TESIS DE LA UNIVERSIDAD
DE ZARAGOZA

2015

56

Miguel Ángel Puente Ajovín

Essays on Economic Geography, Regional and Urban Economics

Departamento

Análisis Económico

Director/es

Sanz Gracia, Fernando
Ramos Gutiérrez, Arturo

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

ISSN 2254-7606

Universidad Zaragoza

1542

Tesis Doctoral

ESSAYS ON ECONOMIC GEOGRAPHY, REGIONAL AND URBAN ECONOMICS

Autor

Miguel Ángel Puente Ajovín

Director/es

Sanz Gracia, Fernando
Ramos Gutiérrez, Arturo

UNIVERSIDAD DE ZARAGOZA

Análisis Económico

2015

Essays on Economic Geography, Regional and Urban Economics

A thesis submitted by
Miguel Ángel Puente Ajovín

Supervised by
Arturo Ramos Gutiérrez and Fernando Sanz Gracia

for the degree of

PhD Programme in Economics

Economic Analysis Department

Universidad Zaragoza

April 2015

Acknowledgements

I would like to acknowledge the support of my two supervisors, Arturo Ramos Gutiérrez and Fernando Sanz Gracia, whose help, both personal and professional, has been invaluable.

My thanks to the members of the research group Adetre, especially to Domingo Pérez-Ximénez-de-Embún, Marcos Sanso-Navarro and María Vera-Cabello, whose friendship is one of the best consequences of this process; and to Daniel Arribas Bel, for his kind welcome in Birmingham.

I also want to thank the members of Econstuff, with whom I have shared my passion for the divulgation of economic ideas all this time.

Finally, I want to express my deepest gratitude to my family for their patience and support and for believing that I could achieve something even when I had my doubts.

Contents

List of Figures

List of Tables

Resumen

La presente Tesis está compuesta de cuatro artículos originales que pertenecen al campo de la Geografía Económica y la Economía Regional y Urbana. Se presenta en dos partes, cada una compuesta, a su vez, por dos papeles.

El primer capítulo presenta un nuevo modelo teórico de la Nueva Geografía Económica en donde se introducen empresas heterogéneas, permitiendo que los trabajadores de una cierta industria puedan, como los empresarios con alto nivel de gestión empresarial, crear su propia empresa. La distinción entre estos dos tipos de emprendedores no se basa únicamente en su nivel de productividad: su capacidad para moverse entre regiones es también diferente. Este hecho nos permite analizar los efectos derivados de la competencia entre diferentes tipos de empresa en un ámbito geográfico y cómo afectan al proceso de concentración o dispersión económica en el espacio.

El modelo concluye que la heterogeneidad empresarial fomenta la concentración de las empresas menos productivas, si bien la localización de estas depende del valor del parámetro que representa la libertad de comercio. Además, una mayor heterogeneidad favorece la localización de la actividad económica (medida en ingresos, beneficios o número de trabajadores) en aquella región que concentra, a su vez, a las empresas más productivas.

El segundo capítulo desarrolla un marco alternativo al proceso de transporte de tipo Iceberg, asumido de forma general en los modelos de la Nueva Geografía Económica para introducir una cierta fricción en el comercio derivada del espacio. Nuestro marco teórico resuelve los problemas más importantes asociados al planteamiento Iceberg, a la vez que mantiene su misma estructura matemática, de forma que los dos modelos son isomórficos, aunque llegan a diferentes conclusiones. En nuestro marco introducimos el coste de manera más realista, presentando los costes de transporte de forma directa en la función de producción y no como una "caja negra", como hace el marco Iceberg. Además, se producen cambios importantes en la forma de interpretar el parámetro de libertad de comercio, que no depende solo del coste de transporte, sino también de su relación con el coste de producción.

El propósito del tercer capítulo es estimar la distribución del tamaño de las ciudades con tres funciones de densidad: la lognormal, la doble Pareto lognormal y la normal-Box-Cox.

La base de datos se compone de cuatro países: Alemania, Francia, Italia y España, estos dos últimos con información de todo el siglo XX. Realizando los contrastes de Kolmogorov-Smirnov y Cramér-Von Mises podemos discriminar entre las funciones que pueden explicar esta distribución; asimismo, a partir de los criterios de información de Akaike y Bayesiano elegimos la mejor opción.

Nuestra principal conclusión es que no hay ninguna función que, claramente, sea mejor que el resto para los cuatro países seleccionados.

El cuarto capítulo profundiza en la temática del anterior, analizando ahora la distribución de las tasas de crecimiento de las ciudades para los mismos cuatro países: Alemania, Francia, Italia y España. En este caso, las funciones consideradas son la normal, la *alpha-Stable* y la t de Student. Utilizando los contrastes de Kolmogorov-Smirnov y Cramér-Von Mises, para ver el ajuste entre la función teórica y la distribución empírica, y los criterios de información de Akaike y Bayesiano concluimos, como en el capítulo anterior, que ninguna función destaca entre las tres de manera regular.

Introduction

The present PhD Thesis is composed of four original papers that belong to the field of Economic Geography, Regional and Urban Economics. It is presented in two parts, each part containing two chapters.

The first part is based on a pure theoretical analysis within the New Economic Geography framework; its purpose is to understand the behavior of the economy through mathematical models that take the space into account as one of its main components. The focus of this approach is mainly on the distribution of economic activities, and how this process is shaped by the concentration and dispersion forces that are created endogenously.

Although the economic studies that used the space dimension can be traced as far as back as Von Thunen (1826), it is the model developed by Krugman (1991) that is used as the base for the New Economic Geography. This model is determined by two important features: The introduction of economies of scale into the production of differentiated goods that foster the concentration of the production process, and the presence of costs of transportation, in which the effects of the space on the localization of agents impose certain restrictions on their economic interactions. The first two chapters of this thesis deal with two main objectives: The first develops a theoretical model that introduces heterogeneous firms, and analyzes its effects on the concentration process; the second proposes an alternative to the Iceberg transportation costs that is more in line with the real world.

The first chapter is associated with the inherent heterogeneity of the agents that compose the economic system. It defines a topic of great interest in the field of the New Economic Geography since it was presented as a necessary new area of research by Behrens and Robert-Nicoud (2011) and Ottaviano (2011). As a result, in recent years, there has been an increase in the number of scientific papers that try to describe the localization of different agents using some kind of heterogeneity. The aim of these studies is to analyze the process of economic concentration over the space and, thus, how regions can specialize not only in one particular economic sector, but also in certain types of firms (a country could concentrate the more productive or the less productive companies, for example). Furthermore, competition forces foster the entry or exit of these heterogeneous firms differently, changing their location decisions which, in turn, also affects the welfare level of the regions (due, for example, to the access to a cheaper range of goods or to a greater income derived by a higher productivity level of the businesses). This has been done, mainly, within the framework developed by Melitz (2003), which introduces differences in the productivity levels of firms. Redding (2011) review the theoretical literature between heterogeneity and trade, and Melitz and Redding (2013) account for the empirical nature of this synthesis.

These previously cited works highlight the conclusion that a trade liberalization process will generate an increase in the average productivity level of the economy due to higher competition that pushes the less competitive firms out of the market, while the more productive ones expand to other regions. However, questions as important as how a greater firm heterogeneity affects the concentration process of these companies, or how these relationships can change within different geographical units (cities, regions, countries), have less consensus.

The first chapter of this thesis develops a new model within the New Economic Geography framework with heterogeneous firms. It introduces a new type of firm heterogeneity, in which the differences between firms are not only present in the productivity level, but also in how these companies are created. This allows us to acknowledge the fact that the workers of a certain industry, in addition to pure entrepreneurs, can create their own firms, even though they are less prepared for the management process than other businessmen. This distinction between the two kinds of entrepreneurs is not based only on the productivities that their companies will achieve: their ability to move between regions is also different. This new interpretation shows that firm heterogeneity fosters the concentration of the least productive firms; however, the location of this agglomeration depends on the value of the freeness of trade parameter. We also conclude that a greater heterogeneity concentrates the economic activity in the core, due to localization of the most productive firms in it.

Continuing with the theoretical part of this thesis, the second chapter develops an alternative to the Iceberg transportation costs that tries to explain, in a more realistic way, the negative effects that the geographical space impose on the economic transactions of goods between regions.

Since Samuelson (1952), the Iceberg approach has been proposed to explain, in a simple manner, the way in which the transportation process occurs: T($>$1) units of a certain commodity have to be sent from one region for one unit to arrive at another region. Krugman (1991) adopted this approach to implement it into his own models, contributing to its wide-spreading and converting it into one of the core elements of the New Economic Geography. But, although the Iceberg transportation costs are a really simple and operative instrument for the acknowledgement of the effects of space, they have important flaws: many of their economic properties are at odds with the transport cost's properties derived from the real world.

In a more detailed way, McCann (2005) explains that, due to its exponential formulation, the Iceberg has three problematic outcomes: First, the transportation costs increase in a convex manner with respect to distance; second, the transport cost per ton-kilometer is invariant in relation to the initial weight of the product being sent; and third, the export price has a unitary elasticity with respect to the domestic price. Finally we introduce a fourth criticism which is associated with the Iceberg approach's consideration of an irreal homogeneous space.

Thus, the main purpose of the second chapter is to develop an alternative framework to resolve the several and important problems pointed out before. In doing so, we maintain the same structure as the Iceberg cost and, therefore, the models are isomorphic, although with different economic underpinnings. This new approach is based on a more realistic analysis of the modern transport process, incorporating the transportation costs directly into the production function and not as a "black box", like the Iceberg approach does. Furthermore, a new and interesting way to understand the freeness of trade parameter is deduced.

In contrast with the first two, purely theoretical, chapters, the second part of this thesis deals with the empirical analysis of city size distribution. This topic has a long research tradition, encouraged by the ubiquitous advance of the urban society. A rigorous analysis of the size and growth distribution of cities can improve our knowledge of the fundamental reasons that explain these urban agglomerations. We can not forget that it is in cities where the main economic activity and the main forces that are theoretically studied in the Economic Geography framework take place; moreover, the nature of the empirical distribution can be used as a benchmark for the analysis of the concentration of other economic variables, such as the distribution of income and the number of firms or employees.

Of course, the continuous interest in these topics is explained by the lack of consensus about the theoretical distribution that could explain the particular equilibrium that each country presents. As a general approximation, one of the more frequently used functions for the size distributions of cities is the rank-size rule, or Zipf's law. Following this function, based on the contribution of Zipf (1949), the second most populated city will have half of the population of the first, the third will have one third, and so on. This rule accurately explains the distribution of population of the bigger cities within a country (upper tail); however, it lacks real explanatory power when all the nuclei are considered.

A new wave of studies has been based on the necessity of taking into account two important remarks. First, the use all the cities of the country with no cutoffs, as Eeckhout (2004) proposes, because selecting only the most populated cities can be understood as a selection bias from which the rule will always be true. Secondly, graphical analysis is not the best way of analyzing the goodness of fit, and, therefore, the additional use of tests is preferable (González-Val et al., 2013b).

Eeckhout (2004), one of the first proposals to analyze an un-truncated database of cities, concludes that city size distribution follows the lognormal function, rather than the Pareto (Zipf is a particular case of the latter). More recent works have stated that the function which best describes the data is the double Pareto-lognormal (Giesen et al., 2010) or the normal-Box-Cox (Schluter and Trede, 2013). There is, then, an on-going debate about the best theoretical functions that explain the empirical nature of city size distribution.

The aim of the third chapter is to estimate the city size distribution with the lognormal, double Pareto lognormal and normal-Box-Cox density functions for Germany, France, Italy and Spain. The Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests are carried out to discriminate among the functions; Akaike and Bayesian Information Criteria are also used to select the best option. Our main conclusion is that there is no clear function that outperforms the others, and that the differences between them allows us to think that there may be a new function which could describe the size distribution of cities with a better fit.

From the wide literature that deals with city size distribution, the analysis of the growth process of cities follows naturally. However, there is a less extensive tradition devoted to the study of the empirical distribution of city size growth rates. Among them, can be found Schluter and Trede (2013) and the non-parametric analysis of González-Val et al. (2013a, 2014). The former concludes that the lognormal is rejected as a function to describe city size distribution. This implies that Gibrat's law (the proposition that city sizes grow independently with respect to their size), whose main conclusion is that city size is distributed as a lognormal, has certain problems.

The fourth chapter continues this path, analyzing the city size growth distribution, estimating the fitness of the normal, α-Stable and Student's-t distribution functions, for the case of Germany, France, Italy and Spain. Again, we use the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests to analyze the goodness of fit of the estimated functions and, again, try to select the best one based on the Akaike and Bayesian Information Criteria. In conclusion, as in the previous chapter, we do not find a theoretical density that outperforms the others, although it seems clear that the normal distribution is the worst for describing the city size growth rates. This outcome, together with the third chapter, seems to perpetuate the lack of consensus of an explanation that accounts for both the distribution of the city sizes and their growth over time. The connection between the results found in these two chapters encourage us to continue following this research line, in search of a new theoretical approach that can improve the previous analysis and explain the process of the distribution of population in levels and in growth rates.

References

Behrens, K. and Robert-Nicoud, F. (2011). Tempora mutantur: In search of a new testament for NEG. *Journal of Economic Geography*, 11(2):215–230.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 94(5):1429–1451.

Giesen, K., Zimmermann, A., and Suedekum, J. (2010). The size distribution across all cities-double Pareto lognormal strikes. *Journal of Urban Economics*, 68(2):129–137.

González-Val, R., Lanaspa, L., and Sanz-Gracia, F. (2013a). Gibrat's law for cities, growth regressions and sample size. *Economics Letters*, 118:367–369.

González-Val, R., Lanaspa, L., and Sanz-Gracia, F. (2014). New evidence on Gibrat's Law for cities. *Urban Studies*, 51(1):93–115.

González-Val, R., Ramos, A., and Sanz-Gracia, F. (2013b). The accuracy of graphs to describe size distributions. *Applied Economics Letters*, 20(17):1580–1585.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–99.

McCann, P. (2005). Transport costs and new economic geography. *Journal of Economic Geography*, 5(3):305–318.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Melitz, M. J. and Redding, S. J. (2013). Heterogeneous firms and trade. CEPR Discussion Papers No. 9317.

Ottaviano, G. I. P. (2011). 'New' new economic geography: firm heterogeneity and agglomeration economies. *Journal of Economic Geography*, 11(2):231–240.

Redding, S. J. (2011). Theories of heterogeneous firms and trade. *Annual Review of Economics*, 3(1):77–105.

Samuelson, P. (1952). The transfer problem and transport costs: the terms of trade when impediments are absent. *Economic Journal*, 62:278–304.

Schluter, C. and Trede, M. (2013). Gibrat, Zipf, Fisher and Tippet: City size and growth distributions reconsidered. Working Paper 27/2013, Center for Quantitative Economics WWU Münster.

Von Thunen, J. (1826). *Der isolierte Staat in Beziehung auf landtschaft under nationalokonomie. Von Thunen's isolated state.* (English translation). Pergamon Press, 1966.

Zipf, G. (1949). *Human behavior and the principle of least effort.* Addison-Wesley Press.

Chapter 1

Economic concentration and firm heterogeneity

1.1 Introduction

One of the traditional assumptions of economic geography models and, in general, of any kind of theoretical model of an economy, is that all companies are equal in size and productivity. However, in reality, the situation is clearly different. Firstly, with regard to the size, as the latest report from the OECD on companies says, "the business population is composed, in any region, of a predominant number of micro-firms, i.e. firms with fewer than ten employees". Thus, large companies, those with more than 250 workers, represent, on average, over all OECD regions, 0.2% of the total number of companies, this average ranging between a minimum of 0.06% in South Korea and a maximum of 1.06% in Russia (La Caixa Research 2014). Despite being few in number, big companies monopolize a much more important percentage of employment. For example, in Spain, companies with over 250 employees represented 26.2% of total employment in 2000, a number that increased to 30.6% in 2013. Furthermore, in 2010, the big companies represented 49.8% of employment in the United States and the 37% in Germany (OECD).

Large firms do not only have more workers, they are also more productive. In other words, size and productivity are not independent variables. Indeed, according to the Eurostat Structural Business Statistics, in 2005, the productivity of the larger firms of Germany, Spain, France, Italy and United Kingdom was higher than the productivity of small firms by a factor of 2.3. Latest data from 2013, although only for Spain, shows that the gross value-added per worker of companies with more than 250 workers is 65 % higher than of firms with fewer than 50 workers (La Caixa Research 2014).

In this context, it is necessary, if we want to approach to the stylized facts described above, to introduce the heterogeneity of firms into the theoretical models of economic geography, as Behrens and Robert-Nicoud (2011) and Ottaviano (2011) claim. In the words of the latter: "future research should look more deeply into finer micro-heterogeneity across people and firms". As a result, the main objective of this work is to reconcile and unite two branches of the theoretical literature in a single model: one part concerning business heterogeneity in the line of Melitz (2003), Melitz and Ottaviano (2008) and Melitz and Redding (2013); the other, related to the new economic geography (NEG) formulated originally by Krugman (1991), and systematized by Fujita et al. (1999) and Baldwin et al. (2003).

Of the various possible models of the NEG, we are going to take the Footloose Entrepeneur Model proposed by Forslid and Ottaviano (2003) as our point of reference. The reason for this decision is twofold: on the one hand, the model is completely solvable analytically; on the other hand, it is especially suitable for introducing the distinction between qualified and less qualified business managers. Other standard models in the NEG, such as the original Core-Periphery Model of Krugman (1991), the Footloose Capital Model of Martin and Rogers (1995) and the Vertical Linkages Model of Krugman and Venables (1995) and Venables (1996), do not present either of the two advantage of the Footloose Entrepeneur Model that we have mentioned.

Certainly, the introduction of firm heterogeneity into these kinds of models has found its way in the literature, and a majority of them use the seminal contribution of Melitz (2003). Okubo (2009) examines how a higher freeness of trade gradually increases the economic agglomeration in the context of firm heterogeneity with intermediate input linkages. Venables (2011) discusses the case in which heterogeneous workers decide their location according to the probability of finding a better match with another worker. Okubo et al. (2010) conclude that the most productive companies tend to locate in the larger region (core) while the small and less productive firms are found first in the periphery; they move towards the region with highest number of consumers with a higher freeness of trade. Finally, Forslid and Okubo (2012), using a multi-region version of the Baldwin and Okubo (2006) model, deduce the interesting result that optimal commercial policies and the level of mobility of capital suitable for each region depends on the size of the region.

For its resemblance to our approach and objectives, the work of Baldwin and Okubo (2006) and, above all, Okubo (2010), deserves special mention.

Baldwin and Okubo (2006) take the Footloose Capital Model as its starting point. The companies have sunk costs which are incurred in each market where they operate. This element divides the spectrum of companies into three categories: those that are not profitable even in the domestic market, those that only sell in the domestic market, and the most productive firms that sell in all the regions. The main conclusions of Baldwin and Okubo (2006) are that, first, companies with higher productivity are more footloose and tend to be located in the largest markets and, second, that greater heterogeneity between levels of business productivity acts as a centrifugal force, that is, in favor of dispersion, through moderation in the processes of agglomeration.

Like our model, Okubo (2010) is based on the Footloose Entrepeneur Model, but incorporates a quasi-linear demand function, as suggested in Pfluger (2004). Companies, as in Baldwin and Okubo (2006), have to overcome a fixed cost to export. One similarity with our approach is that differences between the productivity of firms are introduced in a discrete way, that is, there are two types of companies (productive and not so productive) and not in a continuous way, as in the original formulation of Melitz (2003) and Baldwin and Okubo (2006). This hypothesis of only two types of companies greatly simplifies the calculations without losing, as we will see later in our model, interpretative richness. The main conclusion of Okubo (2010) is that the heterogeneity of companies acts as a centripetal force, i.e. it is more plausible that a symmetric equilibrium becomes unstable and, therefore, that agglomeration occurs.

Against this background, our bi-regional model aims to shed light on the effect of heterogeneity on the binomial business agglomeration-dispersion from, as already mentioned, an initial approach based on the Footloose Entrepreneur Model from Forslid and Ottaviano (2003). The fundamental differences with Baldwin and Okubo (2006) and Okubo (2010) define our contribution and explain why our results are not coincident with theirs and, therefore, represent a novelty.

Firstly, we do not consider that companies suffer from a fixed cost to enter the foreign market, in the belief that the cost of transport, which only affects exports and not domestic sales, already gathers, at least in part, this fact. Secondly, these models are different to ours: Baldwin and Okubo (2006) uses the Footloose Capital Model as its starting point and Okubo (2010) uses the version of Pfluger (2004) of the Footloose Entrepeneur model. Thirdly, in these two papers, all companies are mobile between regions; in our approach, only companies that are managed by the most qualified and productive workers can move between regions, while the companies managed by less productive workers can not. This hypothesis is a simplification of the robust empirical evidence that the degree of mobility of the labor force increases with the level of qualification: Docquier and Marfouk (2006), Grogger and Hanson (2011) Docquier et al. (2014) and Artuç et al. (2015). Fourthly, the less productive workers will be able to work in the manufacturing sector, in the agricultural sector or, and this is the novelty, they will be able to create and manage their own manufacturing company, which will be, by definition, a less productive firm. This approach intends to capture the fact that, in certain sectors, where the process of creation of a company is not excessively expensive, the workers in this sector, with experience and information about the product on sale, can become managers, creating their own company, in what constitutes a theoretical approach to self-employment.

Finally, and contrary to Baldwin and Okubo (2006), in our model, the productivity of companies does not follow a distribution function, but is introduced, as in Okubo (2010), in a dichotomous way (companies that are more productive and less productive). Also, unlike Okubo (2010), in which the number of companies of both types is exogenously given, in the model that we propose, the number of less productive firms is endogenous, so that they can enter and exit the market according to the degree of competition; in other words, the number of less productive firms is the endogenous variable that leads to the equilibrium in the system.

We obtain four main results. One, the more productive firms are larger and are concentrated in the biggest region. Two, greater heterogeneity acts as a centripetal force, i.e., it increases the concentration of the less productive companies. Three, the freeness of trade increases the number of less productive companies in the region where the most productive companies are concentrated. In this way, the region in which the less productive firms are more concentrated will depend on the value of the freeness of trade parameter. Four, for low enough values of the cost of transportation, either the large region (the one that concentrates the most productive companies) will specialize in the industrial (differentiated) sector, or the small region will specialize in the agricultural (homogeneous) sector. The intuition and explanation associated with these findings are described in detail in the rest of the work.

The chapter is structured as follows. Section 1.2 presents the model and solves its equilibrium. Section 1.3 explains the effect of heterogeneity in the concentration or dispersion of firms. Section 1.4 describes the resulting economic landscape in terms of specialization and agglomeration. Section 1.5 carries out an analysis of welfare. The chapter closes with the conclusions in Section 1.6.

