 (1957) to 171 (2000)	R	par (purt)
Petrakos et al. (2000)	Greece	1981-1991	Urban centres > 5,000 inhabitants	150	R	par (gr reg)
Henderson and Wang (2007)	World	1960-2000	Metro areas > 100,000 inhabitants	1220 (1960) to 1644 (2000)	R	par (purt)
Bosker et al. (2008)	West Germany	1925-1999	Cities > 50,000 inhabitants	62	M	par (purt); non par (ker)
Anderson and Ge (2005)	China	1961-1999	Cities > 100,000 inhabitants	149	M	par (rank reg); non par (tr mat)

Gibrat's Law: GL	EcIss: Econometric Issues	gr reg: growth regressions	corr: coefficient of correlation (Pearson)
A: Accepted	par: parametric methods	ker: kernels	lz: Lorenz curves
R: Rejected	non par: non parametric methods	rank reg: rank regressions	
M: Mixed Results	purt: panel unit root tests	tr mat: transition matrices	

and unit root tests in the analysis of urban growth. This is also the methodology which Resende (2004) applies to his sample of 497 Brazilian cities. However, Henderson and Wang (2007) strongly reject Gibrat's Law and a unit root process in their worldwide data set on all metro areas over 100,000 from 1960 to 2000.

For the case of the US, there are also several works accepting statistically the fulfilment of Gibrat's Law, whether at the level of cities (Eeckhout, 2004, is the first to use the entire sample without size restrictions), or with MSAs (Ioannides and Overman, 2003, whose results reproduce Gabaix and Ioannides, 2004). Also for the US, however, Black and Henderson (2003) reject Gibrat's Law for any sample section, although their database of MSAs is different[19] to that used by Ioannides and Overman (2003).

Other works exist rejecting the fulfilment of Gibrat's Law. Thus, Guérin-Pace (1995) finds that in France for a wide sample of cities with over 2,000 inhabitants during the period 1836-1990 there appears to be a fairly strong correlation between city size and growth rate, a correlation which is accentuated when the logarithm of the population is considered. This result goes against that obtained by Eaton and Eckstein (1997) when considering only the 39 most populated French cities. Soo (2007) and Petrakos et al. (2000) also reject the fulfilment of Gibrat's Law in Malaysia and Greece, respectively.

For the case of China, Anderson and Ge (2005) obtain a mixed result with a sample of 149 cities of more than 100,000 inhabitants: Gibrat's law appears to describe the situ-

[19] The standard definitions of metropolitan areas were first published in 1949 by what was then called the Bureau of the Budget, predecessor of the current Office of Management and Budget (OMB), with the designation Standard Metropolitan Area. This means that if the objective is making a long term analysis it will be necessary to reconstruct the areas for earlier periods, in the absence of a single criterion.

ation well prior to the Economic Reform and One Child Policy period, but later Kalecki's reformulation seems to be more appropriate.

What we wish to emphasize is that, with the exception of Eeckhout (2004), none of these studies considers the entire distribution of cities, as all of them impose a truncation point, whether explicitly, by taking cities above a minimum population threshold or implicitly, by working with MSAs[20]. This is usually due to a practical reason of data availability. For this reason most studies focus on analysing the most populous cities, the upper tail distribution. There are two very reasonable justifications for this approach. First, the largest cities represent most of the population of a country. And second, the growth rate of the biggest cities has less variance than the smallest ones (scale effect).

However, it should be pointed out that any test done on this type of sample will be local in character, and the behaviour of large cities cannot be extrapolated to the entire distribution. This type of deduction can lead to erroneous conclusions, as it must not be forgotten that what is being analysed is the behaviour of a few cities, which as well as being of a similar size, can present common patterns of growth. Therefore, we might conclude that Gibrat's Law is fulfilled when in fact we have focused our analysis on a club of cities which cannot be representative of all urban centres.

[20] In the US, to qualify as a MSA a city needs to have 50,000 or more inhabitants, or the presence of an urbanised area of at least 50,000 inhabitants, and a total metropolitan population of at least 100,000 (75,000 in New England), according to the OMB definition. In other countries similar criteria are followed, although the minimum population threshold needed to be considered a metropolitan area may change.

3.3 The databases

We use city population data from three countries: the US, Spain and Italy. The US is an extremely interesting country in which to analyse the evolution of urban structure, as it is a relatively young country whose inhabitants are characterised by high mobility. On the other hand we have the European countries, with a much older urban structure and inhabitants who present greater resistance to movement; specifically, Cheshire and Magrini (2006) estimate mobility in the US is fifteen times higher than in Europe.

Considering these two types of country gives us information about different urban behaviours, as while Spain and Italy have an already consolidated urban tissue and new cities are rarely created (urban growth is produced by population increase in existing cities), in the US urban growth has a double dimension: as well as increases in city size, the number of cities also increases, with potentially different effects on city size distribution. Thus, the population of cities (incorporated places) goes from representing less than half the total population of the US in 1900 (46.99%) to 61.49% in 2000; at the same, time the number of cities increases by 82.11%, from 10,596 in 1900 to 19,296 in 2000.

The data for the US we are using are the same as those used by González-Val (2010). Our database, created from the original documents of the annual census published by the US Census Bureau (www.census.gov), consists of the available data of all incorporated places without any size restriction, for each decade of the twentieth century. The US Census Bureau uses the generic term "incorporated place" to refer to the governmental unit incorporated under state law as a city, town (except in the states of New England, New

York and Wisconsin), borough (except in Alaska and New York), or village, and which has legally established limits, powers and functions.

The number of cities (in brackets) corresponding to each period is: 1900 (10,596 cities), 1910 (14,135), 1920 (15,481), 1930 (16,475), 1940 (16,729), 1950 (17,113), 1960 (18,051), 1970 (18,488), 1980 (18,923), 1990 (19,120), and 2000 (19,296).

Two details should be noted. First, that all the cities corresponding to Alaska, Hawaii, and Puerto Rico for each decade are excluded, as these states were annexed during the 20th century (Alaska and Hawaii in 1959, and the special case of Puerto Rico was annexed in 1952 as an associated free state), and data are not available for all periods. Their inclusion would produce geographical inconsistency in the sample, which would not be homogenous in geographical terms and thus could not be compared. And, second, for the same reason we also exclude all the unincorporated places (concentrations of population which do not form part of any incorporated place, but which are locally identified with a name), which began to be accounted after 1950. However, these settlements did exist earlier, so that their inclusion would again present a problem of inconsistency in the sample. Also, their elimination is not quantitatively important; in fact, there were 1,430 unincorporated places in 1950, representing 2.36% of the total population of the US, which by 2000 were 5,366 places and 11.27%.

For Spain and Italy the geographical unit of reference is the "municipality" and the data come from the official statistical information services. In Italy this is the Servizio Biblioteca e Servizi all'utenza, of the Direzione Centrale per la Diffusione della Cultura e dell'informazione Statistica, part of the Istituto Nazionale di Statistica (www.istat.it), and for Spain we have taken the census of the Instituto Nacional de Estadística[21], INE (www.ine.es). The de facto resident population has been taken for each city.

We have taken the data corresponding to the census of each decade of the 20th century. For Italy data for the following years have been considered (in brackets, the number of cities for each year): 1901 (7,711), 1911 (7,711), 1921 (8,100), 1931 (8,100), 1936 (8,100), 1951 (8,100), 1961 (8,100), 1971 (8,100), 1981 (8,100), 1991 (8,100), and 2001 (8,100). No census exists in Italy for 1941, due to its participation in the Second World War, so we have taken the data for 1936. For Spain the following years are considered: 1900 (7,800), 1910 (7,806), 1920 (7,812), 1930 (7,875), 1940 (7,896), 1950 (7,901), 1960 (7,910), 1970 (7,956), 1981 (8,034), 1991 (8,077), and 2001 (8,077).

3.4 Gibrat's Law in the short term

In this section we offer a first approach to the behaviour of city growth from a short term perspective, i.e., considering each decade individually. Following Gabaix and Ioannides (2004), Gibrat's law states that the growth rate of an economic entity (firm, mutual fund, city) of size S has a distribution function with mean and variance that are independent of S. Therefore, if S_{it} is the size of city i at the time t and g is its growth rate, then $S_{it} = S_{it-1}(1+g)$. Taking logarithms and considering additionally that the rate could depend

[21] The official INE census have been improved in an alternative database, created by Azagra et al. (2006), reconstructing the population census for the twentieth century using territorially homogeneous criteria. We have repeated the analysis using this database and the results are not significantly different, so we have presented the results deduced from the official data.

on the initial size, we have the following general expression for the growth equation[22,23]:

$$\ln S_{it} - \ln S_{it-1} = \mu + \beta \ln S_{it-1} + u_{it}, \tag{3.58}$$

where $\mu = \ln(1+g)$ and u_{it} is a random variable representing the random shocks which the growth rate may suffer, which we shall suppose to be identically and independently distributed for all cities, with $E(u_{it}) = 0$ and $Var(u_{it}) = \sigma^2 \; \forall i, t$. If $\beta = 0$ Gibrat's Law holds and we obtain that growth is independent of the initial size.

In such case ($\beta = 0$), it is easy to prove that the expected value of the size of city i at the time t depends only on the number of periods which have passed and on the size in the first period:

$$E(\ln S_{it}) = \mu \cdot t + \ln S_{i0}, \tag{3.59}$$

while the variance would be given by:

$$Var(\ln S_{it}) = t \cdot \sigma^2. \tag{3.60}$$

Consequently, the mean grows over time, and variance does too. The increased variance over time is consistent with the prediction of a Brownian motion: proportionate growth leads to a lognormal distribution with a standard deviation that is increasing in time t.

We adopt the Eaton-Eckstein terminology of convergent, parallel, vs. divergent city growth processes. Remember that if $\beta = 0$ city growth is parallel, as it does not depend

---

22 The size of a city can be defined, according to the literature, in three ways: in levels (S_{it}), in relative values ($\frac{S_{it}}{\bar{S}_t}$, $\bar{S}_t$ being the mean size) or in shares ($\frac{S_{it}}{\sum_i S_{it}}$). The crucial parameter in (3.58) is β, which determines whether Gibrat's Law holds. The specification (3.58) in logs makes the estimation of β robust to the three different definitions of city size.

23 Taking logarithms we reduce the distortions that may occur in the mean and variance of the growth rate due to changes in the variable.

on initial size. Thus, if the estimation of β is significantly different to zero we will reject the fulfilment of Gibrat's Law. In the case of being greater than zero, we will have divergent growth, because city growth would depend directly and positively on initial size. A sustained process of divergent growth of this kind would result in an increasingly asymmetrical distribution, with small cities getting further and further away from large ones. Finally, if β is negative, urban growth would be convergent, as the growth-size ratio would be negative; a larger initial population would mean less growth and vice versa, so that in the long term the distribution would tend to be concentrated around a median value. It is simple to prove that when $\beta \neq 0$ the expressions (3.59) and (3.60) change, becoming

$$E\left(\ln S_{it}\right) = \mu \cdot \frac{\left(\beta + 1\right)^{t} - 1}{\beta} + \left(\beta + 1\right)^{t} \ln S_{i0}, \tag{3.61}$$

$$Var\left(\ln S_{it}\right) = \sigma^{2} \cdot \frac{\left(\beta + 1\right)^{2t} - 1}{\beta^{2} + 2\beta}, \tag{3.62}$$

and it can be demonstrated (see Appendix) that when $t > 1$ and growth is divergent $(\beta > 0)$ the variance in (3.62) grows even faster than that in (3.60), while if city growth were convergent $(\beta < 0)$ the variance in (3.62) would be less than that in (3.60).

The first result we wish to present is the estimation of equation (3.58). We will focus on the analysis of the estimation of parameter β, as whether Gibrat's Law is fulfilled or not depends on its significance and its sign. Table 3.3 shows the results of the OLS estimation of β for each decade in the three countries considering all the cities, without size restrictions. The results of these regressions are usually heteroskedastic, so we have calculated the t-ratios using White's (1980) Heteroskedasticity-Consistent Standard Errors.

Table 3.3. ESTIMATED BETA COEFFICIENTS FOR THE ENTIRE SAMPLE SIZE

US										
Period	1900-1910	1910-1920	1920-1930	1930-1940	1940-1950	1950-1960	1960-1970	1970-1980	1980-1990	1990-2000
β estimated	0.008	0.022	0.042	0.009	0.048	0.051	0.027	-0.005	0.042	0.015
t-ratio	2.875	10.543	21.568	7.958	30.998	28.674	18.029	-3.934	35.152	12.352
Spain										
Period	1900-1910	1910-1920	1920-1930	1930-1940	1940-1950	1950-1960	1960-1970	1970-1981	1981-1991	1991-2001
β estimated	0.010	0.020	0.023	0.028	0.012	0.048	0.115	0.115	0.047	0.013
t-ratio	7.424	14.282	15.512	18.641	6.647	31.734	44.743	49.699	24.524	7.259
Italy										
Period	1901-1911	1911-1921	1921-1931	1931-1936	1936-1951	1951-1961	1961-1971	1971-1981	1981-1991	1991-2001
β estimated	0.010	0.022	0.019	0.014	0.033	0.042	0.066	0.046	0.025	0.017
t-ratio	5.907	18.507	7.757	10.787	24.031	25.329	33.739	32.319	16.858	16.541

Note: t-ratios calculated using White Heteroskedasticity-Consistent Standard Errors.

The first conclusion we obtain is that when the entire sample of cities is considered, β is always significantly different to zero, for any period and in the three countries. This result is robust as, while the literature usually admits the possibility of occasional deviations from Gibrat's Law in the short term (with some periods in which urban growth may be convergent or divergent), we are rejecting the fulfilment of Gibrat's Law for each decade of the 20th century and for three nations. But the really surprising finding is that, despite their different urban structures and histories, the estimated parameter is always positive (except in the period 1970-1980 in the US) for the three countries, so that all of them exhibit divergent behaviour throughout the 20th century.

The only exception to this process of divergence is the estimation obtained for the US in the decade 1970-1980. The fact that this parameter is negative shows that during this decade the most populous cities grew more slowly. However, this result is atypical, and reflects two demographical circumstances in the United States during this period. First, between 1960 and 1990 there was a decline in the growth of the total population of the US, going from a growth rate of 18.5% in 1950-1960 to 9.8% in 1980-1990[24]. Then, that the total population grew by only 11.4% in 1970-1980, the third lowest growth rate in the history of the US since the first census was published in the late 18th century. And in this context of low growth of the total population, the percentage of urban population also fell (understood now as the percentage of the population associated with incorporated places), going from 64.51% of the total population in 1970 to 61.78% in 1980, which is by far the biggest fall in the 20th century. The fact that our estimation of β is negative would

24 Source: http://www.census.gov/population/censusdata/table-4.pdf.

reflect that the cities in the upper half of the distribution experienced the highest fall in their growth rate.

We have obtained that, in the short term, the city growth process was divergent in the three countries. However, this conclusion can change in the long term. But before we will analyse in section 3.5 the consequences on city size distribution of the divergent tendency we have observed.

3.5 What about city size distribution? Lognormality is maintained

In the section above, it has been shown that the overall result in the short term when the whole distribution is used is divergence. Also, as $\beta > 0$, the variance grow more than linearly (equation (3.62)), so that the growth process would be explosive, generating a city size distribution increasingly asymmetrical. But our results show that the growth process lead to a lognormal distribution with a standard deviation that is increasing in time t (as a Brownian motion would predict) in the three countries.

We carried out the Wilcoxon's lognormality test (rank-sum test), which is a non-parametric test for assessing whether two samples come from the same distribution. The null hypothesis is that the two samples are drawn from a single population, and therefore that their probability distributions are the same, in our case, the lognormal distribution. Wilcoxon's test has the advantage of being appropriate for any sample size. The more frequent normality tests –Kolmogorov-Smirnov, Shapiro-Wilks, D'Agostino-Pearson– are designed for small samples, and so tend to reject the null hypothesis of normality for large sample sizes, although the deviations from lognormality are arbitrarily small.

Table 3.4 shows the results of the test. The conclusion is that the null hypothesis of lognormality is accepted at 5% for all periods of the 20th century in Spain and Italy. In the US a temporal evolution can be seen; in the first decades lognormality is rejected and the p-value decreases over time, but from 1930 the p-value begins to grow until lognormal distribution is accepted at 5% from 1960 onwards (the same conclusion is reached by González-Val (2010) through a graphical examination of the adaptive kernels corresponding to the estimated distribution of different decades). In fact, if instead of the 5% we take a significance level of the 1%, the null hypothesis would only be rejected in 1920 and 1930.

However, the shape of the distribution in the US for the period 1900-1950 is not far from lognormality, either. Figure 3.14 shows the empirical density functions estimated by adaptive Gaussian kernels for 1900 and for 1950 (the last year in which lognormality is rejected). The motive for this systematic rejection appears to be an excessive concentration of density in the central values, higher than would correspond to the theoretical lognormal distribution (in black). Starting in 1900 with a very leptokurtic distribution, with a great deal of density concentrated in the mean value, from 1930 (not shown), when the growth of urban population slows, the distribution loses kurtosis and concentration decreases, accepting lognormality statistically at 5% from 1960.

To sum up, both the test carried out and the visualisation of the estimated empirical density functions seem to corroborate that city size distribution can be approximated correctly as a lognormal (in Spain and Italy during the entire 20th century, and in the US for

Table 3.4. WILCOXON RANK-SUM TEST OF LOGNORMALITY

US											
Year	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
p-value	0.0252	0.017	0.0078	0.0088	0.0208	0.0464	0.1281	0.1836	0.2538	0.323	0.4168
Spain											
Year	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001
p-value	0.5953	0.6144	0.6233	0.6525	0.4909	0.5792	0.6049	0.522	0.5176	0.622	0.7212
Italy											
Year	1901	1911	1921	1931	1936	1951	1961	1971	1981	1991	2001
p-value	0.2081	0.2205	0.2352	0.291	0.2864	0.3118	0.2589	0.272	0.382	0.4671	0.5287

Note: Ho: The distribution of cities follows a lognormal.

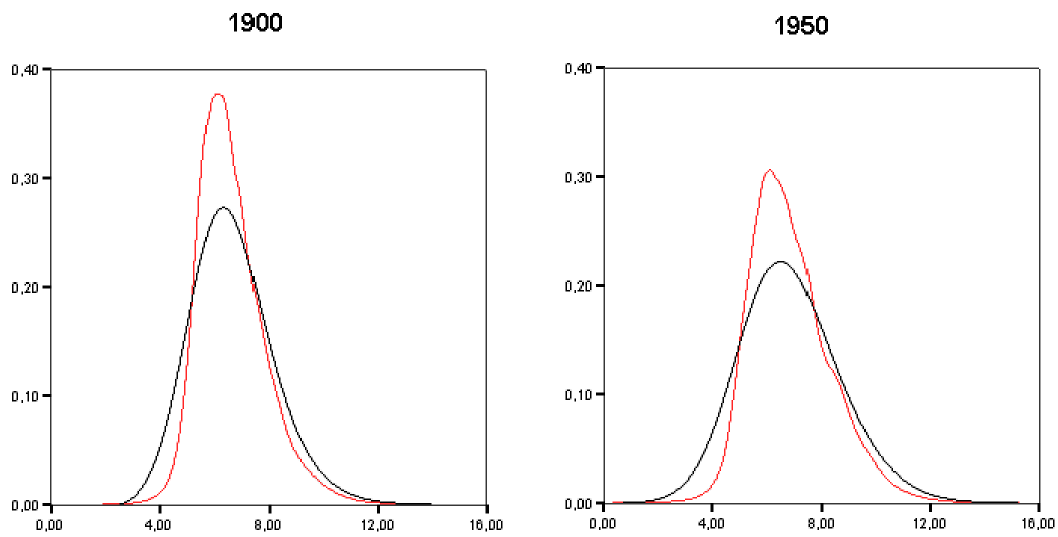


Fig. 3.14. Comparison of the estimated density function (ln scale) and the theoretical log-normal in black (US).

most decades, depending on the significance level), despite the urban growth was divergent every decade over the entire 20th century for the three countries (with the single exception of the period 1970-1980 in the US).

3.6 Gibrat's Law in the long term

In this section, we change our temporal perspective to the long term (the entire twentieth century). In order to carry out this analysis, we transform city population (S_{it}) to city relative size (s_{it}), defined as $s_{it} = \frac{S_{it}}{\bar{S}_t} = \frac{S_{it}}{\frac{1}{N}\sum_{i=1}^{N} S_{it}}$, as in a long term temporal perspective of steady state distributions it is necessary to use a relative measure of size.

This approach is more interesting, as the phenomenon under study (Gibrat's Law) is, by definition, a long term result. For this we combine parametric methods (the panel dimension of our data has been exploited in order to test for a unit root) with non-parametric ones, enabling us to study the relationship of growth and the variance of growth with city size.

3.6.1 Parametric analysis: panel unit root testing

Clark and Stabler (1991) suggested that testing for Gibrat's Law is equivalent to testing for the presence of a unit root. This idea has also been emphasized by Gabaix and Ioannides (2004), who expect "*that the next generation of city evolution empirics could draw from the sophisticated econometric literature on unit roots*". In line with this suggestion, most studies now apply unit root tests (see Table 3.2).

Some authors (Black and Henderson, 2003; Henderson and Wang, 2007; Soo, 2007) test the presence of a single root by proposing a growth equation similar to our equation (3.58), which they estimate using panel data. Nevertheless, as pointed out by Gabaix and Ioannides (2004) and Bosker et al. (2008), this methodology presents some drawbacks. First, the periodicity of our data is by decades, and we have only 11 temporal observations (decade-by-decade city sizes over a total period of 100 years), when the ideal would be to have at least annual data. And second, the presence of cross-sectional dependence across the cities in the panel can give rise to estimations which are not very robust. It has been well established in the literature that panel unit root and stationarity tests that do not explicitly allow for this feature among individuals present size distortions (Banerjee et al. 2005).

Therefore, we use one of the tests especially created to deal with this question when testing unit root. Pesaran's (2007) test for unit roots in heterogeneous panels with cross-section dependence is calculated on the basis of the CADF statistic (cross-sectional augmented ADF statistic). To eliminate the cross dependence, the standard Dickey-Fuller (or Augmented Dickey-Fuller, ADF) regressions are augmented with the cross section averages of lagged levels and first-differences of the individual series, such that the influence of the unobservable common factor is asymptotically filtered.

The test of the unit root hypothesis is based on the t-ratio of the OLS estimate of b_i $\left(\widehat{b}_i\right)$ in the following cross-sectional augmented DF (CADF) regression:

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + c_i \bar{y}_{t-1} + d_i \Delta \bar{y}_t + e_{it}. \tag{3.63}$$

We will test for the presence of a unit root in the natural logarithm of city relative size $(y_{it} = \ln s_{it})$ taking this into account. The null hypothesis assumes that all the series are non-stationary, and Pesaran's CADF is consistent under the alternative that only a fraction of the series is stationary.

However, the problem with Pesaran's test is that it is not designed to deal with such large panels (22,078 cities in the US, 8,077 in Spain and 8,100 in Italy), especially when so few temporal observations are available $(N \to \infty, T = 11)$. For this reason, we must limit our analysis to the largest cities (although the next section offers a long term analysis of the entire sample).

Table 3.5 shows the results of Pesaran's (2007) test, both the value of the test statistic and the corresponding p-value, applied to the upper tail distribution until the 500 largest cities in the initial period have been considered. All statistics are based on univariate AR(1) specifications including constant and trend.

Table 3.5. PANEL UNIT ROOT TESTS, PESARAN'S CADF STATISTIC

Cities (N)	US	Spain	Italy
50	-0.488 (0.313)	-0.915 (0.180)	4.995 (0.999)
100	0.753 (0.774)	0.050 (0.520)	5.983 (0.999)
200	1.618 (0.947)	-2.866 (0.002)	-1.097 (0.136)
500	1.034 (0.849)	-12.132 (0.000)	5.832 (0.999)

Note: test-statistic (p-value)
Pesaran's CADF test: standarized Ztbar statistic, $Z\left[\bar{t}\right]$
Variable: Relative size (in natural logarithms)
Sample size: (N, 11)

The null hypothesis of a unit root is not rejected in the US or Italy for any of the sample sizes considered, providing evidence in favour of the long term validity of Gibrat's law. Spain's case is different, as when the sample size is more than the 200 largest cities, the unit root is rejected, indicating a relationship between relative size and growth rate even for the largest cities.

3.6.2 Non-parametric analysis: kernel regression conditional on city size

This section on the nonparametric analysis follows closely the analysis in Ioannides and Overman (2003), and Eeckhout (2004). It consists of taking the following specification:

$$g_i = m\left(s_i\right) + \epsilon_i, \tag{3.64}$$

where g_i is the growth rate $\left(\ln s_{it} - \ln s_{it-1}\right)$ normalised (subtracting the mean and dividing by the standard deviation) and s_i is the logarithm of the i-th city relative size. Instead of making suppositions about the functional relationship m, $\hat{m}\left(s\right)$ is estimated as a local mean around the point s and is smoothed using a kernel, which is a symmetrical, weighted and continuous function around s.

To analyse all the 20th century we build a pool with all the growth rates between two consecutive periods. This enables us to carry out a long term analysis. The Nadaraya-Watson method is used, exactly as it appears in Härdle (1990), based on the following expression[25]:

$$\hat{m}(s) = \frac{n^{-1}\sum_{i=1}^{n} K_h(s - s_i)\, g_i}{n^{-1}\sum_{i=1}^{n} K_h(s - s_i)}, \tag{3.65}$$

where K_h denotes the dependence of the kernel K (in this case an Epanechnikov) on the bandwidth h. We use the same bandwidth (0.5) in all estimations in order to allow comparisons between countries.

Starting from this calculated mean $\hat{m}(s)$, the variance of the growth rate g_i is also estimated, again applying the Nadaraya-Watson estimator:

$$\hat{\sigma}^2(s) = \frac{n^{-1}\sum_{i=1}^{n} K_h(s - s_i)\,(g_i - \hat{m}(s))^2}{n^{-1}\sum_{i=1}^{n} K_h(s - s_i)}. \tag{3.66}$$

The estimator is very sensitive, both in mean and in variance, to atypical values. For this reason we decide to eliminate from the sample the 5% smallest cities, as they usually have much higher growth rates in mean and in variance. This is logical; they are cities of under 200 inhabitants, where any small increase in their population becomes very large in percentage terms.

Gibrat's Law implies that growth is independent of size in mean and in variance. As growth rates are normalised, if Gibrat's Law in mean were strictly fulfilled, the nonparamet-

---

[25] The calculation was done with the KERNREG2 Stata module, developed by Nicholas J. Cox, Isaias H. Salgado-Ugarte, Makoto Shimizu and Toru Taniuchi, and available online at: http://ideas.repec.org/c/boc/bocode/s372601.html.

ric estimate would be a straight line on the zero value. Values different from zero involve deviations from the mean. In turn, the estimated variance of the growth rate would also be a straight line in the value one, which would mean that the variance does not depend on the size of the variable analysed. To be able to test these hypotheses, we have constructed bootstrapped 95-percent confidence bands (calculated from 500 random samples with replacement).

Figure 3.15 shows the nonparametric estimates of the growth rate of a pool for the entire 20th century for the US (1900-2000, 152,475 observations), Spain (1900-2001, 74,100 observations) and Italy (1901-2001, 73,260 observations). For the US the value zero appears always inside the confidence bands, so that it cannot be rejected that the growth rates are significantly different for any city size. For Spain and Italy the estimated mean grows with the sample size, although it is significantly different to zero only for the largest cities[26]. One possible explanation is historical: both Spain and Italy suffered wars on their territories during the 20th century, so that for several decades, the largest cities attracted most of the population[27]. Therefore, we find evidence in favour of Gibrat's Law for the US throughout all the 20th century. Also for Spain and Italy, although the largest cities would present some divergent behaviour.

---

[26] In the case of Spain, this divergent behaviour could be the explanation for the rejection obtained in the previous section of the null hypothesis of a unit root.

[27] This result can be related with the "safe harbour effect" of Glaeser and Shapiro (2002), which is a centripetal force which tends to agglomerate the population in large cities when there is an armed conflict.