1.2 The Model

1.2.1 Basic concepts

The economy consists of two regions, 1 and 2. There are two factors of production, which we call productive and non-productive workers, where H^T is the total amount of the former and L^T that of the latter. To talk about productive versus non-productive firms and workers is really an abuse of language. In reality, strictu sensu, both types are productive, although with different intensities. $H^T = H + H^*; L^T = L + L^*$. We use the superscript * to refer to region 2. Productive workers can be associated with business people who have acquired expertise in the management and creation of companies and can move freely between regions. They are responsible for creating productive firms. Non-productive workers can not migrate, but may change sector, so that they are able to work in the industrial sector, both as workers or managers of their own firm, and in the agricultural sector. This allows the presence of the two types of companies in the industrial sector, led by productive and non-productive workers.

1.2.1.1 Demand

The utility is defined with two goods, a horizontally differentiated good that we associate with the industrial sector and a homogeneous agricultural commodity. The utility function of a consumer in region i = 1, 2, is:

$$U_i{=}X_i^{\mu}A_i^{1-\mu} \tag{1.1}$$

$$X_i{=}\left(\int_{s\in N+N^*} d_i(s)^{\frac{\sigma-1}{\sigma}}ds\right)^{\frac{\sigma}{\sigma-1}} \tag{1.2}$$

where $\mu{\in}(0,1)$, is a parameter that expresses the proportion of income used to consume industrial goods, X_i is the consumption of the composite industrial good, A_i is the consumption of agricultural goods, $d_i(s)$ is the consumption of the variety s in i of good X. N is the number of companies in region 1 (N = H + n)(where H is the total amount of productive firms, each managed by a productive businessman, and n the total number of non-productive firms in the region). N^* is the number of companies in region 2, where $N^* = H^* + n^*$; $\sigma > 1$ is the elasticity of demand of a variety and the elasticity of substitution between two varieties. The demand for a variety s of the industrial sector in region i will be determined by the following function, which is the result of the standard maximization of the utility (1.1)

$$d_i(s){=}\frac{p_i(s)^{-\sigma}}{P_i^{1-\sigma}}\mu Y_i, \quad i = 1, 2 \tag{1.3}$$

$p_i(s)$ being the price of the variety s in region i and P_i the price of the composite industrial good in region i.

$$P_i{=}\left[\int_{s\in H+n} p_{ki}(s)^{1-\sigma}ds{+}\int_{s\in H^*+n^*} p_{kj}(s)^{1-\sigma}ds\right]^{\frac{1}{1-\sigma}} \tag{1.4}$$

where k = n, H indicates the type of company: n if it is managed by non-productive workers, H if it is managed by a productive worker. A region's income will be determined by the sum of the wages of all workers

$$Y_i{=}w_iH_i{+}w_i^L L_i, \quad i = 1, 2 \tag{1.5}$$

$w_i(w_i^L)$ being the wages of the (non-)productive workers.

1.2.1.2 Supply

The agricultural sector operates under perfect competition and constant returns to scale, producing a homogeneous good. Without loss of generality, we choose units so that producing an agricultural good requires one unit of non-productive labor. The price of the good, therefore, will be equal to the marginal cost of producing it, which is the worker's salary: $p_i^A = w_i^L$ =1, normalized to the unit price and, since the goods are traded without transport costs, the price will be the same in both regions.

The industrial sector produces horizontally differentiated goods (a variety for each existing firm), with increasing returns to scale. Companies in the industrial sector may be formed by productive workers (needing one productive worker per firm) or non-productive workers ($t > 1$ of these being required to manage the firm). We can associate these differences with the lower preparation of the workers for the fixed activities of the firm (independent of the total production) and which are related by definition to the managerial functions of the companies. The way of introducing heterogeneity, unlike Melitz (2003) and Baldwin and Okubo (2006), is discretely, as in Okubo (2010). Each company, k =(H,n), uses $\beta_k x_i$ non-productive workers to produce x_i output units, being $\beta_H < \beta_n$. In this way, producing the same amount will require a greater number of workers in the non-productive firms. We can associate this difference with a poorer quality of matching between manager and employee, or directly with a lower productivity of the firm, consequence of a lower formation of the manager as an entrepreneur.

The number of productive firms is determined by the total number of productive workers that reside in the region, H in region 1 and H^* in region 2. In both cases, the value of H represents both the number of productive workers and the number of productive firms. The number of non-productive firms is an endogenous variable, which is set as an equilibrium variable in the labor market. Since there is perfect mobility between sectors, the wage that a non-productive worker receives in the agricultural sector and in the industrial sector is the same, $w^L = 1$.

A non-productive worker will create his own company when he notices that the utility obtained as the manager of a firm is greater than or equal to the wage he is receiving as a worker. Obtaining the indirect (standard in the Dixit-Stiligtz framework) utility function, this will happen whenever:

$$U_{manager} - U_{worker} = \mu^{\mu}(1-\mu)^{(1-\mu)} \left(\frac{\prod_{ni}}{tP_i^{\mu}} - \frac{w^L}{P_i^{\mu}} \right) > 0 \tag{1.6}$$

$\prod_{ni}$ being the profit that a non-productive company obtains. We assume that the profit is divided in equal parts between the t managers. The equilibrium condition (there are no incentives for workers to create their own companies) is that $U_{manager} = U_{worker}$. Meaning that the system is in equilibrium when real incomes are equal, and, therefore, when $\prod_{ni} = t$.

Non-productive workers will create non-productive companies until the profit obtained is equal to the wage they would get as a worker, so that there are no incentives to change of employment status. In this way, the number of non-productive firms will be adjusted until the profit obtained by the non-productive firms is equal to t and, in equilibrium, the incomes of workers will be the same whether in agriculture, in the industrial sector or as managers of their own businesses.

At the same time, industrial goods are traded with the common iceberg transportation cost, where a firm should send $\tau \in [1,\infty)$ units of the good so that one unit is received in the other region.

With all of the above, the profit of a company, whatever its type, equals revenues minus the costs (which, in this case, only depend on the wages of the workers) of producing the amount demanded:

$$\prod\nolimits_k (s) = p_{ki}(s) d_{ki}(s) + p_{kj}(s) d_{kj}(s) - w^L \beta_k \left[d_{ki}(s) + \tau d_{kj}(s) \right], k = H, n \tag{1.7}$$

where the subscript i refers to the domestic region and j to the foreign region. Companies choose the price that maximizes their profit. We optimize both profit equations and get:

$$p_{ki}(s) = w^L \beta_k \frac{\sigma}{\sigma - 1} = \beta_k \frac{\sigma}{\sigma - 1} \tag{1.8}$$

$$p_{kj}(s) = w^L \tau \beta_k \frac{\sigma}{\sigma - 1} = \tau \beta_k \frac{\sigma}{\sigma - 1} \tag{1.9}$$

First of all, we get that the price of the good increases with the marginal cost of producing it, β_k, and decreases with the price elasticity of demand (with less differentiation the market is closer to perfect competition, decreasing the markups). Second, the price is more expensive in regions where the goods have to be exported. In this case, the consumer is paying for the units that he is going to consume, but also for those that the company has had to produce and have been lost in the transportation process (the iceberg transportation cost). Thirdly, the prices of goods produced in non-productive companies, with higher marginal costs, are more expensive, which means that they will have a lower demand and, therefore, a lower profit. At the same time, the price index for region i is determined by the total set of prices of the industrial goods that can be purchased in that region. Bringing (1.8) and (1.9) to (1.4):

$$P_i = \left(\beta_n \frac{\sigma}{\sigma - 1} \right) \left(H\varphi + n + H^* \varphi \phi + n^* \phi \right)^{\frac{1}{1-\sigma}} \tag{1.10}$$

where $\phi = \tau^{1-\sigma} \in (0,1]$ is a parameter derived from the transport costs that expresses the freeness of trade, taking a value of 0 when there is no trade (when transport costs are prohibitive) and a value of 1 when trade is completely free (there are no transport costs). In turn, $\varphi = \left(\frac{\beta_H}{\beta_n} \right)^{1-\sigma} = \left(\frac{\beta_n}{\beta_H} \right)^{\sigma - 1} > 1$, is a key parameter in this model, resulting from the ratio between the productivity or marginal costs between firms. The higher this parameter is, the higher the difference between the two types of companies of the economy. The parameter takes a value equal to 1 when there is no difference between the two productivities.

Knowing how the price is determined, we can deduce the profit that each company obtains according to their demand:

$$\prod_{Hi}(s)=\frac{\beta_H x_{Hi}}{\sigma-1} \tag{1.11}$$

$$\prod_{ni}(s)=\frac{\beta_n x_{ni}}{\sigma-1}=tw_i^L=t \tag{1.12}$$

where, in (1.12), we have applied the arbitrage condition which requires that the real income of the non-productive workers is the same if they are workers or managers. Taking into account that the profit of non-productive firms is equal to t and substituting the equations of the demand for each variety (1.3) in (1.11) and (1.12), we obtain the complete equations of the profit:

$$\prod_{ni}(s)=\frac{\beta_n[d_{ni}(s)+\tau d_{nj}(s)]}{\sigma-1}=\frac{\beta_n}{\sigma-1}\left[\frac{p_{ni}^{-\sigma}}{P_i^{1-\sigma}}\mu Y_i+\tau\frac{p_{nj}^{-\sigma}}{P_j^{1-\sigma}}\mu Y_j\right]=$$

$$=\frac{\beta_n}{\sigma-1}\left[\frac{\left(\beta_n\frac{\sigma}{\sigma-1}\right)^{-\sigma}}{\left(\beta_n\frac{\sigma}{\sigma-1}\right)^{1-\sigma}(H\varphi+n+H^*\varphi\phi+n^*\phi)}\mu Y_i+\tau\frac{\left(\tau\beta_n\frac{\sigma}{\sigma-1}\right)^{-\sigma}}{\left(\beta_n\frac{\sigma}{\sigma-1}\right)^{1-\sigma}(H\varphi\phi+n\phi+H^*\varphi+n^*)}\mu Y_j\right]=$$

$$=\left[\frac{1}{\sigma(H\varphi+n+H^*\varphi\phi+n^*\phi)}\mu Y_i+\phi\frac{1}{\sigma(H\varphi\phi+n\phi+H^*\varphi+n^*)}\mu Y_j\right]=t$$

$$\prod_{Hi}(s)=\frac{\beta_H}{\sigma-1}\left[\frac{p_{Hi}^{-\sigma}}{P_i^{1-\sigma}}\mu Y_i+\tau\frac{p_{Hj}^{-\sigma}}{P_j^{1-\sigma}}\mu Y_j\right]=$$

$$=\frac{\beta_H}{\sigma-1}\left[\frac{\left(\beta_H\frac{\sigma}{\sigma-1}\right)^{-\sigma}}{\left(\beta_n\frac{\sigma}{\sigma-1}\right)^{1-\sigma}(H\varphi+n+H^*\varphi\phi+n^*\phi)}\mu Y_i+\tau\frac{\left(\tau\beta_H\frac{\sigma}{\sigma-1}\right)^{-\sigma}}{\left(\beta_n\frac{\sigma}{\sigma-1}\right)^{1-\sigma}(H\varphi\phi+n\phi+H^*\varphi+n^*)}\mu Y_j\right]=$$

$$=\left(\frac{\beta_H}{\beta_n}\right)^{1-\sigma}\left[\frac{1}{\sigma(H\varphi+n+H^*\varphi\phi+n^*\phi)}\mu Y_i+\phi\frac{1}{\sigma(H\varphi\phi+n\phi+H^*\varphi+n^*)}\mu Y_j\right]=\varphi\prod_{ni}(s)=\varphi t$$

In short, an important result that we obtain is:

$$\prod_{Hi}(s)=\varphi\prod_{ni}(s) \tag{1.13}$$

In this way, the profit of the productive firms is greater than that of the non-productive companies to a factor φ, so that the larger the difference between the productivities, the greater the divergence in their profits.

Unlike the original model of Forslid and Ottaviano (2003), where the exogenous variable is the number of companies and the endogenous variable is the profit of the companies (or the wage of their managers), the model proposed here has the number of productive firms as its exogenous variable and endogenously deduces the number of non-productive firms. Knowing the individual remuneration of the productive workers, ($w^i=\prod_{Hi}=t\varphi$), we can obtain the income of a given region from (1.5):

$$Y_i=\varphi tH_i+L_i \tag{1.14}$$

We can also deduce the number of non-productive workers that each type of company employs:

$$\beta_n x_{ni} = t(\sigma - 1) \tag{1.15}$$

$$\beta_H x_{Hi} = \varphi t(\sigma - 1) \tag{1.16}$$

Productive firms have a greater number of non-productive workers and, furthermore, each of them is more productive, so the ratio between the output of each type of company is greater than the ratio of productivities:

$$\frac{x_{Hi}}{x_{ni}} = \frac{\beta_n}{\beta_H}\varphi = \frac{\beta_n}{\beta_H}\left(\frac{\beta_n}{\beta_H}\right)^{\sigma-1} = \left(\frac{\beta_n}{\beta_H}\right)^{\sigma} > \varphi \tag{1.17}$$

The more differentiated the product, the smaller the companies will be (each will use fewer workers) and smaller will be the divergences in the quantities that each type of company produces. With less differentiation (greater σ), each firm will employ more workers, and there will be a greater difference between the outputs of the two types of firms.

Now we are in a position to establish the basic conclusions of this model. With knowledge in business and through quality management, productive workers increase the productivity of their employees. Their companies are bigger and they get an income greater than that of their workers. Furthermore, the companies managed by non-productive workers, with experience in the sector, able to sell a variety of differentiated products, but with less knowledge in economics and business management, are smaller, and may identify themselves, as SMEs (Small and medium-sized enterprises). In these smaller companies, the managers will earn the same income as their employees.

1.2.2 Equilibrium

Although we apparently have four equations to determine the equilibrium (two for each type of company and two for each region), as the profit of productive firms is proportional to the profit of the non-productive companies, we really have only two equations. We use the profits of non-productive firms in regions 1 and 2.

$$\prod\nolimits_{\mathrm{n}}(s) = \frac{\mu}{\sigma}\left[\frac{1}{(H\varphi + n + H^*\varphi\phi + n^*\phi)}(Ht\varphi + L) + \right.$$
$$\left. + \phi\frac{1}{(H\varphi\phi + n\phi + H^*\varphi + n^*)}(H^*t\varphi + L^*)\right] = t \tag{1.18}$$

$$\prod\nolimits_{\mathrm{n}^*}(s) = \frac{\mu}{\sigma}\left[\phi\frac{1}{(H\varphi + n + H^*\varphi\phi + n^*\phi)}(Ht\varphi + L) + \right.$$
$$\left. + \frac{1}{(H\varphi\phi + n\phi + H^*\varphi + n^*)}(H^*t\varphi + L^*)\right] = t \tag{1.19}$$

Solving the system of two equations for the two unknowns (n and n^*) we obtain:

$$n{=}\frac{\mu L{-}\mu\phi L^*{+}t\varphi\left(H\left(\mu{-}\sigma\left(1{-}\phi\right)\right){-}\mu\phi H^*\right)}{t\sigma\left(1{-}\phi\right)} \tag{1.20}$$

$$n^*{=}\frac{\mu L^*{-}\mu\phi L{+}t\varphi\left(H^*\left(\mu{-}\sigma\left(1{-}\phi\right)\right){-}\mu\phi H\right)}{t\sigma\left(1{-}\phi\right)} \tag{1.21}$$

In this way, we know the number of non-productive firms that are created for the labor market to reach its equilibrium, that is, the wages of non-productive workers as workers in any of the two sectors are the same as those they get as managers of their own businesses.

We can do a static comparative analysis and observe how the number of firms increases or decreases with the variables that determine it. In the majority of cases, the sign of the effect is indeterminate and depends on the size of the regions (L and L^*) and the distribution of productive workers (H and H^*). We accompany each derivative from the corresponding sign, assuming that region 1 is of equal or greater size than region 2 ($L \geq L^*$ and $H \geq H^*$):

$$\frac{\partial n}{\partial L} = \frac{\mu}{t\sigma(1-\phi)} \geq 0.$$
$$\frac{\partial n}{\partial \mu} = \frac{(L-\phi L^*+\varphi(H-\phi H^*)}{\sigma t(1-\phi)} \geq 0.$$
$$\frac{\partial n}{\partial \sigma} = -\mu\frac{(L+\varphi tH)-\phi(L^*+\varphi tH^*)}{\sigma^2 t(1-\phi)} \leq 0.$$
$$\frac{\partial n}{\partial t} = -\frac{\mu(L-\phi L^*)}{\sigma t^2(1-\phi)} \leq 0.$$
$$\frac{\partial n}{\partial \varphi} = \frac{H(\mu-\sigma(1-\phi))-\mu\phi H^*}{\sigma(1-\phi)} \geq (\leq) 0 \text{ if } \phi \geq (\leq)\frac{H(\sigma-\mu)}{H\sigma-H^*\mu}.$$
$$\frac{\partial n}{\partial H} = \varphi\left(\frac{\mu(1+\phi)}{\sigma(1-\phi)}-1\right) \geq (\leq) 0 \text{ if } \phi \geq (\leq)\frac{(\sigma-\mu)}{\sigma+\mu}.$$
$$\frac{\partial n}{\partial \phi} = \mu\frac{L-L^*+(H-H^*)t\varphi}{(1-\phi)^2\sigma t} \geq 0.$$

The less differentiated the industrial sector (higher σ), the lower n will be. At the same time, as seems logical, the number of managers needed to start the business, t, also has a negative influence, while the percentage of income that is spent on manufactured goods, μ, has a positive one. The effects of a greater freeness of trade, ϕ, and a larger work force, L, are also positive. On the other hand, the effect by the variable that measures the business heterogeneity (φ), and the number of productive firms, H, on n is still undetermined and will depend on the level of freeness of trade.

With regards to the effect of H, on the one hand, we find a competitive effect that implies that a greater number of productive firms in a region will foster competition and lower the output per company, reducing the profits for all firms and forcing the non-productive firms out of the market until the market regains its equilibrium. This is the market crowding effect. On the other hand, more productive firms increases demand and income in the region, increasing the output per company and, therefore the profit, encouraging the creation of non-productive companies until the market again reaches its equilibrium.

In addition, an increase in the number of productive firms also increases the demand from the foreign region. This is the home market effect. The intensity of each of the effects depends on the magnitude of the costs of transportation. If we are in a state of autarky, the effect of an increase in the number of productive firms on the number of non-productive firms will be negative, (the competitive effect will be greater than the effect via demand). In a state of free trade, the effect will be positive.

With respect to how a greater heterogeneity (a greater φ) affects the number of SMEs, the result is similar, in terms of its indeterminacy. A higher φ negatively affects the number of SMEs since they have less capacity to compete in the market with productive firms (lower output, fewer profits). On the other hand, a greater heterogeneity positively affects n as it increases the income of the region and attracts more demand from abroad (home market effect).

These two forces (home market effect and market crowding effect) alter the profits via competition and via demand and directly affect the entry and exit of non-productive firms, which means that market adjustment will be made by the number of non-productive firms. We can associate this fact with greater market power of large companies that gives them flexibility when assuming the shocks and the changing conditions of the market. Small businesses are not only affected quantitatively (with a smaller or bigger profit), but qualitatively, that is, by changing its decision to remain in the market as managers or workers.

Productive firms assume that small businesses will leave or enter the market, and decide their location knowing that their profit is not going to vary once equilibrium is reached (see the Appendix for an analytic demonstration of this process). In other words, the entry or exit of SMEs to and from the market neutralizes the effect of the home market effect and the market crowding effect on the productive firm's profit, which will be thus invariant. In consequence, location in one area or another of these productive companies will depend primarily on a third standard effect on the models of the NEG, namely the price index effect, which, as we shall see, will favor the total concentration of companies in one of the regions (the larger). In this context, Holmes and Stevens (2002) found empirical evidence that explains how, within the same sector, large firms tend to concentrate more than small ones.

A productive worker of region 2 will always move into region 1 if the utility (or, as we see in (1.22), profit or income in real terms) that he obtains in region 1 is higher than the one he gets in region 2. In this way, and using the indirect utility function we get, with obvious notation:

$$M_H = U_H - U_{H^*} = \mu^{\mu}(1-\mu)^{(1-\mu)} t\varphi \left(\frac{1}{P^{\mu}} - \frac{1}{P^{*\mu}} \right) \tag{1.22}$$

To find out if an equilibrium is stable (where $M_H = 0$ and no companies have incentives to change their region), we must see how M_H varies with respect to H. If the derivative is positive, any movement towards region 1 will create more incentives for companies of region 2 to move to region 1. Otherwise, for the equilibrium to be stable, $\frac{\partial M_H}{\partial H} < 0$. Replacing (1.10) in (1.22) as well as (1.20) and (1.21) in (1.10), we get M_H as a function of H, and deriving:

$$\frac{\partial M_H}{\partial H} = \frac{\mu^{(1+\mu)}(1-\mu)^{(1-\mu)}t^2\varphi^2}{\sigma-1}\left(\frac{\left(\frac{\beta\sigma\left(\frac{(1+\phi)\mu(L+Ht\varphi}{\sigma t}\right)^{\frac{1}{1-\sigma}}}{\sigma-1}\right)^{-\mu}}{L+Ht\varphi}+\frac{\left(\frac{\beta\sigma\left(\frac{(1+\phi)\mu(L^*+H^*t\varphi}{\sigma t}\right)^{\frac{1}{1-\sigma}}}{\sigma-1}\right)^{-\mu}}{L^*+H^*t\varphi}\right) > 0 \tag{1.23}$$

With this we demonstrate that, based on a situation of equilibrium, any movement by productive firms creates a chain reaction that ends with all the companies in one of the two regions. That is, there is no stable equilibrium.

The equilibrium can also break if either of the two regions increases in size. In particular, productive firms will move to region 1 if:

$$P<P^* \longleftrightarrow \frac{\mu\left(1+\phi\right)\left(L-L^*+t\varphi\left(H-H^*\right)\right)}{t\sigma}>0 \tag{1.24}$$

Suppose, for simplicity, that $H = H^*$. In this case, an increase in L (above L^*) will lower the price index for region 1 (as it increases the number of SMEs) and will raise the real income of productive workers that are located there, encouraging the entry of more productive workers from region 2 which, in turn, will decrease the price index of region 1 even further. We can say, therefore, that productive companies have incentives to locate in the larger areas.

In summary, the first major conclusion of the structure of the model is that all productive firms will finally be concentrated in a single region. Without loss of generality, we assume that this region is 1.