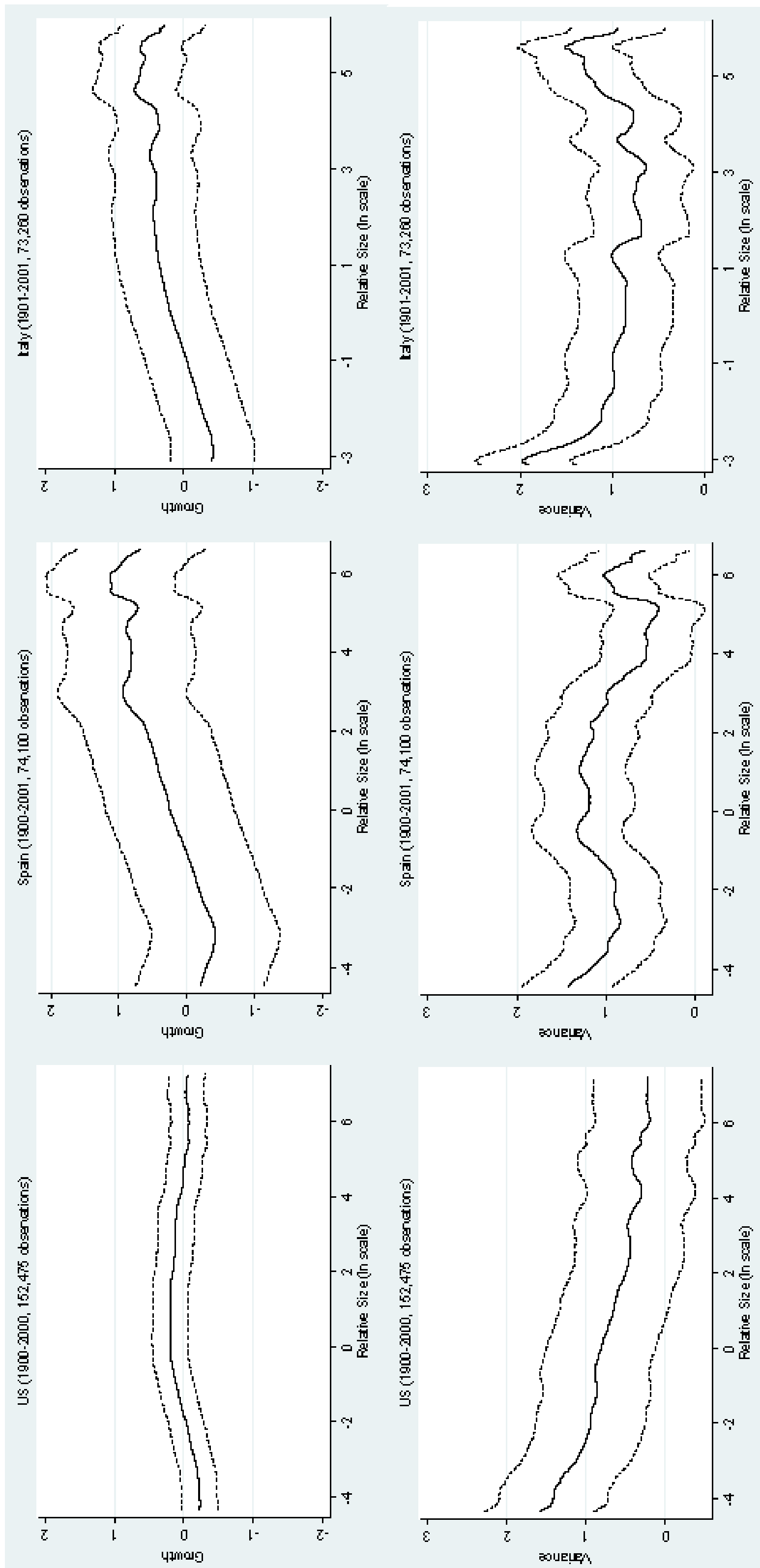


Fig. 3.15. Nonparametric estimates (bandwidth 0.5).

Figure 3.15 also shows the nonparametric estimates of the variance of growth rate of a pool for the entire 20th century for the US, Spain and Italy. As expected, while for most of the distribution the value one falls within the confidence bands, indicating that there are no significant differences in variance, the tails of the distribution show differentiated behaviours. In the US the variance clearly decreases with the size of the city, while in Spain and Italy the behaviour is more erratic and the biggest cities also have high variance.

Our results, obtained with our sample of all incorporated places without any size restriction, are similar to those obtained by Ioannides and Overman (2003), with their database of the most populous MSAs. To sum up, the nonparametric estimates show that while average growth seems to be independent of size in the three countries (although in Spain and Italy the largest cities present some divergent behaviour), variance in growth does depend negatively on size: the smallest cities present clearly higher variance in all three countries (although in Spain and Italy the behaviour is more erratic and the biggest cities also have high variance).

This points to Gibrat's Law holding weakly (growth is proportional on average but not in variance). Gabaix (1999) contemplates this possibility, that Gibrat's Law might not hold exactly, and examines the case in which cities grow randomly with expected growth rates and standard deviations that depend on their sizes. Therefore, the size of city i at time t varies according to[28]:

$$\frac{dS_t}{S_t} = \mu\left(S_t\right)dt + \sigma\left(S_t\right)dB_t,$$

[28] Equation (11) in Gabaix (1999).

where $\mu(S)$ and $\sigma^2(S)$ denote, respectively, the instantaneous mean and variance of the growth rate of a size S city, and B_t is a standard Brownian motion. Córdoba (2008) also introduces a parsimonious generalization of Gibrat's Law that allows size to affect the variance of the growth process but not its mean.

Nevertheless, we must distinguish between the American and European cases, as Gibrat's Law assumes a fixed and invariant number of locations. The number of cities remains constant in Spain and Italy, but this is not true for the US; from the beginning of the period considered to the end, the number of cities doubles. And while a Brownian motion can be adjusted to include new entrants, the distribution from which the entrants are drawn and the magnitude of entrants will affect the distribution. In particular, in the presence of a drift (as in this case where there is average city growth), the distribution from which new entrants are drawn is unlikely to be stationary if one wants to obtain the result that growth is proportionate.

Figure 3.16 shows the nonparametric estimates of the growth rate and its variance from a pool for the entire 20th century for the US (1910-2000, 59,865 observations) considering only the new entrant cities since 1910. Bootstrapped 95-percent confidence bands are also presented. The estimations show how the cities entering the sample from 1910 usually had growth rates higher on average and in variance than the average of the entire sample (dotted blue line), although the bands do not allow to reject that they are significantly different. The differences in variance indicate that part of the increased variance at the bottom of the size distribution can be explained by the cities which entered the distribution throughout the twentieth century.

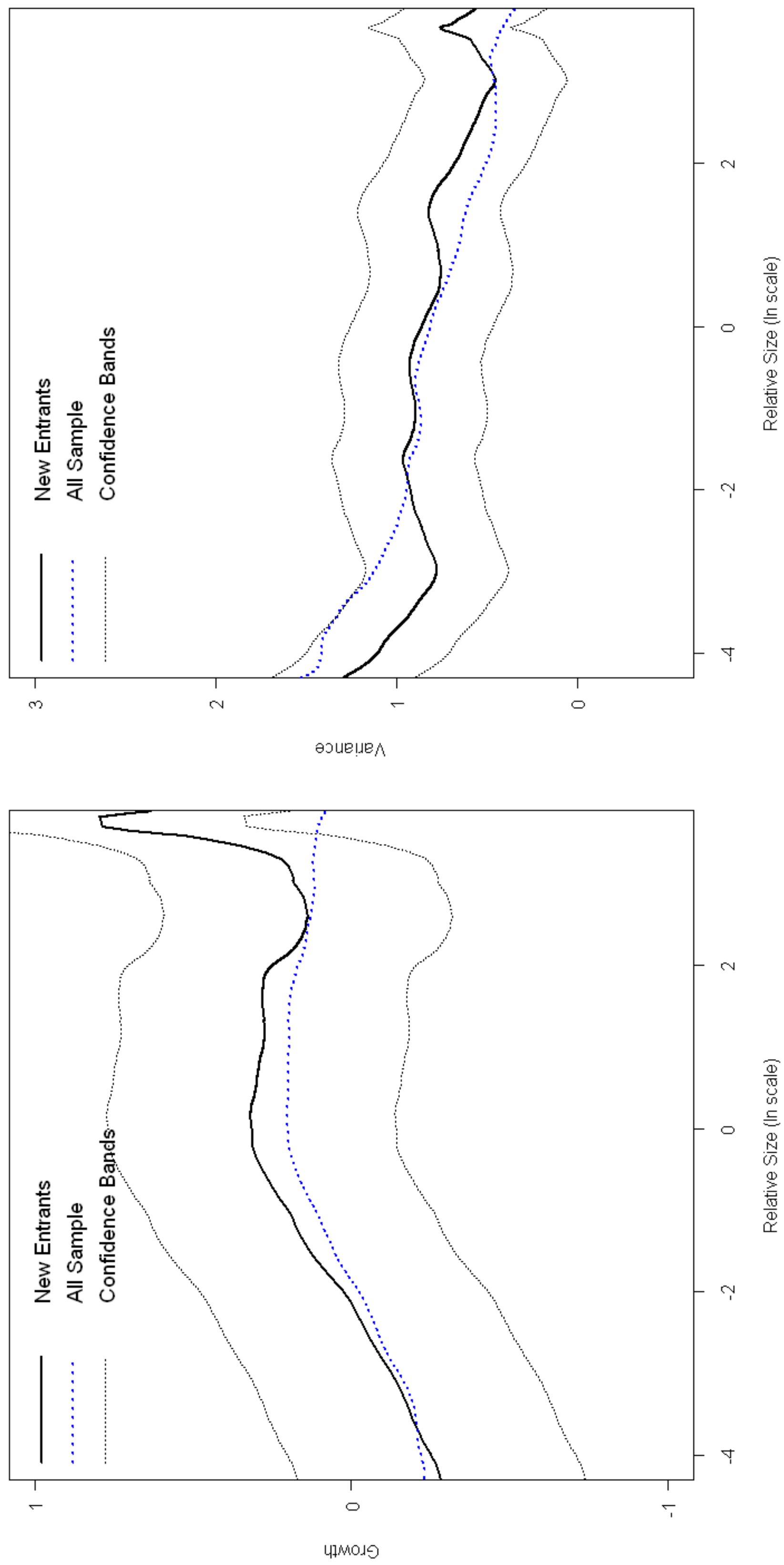


Fig. 3.16. New entrants nonparametric estimates (bandwidth 0.5), (US, 1910-2000), 59,865 observations.

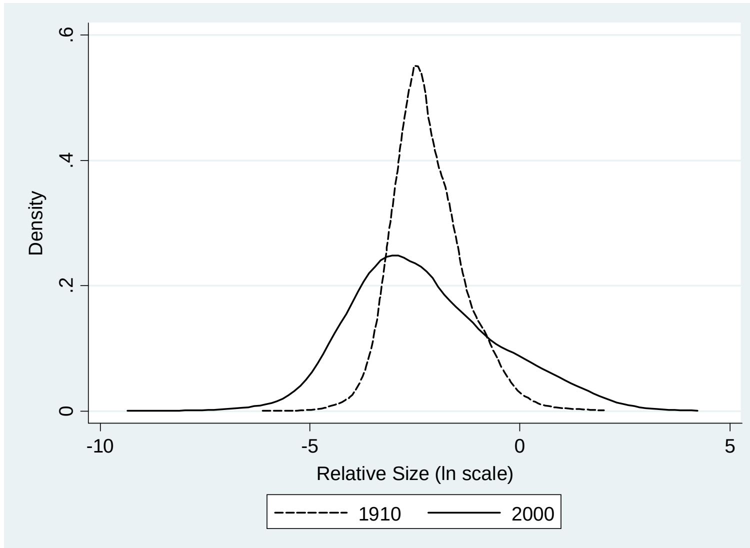


Fig. 3.17. Empirical density functions of the new entrants.

Moreover, the Figure 3.17, representing the empirical estimated distributions of entrant cities in 1910 and 2000 (normalized by the average size of the cohort of the entire distribution), shows the change in distribution of entrant cities. Starting from a very leptokurtic distribution in 1910 (more leptokurtic than the distribution of the whole sample) concentration decreases until the 2000 distribution, very close to a lognormal.

3.7 Conclusions

This chapter contributes to the literature empirically testing the validity of Gibrat's Law in the growth of cities using data of the complete distribution of cities (without any size restrictions) for all the twentieth century in three countries: the US, Spain and Italy. To do so, we use different techniques (parametric and non-parametric methods), obtaining mixed evidence. Our results confirm that, as Gabaix and Ioannides (2004) suggest, Gibrat's law for means holds only as a long-run average.

In the short term, considered decade by decade, we find that growth was divergent in all three countries. Despite being three countries with very different urban structures and histories, we find a positive relationship between the growth rate of cities and their initial size throughout the 20th century (except in the period 1970-1980 in the US). However, the empirical evidence shows that this does not impede city size distribution being approximated as a lognormal distribution.

In the long term, panel data unit root tests confirm the validity of Gibrat's Law in the upper tail distribution. The use of non-parametric methods which relate the growth rate with city size through the estimation of local means enable us to observe that, in the long term, the evidence in favour of a weak Gibrat's Law increases (size affects the variance of the growth process but not its mean).

The case of the US is different because number of cities doubles over the twentieth century. The new entrant cities present higher growth rates on average and in variance than the average for the whole sample, although we cannot reject that they are significantly different. The differences are greater in variance, indicating that part of the increased variance at the bottom of the size distribution can be explained by the cities which entered the distribution throughout the twentieth century.

3.A Appendix: Variance and city growth processes

We have two expressions (3.60 and 3.62):

$$Var\left(\ln S_{it}\right) = t \cdot \sigma^2, \tag{3.60}$$

$$Var\left(\ln S_{it}\right) = \sigma^2 \cdot \frac{\left(\beta+1\right)^{2t} - 1}{\beta^2 + 2\beta}. \tag{3.62}$$

If Gibrat's Law is fulfilled ($\beta = 0$), and applying L'Hôpital's rule we obtain that (3.62) converges to (3.60): $\lim\limits_{\beta \to 0} \left(\sigma^2 \cdot \frac{2t(\beta+1)^{2t-1}}{2\beta+2}\right) = \frac{2t\sigma^2}{2} = t\sigma^2$.

Let's see what happens if $\beta > 0$ or $\beta < 0$:

$$\begin{aligned}\text{(3.60)-(3.62)} &= t \cdot \sigma^2 - \sigma^2 \cdot \frac{\left(\beta+1\right)^{2t} - 1}{\beta^2 + 2\beta} = \frac{\sigma^2}{\beta\left(\beta+2\right)}\left[t\left(\beta^2 + 2\beta\right) - \left(\beta+1\right)^{2t} + 1\right] = \\ &= \frac{\sigma^2}{\beta\left(\beta+2\right)}\left[f\left(\beta\right)\right].\end{aligned}$$

Considering time t as a continuum beginning in zero, the expression between brackets $f\left(\beta\right)$ is only defined if $-1 < \beta$. Also, if $\beta > 0$ then $\frac{\sigma^2}{\beta(\beta+2)} > 0$, while if $-1 < \beta < 0$ then $\frac{\sigma^2}{\beta(\beta+2)} < 0$.