This endogenous result ($H^* = 0$) allows us to greatly simplify the expressions of n and n^*, with the assumption that the number of non-productive workers is still the same in the two regions ($L = L^*$):

$$n=\frac{L\mu\left(1-\phi\right)+Ht\varphi\left(\mu-\sigma\left(1-\phi\right)\right)}{t\sigma\left(1-\phi\right)} \tag{1.25}$$

$$n^* = \frac{L\mu\left(1-\phi\right)-Ht\phi\varphi\mu}{t\sigma\left(1-\phi\right)} \tag{1.26}$$

Now ($H^* = 0$ and $L = L^*$), the signs of the derivatives in page 16 are the same, and, as the most important case, we have:

$$\frac{\partial n}{\partial \phi}=\frac{H\mu\varphi}{\sigma(1-\phi)^2}>0 \tag{1.27}$$

$$\frac{\partial n^*}{\partial \phi}=-\frac{H\mu\varphi}{\sigma(1-\phi)^2}<0 \tag{1.28}$$

As we can see, an increase in freeness of trade (a decrease in transportation costs)positively affects the number of non-productive firms located in the region that concentrates the productive firms, while negatively affects the region which has only non-productive companies.

The explanation is as follows. Let us start with a state of autarky and infinite transport costs. Since all productive companies are located in region 1, its price index will be lower than that of region 2. As transport costs go down, the exports of each company will increase. Then, as the the price index of region 1 is lower than region 2, the goods of region 2 are more expensive for individuals of region 1 than the goods of region 1 are for individuals of region 2. This makes the agents of region 1 consume more products of their own region, importing less than region 2 imports from region 1, which causes the demand in region 1 to increase with the consequent positive effect positive on n.

In this line of reasoning, if we obtain the derivative of the profit of SMEs with respect to ϕ and then replace the value of n and n^* by their expressions in the equilibrium in (1.25) and (1.26) we also obtain some equations with a non-ambiguous sign[1]:

$$\frac{\partial \prod_n}{\partial \phi} = \frac{Ht^2\varphi\left(L\left(1-\phi\right)-Ht\varphi\phi\right)}{L\left(1-\phi\right)\left(1+\phi\right)^2\left(L+Ht\varphi\right)} > 0 \tag{1.29}$$

$$\frac{\partial \prod_{n^*}}{\partial \phi} = -\frac{Ht^2\varphi\left(L\left(1-\phi\right)+Ht\varphi\phi\right)}{L\left(1-\phi\right)\left(1+\phi\right)^2\left(L+Ht\varphi\right)} < 0 \tag{1.30}$$

Lower transport costs increase the profits of SMEs in region 1, creating incentives for workers to create their own company. At the same time, they lower the demand and profits in region 2. The fact that companies are created in region 1 and leave the market in region 2, causes the price index in region 1 to decrease and that of region 2 to increase even more. Therefore, a decrease in the costs of transportation has an amplifying effect.

In conclusion, more competitive regions attract more foreign demand with more free trade, which increases domestic economic activity in the form of a greater number of companies and varieties.

1.3 Effects of heterogeneity on the location of companies

Heterogeneity is one of the key elements of our model. Therefore, since we have already deduced that the most productive firms are concentrated entirely in one of the regions, we

[1] In equilibrium $\prod_n = \prod_{n^*} = t$ =constant. These derivatives should explain how profits vary with ϕ before the number of non-productive firms meets the arbitrage condition

will study in more detail how greater heterogeneity affects the distribution of the less productive companies between the two regions (in the following section we analyze the distribution of the economic activity as a whole).

The proportion of small businesses that perform their economic activity in region 1, d_n is, replacing the number of firms in each region by equations (1.25) and (1.26), given by:

$$d_n = \frac{n}{n+n^*} = \frac{(1-\phi)\,L\mu + H\,(\mu - (1-\phi)\,\sigma)\,t\varphi}{(1-\phi)\,(2L\mu + H\,(\mu-\sigma)\,t\varphi)} \tag{1.31}$$

The effect of more heterogeneity between firms (φ) in the proportion of non-productive firms in region 1 is:

$$\frac{\partial d_n}{\partial \varphi} = \frac{\partial\left(\frac{n}{n+n^*}\right)}{\partial \varphi} = \frac{HL\mu\,((1+\phi)\,\mu - (1-\phi)\,\sigma)\,t}{(1-\phi)\,(2L\mu + H\,(\mu-\sigma)\,t\varphi)^2} \tag{1.32}$$

It can be checked that $d_n > (<)\ 0.5$ if and only if $\phi > (<) \frac{\sigma-\mu}{\sigma+\mu}$ and that $\frac{\partial d_n}{\partial \varphi} > (<)\ 0$ if and only if $\phi > (<) \frac{\sigma-\mu}{\sigma+\mu}$.

That is, if the parameter of freeness of trade is less than the threshold $\frac{\sigma-\mu}{\sigma+\mu}$, on the one hand, SMEs are more concentrated in region 2 (it has more than 50% of all non-productive firms) and, on the other hand, a greater heterogeneity increases the percentage of SMEs in region 2. But if the parameter of freeness of trade is higher than the threshold, then

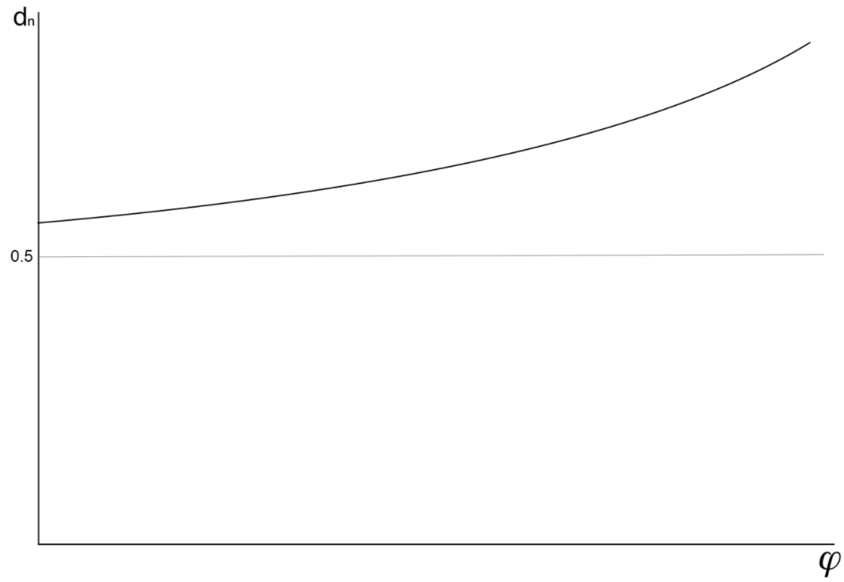

Figure 1.1: Percentage of SMEs in region 1 when $\phi > \frac{\sigma-\mu}{\sigma+\mu}$

region 1 has more than 50% of the non-productive firms or SMEs and, in addition, greater heterogeneity raises this percentage (see Figures 1.1 and 1.2). In both cases we see that, whatever the variable of freeness of trade, greater heterogeneity increases the proportion of SMEs in the region that has more than 50% of the SMEs. We, conclude, therefore, that

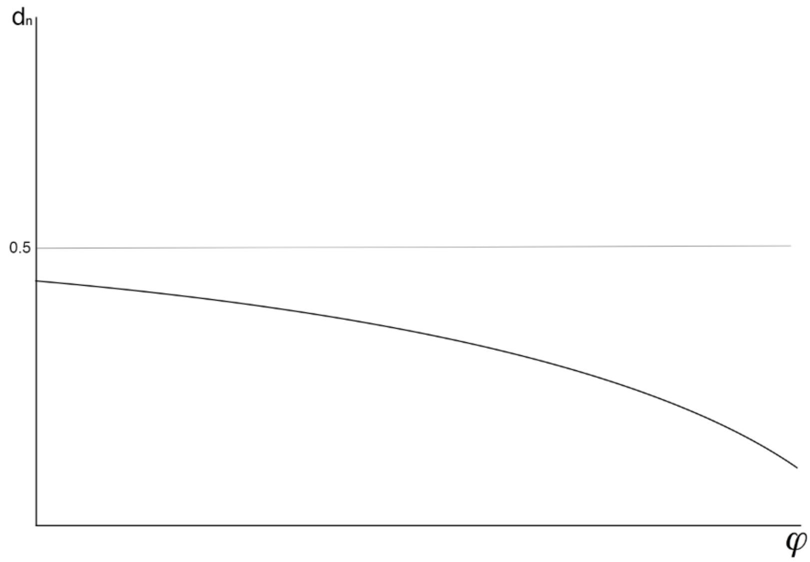

Figure 1.2: Percentage of SMEs in region 1 when $\phi < \frac{\sigma-\mu}{\sigma+\mu}$

business heterogeneity fosters the concentration of non-productive companies, even though the region where this agglomeration will take place depends on the value of the freeness of trade: If it is high (low), will be in the region that has (does not have) the productive firms.

This is because greater heterogeneity reinforces the effect generated by productive firms on the non-productive firms that we saw in Section 1.2.2. In this way, with high transport costs, the market crowding effect was higher than the home market effect, in such a way that the effect of productive firms on non-productive firms in the same region was negative.

Greater heterogeneity reinforces these effects, reducing the number of non-productive firms in region 1, and decreasing the percentage of non-productive firms that are located in region 1 (since the effect, although it is also negative for firms in region 2, it is to a lesser extent, as the magnitude of the market crowding effect is lower). On the other hand, with low transport costs, productive firms attract demand from abroad, capable of generating income and market potential for the creation of non-productive companies. An increase in heterogeneity reinforces this effect, augmenting the market potential for non-productive firms and, therefore, increasing both the number and the percentage of non-productive firms in region 1.

1.4 Agglomeration and specialization

For the following analysis, we will continue assuming that the number of non-productive workers is the same in both regions (that is, $L = L^*$). We have also inferred that productive firms are concentrated (by the mechanism described in the previous section) in region 1 ($H^* = 0$). As we have seen, lower transportation costs increase the number of non-productive companies in region 1 and lower it in region 2. It is convenient to establish three restrictions that will maintain economic common sense with respect to the values that can be adopted by the endogenous variables n and n^*.

The first restriction will imply that the number of non-productive firms in region 1 will always be positive. An increase in transport costs decreases the number of non-productive companies in region 1. We want to make sure that, with infinite costs of transportation (with no trade between the two regions), the number of these companies, which will be the minimum possible, is at least zero. From (1.25)

$$n_{min} = n\,(\phi{=}0) {=} \frac{L\mu{-}Ht\varphi(\sigma{-}\mu)}{t\sigma} {>} 0 \tag{1.33}$$

Secondly, as the costs of transportation are reduced, the number of non-productive firms in region 1 increases. However, the total population of region 1 is fixed (L), and the number of workers that uses each company is also fixed ((1.15) and (1.16)). Therefore, there is a maximum amount of companies who can enter the market. In particular, discounting the population working for productive firms ($H\varphi(\sigma t - 1)$), the total amount of non-productive labor force that is available to work in non-productive firms is $L - H\varphi(\sigma t - 1)$. Taking into account that each SME uses $t(\sigma - 1) + t = t\sigma$ non-productive workers (considering also the necessary number of worker to manage the company), we can establish the minimum transport cost that will result in the maximum number of firms in region 1.

$$n_{max} = \tfrac{L - Ht\varphi(\sigma-1)}{t\sigma} = \{(1.25)\} = \tfrac{L\mu(1-\phi)+Ht\varphi(\mu-\sigma(1-\phi))}{t\sigma(1-\phi)} \longrightarrow$$

$$\longrightarrow \phi_M = \frac{(1-\mu)(L + Ht\varphi)}{(1-\mu)L + Ht\varphi} \in (0,1) \tag{1.34}$$

For values greater than the parameter ϕ_M the number of companies in 1 will not increase, because it will have reached the maximum possible (we will see later in detail what happens for greater values of ϕ).

The third restriction refers to the number of companies in region 2. As transport costs decrease, the total number of firms is lowered, but it is also impossible for that number to be negative. Namely, there is a minimum number of companies, 0, which is found when $\phi \geq \phi_P$.

$$n^*_{min} = 0 = \{(1.26)\} = \frac{L\mu(1-\phi) - Ht\mu\varphi\phi}{t\sigma(1-\phi)} \longrightarrow \phi_P = \frac{L}{L + Ht\varphi} \in (0,1) \tag{1.35}$$

For values greater than the parameter ϕ_P, the profit gained by the SMEs would be less than t. In these circumstances, non-productive workers would not want to start a business and, in the end, they would all end up working in the agricultural sector.

When analyzing the resulting economic landscape for a high enough freeness of trade, we have two possible outcomes if ϕ_M is greater than or less than ϕ_P, because the system of equations will change depending on the case.

1.4.1 Total specialization of region 1 in the industrial sector if $\phi \geq \phi_M$

This is the case in which $\phi_M < \phi_P$. This occurs whenever:

$$\phi_M = \frac{(1-\mu)(L+Ht\varphi)}{(1-\mu)L+Ht\varphi} < \frac{L}{L+Ht\varphi} = \phi_P \rightarrow \mu > \frac{L+Ht\varphi}{2L+Ht\varphi} = S_e > 0.5 \tag{1.36}$$

That is, if the proportion of the income to consume industrial goods (μ) is greater than the variable $S_e = \frac{Y}{Y+Y^*} \in (0,1)$ (which represents the percentage of income that region 1 has over the total amount of both regions), region 1 specializes in the industrial sector.

For a parameter of free trade greater than ϕ_M, the system of equations that determines the number of firms (1.18 and 1.19) changes because, when all the workers of region 1 are in the industrial sector, the number of industrial firms cannot increase, since it is exogenously determined by the maximum number of companies, n_{max}. Not having an agricultural sector in region 1 implies that there is no arbitrage condition that limits the salaries of non-productive workers, i.e., w^L does not have to be equal to 1 but is determined endogenously via market equilibria (the endogenous variable changes from n to w^L). From (1.6) $\prod_n = tw^L$; in addition, (1.8) and (1.9) imply that prices are rising at the same rate as w^L and from (1.13), $\prod_H = \varphi \prod_n$ and, therefore, $\prod_H = \varphi t w^L$.

The new system of equations of the business profits with $\phi > \phi_M$ is not solvable, in the style of the original model of Krugman, so we can not obtain a solution for n^* or for w^L. Replacing (1.3) in (1.11) and (1.12) and taking into account the new prices (which are now multiplied by the factor w^L) and the new income of productive and non-productive workers of region 1 (which are also multiplied by w^L) the system of equations is given by:

$$\prod_{\mathrm{n}}(s) = \frac{\mu}{\sigma}\left[\frac{(w^L)^{1-\sigma}}{(w^L)^{1-\sigma}(H\varphi+n)+n^*\phi} w^L(H\varphi t+L) + \right.$$
$$\left. + \phi \frac{(w^L)^{1-\sigma}}{\phi(w^L)^{1-\sigma}(H\varphi+n)+n^*}(L)\right] = tw^L \tag{1.37}$$

$$\prod_{\mathrm{n^*}}(s) = \frac{\mu}{\sigma}\left[\phi \frac{1}{(w^L)^{1-\sigma}(H\varphi+n)+n^*\phi} w^L(H\varphi t+L) + \right.$$
$$\left. + \frac{1}{\phi(w^L)^{1-\sigma}(H\varphi+n)+n^*}(L)\right] = t \tag{1.38}$$

where, with all non-productive workers in region 1 in the industrial sector:

$$n(\phi > \phi_M){=}n_{\max}{=}\frac{L{-}Ht\varphi(\sigma{-}1)}{t\sigma} \tag{1.39}$$

However, we can know the value at the extremes of this restriction, that is, at $\phi = \phi_M$ (from which this restriction takes effect) and $\phi = 1$ (the maximum freeness of trade possible). In this case:

$$w^L = 1 \tag{1.40}$$

$$n^*(\phi = \phi_M \text{ and } \phi = 1) = \frac{L(2\mu - 1) - Ht\varphi(1 - \mu)}{t\sigma} \tag{1.41}$$

The number of companies that are in region 2 in a full freeness of trade state depends on the marginal propensity to consume industrial goods. Thus, in the extreme case that all income will be spent on industrial goods, $\mu = 1$, the population of region 2 will also be devoted entirely to the production of the industrial sector (see (1.41)).

The following graphs illustrate the process of economic concentration when $\phi_M < \phi_P$. We have the freeness of trade parameter on the horizontal axis and on the vertical axis the total number of non-productive companies.

Figure 1.3 shows the evolution of the number of SMEs in region 1, which increases more and more as the cost of transport is reduced, until reaching $\phi = \phi_M$, after which region 1 devotes its entire population to the production of industrial goods and n is maximum.

The second graph is for region 2. In this case, the number of SMEs falls increasingly as the freeness of trade increases, reaching ϕ_M. From there on, the number of companies remains practically constant, as we have seen when we obtained their values by a simulation process. Note how, over to the right of ϕ_P, n^* is not equal to zero, as would be expected. This is because, in each of the two possible scenarios (Sections 1.4.1 and 1.4.2), only the restriction which affects the minimum of (ϕ_M, ϕ_P) is operational and the other is not. The explanation is as follows. Suppose that we are in the case in which $\phi_M < \phi_P$ (for the opposite situation, the reasoning is analogous, mutatis mutandis). When $\phi = \phi_M$, n adopts a specific value that will not change, this makes equations (1.18 and 1.19) become (1.37 and 1.38), so (1.21 and 1.26) are no longer valid (remember that ϕ_P is deduced from (1.26)). In summary, as shown in Figure 1.4, with values of ϕ over ϕ_P, n^* is not zero.

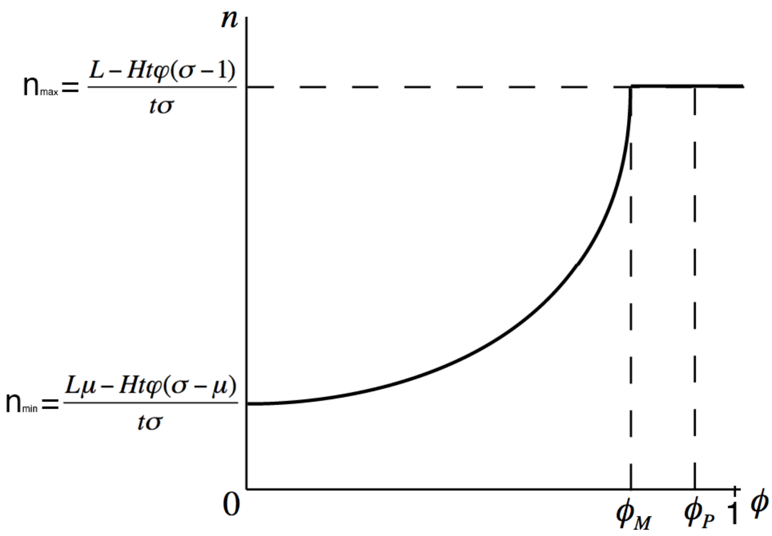

Figure 1.3: n as a function of ϕ ($\phi_M < \phi_P$)

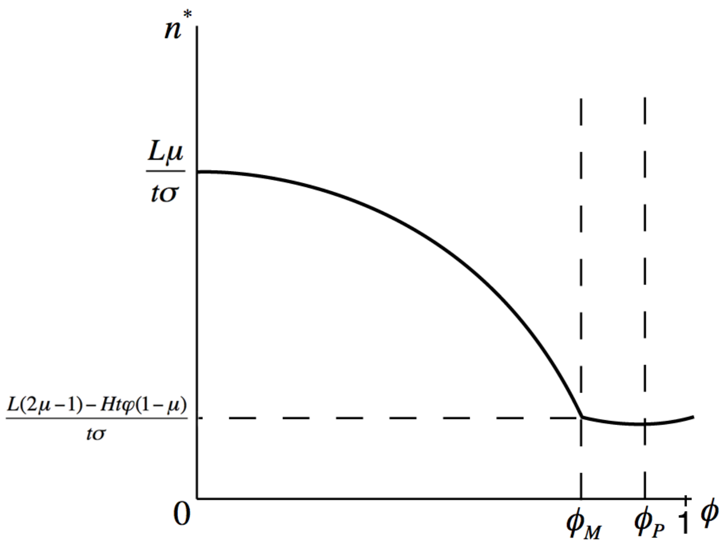

Figure 1.4: n^* as a function of ϕ ($\phi_M < \phi_P$)

1.4.2 Total specialization of region 2 in the agricultural sector if $\phi \geq \phi_P$

Now $\phi_P < \phi_M$. Lower transport costs from ϕ_P imply that region 2 is fully specialized in the agricultural sector because it has no industrial firms ($n^* = 0$). This fact greatly simplifies the new system of equations since we will only have one equation, namely the corporate profits of the non-productive companies of region 1:

$$\prod\nolimits_{\text{n}} (s) {=} \tfrac{\mu}{\sigma} \left[\tfrac{1}{(H\varphi+n)} (H\varphi t{+}L){+} \ \phi \tfrac{1}{\phi(H\varphi+n)} (L) \right] {=} t$$

$$n {=} \frac{2L\mu {-} Ht\varphi(\sigma{-}\mu)}{t\sigma} \ \forall \phi \geq \phi_p \tag{1.42}$$

Figure 1.5 shows the increasing evolution of the number of SMEs in region 1 as the transport costs are reduced. However, for any value of free trade greater than ϕ_P, the number of SMEs remains constant, due to the disappearance of all industrial SMEs in region 2 (the increase in demand due to the reduction of transport costs is offset by the reduction of the production needed for one unit to arrive at region 2, maintaining the total production and profits constant for SME's in region 1). Figure 1.6 provides the behavior of region 2, where the number of SMEs decays reaching its minimum with $\phi \geq \phi_P$.

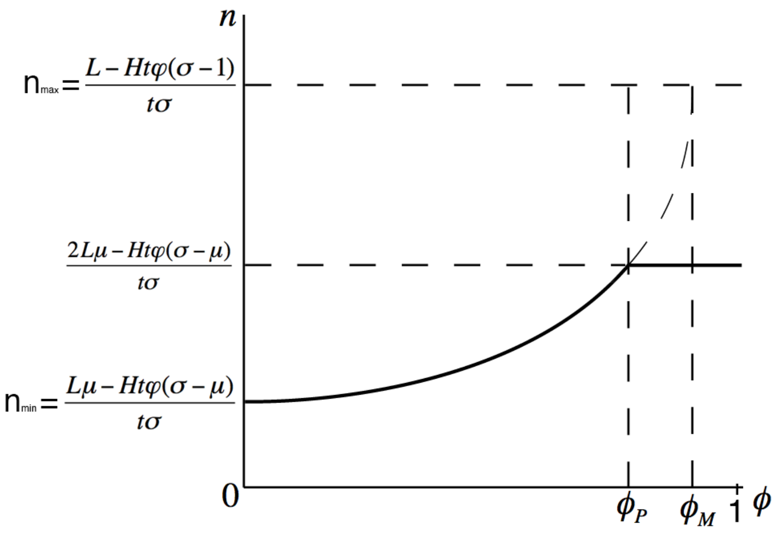

Figure 1.5: n as a function of ϕ ($\phi_P < \phi_M$)

The special case in which $\phi_M = \phi_P$ implies that both restrictions are met simultaneously: any greater ϕ will imply that both regions 1 and 2 will be specialized in the industrial and the agricultural sector, respectively. This special case occurs when $\mu = \frac{L+Ht\varphi}{2L+Ht\varphi}$.

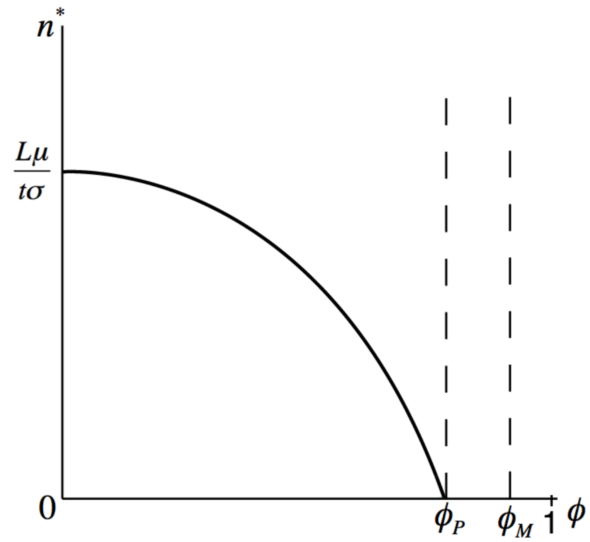

Figure 1.6: n^* as a function of ϕ ($\phi_P < \phi_M$)

1.4.3 Industrial concentration

We have seen the evolution of the number of SMEs, but to study the economic concentration in an industry we must take into account all businesses (large and small ones). As we know, as transport costs decrease, industry will increase (due to the creation of SMEs) where the productive firms are located (in our case, region 1). To study the evolution of the industry, we believe that it is not convenient to directly add up the number of firms, since they are different entities and it is preferable to consider their economic weight.

Therefore, we define S_I as the percentage of revenues of the industrial sector in region 1 over the total amount of both regions (it also represents the percentage of profits earned by the industrial sector in region 1 over the total amount and the percentage of non-productive workers dedicated to the industrial sector, without counting the business managers): $S_I = \frac{H\varphi+n}{H\varphi+n+n^*}$. Substituting n and n^* by (1.25) and (1.26) and taking $S_e = \frac{Y}{Y+Y^*} = \frac{Ht\varphi+L}{Ht\varphi+L+L^*}$, we reach:

$$S_I = \frac{1}{2} + \frac{1+\phi}{1-\phi}\left(S_e - \frac{1}{2}\right) \tag{1.43}$$

These equations are only met when ϕ is less than ϕ_M and ϕ_P. In another case, when region 1 is fully specialized in the industrial sector and is unable to continue to concentrate industrial productive capacity or when region 2 fully specializes in the agricultural sector and the entire industry is now concentrated in region 1, we have seen that the behavior of n and n^* changes and the equations that describe the business concentration process cannot be set analytically (although we can make numerical approximations to see what is really happening).

Figures 1.7 and 1.8 represent, as is customary in the new economic geography models, the fork diagram, showing the industrial concentration according to the equation (1.43) for the region that concentrates the productive firms (represented by the top line) and the region that only has non-productive companies (represented by the bottom line).

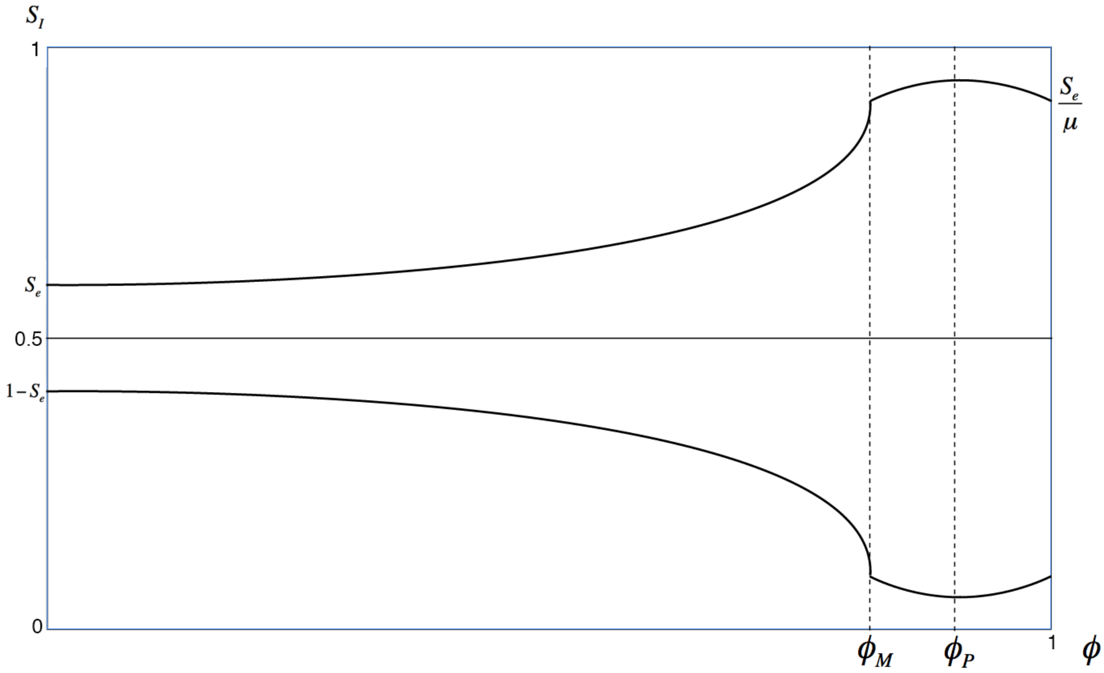

Figure 1.7: Industrial concentration as a function of ϕ ($\mu > S_e$, $\phi_M < \phi_P$)

On the horizontal axis, we represent the different values of the freeness of trade and, on the vertical axis, the industrial concentration, measured as a percentage of the industry in each region over the total. Thus, in the case of a null freeness of trade, the percentage of industry in each region is exactly equal to the percentage of income in that region. However, as the transport costs decrease, the number of non-productive firms in the region with the highest percentage of income and industry increase. This raises production, revenues, profits and the number of workers (the last three at the same rate) increasing the weight of the industry in this region.

Figure 1.7 illustrates the case where $\mu > S_e$ which implies that $\phi_M < \phi_P$. In this situation, when the parameter of free trade (ϕ) is greater than ϕ_M, we are in the case in which region 1 is unable to increase the number of non-productive companies, which encourages an increase in prices, wages, income and corporate profits, leading to an increase in industry concentration but less intensively (due to the loss of competitiveness that the increase in prices implies).

Above a certain level of freeness of trade, this process is reversed, as the loss of competitiveness is too large to continue increasing the profits, so it produces the opposite

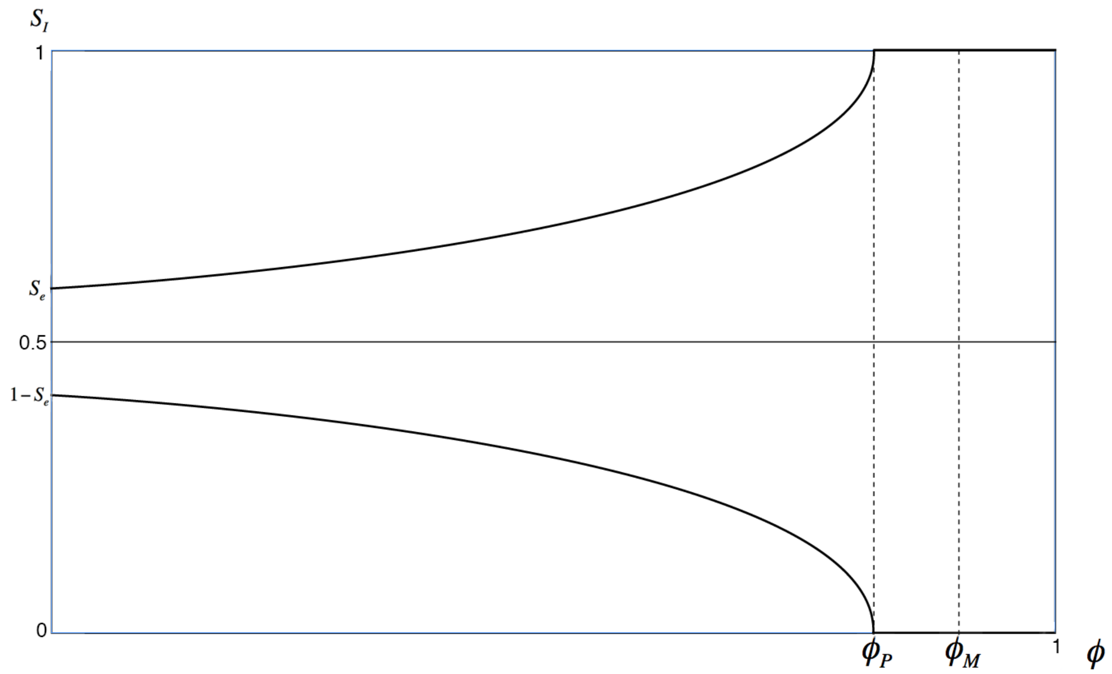

Figure 1.8: Industrial concentration as a function of ϕ ($\mu < S_e$, $\phi_M > \phi_P$)

effect, a reduction in prices, salaries, wages and profits up to precisely the level prior to the increase of these monetary factors, $\frac{S_e}{\mu}$.

Figure 1.8 illustrates the case in which $\mu < S_e$, so that $\phi_P < \phi_M$. In this scenario, when the parameter of free trade is greater than ϕ_P, region 2 is unable to further reduce its industrial production, because it cannot be negative. In this case, region 1 is the only one that has an industrial sector, thus making up 100% of the industry between the two regions, percentage that remains while $\phi > \phi_P$.

1.5 Welfare

The indirect utility function is:

$$U_{si}{=}\mu^{\mu}(1{-}\mu)^{1-\mu}\frac{w_s}{P_i^{\mu}}, \quad s = L, L^*, H \quad i = 1, 2. \tag{1.44}$$

where $w_s, s = L, L^*, H$, represents the associated income of both types of workers in the corresponding region. Taking (1.10), (1.25) and (1.26), with $\psi = \frac{\mu^{\mu}(1-\mu)^{(1-\mu)}}{(\beta_n \frac{\sigma}{\sigma-1})^{\mu}}$ we get:

$$U_L = \psi \frac{1}{(H\varphi + n + \phi n^*)^{\frac{\mu}{1-\sigma}}} = \psi \left(\frac{\mu(1+\phi)(L + Ht\varphi)}{t\sigma} \right)^{\frac{\mu}{\sigma-1}} \tag{1.45}$$

$$U_{L^*} = \psi \frac{1}{(\phi H\varphi + \phi n + n^*)^{\frac{\mu}{1-\sigma}}} = \psi \left(\frac{\mu(1+\phi)L}{t\sigma} \right)^{\frac{\mu}{\sigma-1}} \tag{1.46}$$

$$U_H = \psi t\varphi \left(\frac{\mu(1+\phi)(L+Ht\varphi)}{t\sigma} \right)^{\frac{\mu}{\sigma-1}} \tag{1.47}$$

All functions increase with respect to the freeness of trade parameter, that is, the lower the transportation costs, the greater the utility level, since prices will be lower and individuals can purchase more industrial goods. At the same time:

$$\frac{U_{\mathrm{L}}}{U_{\mathrm{L}^*}} = \frac{Y}{Y^*} = \frac{L+Ht\varphi}{L} > 1 \tag{1.48}$$

$$\frac{U_H}{U_L} = t\varphi > 1 \tag{1.49}$$

Productive workers have a greater utility than non-productive workers. These in turn enjoy a greater utility if they are located in region 1. This occurs until the transport costs are so low that either of the two restrictions are met, ϕ_M or ϕ_P. From then on, the utility of the individuals in region 2 increases at a faster pace than the individuals of region 1 so that, in a state of complete freeness of trade, $\phi = 1$, the value of the utility of non-productive workers is the same in both regions. In fact, evaluating the price index (1.10) with (1.41) and (1.39) (if $\mu > S_e$) or $n^* = 0$ and (1.42) (if $\mu < S_e$) we have:

$$U_L = U_{L^*} = \psi \left(\mu \frac{2L + Ht\varphi}{t\sigma} \right)^{\frac{\mu}{\sigma-1}} \tag{1.50}$$

This convergence can be explained as that once a region is specialized, the improvement in the trade involves only a greater accessibility to products from both regions. Since region 2 is at a disadvantage, the improvement of trade more positively affects this region, allowing access to more varieties at a lower price. Region 1 is also affected positively, but with less intensity. With free trade the utility of non-productive workers is, therefore, the same, regardless of the region in which they are located. This conclusion differs from that obtained by Okubo (2010) where, even with full freeness of trade, the region that concentrated the industry obtains greater welfare for their non-productive workers.

1.6 Conclusions

This work has introduced business heterogeneity in the level of productivity into a standard model of the New Economic Geography using the Footloose Entrepeneur Model of Forslid and Ottaviano (2003). The most important differences with the previous literature in this line, primarily the works of Baldwin and Okubo (2006) and Okubo (2010), are three in number. First, we do not consider that companies suffer from a fixed cost to enter foreign markets. Second, in the two cited papers, firms are mobile between regions while, in our approach, only companies that are managed by the most qualified and productive workers can move and companies managed by the less productive workers are immobile. Third, the less productive workers will be able to work in the manufacturing sector, in the agriculture sector or, and this is the novelty, will be able to create and manage their own manufacture company, that will be, by definition, a less productive firm.

The main conclusions that we can deduce from this model are the following.

One, the symmetrical equilibrium of productive firms is unstable, so the largest region will concentrate all productive firms (a result also obtained by Nocke (2006) and, although gradually, by Baldwin and Okubo (2006)). In this same line of reasoning, from a situation in which each region has 50% of productive firms, any industrial policy that produces a movement of a productive firm from one region to another (through, for example, tax incentives or the implementation of advantages that attract more skilled workers who can create these productive companies) means that all productive companies will be located in this region.

Two, in this model, the key endogenous variable is the number of SMEs in each region. How does business heterogeneity affects these less productive companies? Greater business heterogeneity leads to a higher concentration of SMEs. The question is, in which of the regions will this concentration occur? If transport costs are high, the SMEs are concentrated in the periphery (the smallest region or the one that does not concentrate the more productive firms). If transport costs are low, concentration will be in the core. What is the economic explanation of this important result? There are two forces acting with different signs. On the one hand, a greater number of productive firms in a region or a greater difference in productivity between the two types of companies acts as a force that drives SMEs out of the market through a mechanism linked to greater competition. On the other hand, more productive firms in a region or a greater difference in productivity between the two types of companies acts as a force that creates non-productive firms through a mechanism linked to a greater income and, therefore, greater demand (there are more productive firms, which generates a higher income). In addition, these effects will be greater the higher the difference between productivities. With high transport costs, the competitive force is superior to the demand-linked force, while the increase in free trade (the freeness of trade parameter gradually approaches to one) will intensify the positive effect until it dominates the negative force, with the result that the number of productive firms or a superior firm heterogeneity favors the creation of SMEs.

Three, a direct consequence of the previous point is that, in the region that does not have the productive companies, the number of non-productive firms decreases as the freeness of trade increases and, therefore, it makes sense for that region to establish protectionist policies to boost its manufacturing sector. If this is true, the trade off is that this policy decreases the welfare of all individuals, including those located in the region whose industry it is trying to protect, because of a reduction in the accessibility to foreign industrial goods produced by more efficient firms that can sell them cheaply.

Four, the resulting economic landscape of the model is, for sufficiently low transportation costs, the following: either the region that owns all productive firms is fully specialized in the industrial sector or the other region is completely specialized in the agricultural sector. For a specific value of μ it is even possible for both events to occur simultaneously, i.e., there would be complete specialization in production at the international level. In our model, each region specializes in the sector in which it has a comparative advantage. The region that has productive firms, which gives it a greater technical efficiency in production, specializes in manufacturing, and the other region in the sector in which it is relatively better, namely the agricultural sector.

Five, from point two, it can be derived that, for a high freeness of trade, productive firms in the region in which they are installed favor the creation of less productive firms, generating a higher income resulting in an increase in demand and, therefore, an increase of opportunities for SMEs opportunities. These positive externalities that productive companies project onto the non-productive appear in Delgado et al. (2010): “There is a strong evidence that the presence of a strong cluster surrounding a region industry accelerates the growth in a start-up activities... Strong regional clusters enhance the range and diversity of entrepreneurial start-up opportunities while also reducing the costs of starting a new business”.

Finally, the reduction of transport costs unequivocally increases the welfare of all agents. Productive workers also have a higher utility than non-productive workers, and those located in the core will have a greater utility than those located in the periphery. Although, for a sufficiently high freeness of trade, the welfare of the workers on the periphery increases more quickly, so that in a state of full freeness of trade, all the non-productive workers enjoy the same level of welfare.

Appendix: "market crowding effect" and the movement of firms

In the model we have supposed that the "market crowding effect" (which decreases the profits of industrial firms due to the increase in the number of companies competing in the same region) first affects the number of SMEs and that they are therefore the first to move when there is a decrease in their profits. This assumption is key to the model.

When a productive firm moves from region 2 to region 1, competition increases, which decreases the profits of all the companies. When deciding if it should move or not, on the other hand, the manager does not consider this initial profit reduction because he knows that the SMEs will leave the market in response to this increased competition, so that, later, the firm’s profit will not change regardless of where it is located. What we are saying, therefore, is that non-productive workers are the first to change their entrepreneurial status when both types of firms (productive and non-productive) are facing greater competition. If it wasn’t so, productive firms having moved into region 1, and due to the increased competition, might decided that they were better in region 2, which could make the symmetrical equilibrium stable.

The above seems to infer that the model rests solely on a hypothesis which, if not adopted, would be a non-trivial change in the outcomes arising from this model. But, fortunately, this is not true: assuming that the SMEs are the first to move in an unfavorable situation is not an imposed condition, it can also be obtained analytically.

In effect, we have to assume that the first to move will be those who have more incentives (or relative earnings), similar to (though not the same as) what occurs in Baldwin and Okubo (2006). Suppose, therefore, that a number of productive companies have been installed in region 1, increasing competition and reducing the profits of all companies by a fraction $h < 1$:

$$\prod_{ni} = ht < t \tag{1.51}$$

$$\prod_{Hi} = h\varphi t < t \tag{1.52}$$

Each non-productive worker as a manager, earns an income equal to $h < 1$. If they change their employment status to industrial workers (they can not move from one region to the other), they would get a salary equal to 1. The expected relative growth of their income (or utility) is:

$$\frac{1-h}{h} \tag{1.53}$$

Each productive workers gets $h\varphi t < \varphi t$. If they change their region, they would get φt, but they also take into account the effect of the price index:

$$\frac{\frac{\varphi t}{P^{*\mu}} - \frac{h\varphi t}{P^{\mu}}}{\frac{h\varphi t}{P^{\mu}}} = \frac{\left(\frac{P}{P^{*}}\right)^{\mu} - h}{h} \tag{1.54}$$

Since $P < P^{*}$, the improvement (or increase in the utility level) derived from the change of status of non-productive individuals is greater than that of the productive individuals, so non-productive workers will have more incentives to exit the market in a situation of increased competition. For productive workers, the effect of competition will, therefore, be neutralized by the movement of non-productive workers, and they can always assume a profit equal to φt.

References

Artuç, E., Docquier, F., Özden, C., and Parsons, C. (2015). A global assessment of human capital mobility: The role of non-OECD destinations. *World Development*, 65:6–26.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G., and Robert Nicoud, F. (2003). *Economic geography and public policy*. Princeton University Press.

Baldwin, R. E. and Okubo, T. (2006). Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. *Journal of Economic Geography*, 6(3):323–346.

Behrens, K. and Robert-Nicoud, F. (2011). Tempora mutantur: In search of a new testament for NEG. *Journal of Economic Geography*, 11(2):215–230.

Delgado, M., Porter, M. E., and Stern, S. (2010). Clusters and entrepreneurship. *Journal of Economic Geography*, 10(4):495–518.

Docquier, F. and Marfouk, A. (2006). International migration by educational attainment (1990-2000). In *International migration, remittances and the brain drain*. Palgrave-Macmillan.

Docquier, F., Peri, G., and Ruyssen, I. (2014). The cross-country determinants of potential and actual migration. *International Migration Review*, 48:37–99.

Forslid, R. and Okubo, T. (2012). On the development strategy of countries of intermediate size—An analysis of heterogeneous firms in a multi-region framework. *European Economic Review*, 56(4):747–756.

Forslid, R. and Ottaviano, G. I. P. (2003). An analytically solvable core-periphery model. *Journal of Economic Geography*, 3(3):229–240.

Fujita, M., Krugman, P., and Venables, A. (1999). *The spatial economy: cities, regions and international trade*. The MIT Press.

Grogger, J. and Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, 95:42–57.

Holmes, T. J. and Stevens, J. J. (2002). Geographic concentration and establishment scale. *The Review of Economics and Statistics*, 84(4):682–690.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–99.

Krugman, P. R. and Venables, A. J. (1995). Globalization and the inequality of nations. *The Quarterly Journal of Economics*, 110(4):857–80.

Martin, P. and Rogers, C. A. (1995). Industrial location and public infrastructure. *Journal of International Economics*, 39(3-4):335–351.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Melitz, M. J. and Ottaviano, G. I. P. (2008). Market size, trade, and productivity. *Review of Economic Studies*, 75(1):295–316.

Melitz, M. J. and Redding, S. J. (2013). Heterogeneous firms and trade. CEPR Discussion Papers No. 9317.

Nocke, V. (2006). A gap for me: entrepreneurs and entry. *Journal of the European Economic Association*, 4(5):929–956.

Okubo, T. (2009). Trade liberalisation and agglomeration with firm heterogeneity: forward and backward linkages. *Regional Science and Urban Economics*, 39(5):530–541.

Okubo, T. (2010). Firm heterogeneity and location choice. Discussion Paper Series DP2010-11, Research Institute for Economics & Business Administration, Kobe University.

Okubo, T., Picard, P. M., and Thisse, J.-F. (2010). The spatial selection of heterogeneous firms. *Journal of International Economics*, 82(2):230–237.