Therefore, to find out the total sign of the difference (3.60)-(3.62) we must study the behaviour of the function $f\left(\beta\right) = t\left(\beta^2 + 2\beta\right) - \left(\beta+1\right)^{2t} + 1$. The maximum or minimum of this function is given by:

$$\frac{df\left(\beta\right)}{d\beta} = f'\left(\beta\right) = 2t\left(\beta + 1 - \left(\beta+1\right)^{2t-1}\right) = 0,$$

from which we deduce that at the extreme $1 = \left(\beta+1\right)^{2t-2}$, which means that $f\left(\beta\right)$ is maximum or minimum in $\beta = 0$. In order to know if $\beta = 0$ is a maximum or a minimum we obtain the second order condition:

$$\frac{d^2 f\left(\beta\right)}{d\beta^2} = f''\left(\beta\right) = 2t\left(1 - \left(2t-1\right)\left(\beta+1\right)^{2t-2}\right),$$

and evaluate the sign in $\beta = 0$: $f'' (\beta = 0) = 4t (1 - t) < 0$ as long as $t > 1$.

Thus, we already know that the function $f (\beta)$ is concave and reaches its maximum in $\beta = 0$ as long as $t > 1$. Considering that $f (0)$=0, this function always takes negative values except in the maximum.

The final sign of the difference (3.60)-(3.62) will be (maintaining the conditions $-1 < \beta$ and $t > 1$):

1. When $\beta > 0$, we have seen that $\frac{\sigma^2}{\beta(\beta+2)} > 0$ is fulfilled and city growth is divergent. The variance of the cities will is higher than if Gibrat's Law were fulfilled:(3.60)<(3.62).

2. When $\beta < 0$, city growth is convergent. The variance of the cities is lower than if Gibrat's Law were fulfilled: (3.60)>(3.62).

3. When $\beta = 0$, (3.60)=(3.62).

References

Anderson, G., and Y. Ge, (2005). The Size Distribution of Chinese Cities. Regional Science and Urban Economics, 35: 756-776.

Azagra, J., P. Chorén, F. J. Goerlich, and M. Mas, (2006). La localización de la población española sobre el territorio. Un siglo de cambios: un estudio basado en series homogéneas (1900-2001). Fundación BBVA.

Banerjee, A., M. Massimiliano and C. Osbat, (2005). Testing for PPP: should we use panel methods? Empirical Economics, Vol. 30, 77-91.

Black, D., and V. Henderson, (2003). Urban evolution in the USA. Journal of Economic Geography, Vol. 3(4): 343-372.

Bosker, M., S. Brakman, H. Garretsen, and M. Schramm, (2008). A Century of Shocks: the Evolution of the German City Size Distribution 1925 – 1999. Regional Science and Urban Economics 38: 330–347.

Brakman, S., H. Garretsen, and M. Schramm, (2004). The Strategic Bombing of German Cities during World War II and its Impact on City Growth. Journal of Economic Geography, 4: 201-218.

Cheshire, P. C., S. and Magrini, (2006). Population Growth in European Cities: Weather Matters- but only Nationally. Regional Studies, 40(1): 23-37.

Clark, J. S., and J. C. Stabler, (1991). Gibrat's Law and the Growth of Canadian Cities. Urban Studies, 28(4): 635-639.

Córdoba, J. C., (2008). A generalized Gibrat's Law for Cities. International Economic Review, Vol. 49(4): 1463-1468.

Davis, D. R., and D. E. Weinstein, (2002). Bones, Bombs, and Break Points: The Geography of Economic Activity. American Economic Review, 92(5): 1269-1289.

Duranton, G., (2006). Some Foundations for Zipf´s Law: Product Proliferation and Local Spillovers. Regional Science and Urban Economics, 36: 542-563.

Duranton, G., (2007). Urban Evolutions: The Fast, the Slow, and the Still. American Economic Review, 97(1): 197-221.

Eaton, J., and Z. Eckstein, (1997). Cities and Growth: Theory and Evidence from France and Japan. Regional Science and Urban Economics, 27(4–5): 443–474.

Eeckhout, J., (2004). Gibrat's Law for (All) Cities. American Economic Review, 94(5): 1429-1451.

Eeckhout, J., (2009). Gibrat's Law for (all) Cities: Reply. American Economic Review, 99(4): 1676–1683.

Gabaix, X., (1999). Zipf's Law for Cities: An Explanation. Quaterly Journal of Economics, 114(3): 739-767.

Gabaix, X., and Ioannides, Y. M. (2004). The Evolution of City Size Distributions. In Handbook of Urban and Regional Economics, vol. 4, ed. John V. Henderson and Jean. F. Thisse, 2341-2378. Amsterdam: Elsevier Science, North-Holland.

Gibrat, R., (1931). Les Inégalités Économiques. París: Librairie du Recueil Sirey.

Glaeser, E. L., and J. M. Shapiro, (2002). Cities and Warfare: The Impact of Terrorism on Urban Form. Journal of Urban Economics, 51: 205-224.

González-Val, R., (2010). The Evolution of US City Size Distribution from a Long Term Perspective (1900-2000). Journal of Regional Science, forthcoming.

Guérin-Pace, F., (1995). Rank-Size Distribution and the Process of Urban Growth. Urban Studies, 32(3): 551-562.

Härdle, W., (1990). Applied Nonparametric Regression. In Econometric Society Monographs. Cambridge, New York and Melbourne: Cambridge University Press.

Henderson, V. and H. G. Wang, (2007). Urbanization and city growth: The role of institutions. Regional Science and Urban Economics, 37(3): 283–313.

Ioannides, Y. M., and H. G. Overman, (2003). Zipf's Law for Cities: an Empirical Examination. Regional Science and Urban Economics, 33: 127-137.

Kalecki, M., (1945). On the Gibrat Distribution. Econometrica, 13(2): 161-70.

Levy, M., (2009). Gibrat's Law for (all) Cities: A Comment. American Economic Review, 99(4): 1672–1675.

Parr, J. B., and K. Suzuki, (1973). Settlement Populations and the Lognormal Distribution. Urban Studies, 10: 335-352.

Pesaran, M. H., (2007). A simple panel unit root test in the presence of cross-section dependence. Journal of Applied Econometrics, 22: 265–312.

Petrakos, G., P. Mardakis, and H. Caraveli, (2000). Recent Developments in the Greek System of Urban Centres. Environment and Planning B: Planning and Design, 27(2): 169-181.

Resende, M., (2004). Gibrat's Law and the Growth of Cities in Brazil: A Panel Data Investigation. Urban Studies, 41(8): 1537-1549.

Santarelli, E., L. Klomp, and R. Thurik, (2006). Gibrat's Law: An Overview of the Empirical Literature. In Entrepreneurship, Growth, and Innovation: the Dynamics of Firms and Industries, ed. Enrico Santarelli, 41-73. New York: Springer.

Soo, K. T., (2007). Zipf's Law and Urban Growth in Malaysia. Urban Studies, 44(1): 1-14.

Sutton, J., (1997). Gibrat's Legacy. Journal of Economic Literature. 35(1): 40-59.

White, H., (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica, 48: 817-38.

Zipf, G., (1949). Human Behaviour and the Principle of Least Effort, Cambridge, MA: Addison-Wesley.

Chapter 4
What Makes Cities Bigger and Richer? Evidence from 1990-2000 in the US

4.1 Introduction

Jacobs (1969) was the first to suggest that the city is the basic economic unit of each country when she stated "cities are also primary economic organs". Later, other writers argued in the same way[29] (Duranton, 2000; Quigley, 1998; Fujita and Thisse, 2002). Indeed, some very special characteristics coincide in the city as an economic unit. First, there is complete freedom of movement in labour and capital among cities (they are completely open economies). Second, it is in cities where knowledge spillovers are most easily generated and transmitted, documented both at the theoretical level (Loury, 1979; Garicano and Rossi-Hansberg, 2006) and empirically (Glaeser et al., 1992; Henderson et al., 1995). Finally, the New Economic Geography highlights that cities are a source of agglomeration economies (Duranton and Puga, 2004).

The starting point for this chapter is the idea that the city has a double nature, on one hand as a population centre and on the other as an engine of economic growth, and that the different external effects generated in cities can potentially have different effects on population growth and per capita income growth. In particular, this chapter analyses the determinants of growth of American cities, understood as growth of either their population

[29] A good commentary on the relationship between cities and national economic growth can be found in Polèse (2005).

or their per capita income, from 1990 to 2000. This empirical analysis uses data from all cities with no size restriction (our sample contains data for 21,655 cities).

We will use a two-steps strategy. First, we analyse if the distribution of population and per capita income of the cities have followed similar paths in the 1990s. The results show that, while population growth in cities appears to be independent of initial size, the growth of per capita income is negatively correlated to initial per capita income: the richest cities grew less in this period. This explains why, while the empirical distribution of city population remains stable in the decade 1990-2000, the empirical distribution of city per capita income changes.

Second, in order to explain these differentiated behaviours, we examine the relationship between the urban characteristics in 1990 and city growth (both in population and in per capita income) using a Multinomial Logit Model. Apart from the initial levels of population and per capita income, we will focus on the role played by employment, including variables reflecting the productive structure (percentage of employment by sector: agriculture, construction, manufacturing, services, etc.) and the unemployment rate. We will also use the median travel time (as a variable reflecting the costs of urban congestion), human capital variables, and geographical variables.

The American case has already been dealt with in the literature, using different econometric techniques and considering different periods and sample sizes. The two most direct precedents are Glaeser et al. (1995) and Glaeser and Shapiro (2003).

Glaeser et al. (1995) examine the urban growth patterns in the 200 most populous cities in the US between 1960 and 1990 in relation to various urban characteristics in 1960. They find that income and population growth are

(1) positively related to initial schooling,

(2) negatively related to initial unemployment, and

(3) negatively related to the initial share of employment in manufacturing.

This behaviour would have continued during the decade 1990-2000, conclude Glaeser and Shapiro (2003), using a slightly larger sample size (they imposed a minimum population threshold of 25,000 inhabitants, considering the 1,000 most populous cities). During this decade the three most relevant variables were the human capital, climate and individuals' transport systems (public or private). The growth of cities was determined by three main trends:

(1) cities with strong human capital bases grew faster than cities without skills,

(2) people moved to warmer, drier places, and

(3) cities built around the automobile replaced cities that rely on public transportation.

Other empirical studies exist analysing American population and per capita income growth, although the geographical unit analysed is not the city. At the county level, Beeson et al. (2001) study the evolution of population from 1840 to 1990, while Young et al. (2008) analyse the evolution of income distribution from 1970 to 1998. Mitchener and McLean (2003) use data beginning in 1880 to study variations among states in labour productivity. Finally, Yamamoto (2008) examined the disparities in per capita income in the period 1955-2003 using different geographical levels (counties, economic areas, states and regions).

The main contribution of this chapter compared to earlier studies is the use of the distribution of all cities, without size restrictions. The reason is that larger cities present very concrete characteristics, which also differentiate them from other cities in the distribution. By focusing only on the most populous cities, part of the story was not being told.

Table 4.6 presents the values of the averages and standard deviations of different variables for the entire distribution of cities in 1990, and for the 1,000 and 200 largest cities. We can see how the most populous cities bear a greater congestion cost, measured by travel time, although its inhabitants enjoy higher levels of education. However, the most interesting differences appear in the productive structure. In the biggest cities, the services sector has a higher weight, while the employment percentage in the agriculture, forestry, fishing, mining, construction, and manufacturing sectors is below the average when considering the whole sample. The most populous cities are also characterised by a higher unemployment rate and a lower economic growth.

However, it could be said that our sample includes places which should not be considered urban, due to their small population. Despite this, the results we obtain with our sample of 21,655 cities are similar to those of Glaeser et al. (1995). Thus, we find that the probability of a city being in the 25% of cities with the highest growth rates in income or population (i.e., the probability of the growth rate of per capita income or population being in the top quartile of the distribution) depends

(1) positively on the initial percentage of inhabitants with higher educational levels (some college or higher degree), although the sign and intensity of the effect change when considering a wider concept of education (high school graduate or higher degree);

Table 4.6. MEANS AND STANDARD DEVIATIONS, CITY VARIABLES IN 1990

	Mean			Stand. dev.		
Variable	All sample	Top 1000	Top 200	All sample	Top 1000	Top 200
Population Growth (ln scale), 1990-2000	0.09	0.10	0.10	0.25	0.16	0.13
Per Capita Income Growth (ln scale), 1989-1999	0.43	0.37	0.36	0.17	0.09	0.08
Congestion cost variables						
Median Travel Time to Work (in minutes)	20.41	21.27	21.16	6.04	6.35	4.13
Human capital variables						
Percent population 18 years and over: Some college or higher degree	27.30	38.23	37.14	13.00	11.80	9.82
Percent population 18 years and over: High school graduate (includes equivalency) or higher degree	53.04	59.03	56.79	11.20	9.71	8.85
Productive structure variables						
Unemployment rate	6.82	6.23	7.11	5.00	2.73	2.66
Percent employed civilian population 16 years and over:						
Agriculture, forestry, fishing, and mining	5.15	1.63	1.66	6.59	1.87	1.84
Construction	6.85	5.60	5.53	4.58	2.08	1.65
Manufacturing (durable and nondurable goods)	19.59	16.75	15.75	11.60	7.56	6.71
Wholesale and Retail trade	21.15	22.50	21.66	15.32	4.49	2.56
Finance, insurance, and real estate	5.14	7.41	7.48	7.90	3.03	2.12
Educational, health, and other professional and related services	22.79	24.64	24.73	22.59	7.99	5.74
Public administration	4.80	5.03	5.38	4.95	4.20	3.43

Source: 1990 and 2000 Census, www.census.gov

(2) negatively on the initial unemployment levels, and

(3) negatively on the initial percentage of employment in the manufacturing sector, although this sector seems to have lost weight, since other economic sectors have a greater influence on probability.

Geography also seems to have a strong influence on cities' per capita income or population growth rate.

The next section studies the evolution of per capita income and population growth in cities in the 1990s. The analysis continues in section 4.3, using a Multinomial Logit Model (MNLM) to examine the relationship between urban characteristics in 1990 and city growth, both in population and in per capita income. The chapter ends with our conclusions.

4.2 City population and city per capita income: Twin paths?

Our first step is to determine whether city population and per capita income distributions followed similar paths in the 1990s. Figure 4.18 shows scatter plots of city per capita income growth and city population growth (in logarithmic scale) against the initial levels in 1989 and 1990, respectively[30]. We use data from the entire distribution of cities without any size restriction: 21,655 places.

We can observe that while in the case of per capita income there is a clear negative relationship between the initial income level and the growth rate, for population growth it is difficult to deduce any relationship between initial size and growth. Thus, while the

[30] Source: 1990 and 2000 Census, www.census.gov.

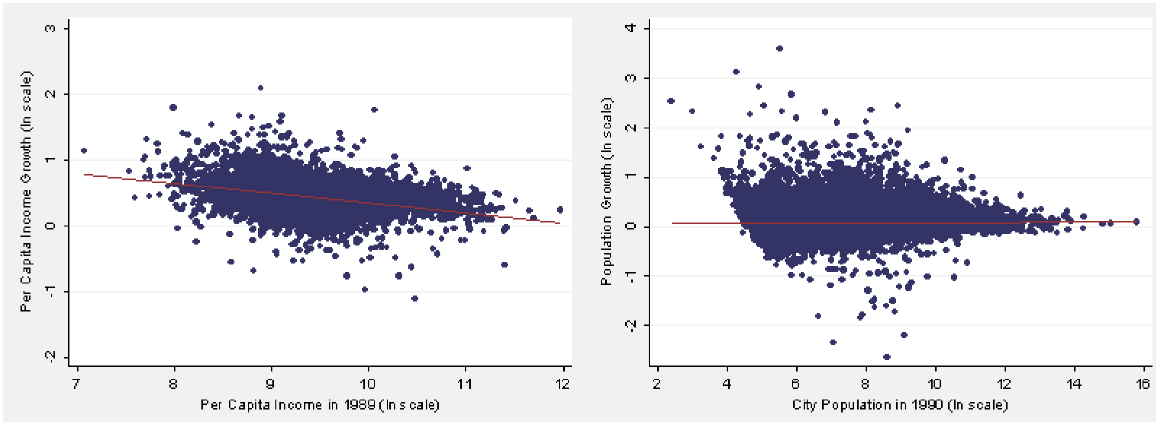


Fig. 4.18. Scatter Plots of City Growth (ln scale) against initial level.

slope β of the line adjusted with OLS[31] in the case of per capita income growth is a clearly significant and negative coefficient (-0.1471), with population growth this coefficient is very close to zero[32] (0.0026) and while it is significantly different to zero at 5%, it is not at 1%. This result, that initial population size does not influence its growth, is not new in urban economics. In fact, proportionate growth is a well-known empirical regularity known as Gibrat's law[33]. Recently Eeckhout (2004) studied the case of American cities during the period 1990-2000, also using data from the entire distribution, and concluded that Gibrat's law was fulfilled in that decade[34].

---

31 Line fitted as $(\ln y_{it} - \ln y_{it-1}) = \alpha + \beta \ln y_{it-1}$.

32 This value does not coincide with that obtained in Chapter 3 (Table 3.3). The differences might be a consequence of including the "unincorporated places."

33 Gibrat (1931) observed that the size distribution (measured in sales or number of employees) of firms tends to be lognormal, and his explanation was that the growth process of firms could be multiplicative and independent of firm size. Starting from the 90s, this proposition has given rise to numerous empirical studies in the field of urban economics, testing its validity for the city size distribution.

34 See Chapter 3 for a detailed analysis of Gibrat's law.

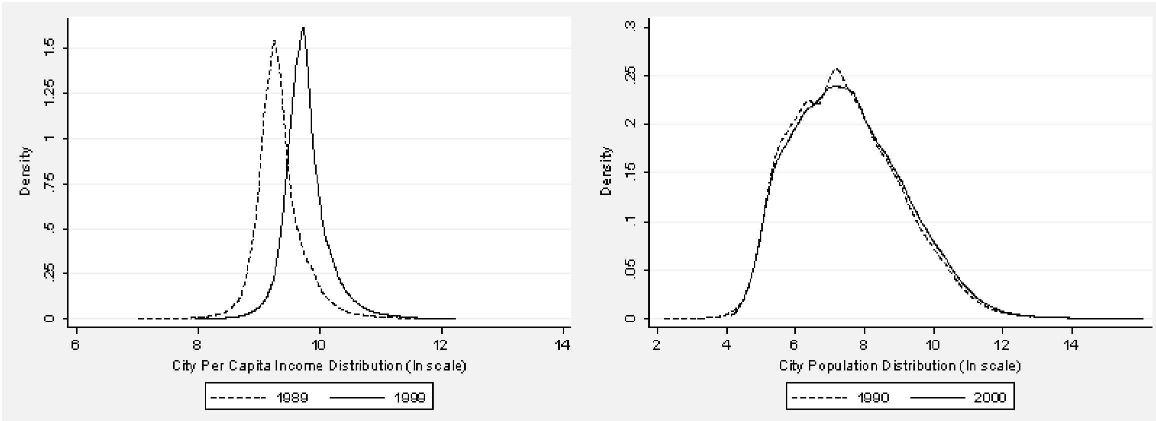


Fig. 4.19. Kernel density estimation (ln scale) of City Per Capita Income and City Population Distributions.

We would expect this different behaviour to have different consequences in the evolution of distributions. Figure 4.19 shows the estimated empirical distributions using an adaptive kernel of city size, both for per capita income and population.

It highlights an important change in the distribution of city per capita income. The negative relationship observed earlier between initial city per capita income and growth, which we can identify with convergent growth, has clearly produced a rightwards displacement of the distribution[35]. Meanwhile, there is hardly any change in the population distribution of the cities, as a consequence of their proportionate population growth.

Finally, Figure 4.20 relates city population growth and city per capita income growth. Have the cities which grew most in terms of population also grown the most in income, or vice versa? The graph shows a cloud of points with no apparent relationship[36], leading us

[35] Everything seems to indicate that this behavior has been produced for decades. Figure 2 of Young et al. (2008), corresponding to the evolution of the Distribution of U.S. Counties' Log Per Capita Incomes from 1970 to 1998, presents a very similar effect to that observed in our estimated kernel of city per capita income distribution from 1989 to 1999.

[36] In this case the adjusted line is not shown because the estimated slope β (-0.0153) is not significantly

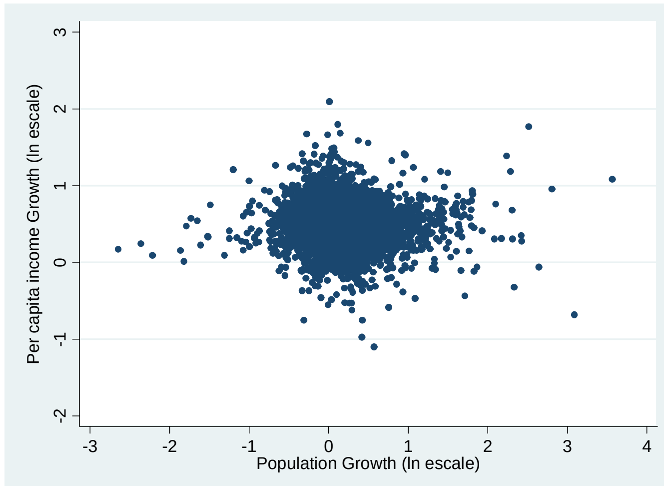


Fig. 4.20. Scatter Plot of City Per Capita Income Growth (ln scale) against City Population Growth (ln scale)

to conclude that during this period there was no relationship between economic growth and population growth in American cities.

However, the differentiated behaviour observed in the growth rates of cities' per capita income and population seems to corroborate our initial idea: the different external effects generated in cities can produce different effects in population growth and per capita income growth. Therefore, the next section analyses the relationship between the city characteristics in 1990 and their growth, both in population and in per capita income.

4.3 Empirical model and results

---

different to zero even at 5%.

4.3.1 Data description

We use data for all cities in the Unites States (21,655), without imposing any minimum population cut-off point, as our proposal is to cover the entire distribution. The data came from the census[37] for 1990 and 2000. We identified cities as what the US Census Bureau calls places. This generic name, since the 2000 census, includes all incorporated and unincorporated places.

The US Census Bureau uses the generic term incorporated place to refer to a type of governmental unit incorporated under state law as a city, town (except the New England states, New York, and Wisconsin), borough (except in Alaska and New York), or village, and having legally prescribed limits, powers, and functions. On the other hand, the unincorporated places (which were renamed Census Designated Places, CDPs, in 1980), designate a statistical entity, defined for each decennial census according to Census Bureau guidelines, comprising a densely settled concentration of population that is not within an incorporated place, but is locally identified by a name. Evidently, the geographical boundaries of unincorporated places may change if settlements move, so that the same unincorporated place may have different boundaries in different census. They are the statistical counterpart of the incorporated places. The difference between them in most cases is merely political and/or administrative. Thus for example, due to a state law of Hawaii there are no incorporated places there; they are all unincorporated.

The explicative variables chosen are similar to those in other studies of city growth in the US and city size, and correspond to the initial 1990 values. The influence of these

[37] The US Census Bureau offers information on a large number of variables for different geographical levels, available on its website: www.census.gov.

variables on city size has been empirically proven by other works studying the largest cities (see Glaeser and Shapiro, 2003). Table 4.6 presents the variables, which can be grouped in four types: congestion cost variables, human capital variables, productive structure variables, and geographical variables. It is apparent that in general, standard deviations are somewhat lower in the biggest cities, which shows that the most populous cities are very similar in their economic structure, while by considering all population centres, we collect more heterogeneous behaviours.

Urban congestion cost variables are basically intended to reflect the effect of city size on urban growth. For this we use two variables: a dummy variable taking value 1 if the city population in 1990 is more than 25,000 inhabitants, enabling us specifically to control the most populous cities of the sample, and the variable Median travel time to work (in minutes), representing the commuting cost borne by workers. This is one of the most characteristic congestion costs of urban growth, explicitly considered in some theoretical models; that is, the idea that as a city's population increases, so do costs in terms of the time taken by individuals to travel from home to work.

Regarding human capital variables, there are many studies demonstrating the influence of human capital on city size, as cities with better educated inhabitants tend to grow more. We took two human capital variables: Percent population 18 years and over: High school graduate (includes equivalency) or higher degree, and Percent population 18 years and over: Some college or higher degree. The former represents a wider concept of human capital, while the latter centres on higher educational levels (some college, Associate degree, Bachelor's degree, and Graduate or professional degree).

The third group of variables, referring to the productive structure, includes the unemployment rate and the distribution of employment by sectors. The distribution of labour among the various productive activities provides valuable information about other characteristics of the city. Thus, the employment level in the primary sector (agriculture; forestry; fishing and hunting; and mining) also represents a proxy of the natural physical resources available to the city (cultivable land, port, etc.). This is also a sector which, like construction, is characterised by constant or even decreasing returns to scale.

Employment in manufacturing informs us about the level of local economies of scale in production, as this is a sector which normally presents increasing returns to scale. The level of pecuniary externalities also depends on the size of the industrial sector. Marshall put forward that (i) the concentration of firms of a single sector in a single place creates a joint market of qualified workers, benefiting both workers and firms; (ii) an industrial centre enables a larger variety at a lower cost of concrete factors needed for the sector which are not traded, and (iii) an industrial centre generates knowledge spillovers. This approach forms part of the basis of economic geography models, along with circular causation: workers go to cities with strong industrial sectors, and firms prefer to locate nearer larger cities with bigger markets. Thus, industrial employment also represents a measurement of the size of the local market. Another proxy for the market size of the city is the employment in commerce, either retail or wholesale.

Information is also included about employment in the most relevant activities in the services sector, which are more important in the most populous cities: Finance, insurance, and real estate, Educational, health, and other professional and related services, and employment in the Public administration.

Finally, we include several dummies which give us information about geographic localisation, and which take the value 1 depending on the region in which the city is located (Northeast Region, Midwest Region, or South Region; the West Region is used as a control category). Figure 4.21 is a map showing which states make up each of these regions, and how places of more than 10,000 inhabitants are distributed spatially[38]. These dummies show the influence of a series of variables for which individual data are not available for all places, and which are directly related to the geographical situation (temperature, rainfall, access to the sea, presence of natural resources, etc.).

[38] Sources:
Wikimedia Commons: http://commons.wikimedia.org/wiki/File:Map_of_USA_showing_regions.png
US Census Bureau: http://www.census.gov/dmd/www/map_1990.pdf

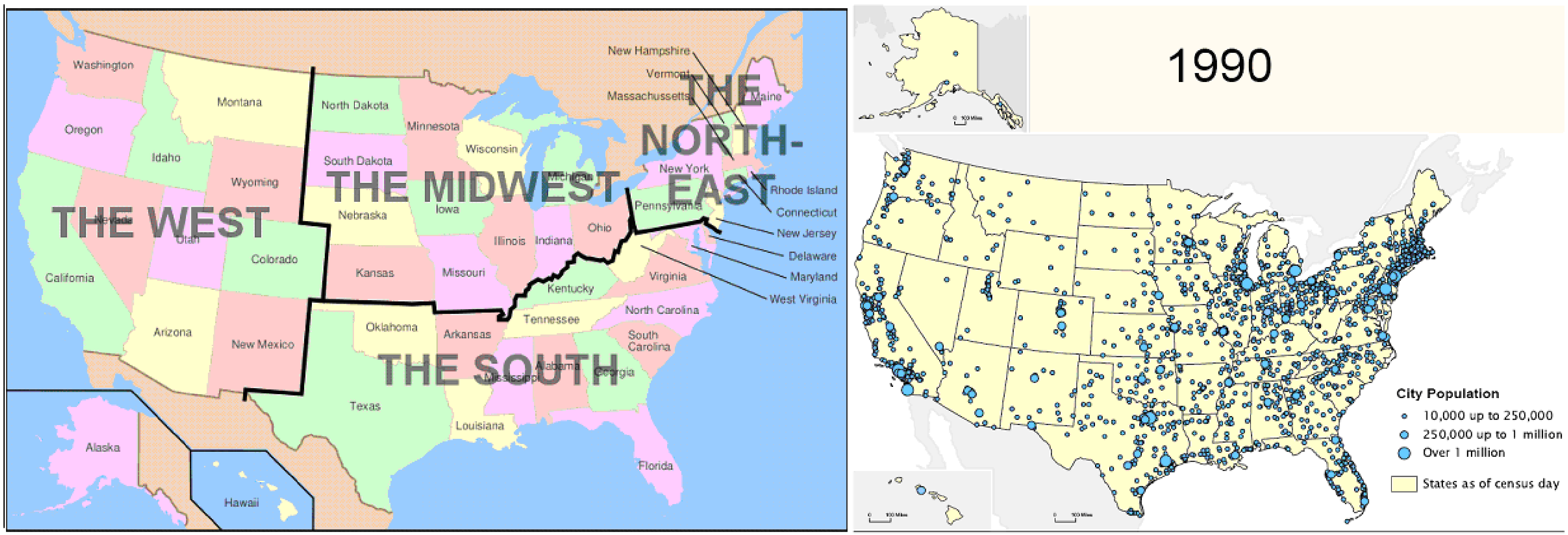


Fig. 4.21. Cities by Region.

4.3.2 Empirical model

In order to explain the different evolution of growth in per capita income and in population among cities in the 90s, we use a Multinomial Logit Model (MNLM), relating cities' probability of being located in any of the distribution quartiles according to growth (both in per capita income and in population) to the urban characteristics in 1990. We propose two separate models, one for the growth of per capita income and another for population growth, although as the explicative variables are the same, we can compare the results of both models.