Ottaviano, G. I. P. (2011). 'New' new economic geography: firm heterogeneity and agglomeration economies. *Journal of Economic Geography*, 11(2):231–240.

Pfluger, M. (2004). A simple, analytically solvable, chamberlinian agglomeration model. *Regional Science and Urban Economics*, 34(5):565–573.

Venables, A. J. (1996). Equilibrium locations of vertically linked industries. *International Economic Review*, 37(2):341–59.

Venables, A. J. (2011). Productivity in cities: self-selection and sorting. *Journal of Economic Geography*, 11(2):241–251.

Chapter 2

Transport costs: A more realistic approach

2.1 Introduction

Space matters, therefore, transport costs matter. However, even though, from an empirical point of view, the magnitude of transport costs or, more generally, trade costs, is very important in practice and "the death of distance is exaggerated" (Anderson and van Wincoop, 2004), considering them explicitly in the theoretical models has not always been a simple matter.

Indeed, the research carried out in the middle of the last century, mainly by Arrow and Debreu (1954), which resolved a key problem in economic theory, the existence of a general competitive equilibrium, is essentially non-spatial. The explanation of why this problem occurs comes from Starrett (1978) and his Spatial Impossibility Theorem: "Assume a two region economy with a finite number of consumers and firms. If space is homogeneous, transport is costly, and preferences are locally non-satiated, there is no competitive equilibrium involving transportation" (Fujita and Thisse, 2002). We had to wait until the New Economic Geography (NEG) emerged in the early 1990s for theoretical models to include transport costs in a way which was operational and at the same time surprisingly simple[1]. This, of course, used a non-competitive market structure characterized by the existence of scale economies and differentiated products.

Space was almost always incorporated into NEG models using the approach known as Iceberg transport costs. What these are, their advantages and, particularly, their disadvantages are described in Section 2.2. Another criticism based on the presence of non-homogeneous transport costs will be detailed in Section 2.5.

The aim of this chapter is to propose an alternative theoretical approach to Iceberg trans-

[1]The importance of how to formulate transport costs in the empirical models of the NEG was demonstrated by Bosker and Garretsen (2010).

port costs. Our proposal presents two main characteristics. One, it is compatible with the equation structure associated with the Iceberg approach and, therefore, shares its simplicity and operational procedure. Two, it is an improvement on, and answers some of the criticisms directed at, Iceberg costs.

This chapter is organized as follows. Section 2.2 gives a brief description of Iceberg costs and their implications. Section 2.3 presents the most important elements of our proposal: in a nutshell, part of the workforce of companies, which is homogeneous, is devoted to producing the good itself and part to transporting it, with the same salary in both cases. Section 2.4 analyzes the relationship between the domestic and the foreign price from both approaches. Section 2.5 explores a new criticism based on non-homogeneous space. Section 2.6 shows the relationship between the new parameters and the standard freeness of trade parameter.

2.2 Iceberg transport costs in the literature: A very short review

This section is based on the excellent article by McCann (2005).

As is well known, Iceberg transport costs were initially proposed by Samuelson (1952). Their main advantage, and the reason for their subsequent widespread use, is that they allow transport costs to be introduced without having to add a specific transport sector, which would make the theoretical models excessively complex. It is essentially very simple: due to being transported, a part of the good evaporates or is lost in its movement from one region to another. Hence the name "Iceberg": like the mass of ice, a portion of output melts away as it travels so that, for a physical unit of the good to arrive at the destination, T units must be sent ($T > 1$).

Despite its success, this approach is not without problems. First, there is no space in the strict sense, as the magnitude of the transport cost, in general, does not depend on distance. Second, de facto, it is equivalent to a level effect similar to that of a tariff: there is a discontinuity between the domestic price and the foreign price, where the difference is due to the cost of transport. Third, and connected to the above, the foreign price is T times the domestic price, so that any increase in the price, even if it has nothing to do with transport, automatically generates an increase in the cost of transport. Fourth, the cost of transport per tonne is independent of the weight/amount of what is transported.

Nearly 40 years later, Krugman (1991, 1992), improved on Samuelson's contribution by considering, always within the Iceberg approach, that geography exists and, therefore, that the magnitude of the transport cost is a function of the distance traveled D, according to the following expression:

$$V_f = V_d e^{-\tau D} \tag{2.1}$$

where V_f is the value of the good which actually reaches the foreign destination, V_d is the value in the original (domestic) location and τ is the Iceberg decay parameter. Thus, the relation between V_f and D is a decreasing function which, due to its convexity, seems to agree with the economies of distance found in reality. However, it can be demonstrated that the relationship between prices, with obvious notation, is given by:

$$p_f = p_d e^{\tau D} = T p_d \tag{2.2}$$

With this transportation cost framework, based on the Iceberg approach, McCann (2005) analyzes three main criticisms.

The first criticism to this standard Iceberg formulation is that, according to (2.2), the relationship between p_f, on the vertical axis, and D is increasing and convex. The second is based on the fact that transport cost per ton-kilometer is perfectly elastic, and not decreasing, in relation to the initial amount (weight of good) transported. These two properties go against the existence of economies of distance, a conclusion that has been empirically obtained (Bayliss and Edwards, 1970; Jansson and Shneerson, 1985; Tyler and Kitson, 1987; Savage, 1997). This conclusion implies that p_f should increase with distance, but in a concave manner, and that the transport cost per ton and per kilometer decrease with respect to the distance and the amount (weight) of the good shipped.

A third criticism is based on the fact that the foreign price (p_f) has a unitary elasticity with respect to the domestic price (p_d), implying that an increase in the domestic price always raises the foreign price by the same proportion., something at least questionable.

In Section 2.5, we will present a fourth criticism derived from a particular conclusion obtained by the Iceberg approach, in a context of non-homogeneous transportation costs, that deals with the outcome that a change in the transport cost in part of the distance covered has the same effect regardless of the size of that part. Our approach resolves the first two criticisms in Section 2.3, the third is resolved in Section 2.4, and the last one in Section 2.5.

2.3 Our proposal and its consequences

Our approach is just as simple and operative as the Iceberg approach; that is, there is no need to introduce a transport sector into the theoretical models and it is not affected by the problems described in Section 2.2.

Let us suppose, without loss of generality[2], that there are only two regions or countries: subindex d refers to the domestic country and subindex f to the foreign country. At the same time, we only refer to the manufacturing good, which in the New Economic Geography is traditionally the differentiated good, produced with increasing returns and subject to Iceberg type transport costs; in its market a structure of monopolistic competition with free entry prevails.

Let us consider the following production function:

$$l = c + \beta q_d + \beta q_f + t + \delta q_f \tag{2.3}$$

where l is the amount of work used in the production process, q_d is the amount of output which goes to the national market, and q_f is the amount of output which is exported. The parameters and their interpretation are as follows: $c > 0$ is the fixed labor needed to start production (there are scale economies), $\beta > 0$ is the variable labor needed per unit of output produced, $t > 0$ is the fixed labor needed to launch the transport process (there are scale economies in transport) and $\delta > 0$ is the variable labor needed per unit of output transported. The new parameters associated with transport are t and δ. In short, part of the workforce of companies, which is homogeneous, is devoted to producing the good itself and part to transporting it, with the same salary in both cases. The workers dedicated to transport perform all the tasks associated with exporting goods: transport itself, the manager who organizes the foreign trade, the vendor who makes the trade agreements, the worker who implements the necessary measures for the product to pass quality tests in the receptor country, etc.

In turn, δ depends on distance as follows:

$$\delta = \ln(1 + \tau D) \tag{2.4}$$

The profit function π of the companies is given, with obvious notation, by the following expression:

$$\pi = p_d q_d + p_f q_f - w(c + \beta q_d + \beta q_f + t + \delta q_f) \tag{2.5}$$

The first order conditions of profit maximization lead to the prices in the domestic market and abroad:

$$p_d = w\beta \frac{\sigma}{\sigma - 1} \tag{2.6}$$

$$p_f = w(\beta + \delta)\frac{\sigma}{\sigma - 1} = p_d + w\frac{\sigma}{\sigma - 1}\ln(1 + \tau D) = p_d\left(1 + \frac{\delta}{\beta}\right) \tag{2.7}$$

[2] All that follows can be extended to the generic case of S countries. Its only effect is to complicate the notation.

where $\sigma > 1$ is the elasticity of substitution between pairs of varieties or the price elasticity of each variety of the product in question (Dixit and Stiglitz, 1977). It can immediately be seen from (2.7) that the relationship between p_f and D is as shown in Figure 2.1.

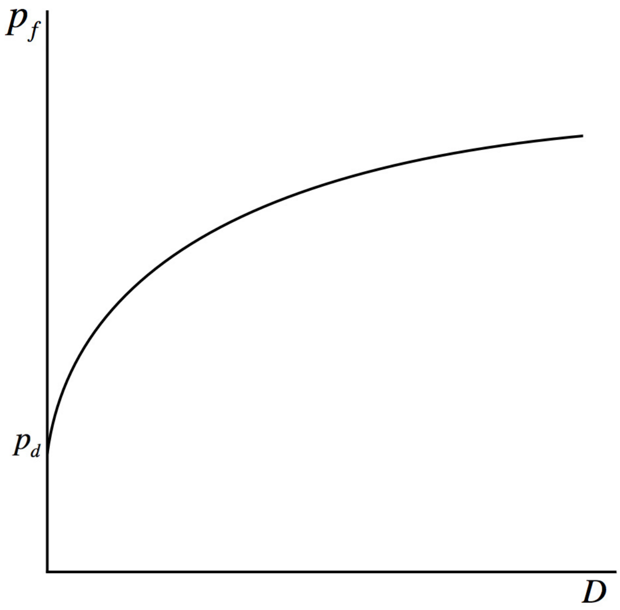

Figure 2.1: Relationship between p_f and distance

Also, if we use Z to denote transport cost per unit (or weight) of the transported good and kilometer traveled, we have:

$$Z = \frac{wt}{q_f D} + w\frac{\ln(1 + \tau D)}{D} \tag{2.8}$$

The $Z - q_f$ and $Z - D$ relationships, respectively, are represented in Figure 2.2.

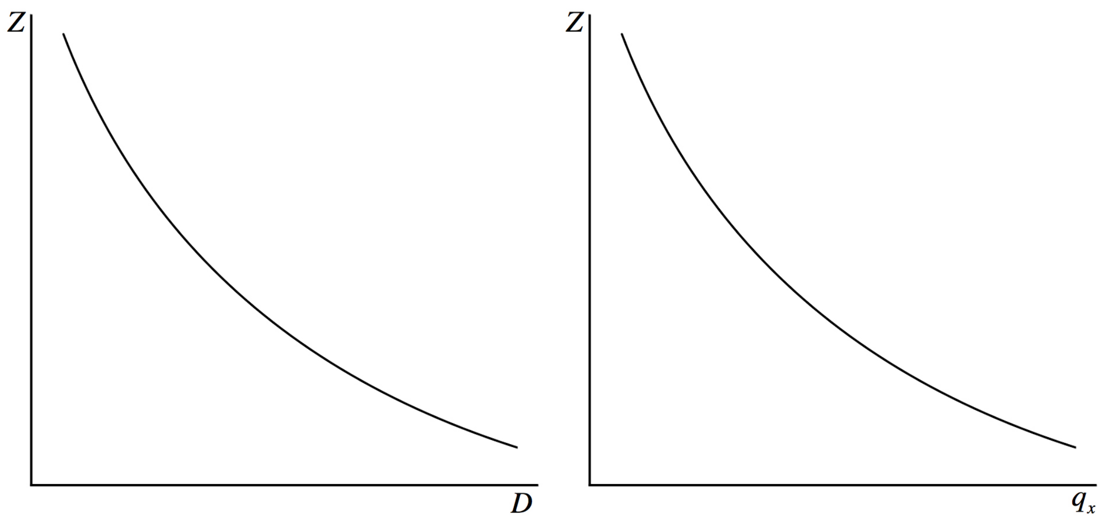

Figure 2.2: Relationship between cost per ton-kilometer and distance (left) and the amount transported (right)

Consequently, our approach is characterized by two main aspects. On one hand, it is perfectly compatible with all the literature relating to Iceberg transport costs, as the sequence of basic equations is equivalent in both approaches (see the Appendix for a detailed demonstration of this statement). On the other, it is a better reflection of the empirical evidence relating to transport costs in the real world and resolves some of the criticisms associated with the Iceberg approach described in Section 2.2. In our framework, the relationship between the foreign price and distance is increasing and concave (Figure 2.1) and the cost per tonne and kilometer is decreasing and convex in relation to the distance traveled (Figure 2.2, left) and in relation to the amount transported (Figure 2.2, right). Therefore the first two criticisms (see the end of Section 2.2) are resolved. The empirical evidence confirming that these three graphs give a better reflection of what happens in the real world is extensive and can be consulted in the work by McCann (2005) cited above.

2.4 The relationship between domestic and foreign prices

As we saw in Section 2.2, the third common criticism of the Iceberg approach, described by Ottaviano and Thisse (2004) as unrealistic, is based on the elasticity of the foreign price over the domestic price. In the standard Iceberg approach, the relationship between the two prices is:

$$p_f = Tp_d = Tw\beta\frac{\sigma}{\sigma-1} \tag{2.9}$$

Based on (2.9), it is trivial to deduce that, if salary increases, the product becomes more differentiated (lower σ) or productivity falls (higher β), an increase in p_d and p_f of the same proportion occurs. Of course, if T varies and p_d is not changed, p_f moves in the same direction and proportion as T (the elasticity of p_f in relation to the parameter representing the transport cost T is equal to one). In our approach things are a little different:

$$p_f = w(\beta+\delta)\frac{\sigma}{\sigma-1} = \left(1+\frac{\delta}{\beta}\right)p_d = Tp_d \tag{2.10}$$

In other words, $T = \left(1+\frac{\delta}{\beta}\right)$. The effects of w and σ on p_f are the same as the Iceberg approach. But this is not the case with the technological parameters β and δ. In fact, the relevant elasticities are now:

$$E_\beta p_d = 1 \qquad E_\beta p_f = \frac{\beta}{\beta+\delta} < 1 \tag{2.11}$$

$$E_\delta p_d = 0 \qquad E_\delta p_f = \frac{\delta}{\beta+\delta} < 1 \tag{2.12}$$

$$E_\beta p_d + E_\delta p_d = E_\beta p_f + E_\delta p_f = 1 \tag{2.13}$$

Consequently, the foreign price p_f responds proportionately lower to variations in parameters β and δ and, thus, resolves the third of the four criticisms presented in Section 2.2. The explanation is simple: as there are now two productive tasks, transport (δ) and production (β), rather than one, the changes in the technological parameters considered individually are translated with amortization to the foreign price, although their combined or simultaneous action maintains the proportional effect ($E_\beta p_f + E_\delta p_f = 1$). This implies that an increase in the production cost, without affecting the transportation cost, will raise the foreign price in a lower proportion than the domestic price, and only an increase in the same proportion of both the production and transport costs will affect the foreign price in the same proportion as the domestic price.

2.5 A criticism based on non-homogeneous transportation costs

The models that deal with the introduction of space into the microeconomics analysis, Krugman (1991), Martin and Rogers (1995) and Forslid and Ottaviano (2003) being the most commonly used (core-periphery model, footloose capital model and footloose entrepreneur model, respectively), always introduce the Iceberg approach in the same manner: as the transportation cost between two separate regions or countries. In this framework, the structure of the space in which the transport occurs is of secondary importance. This makes the models more simple, without having to deal directly with the distance, D, since all the prices associated with transportation are summarized in the 'black box' variable T (in other words, $p_f = Tp_d$).

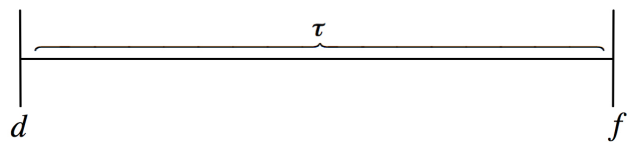

Figure 2.3: Pure homogeneous transportation costs framework

In Figure 2.3 we can see that the distance between the domestic and the foreign countries, D, will be affected by τ, so that $T = e^{\tau D}$ in the Iceberg approach. As D is always the same (physically), the only variable with economic interpretation is τ or, simplifying, just T.

However, we believe, as McCann (2005) does, that even if the simplification is used as an instrumentalist perspective (Blaug, 1993), its economic fundamentals have to be essentially generalizable as realistically as possible, to overcome possible future research problems that deal with the real world in order to foster the introduction of more sophisticated (but nonetheless useful and practical) instruments. One of the new insights through which we can observe the effects of the transportation cost over the space is by considering non-homogeneous transportation costs. This would imply that the same space can be divided into parts that affect to the transportation process differently.

We can decompose distance D into two different parts, D_1 (the distance from d to m) and D_2 (the distance from m to f), in which we will suppose that D_2 is a more complex part so transporting goods across it will be more costly than over the previous terrain: $\tau_2 > \tau_1$. In this case, the space not only affects extensively (more distance implying a

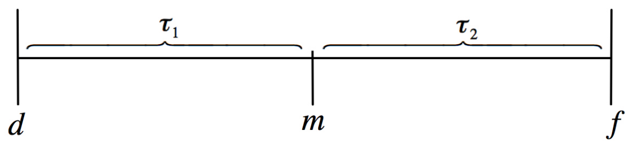

Figure 2.4: A very simple non-homogeneous transportation costs framework

higher cost), but also intensively. This could consist of a more difficult terrain, poor infrastructure or certain parts with more traffic congestion that increase transportation costs.

In accordance with the way the Iceberg cost is constructed, it will be necessary to send e^{τ_2} units from m to f (m being, as shown in the Figure 2.4, the point at which the space changes, and f the foreign region) for one unit to arrive at the end of the transportation process. And it is necessary to send e^{τ_1} units from d (domestic region) for one unit to arrive to m. Therefore, the whole transportation cost, T, will be:

$$T = e^{\tau_1 D_1} e^{\tau_2 D_2} = e^{\tau_1 D_1 + \tau_2 D_2} \tag{2.14}$$

Of course, if the space is homogeneous, that is, if $\tau_1 = \tau_2 = \tau$ we have the common Iceberg cost, $T = e^{\tau D_1 + \tau D_2} = e^{\tau(D_1+D_2)} = e^{\tau D}$, with $D = D_1 + D_2$.

The problem that arises with the Iceberg approach, in a situation like the one defined in Figure 2.4, is that the proportional effect of an increase in the transport cost of one of the parts (let's say D_2) is the same, regardless the relative size of that part:

$$E_{\tau_2} T = \frac{\partial T}{\partial \tau_2} \frac{\tau_2}{T} = \tau_2 D_2 \tag{2.15}$$

which only depends on D_2. In our proposal, for the introduction of the two types of space we can use $\delta = \ln(1+\tau_1 D_1 + \tau_2 D_2)$. If the two distances covered have the same ruggedness, we will have the simplified version: $\delta = \ln(1 + \tau D_1 + \tau D_2) = \ln(1 + \tau D)$. The effect of an increase in τ_2 will be (with $T = 1 + \frac{\delta}{\beta} = 1 + \frac{\ln(1+\tau_1 D_1 + \tau_2 D_2)}{\beta}$)

$$E_{\tau_2} T = \frac{\partial T}{\partial \tau_2} \frac{\tau_2}{T} = \frac{\tau_2 D_2}{(1 + \tau_1 D_1 + \tau_2 D_2)(\beta + \ln(1 + \tau_1 D_1 + \tau_2 D_2))} \tag{2.16}$$

which depends negatively on D_1 and positively on D_2. This seems more realistic, as an increase in the transportation process over a part of the distance covered will have a proportional effect that will be bigger the larger that part over the total distance D, and will be lower the larger the section (D_1) not affected by the increase in transport costs. This result is based on the existence of scale and distance economies in the transportation process, but is only observable with a non-homogeneous transportation cost framework. It also implies that the sudden appearance of a trade barrier (that can be considered as a deterioration of a small part of the space) or even a border effect will affect trade partners that are geographically closer to the new barrier more proportionately, something which seems very plausible. So the proportional effect of an increase of τ_2 in T will be bigger if the goods are originally sent from m than if they are sent from d.

2.6 A new interpretation of the freeness of trade parameter

So far, we have seen how a new conceptual transportation process changes the properties of the transportation costs and how their relationship with certain economic variables (such as the distance, the initial weight, the original price and the heterogeneous structure of the space) can be explained in a more realistic manner. In a nutshell, we have been changing the properties of T so that it better reflects what the empirical studies seem to find. But, as the mathematical interpretation of T changes, the definition of the freeness of trade parameter, which is based on T, will also change.

In the Iceberg formulation, the relationship between domestic and foreign prices is $p_f = Tp_d$, where T, as we state above, is greater than one. In short, this equation can be expressed as $p_f = (1+\varphi)p_d$, with $\varphi > 0$. It is also standard in the literature to define the freeness of trade parameter as $\phi = T^{1-\sigma}$. This parameter ranges from 0 (infinite or prohibitive transport cost) to 1 (absence of transport cost) and is a measure of how easy it is to move goods between countries. In our case, we find directly, based on (2.7), that $T = \left(1 + \frac{\delta}{\beta}\right)$, so our freeness of trade parameter is given by:

$$\phi = \left(\frac{1}{1+\varphi}\right)^{\sigma-1} = \left(\frac{\beta}{\beta+\delta}\right)^{\sigma-1} \tag{2.17}$$

The new parameter $\varphi = \frac{\delta}{\beta}$ is especially useful because it is a relative measurement of transport technology in relation to the production technology of the good, given that it is the quotient between the labor requirements needed for each activity. It can also be interpreted as the amount of goods which could be obtained if the workers devoted to transport were to change to purely productive tasks. The relationship between the freeness of trade parameter and the quotient between transport and production technologies is given in Figure 2.5. The more efficient the transport activity is in relation to the production activity (lower φ), the higher, logically, is the freeness of trade parameter, and vice versa. The fact that ϕ depends, in relative terms, on transport and production technologies is found only in our approach. A sector with greater absolute transport costs (greater δ) than another sector can have greater freeness of trade as long as its relationship with production costs in that industry $\left(\frac{\delta}{\beta}\right)$ is lower.