The MNLM consists of transforming our dependent variable (the growth of city per capita income or of city population) into categories, which, to facilitate interpretation (and to ensure the groups are as homogeneous as possible in size), we make them coincide with the sample quartiles. This allows the results of the estimations to give us information about the probability (but not causality) of each variable affecting each category.

Thus, we rank the cities in descending order according to growth, and assign a value 1, 2, 3 or 4 according to which quartile the city's growth rate falls in, with 1 and 4 corresponding to 25% of cities with least and most growth, respectively.

Figure 4.22 shows the box plots representing these quartiles graphically, and Table 4.7 shows the concrete values separating some quartiles from others.

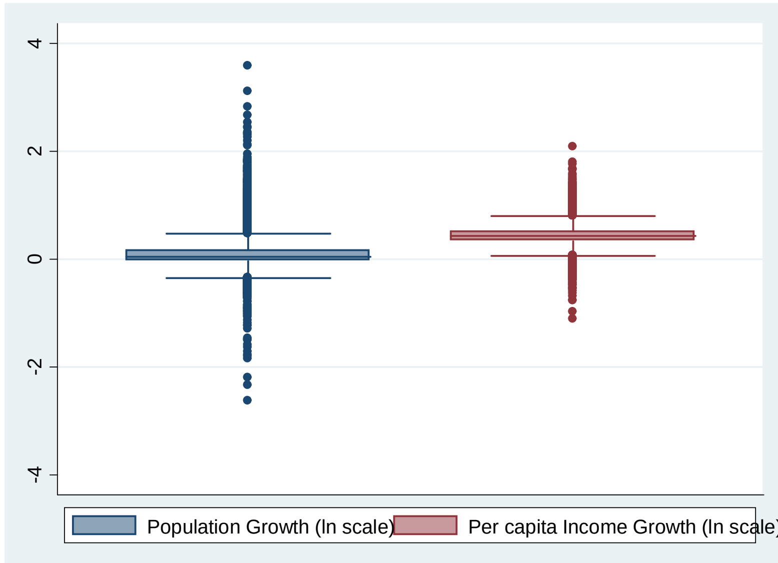


Fig. 4.22. Box Plots of City Per Capita Income Growth (ln scale) and City Population Growth (ln scale).

Table 4.7. CITY PER CAPITA INCOME GROWTH AND POPULATION GROWTH (LN SCALE): SAMPLE QUARTILES

Percentile	Population Growth	Per Capita Income Growth
25%	-0.0383	0.3378
50%	0.0471	0.4263
75%	0.1672	0.5206

Source: 1990 and 2000 Census, www.census.gov

It will be seen that the distribution of income growth is much more concentrated than population growth, which at the tails shows values very far from the median[39]. To complete the information on the quartiles, Table 4.8 relates both distributions.

[39] One advantage of this methodology is that by transforming growth rates into categories we eliminate the large variance they present (which could be the main problem when working with all population centres).

Table 4.8. CITIES BY SAMPLE QUARTILES

			Per Capita Income Growth Quartiles			
			1	2	3	4
Population Growth	Quartiles	1	5.23%	6.05%	6.11%	7.61%
		2	6.23%	7.25%	6.42%	5.10%
		3	6.73%	6.64%	6.28%	5.35%
		4	6.82%	5.05%	6.18%	6.95%

Source: 1990 and 2000 Census, www.census.gov

The first conclusion to be extracted is that, as shown in Figure 4.20, there is no clear relationship between growth in city per capita income and in city population, as none of the groups is over 8%. It is worth pointing out, however, that the most numerous group, 7.61%, indicates that most of the cities with most income growth are those with least population growth.

With the MNLM we estimate a separate binary logit for each pair of categories of the dependent variable. Formally, the MNLM can be written as:

$$\ln \phi_{m|b} = \ln \frac{\Pr(K = m \mid \mathbf{x})}{\Pr(K = b \mid \mathbf{x})} = \mathbf{x}'\beta_{m|b}, \quad \text{for } m = 1 \text{ a } J, \tag{4.67}$$

where b is the base category (in our case this will be category 1, the quartile containing the 25% of cities in the distribution with the lowest growth rates), $J = 4$ and $\mathbf{x}$ is the vector of the explicative variables, reflecting urban congestion costs, human capital, productive structure or geographical situation[40].

[40] The MNLM makes the assumption known as the independence of irrelevant alternatives (IIA). In this model: $\ln \frac{\Pr(K=m|\mathbf{x})}{\Pr(K=n|\mathbf{x})} = e^{\mathbf{x}'\left(\beta_{m|\mathbf{b}} - \beta_{n|\mathbf{b}}\right)}$, where the odds between each pair of alternatives do not depend on other available alternatives. Thus, adding or deleting alternatives does not affect the odds between the remaining alternatives. The assumption of independence follows from the initial assumptions that the disturbances are independent and homoscedastic. We have considered one of the commonest tests developed for testing the validity of the assumption, the Small-Hsiao (1985) test, and we cannot reject the null hypothesis, that is, the odds are independent of other alternatives, indicating that the MNLM is appropriate. The model corresponding to city per capita income growth also passes the Hausman test (Hausman and McFadden, 1998), for

We propose studying how these explicative variables affect the odds of a city being located in one category (quartile) or another, focusing in particular on quartiles 1 and 4, representing the cities (25% of the distribution) which grew least and most, respectively. For example, if the percentage of individuals with higher level education (Percent population 18 years and over: Some college or higher degree) increases, does the probability of the city belonging to that 25% of cities with highest growth also increase?

To deal with these questions we use odds ratios (also known as factor change coefficients). Maintaining the other variables constant, the change in the odds of the outcome m against outcome n, when x_i increases by δ, equals:

$$\frac{\phi_{m|b}\left(\mathbf{x}, x_i + \delta\right)}{\phi_{n|b}\left(\mathbf{x}, x_i\right)} = e^{\beta_{i,m|n}\delta}. \tag{4.68}$$

Thus, if $\delta = 1$ the odds ratio can be interpreted as follows: for each unitary change in x_i it is expected that the odds of m versus n change by a factor $e^{\beta_{i,m|n}}$, maintaining the other variables constant.

4.3.3 Results

This model includes many coefficients, making it difficult to interpret the effects for all pairs of categories. To simplify the analysis, odds-ratio plots were developed, shown in Figures 4.23, 4.24 and 4.25 for different groups of variables. To analyse the marginal effect of each variable in the change in the probability of a city being in one quartile or another, Tables 4.9 and 4.10 are presented, relative to the models of growth of per capita income and

---

the same null hypothesis.

of population, respectively, showing the marginal effects for each category and the absolute average change in probability.

In an odds ratio plot, each independent variable is represented in a separate row, and the horizontal axis indicates the relative magnitude of the coefficients β associated with each outcome[41]. The numbers which appear (1, 2, 3 or 4) are the four possible outcomes: the categories (coinciding with the sample quartiles) which we previously constructed.

These graphs reveal a great deal of information (for more details, see Long and Freese, 2006). To begin, if a category is to the right of another, this indicates that increases in the independent variable make the outcome to the right more likely. Also, the distance between each pair of numbers indicates the magnitude of the effect. And when a line connects a pair of categories, this indicates a lack of statistical significance for this particular coefficient, suggesting that these two outcomes are tied together. The three graphs take outcome 1 as the base category. We are especially interested in categories (quartiles) 1 and 4, corresponding to the tails of the distribution, the 25% of cities with the lowest and highest growth, respectively.

Initial levels

Regarding the effect of initial levels of per capita income and population, Table 4.9 shows that in the model corresponding to income growth the variable presenting the greatest absolute average change in probability (0.3498) is the initial per capita income in 1989. Also, the signs of the coefficients clearly indicate that the cities with the highest initial per capita income have a greater probability of ending up in quartiles 1 and 2 (below median

[41] The values of the coefficients β are shown in Tables 4.12 and 4.13 in the Appendix.

Table 4.9. CITY PER CAPITA INCOME GROWTH: MARGINAL EFFECTS FOR EACH CATEGORY AND THE AVERAGE ABSOLUTE CHANGE IN THE PROBABILITY

	Categories (quartiles)				
Initial levels	1	2	3	4	Total average
City Population (ln scale) in 1990	-0.0250***	0.0493***	0.0286***	-0.0530***	0.0390***
Per Capita Income (ln scale) in 1989	0.6171***	0.0826***	-0.1935***	-0.5062***	0.3498***
Congestion cost variables					
Median Travel Time to Work (in minutes)	-0.0018***	-0.0008	-0.0004	0.0030***	0.0015***
Big city dummy variable (population in 1990>25,000)	0.2221***	-0.0028***	-0.0869***	-0.1324***	0.1110***
Human capital variables					
Percent population 18 years and over: Some college or higher degree	-0.0096***	-0.0038***	0.0031***	0.0103***	0.0067***
Percent population 18 years and over: High school graduate (includes equivalency) or higher degree	-0.0007	0.0043***	0.0002	-0.0037***	0.0022***
Productive structure variables					
Unemployment rate	0.0054***	0.0005***	-0.0028***	-0.0031***	0.0030***
Percent employed civilian population 16 years and over:					
Agriculture, forestry, fishing, and mining	0.0047***	0.0024***	-0.0024***	-0.0046***	0.0035***
Construction	-0.0022**	-0.0006	0.0011**	0.0017***	0.0014**
Manufacturing (durable and nondurable goods)	-0.0001	0.0014**	0.0006	-0.0018***	0.0010***
Wholesale and Retail trade	0.0001	-0.0008	0.0004	0.0004	0.0004
Finance, insurance, and real estate	-0.0047***	-0.0010**	0.0016***	0.0041***	0.0029***
Educational, health, and other professional and related services	0.0025***	0.0012**	-0.0005***	-0.0031***	0.0018***
Public administration	0.0042***	0.0022*	-0.0017***	-0.0047***	0.0032***
Geographical dummy variables					
Northeast Region	-0.0440***	0.0930***	0.0160***	-0.0649	0.0545***
Midwest Region	-0.1736***	0.0321***	0.0992***	0.0423***	0.0868***
South Region	-0.0592***	0.0293***	0.0295***	0.0005***	0.0296***

Note: ***Significant at the 1% level **Significant at the 5% level *Significant at the 10 % level

Table 4.10. CITY POPULATION GROWTH: MARGINAL EFFECTS FOR EACH CATEGORY AND THE AVERAGE ABSOLUTE CHANGE IN THE PROBABILITY

	Categories (quartiles)				
Initial levels	1	2	3	4	Total average
City Population (ln scale) in 1990	-0.0285***	0.0371***	0.0188***	-0.0275	0.0280***
Per Capita Income (ln scale) in 1989	-0.1209***	0.0684***	0.0474***	0.0051***	0.0605***
Congestion cost variables					
Median Travel Time to Work (in minutes)	-0.0068***	-0.0022***	0.0020***	0.0070***	0.0045***
Big city dummy variable (population in 1990>25,000)	-0.0058	-0.0483	0.0252	0.0289	0.0271***
Human capital variables					
Percent population 18 years and over: Some college or higher degree	-0.0030***	-0.0036	0.0002***	0.0064***	0.0033***
Percent population 18 years and over: High school graduate (includes equivalency) or higher degree	0.0029***	0.0019	-0.0013***	-0.0035***	0.0024***
Productive structure variables					
Unemployment rate	0.0035***	0.0017*	-0.0034***	-0.0019***	0.0026***
Percent employed civilian population 16 years and over:					
Agriculture, forestry, fishing, and mining	0.0044***	0.0020**	-0.0014***	-0.0050***	0.0032***
Construction	-0.0038***	-0.0028	0.0031***	0.0034***	0.0033***
Manufacturing (durable and nondurable goods)	0.0003	0.0017	0.0005	-0.0025***	0.0012***
Wholesale and Retail trade	-0.0001	0.0002	-0.0002	0.0000	0.0001
Finance, insurance, and real estate	0.0000	0.0014	-0.0005	-0.0009	0.0007
Educational, health, and other professional and related services	0.0021***	0.0035	0.001	-0.0066***	0.0033***
Public administration	0.0027***	-0.0003**	-0.0012***	-0.0012***	0.0014**
Geographical dummy variables					
Northeast Region	0.3255***	0.0751***	-0.1408***	-0.2598***	0.2003***
Midwest Region	0.1547***	0.0889***	-0.0366***	-0.2070***	0.1218***
South Region	0.1823***	0.0141***	-0.0648***	-0.1316***	0.0982***

Note: ***Significant at the 1% level **Significant at the 5% level *Significant at the 10 % level growth); i.e., the richest cities grew less in this period, relating directly to the negative relationship observed in Figure 4.18. In contrast, the effect of initial population on income growth is not so clear, as the most likely categories are 2 and 3, simply indicating that, with a greater population in 1990, the most likely outcome in the year 2000 is the centre of the distribution. In the case of the model corresponding to population growth (Table 4.10), the effect of both variables is much less important.

Congestion cost variables

In principle, the bigger the city, the greater the median travel time borne by workers. Figure 4.23 points to category (quartile) 4 in both models as the most likely, which would indicate that indeed, where there is an increase in a unit of median travel time, the most likely outcome is that the city belongs to the 25% of cities with the highest growth, either in per capita income or in population. In other words, increases in travel time correspond to the cities which grew most, in population or in income, although the effect is greater in the case of population growth.

The other congestion variable is a dummy which takes the value 1 if the population of the city in 1990 is more than 25,000 inhabitants, enabling us to control specifically the most populous cities of the sample. Figure 4.25 indicates that in the case of population growth none of the odds ratios is significant, indicating the existence of a proportionate growth and the absence of a significant relationship between the initial population and growth (see Figure 4.18). On the contrary, the relationship with income growth appears to be negative: if a city had more than 25,000 inhabitants in 1990 it is most likely that it did not grow much

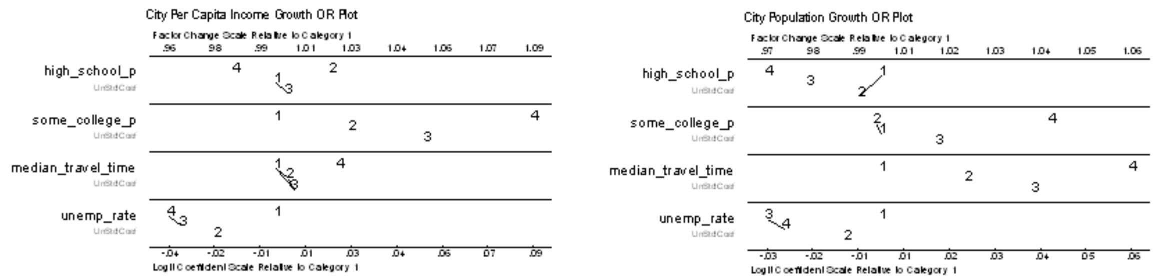


Fig. 4.23. Odds ratio plots of human capital variables, median travel time and unemployment rate.

in per capita income (in fact, the most likely outcome is quartile 1, the 25% of cities with the least income growth).

Human capital variables

The results show an opposite behaviour for the two human capital variables we introduced, both in population growth and in per capita income growth. Thus, if we focus on category 4, representing the 25% of cities which grew most in population or income, Figure 4.23 shows that increases in the percentage of the population with the highest education (some college or higher degree) have a positive impact on growth, since the most likely outcome is that the city will end up in quartile 4, while increases in the percentage of the population with a wider measure of human capital (high school graduate or higher degree) make the presence in outcome (quartile) 4 the least likely.

These results coincide with those of other studies analysing the influence of education in city growth. Glaeser and Shapiro (2003) also find that workers have a different impact depending on their education level[42] (high school or college). Simon and Nardinelli (2002)

42 In their sample of cities the different effect is completely due to the impact of California.

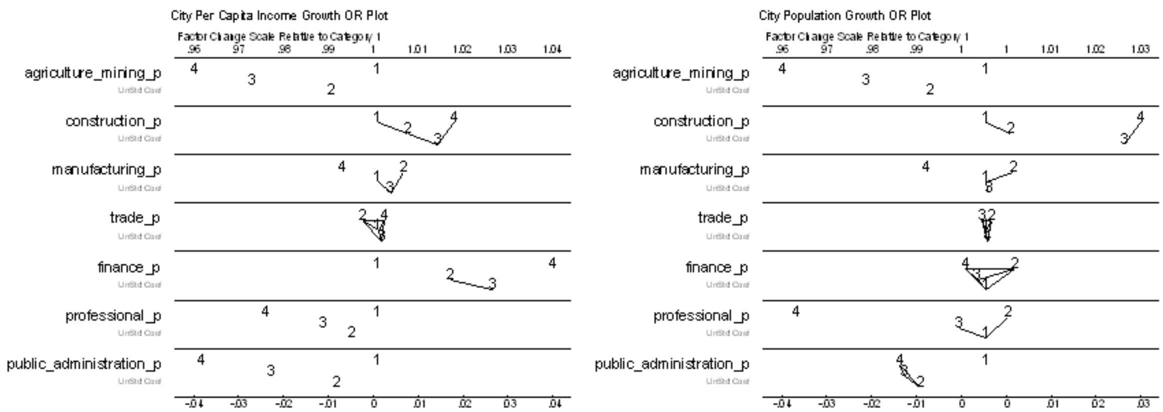


Fig. 4.24. Odds ratio plots of productive structure variables.

analyse the period 1900-1990 for the USA and conclude that the cities with higher average levels of human capital grew faster over the 20th century, and Glaeser and Saiz (2003), in their analysis of the period 1970-2000, show that the reason is that skilled cities are more economically productive (relative to less skilled cities).

Productive structure variables

In general, productive structure variables appear to have a very similar effect on the per capita income and population growth.

Figure 4.24 shows that both per capita income and population growth depend negatively on the initial unemployment rate. Thus, with an increase of 1% in the unemployment rate, the most likely outcome in both models is quartile 1, the 25% of cities with the lowest growth.

Regarding the distribution of employment by sectors, Table 4.9 shows that in the model corresponding to the growth of per capita income, the sector presenting the greatest average absolute change in probability (0.0035) is the primary sector (agriculture, forestry, fishing, and mining). If we interpret this variable as a proxy for the natural physical resources available to the city (cultivable land, access to the sea, etc.), Figure 4.24 points to by far the most likely outcome being category (quartile) 1. In other words, higher employment in the primary sector means a higher probability that the growth rate of the city will be in the lowest quartile, the 25% of cities with the lowest income growth. This negative effect is due to the fact that the primary sector usually presents constant or even decreasing returns to scale. The effect on population growth seems to be the same, with quartile 1 being the most likely outcome.

In contrast, employment in construction has a positive effect on growth, since Figure 4.24 shows 4 as the most likely category (quartile). The larger the percentage of employment in construction, the higher the probability that the city belongs to the 25% of the sample with the highest growth rate, either in per capita income or in population, with the average absolute change in probability being greater when explaining population growth (Tables 4.9 and 4.10).

The probability of per capita income or population growth being in the top quartile of the distribution (category 4) depends negatively on the initial percentage of employment in the manufacturing sector (Figure 4.24). This result coincides with that obtained by Glaeser et al. (1995) for the period 1960-1990, and its explanation is related to the depreciation of capital, suggesting that cities followed the fortunes of the industries that they were exposed to initially. However, in the 1990s the manufacturing sector seems to have lost importance, as the other sectors of activity had a greater influence on probability (Tables 4.9 and 4.10).

In the case of the services sector, only employment in finance, insurance, and real estate have a positive effect (the most likely outcome is category 4) on the growth rate of per capita income. Employment in professional services has a negative effect (the most likely outcome is category 1) and employment in wholesale and retail trade does not have a significant effect (the odds ratios are not significant). The influence of the services sector on the population growth rate seems to be much lower, since almost all the odds ratios are not significant.

The role of Geography

Until now the variables analysed seem to have a very similar effect on the growth both of per capita income and of population, since Figures 4.23 and 4.24 show a similar ordering of the categories in both models. Therefore, none of these variables is much help in explaining the divergence observed in the behaviour of the distributions of per capita income and population in cities.

If we return to Tables 4.9 and 4.10, the variables presenting the greatest average absolute change in probability (after the initial levels) in both models are the dummies corresponding to geographical location, which would indicate that the location of cities in one region or another is one of the most influential factors in the growth rate of per capita income and the population of a city. Also, the odds ratio plot (Figure 4.25) shows a completely different order between the two models, which indicates that the effect on the growth of per capita income and of population is different.

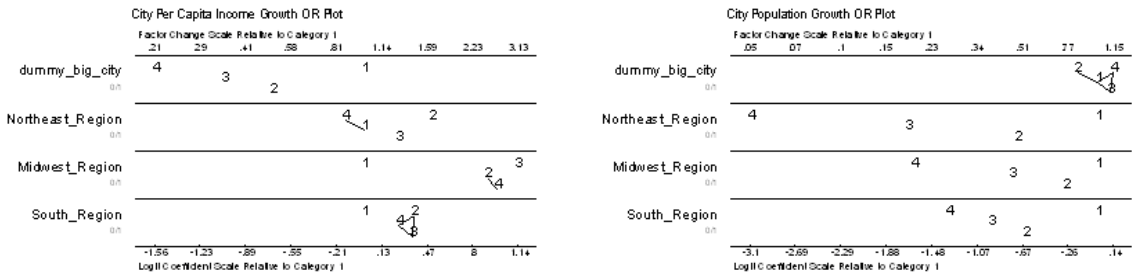


Fig. 4.25. Odds ratio plots of geographical dummy variables.

Remember that this dummy was used to capture the influence of a series of variables for which no individual data was available for all the places, directly related to the geographical situation: temperature, rainfall, access to the sea, the presence of natural resources, the availability of farming land, and even differences in economic and productive structures.

The influence of these variables has already been proven in previous works. Glaeser and Shapiro (2003) find that in the 1990s people moved to warmer, dryer places. Black and Henderson (1998) conclude that the extent of city growth and mobility is related to natural advantage, or geography. Beeson et al. (2001) show that access to transportation networks, either natural (oceans) or produced (railroads) was an important source of growth over the period 1840-1990, and that weather is one of the factors promoting population growth. Access to the sea seems to influence not only the growth rate of cities, but their location itself. In Figure 4.21 we can see how many cities are located on the coast. Finally, Mitchener and McLean (2003) find that some physical geography characteristics account for a high proportion of the differences in state productivity levels.

While the variable we introduced to control geography is a dummy at the regional level, the differences between cities at this geographical level are important.

Table 4.11 shows the average values for the different variables by regions, from which we can observe large regional differences. Thus, on average the West Region is where cities grew most in population, while the cities of the Midwest Region grew most in per capita income. In contrast, the cities of the Northeast Region experienced the lowest growth, both in per capita income and in population. The cities also present differences in their productive structures. The cities of the West Region present the highest unemployment rate, as well as a higher proportion of employment in the primary sector (agriculture, forestry, fishing, and mining), in construction and in the public administration. Employment in agriculture should indicate greater availability of land (as we can see in the map of Figure 4.21) in this region. In turn, the cities of the Midwest Region have a higher proportion of employment in manufacturing and wholesale and retail trade. Finally, the cities of the Northeast Region have the inhabitants with the highest levels of human capital, and a higher proportion of employment in the services sector.

These different economic structures, and geographical characteristics, seem to be the key to explaining the different behaviour of per capita income and population growth in the cities in the 1990s.

4.4 Conclusions

This chapter analyses the determinants of growth of American cities, understood as the growth of either the population or per capita income, from 1990 to 2000. This empirical

Table 4.11. CITY VARIABLES IN 1990: MEANS BY REGION

Variable	Mean				
	All sample	Northeast Region	Midwest Region	South Region	West Region
Population Growth (ln scale), 1990-2000	0.09	0.01	0.06	0.11	0.20
Per Capita Income Growth (ln scale), 1989-1999	0.43	0.38	0.46	0.44	0.41
Congestion cost variable					
Median Travel Time to Work (in minutes)	20.41	21.81	19.91	21.10	19.61
Human capital variables					
Percent population 18 years and over: Some college or higher degree	27.30	32.24	25.32	25.51	31.25
Percent population 18 years and over: High school graduate (includes equivalency) or higher degree	53.04	60.15	53.89	48.91	53.04
Productive structure variables					
Unemployment rate	6.82	5.86	6.33	7.17	8.25
Percent employed civilian population 16 years and over:					
Agriculture, forestry, fishing, and mining	5.15	1.99	4.87	5.52	8.28
Construction	6.85	6.43	6.37	7.25	7.55
Manufacturing (durable and nondurable goods)	19.59	19.28	21.81	20.38	12.58
Wholesale and Retail trade	21.15	21.49	21.89	20.42	20.65
Finance, insurance, and real estate	5.14	6.61	5.00	4.80	4.71
Educational, health, and other professional and related services	22.79	25.42	22.93	21.41	22.88
Public administration	4.80	4.50	3.81	5.26	6.51
Sample size	21,655	3,276	7,922	7,278	3,179

Source: 1990 and 2000 Census, www.census.gov analysis uses data from all cities with no size restriction (our sample contains data for 21,655 cities). Our results show that while population growth of the cities appears to be independent of initial size (the empirical regularity known as Gibrat's law), the growth of per capita income is negatively correlated to initial per capita income: the richest cities grew less in this period. This explains why, while the empirical distribution of city population remains stable in the decade 1990-2000, the distribution of per capita income changes.

In order to explain these differentiated behaviours, we examine the relationship between the urban characteristics in 1990 and the city growth (both in population and in per capita income) using a Multinomial Logit Model. Apart from the initial levels of population and per capita income, we have considered variables for congestion costs, human capital, and the productive structure, as well as geographical variables. The results we obtained with our sample of all cities are similar to those of other studies which focused only on the most populous cities. Thus, we find that the probability of a city being in the 25% of cities with most growth in income or population (i.e., the probability of the growth rate of per capita income or population being in the top quartile of the distribution) depends

(1) positively on the initial percentage of inhabitants with higher educational levels (some college or higher degree), although the sign and intensity of the effect change when considering a wider concept of education (high school graduate or higher degree);

(2) negatively on initial unemployment levels, and

(3) negatively on the initial percentage of employment in the manufacturing sector, although this sector seems to have lost weight, as other economic sectors have a greater influence on probability.

Also, the location of cities on one region or another is one of the most influential factors on the growth rate of a city's per capita income or population.

4.A Appendix: Estimated multinomial logit coefficients

Tables 4.12 and 4.13 show the results of the estimated MNLM. The values of the β coefficients are relative to category (quartile) 1, which is the base outcome.

Table 4.12. City Per Capita Income Growth: Multinomial Logit coefficients relative to Category (quartile) 1

	Categories (quartiles)		
Initial levels	2	3	4
City Population (ln scale) in 1990	0.2873***	0.2070***	-0.1440***
Per Capita Income (ln scale) in 1989	-2.2360***	-3.2478***	-4.9018***
Congestion cost variables			
Median Travel Time to Work (in minutes)	0.0042	0.0058	0.0214***
Big city dummy variable (population in 1990>25,000)	-0.6839***	-1.0468***	-1.5649***
Human capital variables			
Percent population 18 years and over: Some college or higher degree	0.0255***	0.0509***	0.0873***
Percent population 18 years and over: High school graduate (includes equivalency) or higher degree	0.0189***	0.0036	-0.0142***
Productive structure variables			
Unemployment rate	-0.0202***	-0.0325***	-0.0368***
Percent employed civilian population 16 years and over:			
Agriculture, forestry, fishing, and mining	-0.0103***	-0.0281***	-0.0409***
Construction	0.0067	0.0132**	0.0170***
Manufacturing (durable and nondurable goods)	0.0056**	0.0028	-0.0081***
Wholesale and Retail trade	-0.0036	0.0009	0.0014
Finance, insurance, and real estate	0.0158**	0.0254***	0.0387***
Educational, health, and other professional and related services	-0.0058**	-0.0122***	-0.0249***
Public administration	-0.0093*	-0.0238***	-0.0393***
Geographical dummy variables			
Northeast Region	0.5063***	0.2518***	-0.1453
Midwest Region	0.9134***	1.1424***	0.9886***
South Region	0.3624***	0.3601***	0.2567***

Note: 1 is the base outcome. ***Significant at the 1% level **Significant at the 5% level *Significant at the 10 % level

Table 4.13. City Population Growth: Multinomial Logit coefficients relative to Category (quartile) 1

	Categories (quartiles)		
Initial levels	2	3	4
City Population (ln scale) in 1990	0.2629***	0.1890***	0.0007
Per Capita Income (ln scale) in 1989	0.7728***	0.6831***	0.5308***
Congestion cost variables			
Median Travel Time to Work (in minutes)	0.0204***	0.0360***	0.0592***
Big city dummy variable (population in 1990>25,000)	-0.1788	0.1138	0.1434
Human capital variables			
Percent population 18 years and over: Some college or higher degree	-0.0014	0.0134***	0.0402***
Percent population 18 years and over: High school graduate (includes equivalency) or higher degree	-0.0049	-0.0171***	-0.0272***
Productive structure variables			
Unemployment rate	-0.0083*	-0.0272***	-0.0230***
Percent employed civilian population 16 years and over:			
Agriculture, forestry, fishing, and mining	-0.0111**	-0.0237***	-0.0402***
Construction	0.0050	0.0274***	0.0306***
Manufacturing (durable and nondurable goods)	0.0054	0.0007	-0.0119***
Wholesale and Retail trade	0.0012	-0.0006	0.0005
Finance, insurance, and real estate	0.0057	-0.0017	-0.0038
Educational, health, and other professional and related services	0.0045	-0.0054	-0.0375***
Public administration	-0.0128**	-0.0159***	-0.0168***
Geographical dummy variables			
Northeast Region	-0.7153***	-1.6892***	-3.0963***
Midwest Region	-0.2752***	-0.7572***	-1.6395***
South Region	-0.6339***	-0.9430***	-1.3269***

Note: 1 is the base outcome. ***Significant at the 1% level **Significant at the 5% level *Significant at the 10 % level

References

Beeson, P. E., D. N. DeJong, and W. Troesken, (2001). Population Growth in US Counties, 1840-1990. Regional Science and Urban Economics, 31: 669-699.