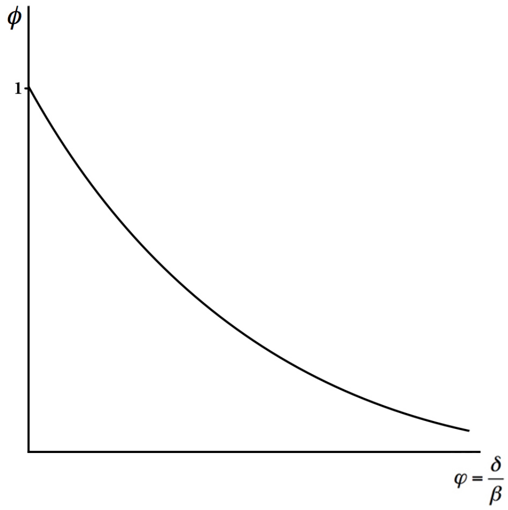

Figure 2.5: ϕ as a function of φ

2.7 Conclusions

Iceberg-type transport costs have been and are used intensely in the theoretical models of the New Economic Geography. Indeed, one of the reasons for the significant development of this type of model is the excellent features of the Iceberg approach. In the words of McCann (2005): "it is clear that new economic geography models incorporating Iceberg transport costs have gone further than any other framework in developing a general equilibrium approach to analysing spatial economic phenomena". However, as the above reference clearly demonstrates, Iceberg costs also present important problems. For instance, when computers, books or cars are transported, no units "evaporate" or "melt" on the way. What the transport activity does is to consume some resources of the firms engaged in it. This is the essence of the alternative approach we suggest: some of the workers of the company carry out production tasks and some are engaged in the process of transporting the goods to different locations.

This approach, more realistic in its basis, presents two important characteristics. On one hand, it is compatible with the entire structure of equations of the Iceberg approach; in other words, it is isomorphic with it. On the other, it overcomes some of the disadvantages associated with Iceberg transport costs. In fact, the c.i.f. price of the product increases with distance, but concavely, and the transport rate per ton-kilometer is decreasing and convex with respect to both distance and the quantity of the good shipped. An increase in the transport cost in some part of the space will affect more proportionately those goods in which this space occupies a bigger proportional part of the whole transportation size to be covered. Furthermore, the standard freeness of trade parameter now depends in a useful way on the relationship between transport and production technologies. All these are empirically tested features of the approach proposed here, which the Iceberg framework does not satisfy.

Appendix: Equivalence between the equations of the Iceberg approach and the approach proposed in this work

We take the book by Fujita et al. (1999) as the standard reference associated with the basic developments of the New Economic Geography and its Iceberg approach, FKV hereafter. We have adapted the notation of FKV to our own and to the two-country case considered. We have to demonstrate that the basic equations of Sections 4.2 ("Multiple locations and transportation costs") and 4.3 ("Producer behavior") in FKV are also valid in our case. Our approach does not introduce any novelty on the demand side.

In FKV, expression (4.14) represents, $p_f = Tp_d$. In our case, see (2.7), we come to the same equation, with $T = \left(1 + \frac{\delta}{\beta}\right)$. At the same time, the equivalent to the production function (4.18) in FKV is given by (2.3). What about the size of a representative firm? In FKV, the size of a manufacturing company is $q^* = q_d + q_f T$ as, to sell q_f units abroad, $q_f T$ must be produced, given that some will evaporate en route. In our approach, nothing is lost en route, so that everything produced is sold. Therefore, we need to demonstrate that our approach leads to the same expression (4.22) which defines q^* in FKV. We start by making the profit in (2.5) equal to zero. After some algebra and replacing p_d and p_f with their expressions in (2.6) and (2.7), we get:

$$\left[q_d + q_f\left(1 + \frac{\delta}{\beta}\right)\right] = q_d + q_f T = \frac{(c+t)(\sigma - 1)}{\beta} \tag{2.18}$$

which is exactly (4.22) in FKV, applying the equivalences between the two notations ($(c+t)$ is the fixed work required to produce the goods which, in FKV, is denoted by F; β is the variable labor needed per unit of output produced, which, in the notation of FKV, is c^M).

At the same time, the associated equilibrium labor input l^* in our approach is:

$$l^* = \beta q_d + \beta q_f + \delta q_f + c + t = \beta\left(q_d + q_f\left(1 + \frac{\delta}{\beta}\right) + c + t\right) \tag{2.19}$$

which, through (2.18), equals:

$$l^* = (c+t)\sigma \tag{2.20}$$

an expression identical to (4.23) in FKV. In turn, we can obtain the important expressions (4.34) and (4.35) in FKV, which close the basic range of equations of the model and which define, respectively, the price index and the wage equation, just by carrying out the normalizations $\beta = \frac{\sigma-1}{\sigma}$ (equivalent to $c^M = \frac{\sigma-1}{\sigma}$, which is (4.29) in FKV) and $c + t = \frac{\mu}{\sigma}$ (equivalent to $F = \frac{\mu}{\sigma}$, which is (4.31) in FKV), where μ is the proportion of income used to consume industrial goods.

References

Anderson, J. E. and van Wincoop, E. (2004). Trade costs. *Journal of Economic Literature*, 42(3):691–751.

Arrow, K. J. and Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, 22(3):265–290.

Bayliss, B. T. and Edwards, S. L. (1970). *Industrial demand for transport*. Her Majesty's Stationery Office.

Blaug, M. (1993). *The methodology of economics*. Cambridge University Press.

Bosker, M. and Garretsen, H. (2010). Trade costs in empirical New Economic Geography. *Papers in Regional Science*, 89(3):485–511.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, 67(3):297–308.

Forslid, R. and Ottaviano, G. I. P. (2003). An analytically solvable core-periphery model. *Journal of Economic Geography*, 3(3):229–240.

Fujita, M., Krugman, P., and Venables, A. (1999). *The spatial economy: cities, regions and international trade*. The MIT Press.

Fujita, M. and Thisse, J. (2002). *Economics of agglomeration: cities, industrial location, and regional growth*. Cambridge University Press.

Jansson, J. and Shneerson, D. (1985). Economies of trade density in liner shipping and optimal pricing. *Journal of Transport Economics and Policy*, 19:7–22.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–99.

Krugman, P. (1992). *Geography and trade*. The MIT Press.

Martin, P. and Rogers, C. A. (1995). Industrial location and public infrastructure. *Journal of International Economics*, 39(3-4):335–351.

McCann, P. (2005). Transport costs and new economic geography. *Journal of Economic Geography*, 5(3):305–318.

Ottaviano, G. and Thisse, J.-F. (2004). Agglomeration and economic geography. In *Handbook of Regional and Urban Economics*, volume 4. Elsevier.

Samuelson, P. (1952). The transfer problem and transport costs: the terms of trade when impediments are absent. *Economic Journal*, 62:278–304.

Savage, I. (1997). Scale economies in United States rail transit systems. *Transportation Research Part A: Policy and Practice*, 31(6):459–473.

Starrett, D. (1978). Market allocations of location choice in a model with free mobility. *Journal of Economic Theory*, 17(1):21–37.

Tyler, P. and Kitson, M. (1987). Geographical variations in transport costs of manufacturing firms in Great Britain. *Urban Studies*, 24:61–73.

Chapter 3

On the parametric description of the French, German, Italian and Spanish city size distributions

3.1 Introduction

Since the contribution of Zipf (1949), the study of city size distribution has been of great importance in the field of Urban Economics. The Zipf distribution, or its more general form of Pareto distribution, has been extensively studied by many authors: Black and Henderson (2003), Ioannides and Overman (2003), Soo (2005), Anderson and Ge (2005) and Bosker et al. (2008). But this chapter is based, above all, in the works of the important contributions of Eeckhout (2004), Giesen et al. (2010) and Ioannides and Skouras (2013).

The first of these last references highlights the need of considering the whole sample of cities when studying their size distribution, and proposes the lognormal distribution (see also Parr and Suzuki (1973)). Giesen et al. (2010) continues a line of research initiated by Reed (2001, 2002, 2003) and Reed and Jorgensen (2004) in which the double Pareto lognormal (dPln) distribution is introduced in the study of city size. This distribution has Pareto tails mixed (by means of a convolution) with a lognormal body and offers a good fit to the data, see Giesen et al. (2010) and González-Val et al. (2013). Ioannides and Skouras (2013) propose two distributions which have a lognormal body and, above a certain exact threshold, a Pareto upper tail mixed or not (by means of a convex linear combination) with the lognormal. These two recently proposed distributions still do not outperform the dPln for US places in the year 2000, as Giesen and Suedekum (2014) indicate.

González-Val et al. (2013a); González-Val et al. (2013) use city population data of France, Italy and Spain without size restrictions. Schluter and Trede (2013) use a dataset of all German municipalities or *Gemeinden* and propose a composition of the normal distribution with a Box-Cox transformation of the population data, with apparently quite good results.

This leads to a distribution which we will call normal-Box-Cox (nBC), to be defined in Subsection 3.3.3.

Our aim in this chapter is to compare the lognormal, the dPln and the nBC distributions for, generally decennial, samples of city size data of France, Germany, Italy and Spain without size restrictions. The main result is that there is no single function that can explain the city size distribution of all countries. Mainly, the dPln works well for the case of France and Italy, but is outperformed by the nBC for the case of Germany and Spain.

The rest of this chapter is organized as follows. Section 3.2 describes the databases used in this paper. Section 3.3 shows the definitions and main properties of the three distributions studied. Section 3.4 shows the detailed results, country by country. Finally, Section 3.5 concludes.

3.2 The databases

There is a lively debate about the proper definition of cities from an economic point of view. The usual datasets comprise administratively defined cities, not always coinciding with the economic sense of a city in terms of, for example, commuting or trade flows. This is a limitation of the usual census datasets, and becomes more relevant perhaps for the biggest agglomerations in a country (Giesen and Suedekum, 2012).

One way to overcome this problem is to define clusters that give economic sense to actual agglomerations, irrespective of their legal borders, as pioneered by Rozenfeld et al. (2008, 2011). The construction of these data sets relies on the availability of previous good information about the geolocalization of urban settlements. There are few of these data sets currently, and the availability of data therefore imposes a clear constraint on the studies that can be carried out.

Another way to deal with this issue is to use recent census data sets that include all of the (administratively defined) cities, and cover (almost) 100% of the population. This is a great advance which allows us to study city size distribution from an un-truncated database, as Eeckhout (2004) advocates.

The description of the database used for the case of France, Germany, Italy and Spain is now presented.

For the case of France, as in González-Val et al. (2013a), we consider the lowest spatial subdivision, the *communes*, as listed by the *Institut national de la statistique et des études é conomiques* (`www.insee.fr`). We have data for the years 1990, 1999 and 2009. For the three samples, we have aggregated the populations of the *arrondisements* of the three biggest cities (which are also *communes*): Paris, Marseille and Lyon. Note that Giesen and Suedekum (2012) use this kind of data for the year 2008.

For the case of Germany, Italy and Spain, the administrative urban unit of the data is the municipality (*Gemeinden* for the case of Germany). For Germany, we take data from two sources. The first is the data used in Schluter and Trede (2013), which has been kindly provided to us by Prof. Trede (the original source is the *Federal German Statistical Office*). We take the data of the years 1996 and 2006 in order to obtain a decennial period similarly to the other countries considered. The second source is, directly, the cited statistical office through its web page `www.Destatis.de`. We use information of the last available year 2011 for comparative purposes. For Italy, the data is obtained from the *Istituto Nazionale di Statistica* (`www.istat.it`), with all the Italian municipalities (*comuni*) for the period 1901-2011. We have used the Italian census for 1936 instead of 1941 because of the participation of Italy in the Second World War. The data for Spain is taken from the *Instituto Nacional de Estadística* (`www.ine.es`). They cover all the municipalities (*municipios*) in the period 1900-2010.

In Table 3.1, we offer the descriptive statistics of the data used for France, Germany, Italy and Spain. The information for Italy and Spain is the same as that in Table 1 of González-Val et al. (2013).

One of the common problems in the analysis of city size distributions is trying to achieve the most homogeneous database for the countries that are going to be analyzed. But countries often have different statistics due to the use of different concepts of city units. From an economic point of view, metropolitan areas are usually considered the proper "economic" base for the analysis of urban units, concentrating not only the core but also the less populated surroundings that have an intense relationship with it.

But for the consideration of an un-truncated database composed of all the cities in a country (one that includes all the population, not only those who live in the more populated areas), the administrative definition used in the different official statistics is more useful. With this in mind, we select the administrative division, based on the *communes* for the case of France, *Gemeinden* for the case of Germany, *comuni* in Italy, and *municipios* in Spain. The consideration of the more detailed data about the distribution of cities, one that allows us to include all the urban units in the four countries from an administrative point of view, provides the database with a certain degree of homogeneity.

Table 3.1: Descriptive statistics of the samples of French, German, Italian and Spanish urban units used

	Urban units	Mean of pop.	SD of pop.	Minimum	Maximum
France					
1990	36,644	1,611	14,157	1	2,175,200
1999	36,643	1,679	14,173	1	2,147,857
2009	36,674	1,793	14,895	1	2,257,981
Germany					
1996	14,559	5,633	40,608	2	3,458,763
2006	12,312	6,686	44,043	7	3,404,037
2011	11,292	7,114	45,415	10	3,326,002
Italy					
1901	7,711	4,275	14,425	56	621,213
1911	7,711	4,648	17,393	58	751,211
1921	8,100	4,864	20,032	58	859,629
1931	8,100	5,067	22,560	93	960,660
1936	8,100	5,234	25,274	116	1,150,338
1951	8,100	5,866	31,138	74	1,651,393
1961	8,100	6,250	39,131	90	2,187,682
1971	8,100	6,684	45,582	51	2,781,385
1981	8,100	6,982	45,329	32	2,839,638
1991	8,100	7,010	42,450	31	2,775,250
2001	8,100	7,021	39,325	33	2,546,804
2011	8,094	7,490	41,505	34	2,761,477
Spain					
1900	7,800	2,282	10,178	78	539,835
1910	7,806	2,452	11,217	92	599,807
1920	7,812	2,622	13,501	82	750,896
1930	7,875	2,892	17,514	79	1,005,565
1940	7,896	3,181	20,100	11	1,088,647
1950	7,901	3,480	26,033	64	1,618,435
1960	7,910	3,802	33,652	51	2,259,931
1970	7,956	4,241	43,972	10	3,146,071
1981	8,034	4,701	45,995	5	3,188,297
1991	8,077	4,882	45,220	2	3,084,673
2001	8,077	5,039	43,079	7	2,938,723
2010	8,114	7,795	47,530	5	3,273,049

3.3 Description of the distributions used

3.3.1 The lognormal distribution (lgn)

The well-known lognormal distribution for the population of cities has been proposed in the field of Urban Economics by Parr and Suzuki (1973) and afterwards by Eeckhout (2004) when considering all the cities. The corresponding density is simply

$$f_{\ln}(x, \mu, \sigma) = \frac{1}{x\,\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \tag{3.1}$$

where $\mu, \sigma > 0$ are, respectively, the mean and the standard deviation (SD) of $\ln x$, x being the population of the urban units under study.

3.3.2 The double Pareto lognormal distribution (dPln)

The second distribution in our study will be the double Pareto lognormal distribution, introduced by Reed (2002, 2003) and Reed and Jorgensen (2004):

$$\begin{aligned} f_{\mathrm{dPln}}(x, \alpha, \beta, \mu, \sigma) &= \frac{\alpha\beta}{2x(\alpha+\beta)} \exp\left(\alpha\mu + \frac{\alpha^2\sigma^2}{2}\right) x^{-\alpha} \left(1 + \mathrm{erf}\left(\frac{\ln x - \mu - \alpha\sigma^2}{\sqrt{2}\sigma}\right)\right) \\ &\quad - \frac{\alpha\beta}{2x(\alpha+\beta)} \exp\left(-\beta\mu + \frac{\beta^2\sigma^2}{2}\right) x^{\beta} \left(\mathrm{erf}\left(\frac{\ln x - \mu + \beta\sigma^2}{\sqrt{2}\sigma}\right) - 1\right) \end{aligned} \tag{3.2}$$

where erf is the error function associated with the normal distribution and $\alpha, \beta, \mu, \sigma > 0$ are the four parameters of the distribution. It has the property that it approximates different power laws in each of its two tails: $f_{\mathrm{dPln}}(x) \approx x^{-\alpha-1}$ when $x \to \infty$ and $f_{\mathrm{dPln}}(x) \approx x^{\beta-1}$ when $x \to 0$, hence the name of double Pareto. The body is approximately lognormal, although it is not possible to exactly delineate the switch between the lognormal and the Pareto behaviors (Giesen et al., 2010). In this last reference, it is shown that the dPln offers a good fit for a number of countries. In this line, see also the work of González-Val et al. (2013) and Giesen and Suedekum (2014).

3.3.3 The normal-Box-Cox (nBC)

Schluter and Trede (2013) propose the idea of composing the normal distribution with the well-known Box-Cox transformation to analyze German city data. We include the distribution so obtained in our study because it turns out that the nBC provides good results in the case of Germany.

The Box-Cox transformation is given by the well-known expression (Box and Cox, 1964)

$$g_\lambda(x) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda} & \text{if} \quad \lambda \neq 0 \\ \ln x & \text{if} \quad \lambda = 0 \end{cases}$$

Its composition with the normal will be $g'_\lambda(x) f_\text{n}(g_\lambda(x), \mu, \sigma)$, where $g'_\lambda(x)$ is the derivative of $g_\lambda(x)$ with respect to x and

$$f_\text{n}(y, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

is the normal density function for the variable y, substituted in the previous expression by $g_\lambda(x)$. The case of $\lambda = 0$ leads to the lognormal, introduced in Subsection 3.3.1 and treated separately. Thus, for the case of $\lambda \neq 0$ we define the normal-Box-Cox (nBC) as the density

$$f_\text{nBC}(x, \mu, \sigma, \lambda) = \frac{x^{\lambda-1}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(\frac{x^\lambda - 1}{\lambda} - \mu\right)^2\right)$$

where x has the same meaning as before. The quantities μ and σ are, respectively, the mean and standard deviation of $\dfrac{x^\lambda - 1}{\lambda}$.

3.4 Results

For the sake of brevity, we will present the results country by country.

3.4.1 Results for France

In Table 3.2, we show the maximum likelihood (ML) estimators of the distributions studied for the 1990, 1999 and 2009 French samples of *communes*. For the lognormal (lgn), the ML estimators are exact and equal to the mean and standard deviation of the log-population data. For the other two distributions (dPln and nBC), we provide the ML estimators and 95% confidence intervals.[1] The estimations appear to be very precise in all cases.

In Table 3.3, we show the results of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for the distributions used. These two tests are very powerful when the sample size is high or very high (Razali and Wah, 2011) as in our French samples, and non-rejections only occur if the deviations (statistics) are really small. We observe that the lgn is strongly rejected in all cases and the nBC is also rejected always, although with lower values of the tests' statistics. The dPln is not rejected by either test in the 100% of the cases. According to these tests, the French *communes* size distribution can be best explained with the dPln. The excellent fit of the dPln for the French *communes* in the year 2008 has been anticipated by Giesen and Suedekum (2012).

In order to choose one of the models, in Table 3.4 we show the results of the Akaike information criterion (AIC) and Bayesian information criterion (BIC), which are especially well suited to the maximum likelihood estimation we have performed before. Both the AIC and BIC favor the distribution with greater maximum likelihood, but there is a penalty for the

[1] We have performed the estimations with MATLAB as in González-Val et al. (2013).

number of parameters used. The distribution with the lowest AIC and/or BIC is preferred.

For the case of the French samples, we observe that the lgn always obtains the greatest values of the AIC and BIC, and that the lowest AIC and BIC occur for the dPln in all cases. This result, together with the outcomes of the KS and CM tests yields that the French *communes* size distribution, can be very well described parametrically by the dPln.

3.4.2 Results for Germany

We now carry out a similar analysis for our 1996, 2006 and 2011 German samples of *Gemeinden*. First, in Table 3.5 we show the estimation results. The estimations obtained for the parameter λ for the years 1996 and 2006 are consistent with the results of Schluter and Trede (2013). In Table 3.6, we show the results of the KS and CM tests. The lgn, the dPln and nBC are (strongly) rejected in all cases.

The results of the AIC and BIC are shown in Table 3.7. The lgn is always the least preferred distribution. Note as well that the nBC is always preferred to the dPln for the German samples. Together with the results of the KS and CM tests, we conclude that the German city size distribution of *Gemeinden*, without size restrictions, is approximately described by the nBC. However, a substantial improvement is still possible.

3.4.3 Results for Italy

We have also performed the ML estimation of the Italian samples of *comuni* in the period 1901-2011. In Table 3.8 we show the estimation results for the lognormal, the dPln and the nBC. The estimations are quite precise in this case as well.

In Table 3.9 we show the results of the KS and CM tests for our three distributions. The lgn is always rejected except in 2011. The dPln is never rejected. The nBC is not rejected for the years 1981, 1991, 2001 and 2011. Thus, it follows that the dPln is a good parametric description of Italian *comuni* size in the period 1901-2011.

In Table 3.10 we show the results of the AIC and BIC for the Italian samples. Out of the three parametric models studied, the dPln is selected in 100% of the cases by both the AIC and BIC. Thus, for the Italian *comuni* without size restrictions we obtain an excellent parametric model, the dPln.

3.4.4 Results for Spain

Again, we have estimated the three distributions studied in this paper by ML for the samples of Spanish *municipios* in the period 1900-2010. In Table 3.11 we show the estimation results for the lognormal, the dPln and the nBC. The estimations are quite precise in this case as well.

In Table 3.12 we show the results of the KS and CM tests. The lgn and the dPln are always strongly rejected. The nBC is rejected in almost all cases, with the exception of the KS test in 1981. In Table 3.13 we show the values of the AIC and BIC. The nBC is always preferred to the dPln for the Spanish *municipios*. In short, Spanish city size in the period 1900-2010 is approximately described by the nBC although, again, a substantial improvement is possible.

Table 3.2: Exact estimators for the lgn and the French samples. Estimators and 95% confidence intervals of the parameters of the dPln and the nBC for the French samples

France	lgn		dPln			
	μ	σ	α	β	μ	σ
1990	6.06	1.34	0.98±0.02	2.89±0.31	5.39±0.04	0.81±0.03
1999	6.11	1.35	0.98±0.02	3.03±0.37	5.42±0.04	0.84±0.03
2009	6.21	1.34	1.02±0.02	3.41±0.19	5.52±0.03	0.89±0.02
	nBC					
	μ	σ	λ			
1990	3.99±0.06	0.54±0.02	-0.14±0.01			
1999	4.05±0.06	0.55±0.02	-0.14±0.01			
2009	4.18±0.07	0.57±0.02	-0.13±0.01			

3.4.5 An informal graphical approximation

The use of graphical tools in assessing the fit of parametric distributions to empirical data has certain shortcomings to be taken into account, see, e.g., González-Val et al. (2013b). In this reference, it has been shown that, when representing the differences between the empirical and estimated ln(1 − cdf)'s, where cdf is the relevant cumulative density function, an amplification effect of the differences of the cdf's is obtained for the upper tail. A similar effect occurs for the ln(cdf)'s and the lower tail. This amplification effect increases as we approach infinity for the upper tail or zero for the lower tail, and it is difficult to quantify.

Furthermore, the goodness-of-fit, as tested by the KS and CM, is strongly dependent on the number of observations in the sample. The graphical fit does not take into account, in an essential way, the number of observations. For completeness, in this subsection we offer some graphs corresponding to the cases studied. For France, we have taken the three samples and the best distribution obtained, the dPln. For Germany we also present the graphs of the three samples used and the best parametric model of the three studied, the nBC. For Italy, we take the sample of 2001 and the dPln. For Spain, we take the sample of 1981, in which the nBC is not rejected by the KS test.

For the French samples and the dPln, the fit at both tails could be improved, although the

Table 3.3: p-values (statistics) of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for French samples and the density functions used. Non-rejections are marked in bold

	lgn		dPln	
France	KS	CM	KS	CM
1990	0 (0.05)	0 (23.07)	**0.064 (0.0080)**	**0.133 (0.30)**
1999	0 (0.04)	0 (21.02)	**0.051 (0.0083)**	**0.105 (0.34)**
2009	0 (0.04)	0 (18.97)	**0.072 (0.0079)**	**0.075 (0.39)**
	nBC			
	KS	CM		
1990	0.002 (0.012)	0.006 (0.95)		
1999	0 (0.012)	0.005 (1.00)		
2009	0.026 (0.009)	0.016 (0.67)		

Table 3.4: Maximum log-likelihoods, AIC and BIC for French samples. The lowest values of AIC and BIC for each sample are marked in bold

	lgn			dPln		
France	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
1990	-284,762	569,529	569,546	-283,460	**566,928**	**566,962**
1999	-286,697	573,398	573,415	-285,487	**570,983**	**571,017**
2009	-290,472	580,947	580,964	-289,437	**578,882**	**578,916**
	nBC					
	log-likelihood	AIC	BIC			
1990	-283,582	567,171	567,196			
1999	-285,580	571,167	571,192			
2009	-289,494	578,993	579,019			

Table 3.5: Exact estimators for the lgn and the German samples. Estimators and 95% confidence intervals of the parameters of the dPln and the nBC for the German samples

Germany	lgn		dPln			
	μ	σ	α	β	μ	σ
1996	7.18	1.49	0.92±0.02	4.74±0.01	6.30±0.01	1.05±0.01
2006	7.43	1.50	1.18±0.03	4.11±0.01	6.82±0.01	1.21±0.01
2011	7.51	1.51	1.34±0.05	3.82±0.01	7.03±0.01	1.29±0.01
	nBC					
	μ	σ	λ			
1996	4.78±0.14	0.62±0.04	-0.12±0.01			
2006	5.61±0.19	0.84±0.06	-0.08±0.01			
2011	6.06±0.22	0.97±0.07	-0.06±0.01			

Table 3.6: p-values (statistics) of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for German samples and the used density functions

	lgn		dPln	
Germany	KS	CM	KS	CM
1996	0 (0.04)	0 (8.82)	0 (0.02)	0 (3.05)
2006	0 (0.03)	0 (3.00)	0 (0.02)	0 (1.51)
2011	0 (0.03)	0 (1.86)	0 (0.02)	0.004 (1.02)
	nBC			
	KS	CM		
1996	0 (0.02)	0 (1.82)		
2006	0.01 (0.01)	0.01 (0.75)		
2011	0.02 (0.015)	0.03 (0.56)		

Table 3.7: Maximum log-likelihoods, AIC and BIC for German samples. The lowest values of AIC and BIC for each sample are marked in bold

	lgn			dPln		
Germany	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
1996	-130,962	261,928	261,944	-130,697	261,402	261,432
2006	-113,895	227,795	227,810	-113,803	227,615	227,645
2011	-105,474	210,952	210,967	-105,426	210,860	210,889
	nBC					
	log-likelihood	AIC	BIC			
1996	-130,634	**261,275**	**261,297**			
2006	-113,775	**227,556**	**227,578**			
2011	-105,411	**210,828**	**210,850**			

Table 3.8: Exact estimators for the lgn and the Italian samples. Estimators and 95% confidence intervals of the parameters of the dPln and the nBC for the Italian samples

Italy	lgn		dPln			
	μ	σ	α	β	μ	σ
1901	7.79	0.92	1.68±0.10	3.04±0.49	7.52±0.05	0.61±0.04
1911	7.84	0.93	1.67±0.10	3.20±0.60	7.56±0.06	0.64±0.04
1921	7.84	0.96	1.62±0.10	3.49±0.72	7.51±0.06	0.68±0.04
1931	7.84	0.99	1.63±0.10	3.93±0.41	7.48±0.04	0.73±0.03
1936	7.84	1.01	1.63±0.10	3.59±0.77	7.51±0.06	0.74±0.04
1951	7.89	1.05	1.59±0.10	3.56±0.36	7.55±0.05	0.78±0.03
1961	7.85	1.10	1.47±0.08	3.84±0.16	7.43±0.04	0.82±0.02
1971	7.79	1.19	1.34±0.07	4.15±0.09	7.28±0.04	0.89±0.02
1981	7.79	1.25	1.36±0.06	4.36±0.02	7.29±0.03	0.98±0.01
1991	7.80	1.28	1.42±0.06	4.17±0.01	7.33±0.02	1.05±0.01
2001	7.80	1.31	1.50±0.06	4.10±0.01	7.38±0.01	1.10±0.00
2011	7.85	1.34	1.58±0.07	3.86±0.01	7.48±0.02	1.15±0.01
	nBC					
	μ	σ	λ			
1901	4.77±0.27	0.31±0.04	-0.14±0.02			
1911	4.77±0.27	0.31±0.04	-0.14±0.02			
1921	4.73±0.25	0.31±0.04	-0.14±0.02			
1931	4.86±0.26	0.35±0.04	-0.13±0.02			
1936	5.02±0.27	0.38±0.05	-0.12±0.02			
1951	5.17±0.27	0.42±0.05	-0.11±0.02			
1961	5.05±0.25	0.42±0.05	-0.12±0.01			
1971	5.13±0.24	0.48±0.05	-0.11±0.01			
1981	5.52±0.26	0.60±0.06	-0.09±0.01			
1991	5.93±0.28	0.72±0.08	-0.07±0.01			
2001	6.27±0.30	0.83±0.09	-0.06±0.01			
2011	6.58±0.32	0.94±0.10	-0.05±0.01			

Table 3.9: p-values (statistics) of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for Italian samples and the used density functions. Non-rejections are marked in bold

	lgn		dPln	
Italy	KS	CM	KS	CM
1901	0 (0.03)	0 (2.42)	**0.40 (0.0106)**	**0.34 (0.167)**
1911	0 (0.03)	0 (2.42)	**0.26 (0.0119)**	**0.42 (0.142)**
1921	0 (0.03)	0 (2.24)	**0.21 (0.0122)**	**0.34 (0.167)**
1931	0 (0.03)	0.02 (1.88)	**0.10 (0.0140)**	**0.29 (0.190)**
1936	0 (0.03)	0 (1.66)	**0.21 (0.0122)**	**0.30 (0.184)**
1951	0 (0.03)	0 (1.59)	**0.11 (0.0140)**	**0.18 (0.254)**
1961	0 (0.03)	0 (2.10)	**0.16 (0.0129)**	**0.20 (0.239)**
1971	0 (0.03)	0 (2.05)	**0.11 (0.0140)**	**0.43 (0.138)**
1981	0 (0.02)	0 (1.52)	**0.52 (0.0094)**	**0.84 (0.056)**
1991	0.002 (0.02)	0.006 (0.94)	**0.83 (0.0072)**	**0.94 (0.039)**
2001	0.005 (0.02)	0.008 (0.84)	**0.94 (0.0061)**	**0.99 (0.024)**
2011	**0.10 (0.014)**	**0.06 (0.43)**	**0.98 (0.0055)**	**0.95 (0.036)**
	nBC			
	KS	CM		
1901	0 (0.02)	0 (0.98)		
1911	0 (0.02)	0.01 (0.80)		
1921	0.01 (0.02)	0.02 (0.65)		
1931	0 (0.02)	0.02 (0.59)		
1936	0.01 (0.019)	0.01 (0.70)		
1951	0 (0.02)	0.01 (0.88)		
1961	0 (0.02)	0.01 (0.88)		
1971	0.02 (0.018)	0.03 (0.55)		
1981	**0.06 (0.015)**	**0.08 (0.38)**		
1991	**0.17 (0.013)**	**0.14 (0.29)**		
2001	**0.44 (0.010)**	**0.27 (0.20)**		
2011	**0.85 (0.007)**	**0.59 (0.10)**		

Table 3.10: Maximum log-likelihoods, AIC and BIC for Italian samples. The lowest values of the AIC and BIC for each sample are marked in bold

	lgn			dPln		
Italy	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
1901	-70,325	140,654	140,668	-70,148.4	**140,305**	**140,333**
1911	-70,871.9	141,748	141,762	-70,698.2	**141,404**	**141,432**
1921	-74,657.4	149,319	149,333	-74,474.5	**148,957**	**148,985**
1931	-74,918.2	149,840	149,854	-74,757.6	**149,523**	**149,551**
1936	-75,091.6	150,187	150,201	-74,942.3	**149,893**	**149,921**
1951	-75,830.9	151,666	151,680	-75,689.6	**151,387**	**151,415**
1961	-75,836.7	151,677	151,691	-75,675.3	**151,359**	**151,387**
1971	-75,951.9	151,908	151,922	-75,798	**151,604**	**151,632**
1981	-76,390.6	152,785	152,799	-76,284.1	**152,576**	**152,604**
1991	-76,653.1	153,310	153,324	-76,583.2	**153,174**	**153,202**
2001	-76,865.2	153,734	153,748	-76,818.1	**153,644**	**153,672**
2011	-77,390.1	154,784	154,798	-77,359.4	**154,727**	**154,755**
	nBC					
	log-likelihood	AIC	BIC			
1901	-70,201.5	140,409	140,430			
1911	-70,743.5	141,493	141,514			
1921	-74,511.5	149,029	149,050			
1931	-74,786	149,578	149,599			
1936	-74,973.1	149,952	149,973			
1951	-75,719.6	151,445	151,466			
1961	-75,702.3	151,411	151,432			
1971	-75,818	151,642	151,663			
1981	-76,297.1	152,600	152,621			
1991	-76,594.1	153,194	153,215			
2001	-76,827.6	153,661	153,682			
2011	-77,365.7	154,737	154,758			

Table 3.11: Exact estimators for the lgn and the Spanish samples. Estimators and 95% confidence intervals of the parameters of the dPln and the nBC for the Spanish samples

Spain	lgn		dPln			
	μ	σ	α	β	μ	σ
1900	6.97	1.06	1.16±0.03	7.71±0.01	6.23±0.01	0.66±0.00
1910	7.01	1.08	1.14±0.03	7.66±0.01	6.26±0.01	0.67±0.00
1920	7.03	1.11	1.11±0.03	7.38±0.02	6.26±0.01	0.69±0.01
1930	7.06	1.14	1.10±0.03	7.04±0.02	6.28±0.01	0.73±0.00
1940	7.06	1.18	1.06±0.03	6.75±0.01	6.26±0.01	0.75±0.00
1950	7.09	1.20	1.04±0.03	6.78±0.01	6.27±0.01	0.76±0.00
1960	7.03	1.27	0.97±0.03	6.43±0.01	6.16±0.01	0.80±0.00
1970	6.83	1.44	0.95±0.03	5.00±0.01	5.97±0.01	0.99±0.00
1981	6.63	1.62	0.9±0.03	4.28±0.01	5.75±0.01	1.18±0.00
1991	6.53	1.71	0.83±0.02	4.11±0.01	5.58±0.01	1.24±0.00
2001	6.54	1.75	0.77±0.02	4.18±0.01	5.48±0.01	1.22±0.00
2010	6.58	1.85	0.75±0.02	3.88±0.01	5.51±0.01	1.32±0.00
	nBC					
	μ	σ	λ			
1900	3.53±0.17	0.22±0.03	-0.22±0.02			
1910	3.56±0.17	0.22±0.03	-0.22±0.02			
1920	3.62±0.17	0.24±0.03	-0.21±0.02			
1930	3.73±0.18	0.26±0.03	-0.20±0.02			
1940	3.85±0.17	0.30±0.03	-0.19±0.02			
1950	3.79±0.17	0.29±0.03	-0.20±0.02			
1960	3.88±0.17	0.33±0.04	-0.19±0.02			
1970	4.38±0.17	0.55±0.05	-0.14±0.01			
1981	4.72±0.16	0.80±0.06	-0.10±0.01			
1991	4.69±0.15	0.86±0.06	-0.10±0.01			
2001	4.55±0.15	0.82±0.06	-0.11±0.01			
2010	4.70±0.15	0.92±0.07	-0.10±0.01			

Table 3.12: p-values (statistics) of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for Spanish samples and the used density functions. Non-rejections are marked in bold

	lgn		dPln	
Spain	KS	CM	KS	CM
1900	0 (0.06)	0 (7.13)	0 (0.03)	0 (1.46)
1910	0 (0.05)	0 (6.42)	0 (0.03)	0 (1.72)
1920	0 (0.06)	0 (7.23)	0 (0.03)	0 (1.76)
1930	0 (0.05)	0 (7.27)	0 (0.03)	0 (2.07)
1940	0 (0.05)	0 (6.75)	0 (0.03)	0 (1.94)
1950	0 (0.06)	0 (7.43)	0 (0.03)	0 (2.01)
1960	0 (0.06)	0 (7.15)	0 (0.03)	0 (2.38)
1970	0 (0.05)	0 (5.48)	0 (0.03)	0 (1.37)
1981	0 (0.05)	0 (4.51)	0.001 (0.02)	0.002 (1.14)
1991	0 (0.05)	0 (4.91)	0 (0.02)	0 (1.39)
2001	0 (0.05)	0 (6.21)	0 (0.03)	0 (2.20)
2010	0 (0.05)	0 (5.17)	0 (0.03)	0 (2.76)
	nBC			
	KS	CM		
1900	0 (0.02)	0.01 (0.85)		
1910	0 (0.02)	0 (1.23)		
1920	0 (0.02)	0.01 (0.91)		
1930	0 (0.02)	0 (1.51)		
1940	0 (0.02)	0 (1.20)		
1950	0 (0.02)	0 (1.23)		
1960	0 (0.02)	0 (1.34)		
1970	0 (0.02)	0.01 (0.72)		
1981	**0.06 (0.015)**	0.04 (0.50)		
1991	0.02 (0.017)	0.01 (0.82)		
2001	0.02 (0.022)	0 (1.29)		
2010	0 (0.03)	0 (1.82)		

Table 3.13: Maximum log-likelihoods, the AIC and BIC for Spanish samples. The lowest values of the AIC and BIC for each sample are marked in bold

	lgn			dPln			nBC		
Spain	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
1900	-65,873.6	131,751	131,765	-65,627.3	131,263	131,290	-65,579.8	**131,166**	**131,186**
1910	-66,413.5	132,831	132,845	-66,169.3	132,347	132,374	-66,119.1	**132,244**	**132,265**
1920	-66,762.6	133,529	133,543	-66,520.7	133,049	133,077	-66,468.5	**132,943**	**132,964**
1930	-67,782.4	135,569	135,583	-67,552.4	135,113	135,141	-67,496.8	**135,000**	**135,021**
1940	-68,291.6	136,587	136,601	-68,042.6	136,093	136,121	-68,003	**136,012**	**136,033**
1950	-68,656.2	137,316	137,330	-68,403.7	136,815	136,843	-68,350.5	**136,707**	**136,728**
1960	-68,762	137,528	137,542	-68,514.3	137,037	137,065	-68,458.6	**136,923**	**136,944**
1970	-68,529.4	137,063	137,077	-68,341.7	136,691	136,719	-68,304.5	**136,615**	**136,636**
1981	-68,568.1	137,140	137,154	-68,424.2	136,856	136,884	-68,398.3	**136,803**	**136,824**
1991	-68,592.2	137,188	137,202	-68,453.7	136,915	136,943	-68,416.8	**136,840**	**136,861**
2001	-68,833.3	137,671	137,685	-68,687.1	137,382	137,410	-68,629.9	**137,266**	**137,287**
2010	-69,911.2	139,826	139,840	-69,795.7	139,599	139,627	-69,729.8	**139,466**	**139,487**

densities apparently present a global close fit, with small discrepancies near the mode of the theoretical distributions.

For the German samples and the nBC, the lower tails present remarkable discrepancies which diminish over time. For the upper tails, there are slight discrepancies. The densities show remarkable discrepancies, visually apparent, which also diminish over time.

In the Italian case of 2001 and the dPln, we observe some slight discrepancies in the lower tail and the six biggest cities in the upper tail deviate slightly from the estimated parametric model. However, the fit of the densities is visually excellent.

For the Spanish sample of 1981 and the nBC, we observe close fits at the tails, but the overall densities present marked deviations.

In short, the graphical approximation in the selected cases by our formal criteria shows that improvements are possible, except maybe in the case of the sample of Italy (2001).

3.5 Conclusions

In this chapter, we have used population data corresponding to the lowest spatial subdivision of four European countries, France, Germany, Italy and Spain in different periods of the last and this century. We have used the data to study the parametric fit of three density functions: the lognormal (lgn) (Parr and Suzuki, 1973; Eeckhout, 2004), the double Pareto lognormal (dPln) (Reed, 2001, 2002, 2003; Reed and Jorgensen, 2004), and the normal-Box-Cox (nBC) (Schluter and Trede, 2013).

We have estimated the three density functions by maximum likelihood for all the samples and have performed Kolmogorov–Smirnov and Cramér–Von Mises tests. We have also studied the distributions according to the AIC and BIC.

The results show that the French city size distribution can be very well described by the dPln. In the German case, the dPln is outperformed by the nBC, but the latter is a parametric model which could be improved, since the statistical tests reject the nBC. In the Italian case, the dPln is again an excellent parametric model and, in the Spanish case, a similar situation to the German case occurs.

This variety of results, in which a clear possibility for the improvement of the parametric description of the respective city size distributions arises for two countries, leads us to think that there might exist another theoretical distribution that would be the best in (almost) all cases, improving all of the results obtained in this chapter. We leave this question open for future research.

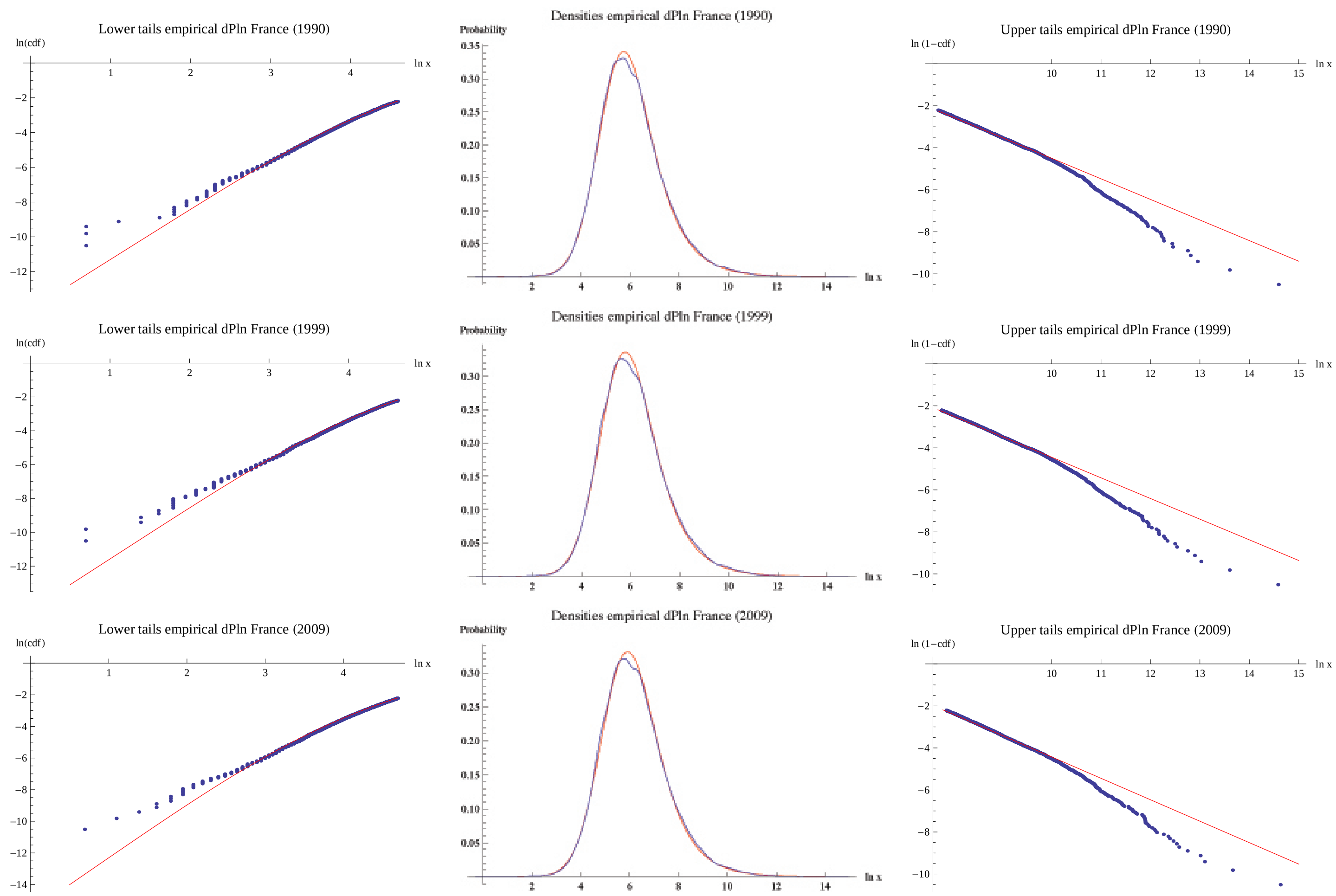

Figure 3.1: For the French city sizes in 1990, 1999 and 2009. Left-hand column: Empirical and estimated dPln ln(cdf) for the lower tail (empirical in blue, estimated in red). Center column: Empirical (Gaussian adaptive kernel density) and estimated dPln density functions (empirical in blue, estimated in red). Right-hand column: Empirical and estimated dPln $\ln(1 - \text{cdf})$ for the upper tail (empirical in blue, estimated in red)