Black, D., and V. Henderson, (1998). Urban evolution in the USA. Brown University Working Paper No. 98-21.

Duranton, G., (2000). Urbanization, urban structure, and growth, in: J.-M. Huriot and J. F. Thisse (Eds) Economics of Cities: Theoretical Perspectives, 290–317. Cambridge: Cambridge University Press.

Duranton, G., and D. Puga, (2004). Micro-foundations of urban agglomeration economies. Handbook of urban and regional economics, Vol. 4, J. V. Henderson and J. F. Thisse, eds. Amsterdam: Elsevier Science, North-Holland, Chapter 48, 2064-2117.

Eeckhout, J., (2004). Gibrat's Law for (All) Cities. American Economic Review, American Economic Association, 94(5): 1429-1451.

Fujita, M. and J.-F. Thisse, (2002). Economics of Agglomeration. Cambridge: Cambridge University Press.

Garicano, L., and E. Rossi-Hansberg, (2006). Organization and inequality in a knowledge economy. The Quarterly Journal of Economics, MIT Press, 121(4): 1383-1435.

Gibrat, R., (1931). Les inégalités économiques, París: Librairie du Recueil Sirey.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer, (1992). Growth in Cities. The Journal of Political Economy, 100(6): 1126-1152.

Glaeser, E. L., and A. Saiz, (2003). The Rise of the Skilled City. Harvard Institute of Economic Research, Discussion Paper number 2025.

Glaeser, E. L., J. A. Scheinkman, and A. Shleifer, (1995). Economic growth in a cross-section of cities. Journal of Monetary Economics, 36: 117-143.

Glaeser, E. L., and J. Shapiro, (2003). Urban Growth in the 1990s: Is city living back? Journal of Regional Science, 43(1): 139-165.

Hausman, J., and D. McFadden, (1984). Specification tests for the multinomial logit model. Econometrica 52, 1219-1240.

Henderson, V., A. Kuncoro and M. Turner, (1995). Industrial development in cities. Journal of Political Economy 103 (5), 1067-1090.

Jacobs, J., (1969). Economy of Cities. Vintage, New York.

Long, J. S., and J. Freese, (2006). Regression Models for Categorical Dependent Variables Using Stata. Stata Press Publication, 2nd ed.

Loury, G. C., (1979). Market Structure and Innovation. The Quarterly Journal of Economics, MIT Press, 93(3): 395-410.

Mitchener, K. J., and I. W. McLean, (2003). The Productivity of US States since 1880. Journal of Economic Growth, 8: 73-114.

Polèse, M., (2005). Cities and national economic growth: A reappraisal. Urban Studies, 42(8): 1429-1451.

Quigley, J. M., (1998). Urban diversity and economic growth. Journal of Economic Perspectives, 12(2): 127–138.

Simon, C. J., and C. Nardinelli, (2002). Human capital and the rise of American cities, 1900-1990. Regional Science and Urban Economics, 32: 59-96.

Small, K. A., and C. Hsiao, (1985). Multinomial logit specification tests. International Economic Review 26, 619-627.

Yamamoto, D., (2008). Scales of regional income disparities in the USA, 1955–2003. Journal of Economic Geography, 8(1): 79-103.

Young, A. T., M. J. Higgins, and D. Levy, (2008). Sigma Convergence versus Beta Convergence: Evidence from U.S. County-Level Data. Journal of Money, Credit and Banking, 40(5): 1083-1093.

Conclusiones

Siguiendo la normativa de la Universidad de Zaragoza, a continuación presentamos un resumen en castellano de las principales conclusiones que han sido ya expuestas en cada capítulo.

Esta tesis doctoral se ha dirigido al estudio de la concentración económica de la actividad y el crecimiento. En primer lugar, planteamos un modelo teórico con el que analizar las consecuencias de la integración comercial entre países que pueden ser distintos, debido a sus distintas dotaciones de riqueza y a sus diferentes políticas comerciales. El objetivo es estudiar como se ve afectada la concentración de la industria, la tasa de crecimiento de la economía y el bienestar. El marco teórico seguido es el de Martin y Ottaviano (1999) completado con la incorporación de asimetrías en los costes de transacción que permiten que los países presenten sesgos o tendencias distintas respecto a los intercambios según se trate de ventas o de compras, es decir, considerando que el coste del comercio internacional es distinto según sea el sentido del intercambio. Para ello, en nuestro modelo distinguimos entre costes comerciales domésticos e internacionales y, dentro de éstos, comunes, de importación y de exportación.

Los resultados que encontramos con esta clasificación coinciden con Martin (1999) en lo que se refiere a los efectos de los costes comerciales domésticos, pero difieren en lo que hace referencia a los costes internacionales. Martin (1999) encontraba una relación positiva directa inequívoca entre menores costes de transacción internacional y crecimiento. En nuestro marco de trabajo, el proceso de integración comercial puede conducir a un aumento de la concentración industrial, de la tasa de crecimiento y del bienestar o no, dependiendo de qué país adopte las medidas necesarias para la integración y de qué costes se reduzcan. Así, si el país rico reduce sus costes comerciales domésticos o de exportación o el país pobre reduce sus costes de importación, la concentración industrial aumenta en el país rico, al tiempo que también aumenta la tasa de crecimiento de la economía y el bienestar en ambos países. Sin embargo, si es el país pobre el que reduce sus costes domésticos o introduce políticas de fomento de sus exportaciones (o el país rico reduce sus costes de importación), la integración conduce a una disminución de la concentración industrial y una disminución de la tasa de crecimiento, pudiendo en este caso el país pobre aumentar su bienestar. Para terminar podemos decir que un proceso de integración comercial no tiene un efecto monótono sobre la concentración industrial ni sobre la tasa de crecimiento de la economía ni sobre el bienestar de los países implicados en el mismo y el resultado final acaba dependiendo de qué país lleve a cabo la actuación política y de qué tipo de coste comercial se vea afectado por esa decisión.

En el segundo capítulo se presenta un modelo que integra características de la Nueva Geografía Económica, la teoría del crecimiento endógeno, y la economía de los recursos naturales. Este marco teórico nos permite estudiar explícitamente el efecto de las "first nature causes" en la concentración de la actividad económica, analizando una de las posibles características naturales geográficas, la presencia de un recurso natural en el territorio. La geografía se introduce en el modelo a través de los costes de transporte, que condicionan la distribución de las empresas, que intentan aprovechar los rendimientos crecientes en un mercado de competencia monopolística. El crecimiento económico se sustenta en un marco endógeno con efectos externos de ámbito nacional en la innovación, lo que provoca que la actividad investigadora se desarrolle en un único país y, por tanto, cuanto mayor es la concentración industrial en ese país mayor es la tasa de crecimiento de la economía. Finalmente, el recurso natural aparece como un input localizado en uno de los dos países, que otorga una ventaja en costes a las empresas situadas en ese país.

Los resultados muestran que, ante un descenso en cualquiera de los costes de transporte, las empresas deciden desplazarse al país con mayor demanda y tamaño del mercado doméstico. A pesar de la ventaja en costes que supone situarse en el Sur debido a la presencia del recurso, las empresas prefieren moverse al Norte, que es el país rico, donde podrán aprovechar mejor los rendimientos crecientes. A su vez, la concentración aumenta la tasa de crecimiento, dada la naturaleza nacional de los efectos externos. Por último, la concentración de las empresas en el Norte también tiene un efecto positivo sobre el stock de recurso natural, que aumenta. A pesar de identificar dos efectos contrapuestos, un efecto localización de las empresas y un efecto crecimiento, el primero es el dominante, ya que a mayor número de empresas en el Norte menos recurso se consume, lo que permite aumentar el nivel del stock en equilibrio estacionario. Esto es debido a que las empresas del Norte reaccionan a la ventaja en costes de las empresas del Sur produciendo menos unidades de bien diferenciado (y consumiendo por tanto menos recurso) pero vendiéndolas a un precio superior.

Esto significa que, en el marco de nuestro modelo, el efecto del mercado doméstico (las "second nature causes"), que actúa como una fuerza centrípeta, tiene mayor peso en la toma de decisiones de las empresas que las ventajas aportadas por las circunstancias geográficas naturales (las “first nature causes”), que actúan como una fuerza centrífuga. No obstante, el Sur podría aumentar la importancia de la “first nature cause” mediante la introducción de políticas públicas que refuercen la ventaja en costes que supone la presencia del recurso natural para las empresas situadas en el Sur. Hemos considerado dos políticas públicas: imponer trabas al comercio internacional del recurso y promover un cambio tecnológico a una tecnología más intensiva en el uso del recurso. En ambos casos, el Sur consigue atraer empresas desde el Norte, lo que provoca que disminuyan tanto la tasa de crecimiento como el stock de recurso natural en equilibrio. El efecto sobre el bienestar queda indeterminado.

Sin embargo, nuestros resultados dependen de las características del recurso natural considerado. Se trata de un recurso natural (i) renovable, (ii) de libre acceso, (iii) que se utiliza únicamente como bien intermedio en la producción de manufacturas, y (iv) cuya explotación requiere solo trabajo. Estos supuestos nos han permitido construir el modelo más sencillo posible analíticamente, que podemos denominar como el modelo básico. Variaciones en cualquiera de estas características pueden dar lugar a extensiones del modelo.

En particular, existen dos posibles extensiones que podrían ampliar nuestro conocimiento sobre la relación entre los recursos naturales y la distribución de la actividad económica. En primer lugar, dado que en la actualidad la mayoría de los recursos naturales que se utilizan en la producción de manufacturas provienen de derivados del petróleo o yacimientos minerales, sería de interés analizar cómo cambia nuestro modelo cuando el recurso natural es no renovable. Otro aspecto de gran interés consiste en cambiar el régimen de los derechos de propiedad del recurso. Si el recurso no fuera de libre acceso, el sector generaría unas rentas adicionales que, dependiendo de quién fuera el propietario de los derechos, podrían repercutir en el Sur. Este efecto renta, sumado al efecto costes que ya aparece en nuestro modelo, aumentaría el peso de las “first nature causes” en la toma de decisiones de las empresas.

En el tercer capítulo, ya empírico, se contrasta el cumplimiento de la Ley de Gibrat en el crecimiento urbano, utilizando datos para todo el siglo veinte de la distribución completa de ciudades (sin ninguna restricción de tamaño) en tres países: Estados Unidos, España e Italia. El primer resultado es que, en el corto plazo, al considerar toda la distribución de ciudades, obtenemos una tendencia a la divergencia. A pesar de ser tres países con estructuras e historias urbanas muy distintas, encontramos una relación positiva entre la tasa de crecimiento de las ciudades y su tamaño inicial durante todo el siglo veinte (excepto en el periodo 1970-1980 en Estados Unidos). Sin embargo, la evidencia empírica muestra que esto no impide que la distribución del tamaño de las ciudades pueda aproximarse como una distribución lognormal.

En el largo plazo, los contrastes de raíz unitaria confirman la validez de la Ley de Gibrat en la cola superior de la distribución. Y el uso de métodos no paramétricos, que relacionan la tasa de crecimiento con el tamaño inicial a través de la estimación de medias locales, nos permite observar que, en el largo plazo, se incrementa la evidencia en favor de un cumplimiento débil de la Ley de Gibrat (el tamaño afecta a la varianza del crecimiento, aunque no a su media).

Además, el caso de Estados Unidos es diferente al de los países europeos debido a la entrada de nuevas ciudades en la distribución. Las ciudades que entran en la distribución presentan tasas de crecimiento superiores en media y varianza a la media de toda la muestra, aunque no podemos rechazar estadísticamente que sean significativamente distintas. Estas diferencias son mayores en varianza, indicando que parte de la varianza superior de la cola inferior de la distribución puede explicarse con las ciudades que van entrando en la distribución a lo largo del siglo veinte.

Por último, el cuarto capítulo de la tesis contiene un estudio de los determinantes del crecimiento urbano americano, entendido como crecimiento en población o en renta per cápita, entre los años 1990 y 2000. Para este análisis empírico se emplean todas las ciudades sin ninguna restricción de tamaño (nuestra muestra contiene datos para 21.655 ciudades). Los resultados muestran que, mientras que el crecimiento de la población de las ciudades parece ser independiente del tamaño inicial (la regularidad empírica conocida como Ley de Gibrat), el crecimiento de la renta per capita está negativamente correlacionado con la renta per cápita inicial: las ciudades más ricas crecieron menos en este periodo. Esto explica por qué mientras que la distribución empírica de la población de las ciudades permanece estable durante la década 1990-2000, la distribución empírica de la renta per cápita cambia.

Para explicar el comportamiento diferenciado de las tasas de crecimiento se examina la relación entre las características urbanas en 1990 y el crecimiento urbano (tanto en población como en renta per cápita) utilizando un Modelo Logit Multinomial. Aparte de los niveles iniciales de población y renta per capita, utilizamos variables de costes de congestión, de capital humano, de estructura productiva y variables geográficas. Los resultados que obtenemos con nuestra muestra de todas las ciudades son similares a los de otros estudios que se centran únicamente en las ciudades más pobladas. Así, encontramos que la probabilidad de que una ciudad se encuentre entre el 25% de las ciudades que más crecen en renta o población (es decir, la probabilidad de que las tasas de crecimiento de la renta per cápita o de la población se encuentren en el cuartil superior de la distribución) depende

(1) positivamente del porcentaje inicial de habitantes con altos niveles educativos, si bien el signo y la intensidad del efecto cambian al considerar un concepto de educación más amplio;

(2) negativamente del nivel de desempleo inicial; y

(3) negativamente del porcentaje inicial de empleo en el sector de manufacturas, aunque dicho sector parece haber perdido importancia ya que el resto de sectores de actividad tienen una mayor influencia en la probabilidad.

Además, la localización de las ciudades en una u otra región aparece como uno de los factores más influyentes en la tasa de crecimiento de la renta per cápita o de la población de una ciudad.