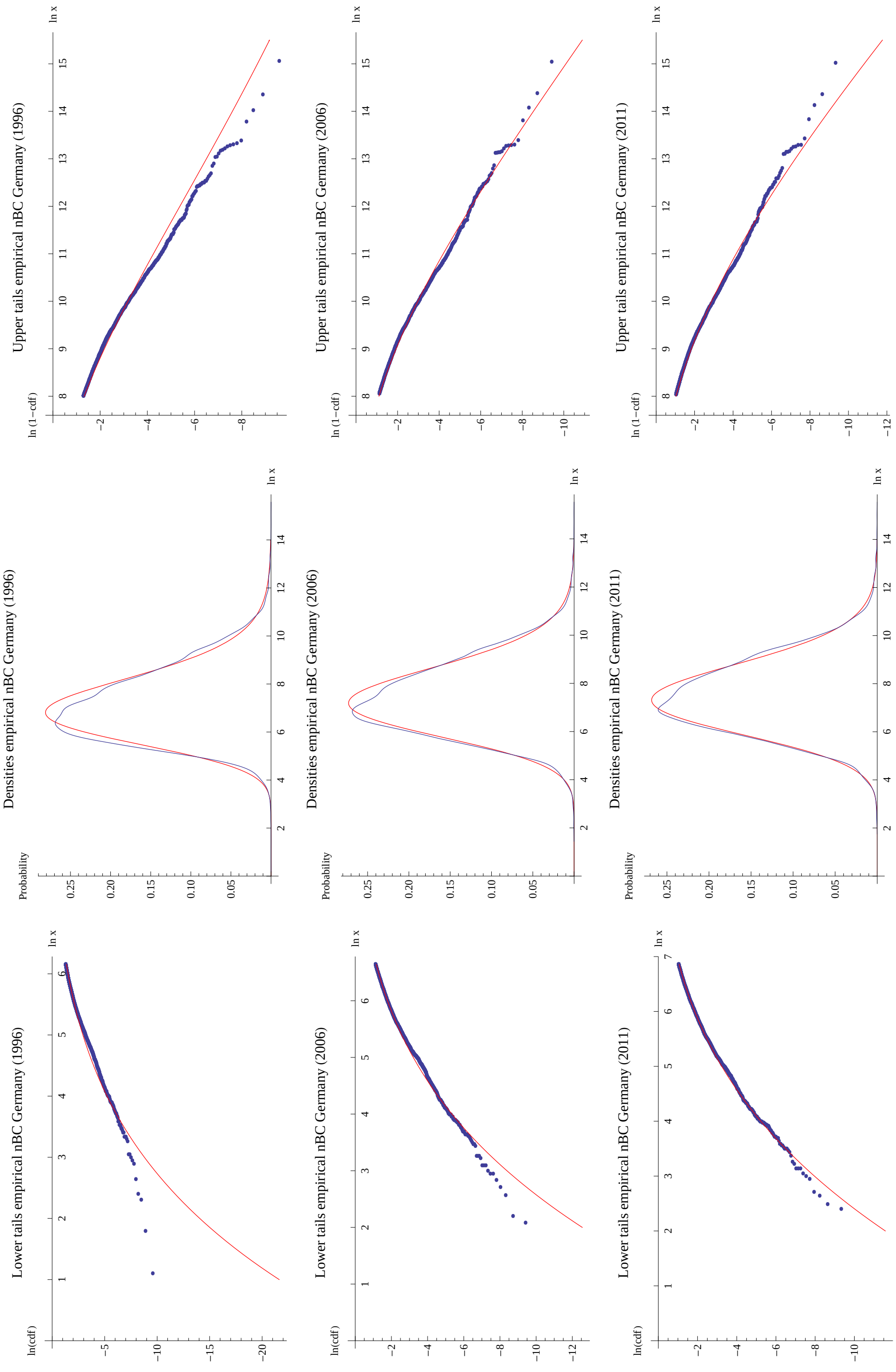

Figure 3.2: For the German city sizes in 1996, 2006 and 2011. Left-hand column: Empirical and estimated nBC ln(cdf) for the lower tail (empirical in blue, estimated in red). Center column: Empirical (Gaussian adaptive kernel density) and estimated nBC density functions (empirical in blue, estimated in red). Right-hand column: Empirical and estimated nBC $\ln(1 - \text{cdf})$ for the upper tail (empirical in blue, estimated in red)

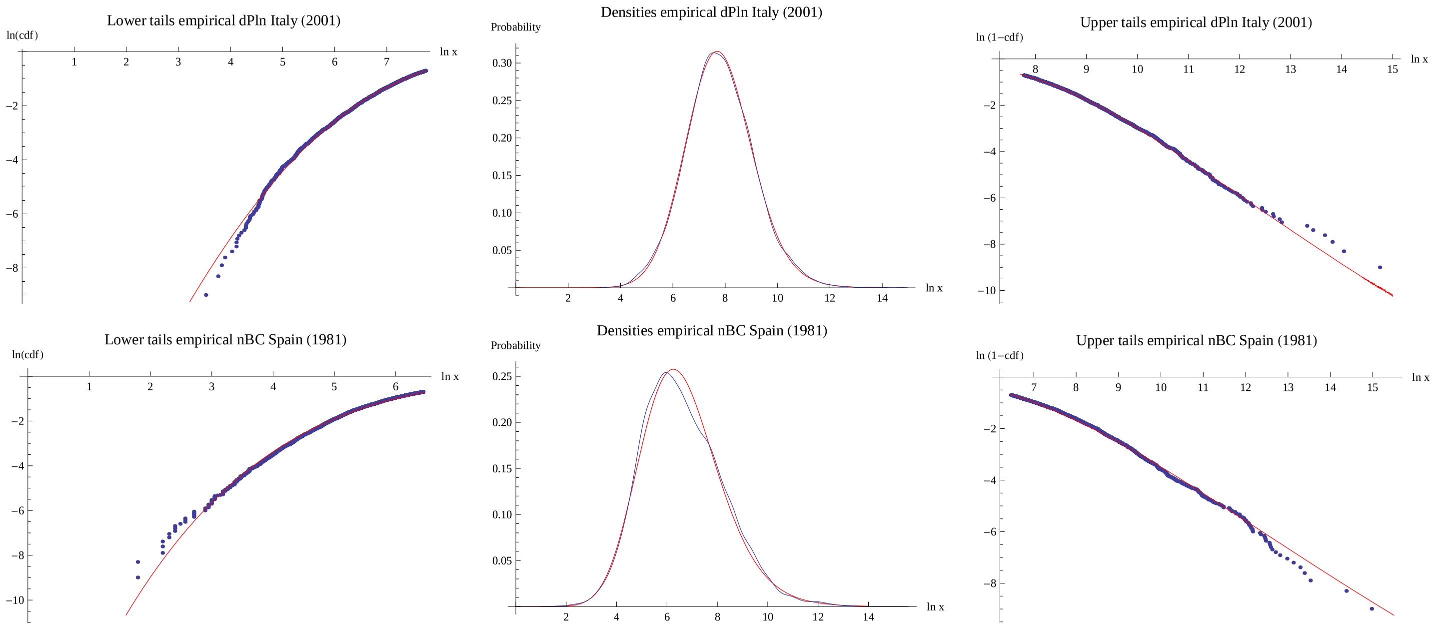

Figure 3.3: For the Italian and Spanish city sizes in 2001 and 1981, respectively. dPln for Italy 2001 and nBC for Spain 1981. Left-hand column: Empirical and estimated ln(cdf) for the lower tail (empirical in blue, estimated in red). Center column: Empirical (Gaussian adaptive kernel density) and estimated density functions (empirical in blue, estimated in red). Right-hand column: Empirical and estimated $\ln(1-\text{cdf})$ for the upper tail (empirical in blue, estimated in red)

References

Anderson, G. and Ge, Y. (2005). The size distribution of Chinese cities. *Regional Science and Urban Economics*, 35(6):756–776.

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

Bosker, M., Brakman, S., Garretsen, H., and Schramm, M. (2008). A century of shocks: The evolution of the German city size distribution 1925-1999. *Regional Science and Urban Economics*, 38(4):330–347.

Box, G. and Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26:211–252.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 94(5):1429–1451.

Giesen, K. and Suedekum, J. (2012). The French overall city size distribution. *Région et Développement*, 36:107–126.

Giesen, K. and Suedekum, J. (2014). City age and city size. *European Economic Review*, 71:193–208.

Giesen, K., Zimmermann, A., and Suedekum, J. (2010). The size distribution across all cities-double Pareto lognormal strikes. *Journal of Urban Economics*, 68(2):129–137.

González-Val, R., Lanaspa, L., and Sanz-Gracia, F. (2013a). Gibrat's law for cities, growth regressions and sample size. *Economics Letters*, 118:367–369.

González-Val, R., Ramos, A., and Sanz-Gracia, F. (2013b). The accuracy of graphs to describe size distributions. *Applied Economics Letters*, 20(17):1580–1585.

González-Val, R., Ramos, A., Sanz-Gracia, F., and Vera-Cabello, M. (2013). Size distributions for all cities: Which one is best? *Papers in Regional Science.* Forthcoming (doi: 10.1111/pirs.12037).

Ioannides, Y. and Overman, H. (2003). Zipf's law for cities: An empirical examination. *Regional Science and Urban Economics*, 33(2):127–137.

Ioannides, Y. and Skouras, S. (2013). US city size distribution: Robustly Pareto, but only in the tail. *Journal of Urban Economics*, 73:18–29.

Parr, J. and Suzuki, K. (1973). Settlement populations and the lognormal distribution. *Urban Studies*, 10(3):335–352.

Razali, N. and Wah, Y. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2:21–33.

Reed, W. (2001). The Pareto, Zipf and other power laws. *Economics Letters*, 74:15–19.

Reed, W. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science*, 42:1–17.

Reed, W. (2003). The Pareto law of incomes–an explanation and an extension. *Physica A*, 319:469–486.

Reed, W. and Jorgensen, M. (2004). The double Pareto-lognormal distribution–a new parametric model for size distributions. *Communications in Statistics-Theory and Methods*, 33(8):1733–1753.

Rozenfeld, H., Rybski, D., Andrade, J., Batty, M., Stanley, H., and Makse, H. (2008). Laws of population growth. *Proceedings of the National Academy of Sciences*, 105(48):18702–18707.

Rozenfeld, H., Rybski, D., Gabaix, X., and Makse, H. (2011). The area and population of cities: new insights from a different perspective on cities. *American Economic Review*, 101:2205–2225.

Schluter, C. and Trede, M. (2013). Gibrat, Zipf, Fisher and Tippet: City size and growth distributions reconsidered. Working Paper 27/2013, Center for Quantitative Economics WWU Münster.

Soo, K. (2005). Zipf's Law for cities: A cross-country investigation. *Regional Science and Urban Economics*, 35(3):239–263.

Zipf, G. (1949). *Human behavior and the principle of least effort.* Addison-Wesley Press.

Chapter 4

On the parametric description of French, German, Italian and Spanish city log-growth rate distributions

4.1 Introduction

Several studies have dealt with the theory of the growth process of cities. However, hardly any of the published papers deal with the study of the parametric description of the distribution of city growth rates. This is possibly due to the lack of good data sets in order to carry out the study until very recently. In the Chapter 3 of this thesis, we have used an ample database covering four big European countries, France (1990-2009), Germany (1996-2011), Italy (1901-2011) and Spain (1900-2010), with, generally, decennial intervals. Using these data sets, the computation of the growth rates is relatively easy so the study of their distribution follows naturally.

This research also has theoretical implications since Gibrat's process, as it is described in Sutton (1997) and references therein, Eeckhout (2004) and Delli Gatti et al. (2005), takes the log-growth rates to be normally distributed. If, empirically, this assumption happens not to hold and, moreover, alternative descriptions are found, then the foundations of Gibrat's process would deserve a reconsideration. In this chapter we succeed in parameterizing the distribution of log-growth rates with two well-known distributions, the α-stable distribution, see, e.g., Zolotarev (1986), Uchaikin and Zolotarev (1999), Nolan (2015) and references therein; and the (non-standardized) Student's-t distribution, see, e.g., Johnson et al. (1995) and references therein. The normal distribution, in turn, reveals itself as a poor description of the log-growth rates. The rest of the chapter is organized as follows. Section 4.2 shows the motivation for our study. Section 4.3 describes the databases. Section 4.4 introduces the parametric distributions used in this chapter. Section 4.5 describes some of the empirical results obtained. Finally, Section 4.6 offers some conclusions.

4.2 Motivation of our approach

Let us denote the population of the urban units under study as $x_{i,t}$, where i is an index for the urban units of a cross-sectional sample, and t is an index for time.[1]

As is well-known, *gross growth rates* are defined as

$$g_{i,t}^{g} = \frac{x_{i,t} - x_{i,t-1}}{x_{i,t-1}}, \tag{4.1}$$

so that $x_{i,t} = x_{i,t-1}(1 + g_{i,t}^{g})$. Taking the natural logarithm of this last expression we have

$$\ln x_{i,t} = \ln x_{i,t-1} + \ln(1 + g_{i,t}^{g})$$

If we define the also well-known *log-growth rates* as

$$g_{i,t}^{l} = \ln x_{i,t} - \ln x_{i,t-1} \tag{4.2}$$

then we can describe the *exact* relation:

$$g_{i,t}^{l} = \ln(1 + g_{i,t}^{g}) \tag{4.3}$$

Under the assumption that, for a given t, the $x_{i,t}$ are independent and identically distributed (i.i.d.) random variables, it is of interest to study the (parametric or nonparametric) distribution that they follow, based on empirical data. This has been treated extensively in the literature of city size distributions.

Likewise, under the assumption that, for a given t, the $g_{i,t}^{g}$ and $g_{i,t}^{l}$ are i.i.d. random variables, it is also of interest to study the parametric distribution they follow, based on empirical data. This is seldom studied in the field of Urban Economics, one exception being Schluter and Trede (2013). There are also non-parametric approaches to the study of city growth rates, see, e.g., González-Val et al. (2013, 2014) and references therein.

From definitions (4.1) and (4.2) and the fact that $x_{i,t} \geq 0$ for all i and $t \geq 0$, it is easy to show that

$$g_{i,t}^{g} \in (-1, \infty), \quad \forall i, t$$

and

$$g_{i,t}^{l} \in (-\infty, \infty), \quad \forall i, t$$

These intervals are, respectively, the supports of the probability density functions of $g_{i,t}^{g}$ and $g_{i,t}^{l}$. Let us fix time t and consider the probability density functions f_g of g^g (subindexes i, t dropped for notational simplicity) and f_l of g^l. Due to (4.3) and the fact that the values

[1] Generally, in our study, the unit of time will be a decade, except for some instances of Italian samples. This irregularity is caused by the participation of Italy in World War II.

of the probability density functions are equal for the corresponding values of the two types of the growth rates, the exact relation

$$f_g(g^g) = f_l(\ln(1+g^g))\frac{dg^l}{dg^g} = \frac{1}{1+g^g} f_l(\ln(1+g^g))$$

holds, so it suffices to study $f_l(g^l)$, which is what we will do in this chapter.

From a practical point of view, our interest is to obtain a very good parametric fit of the log-growth rate distributions. To do so, we have first tried several distributions well-known in the economics literature: the normal; the asymmetric exponential power (AEP) of Bottazzi and Secchi (2011), which generalizes the Laplace distribution of, e.g., Johnson et al. (1995), Stanley et al. (1996) and references therein; the α-stable distribution (see, e.g., Zolotarev (1986), Uchaikin and Zolotarev (1999) and references therein and Nolan (2015)); the generalized hyperbolic distribution (Barndorff-Nielsen (1977), Barndorff-Nielsen and Halgreen (1977), Barndorff-Nielsen and Stelzer (2005)); and the (non-standardized) Student's-t distribution, see, e.g., Johnson et al. (1995) and references therein. For the sake of brevity and comparison, we will present only the results of the normal (which is the distribution that arise from the proportionate growth process of cities, e.g., in Eeckhout (2004)) and the best or competing distributions for each case.[2]

4.3 The databases

In this chapter we use the same databases as in the Chapter 3 of this thesis, with one exception: for Germany, in order to use a generally decennial period, as in all of the other studied countries, we take the samples for 1996 and 2006, disregarding the one for 2011. There is another reason for this: the number and structure of German *Gemeinden* varies notably over time, and the construction of correlated data is very difficult.

The descriptive statistics of the data used can be seen in Table 4.1.

[2]The results for the distributions not presented here are available from the author upon request.

Table 4.1: Descriptive statistics of the samples of French, German, Italian and Spanish urban units used

	Urban units	Mean of log-growth rates	SD of log-growth rates	Minimum	Maximum
France					
1990-1999	36673	0.046	0.127	-1.386	1.786
1999-2009	36643	0.099	0.150	-2.060	2.692
Germany					
1996-2006	12309	0.007	0.112	-0.827	1.006
Italy					
1901-1911	7711	0.054	0.112	-0.957	1.890
1911-1921	7711	0.024	0.095	-0.802	0.683
1921-1931	8100	0.003	0.134	-0.933	3.447
1931-1936	8100	0.003	0.091	-0.704	2.488
1936-1951	8100	0.053	0.126	-0.531	2.324
1951-1961	8100	-0.047	0.161	-0.861	1.873
1961-1971	8100	-0.060	0.200	-1.075	2.234
1971-1981	8100	0.004	0.145	-0.900	1.108
1981-1991	8100	0.003	0.132	-3.098	3.835
1991-2001	8100	0.007	0.110	-1.384	1.366
2001-2011	8081	0.043	0.117	-0.580	3.303
Spain					
1900-1910	7800	0.047	0.117	-0.689	1.493
1910-1920	7806	0.012	0.126	-1.504	2.143
1920-1930	7812	0.034	0.143	-1.304	1.804
1930-1940	7875	0.008	0.144	-3.313	1.330
1940-1950	7896	0.023	0.127	-1.382	2.411
1950-1960	7901	-0.053	0.176	-1.360	1.580
1960-1970	7910	-0.204	0.311	-2.104	2.619
1970-1981	7956	-0.198	0.306	-2.416	2.396
1981-1991	8034	-0.102	0.235	-2.351	3.131
1991-2001	8077	0.007	0.238	-1.985	2.529
2001-2010	8074	0.038	0.244	-1.458	3.258

4.4 Description of the presented distributions

In this section we will introduce the distributions used in the chapter for the log-growth rates, denoted by g for simplicity.

4.4.1 Normal distribution

The well-known normal distribution is, simply,

$$f_{\mathrm{n}}(g,\mu,\sigma) \quad = \quad \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(g-\mu)^2}{2\sigma^2}\right)$$

where $\mu, \sigma > 0$ are respectively the mean and the standard deviation of g for the normal density f_{n}.

4.4.2 α-stable distribution

The general α-stable distribution is described, e.g., in Zolotarev (1986), Uchaikin and Zolotarev (1999) and Nolan (2015), this last reference being the one that we will follow for the parametrization chosen. This distribution does not admit a closed-form expression for the probability density function, but the characteristic function does. Thus we will consider that g follows an α-stable distribution with parameters $\alpha \in (0,2]$ (index of stability), $\beta \in [-1,1]$ (shape or skewness parameter), $\gamma \geq 0$ (scale parameter), $\delta \in \mathbb{R}$ (location parameter) if g is distributed with the characteristic function

$$E(\exp(iug)) \\ = \begin{cases} \exp(-\gamma^\alpha|u|^\alpha[1+i\beta(\tan\frac{\pi\alpha}{2})(\operatorname{sign}u)(|\gamma u|^{1-\alpha}-1)]+i\delta u) & \alpha \neq 1 \\ \exp(-\gamma|u|[1+i\beta\frac{2}{\pi}(\operatorname{sign}u)\log(\gamma|u|)]+i\delta u) & \alpha = 1 \end{cases}$$

The α-stable laws include the normal distribution ($\alpha = 2$, $\beta = 0$), the Cauchy distribution ($\alpha = 1$, $\beta = 0$) and the Lévy distribution ($\alpha = 1/2$, $\beta = 1$) as special cases, and they are known to be associated with Lévy processes, see the cited references for details.

4.4.3 Student's-t distribution

The (non-standardized) Student's-t distribution, see, e.g., Johnson et al. (1995) and references therein, is given by the following probability density function

$$f_{\mathrm{Stu}}(g,\mu,\sigma,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma}\left(1+\frac{1}{\nu}\left(\frac{g-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}$$

where $\mu \in \mathbb{R}$ (location parameter), $\sigma > 0$ (scale parameter), $\nu > 0$ is the parameter of degrees of freedom, and $\Gamma(\cdot)$ denotes the Gamma function. Particular cases of this distribution are the Cauchy distribution ($\nu = 1$) and the normal distribution ($\nu = \infty$).

4.5 Results

4.5.1 The log-growth data

We have calculated the log-growth rates between each two consecutive cross-sections of our data. In order to avoid infinite values, we have removed the observations for which at least one of the population values is zero.

4.5.2 Results for France

Maximum likelihood (ML) is a standard technique which allows the estimation of the parameters of a distribution, given a sample of data. For the case of the normal density function, the corresponding ML estimators can be found easily in an exact closed form (the μ and σ are then the mean and the standard deviation (SD) of the data). However, for the α-stable and Student's-t distributions used in this chapter one must resort to numerical optimization methods in order to find the ML estimators.[3]

Although we estimate the log-growth rate distribution of cities by the standard maximum log-likelihood, it should be noted that Schluter and Trede (2013) estimate their proposed Student's-t distribution by the Expectation-maximization algorithm, EM, (see, e.g., Bickel and Doksum (2001) and references therein). As mentioned, we will present only the results for the normal distribution and the best (competing) distributions.

In Table 4.2 we show the ML estimators of the distribution of the French samples studied. For the normal distribution, the ML estimators are exact and equal to the mean and standard deviation of the log-growth data. For the other two distributions (α-stable and Student's-t), we provide the ML estimators and 95% confidence intervals. The estimators show a slow variation of the values in the two samples, except for the increase in the mean of the normal and Student's-t distribution (μ) and the skewness parameter for the case of the α-stable distribution (β).

In Table 4.3 we show the results of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for the three distributions. These two tests are powerful when the sample size is high or very high, and non-rejections only occur if the deviations are really small. We see that the three distributions are rejected in all cases. According to these tests, there is no clear distribution that can be taken for the French communes.

To select one of the models, we show in Table 4.4 the results of the Akaike information criterion (AIC) and Bayesian information criterion (BIC), which are specially well suited to the maximum likelihood estimation that we have performed. These criteria favour the distribution with a greater maximum likelihood, but with a penalty for the number of

[3]For the α-stable distribution we have used the STABLE software of Robust Analysis Inc., see http://www.robustanalysis.com/ and for the Student's-t distribution we have used MATLAB in order to perform the ML estimations, as in Chapter 3.

parameters that have been used. The distribution with the lowest AIC / BIC is the one preferred.

We observe that the α-stable distribution obtains the lowest values for both criteria in the sample of 1999-2009 growth. For the 1990-1999 sample, the AIC selects the Student's-t distribution, and the BIC the α-stable. These results, yield that the most preferred distribution would be the α-stable, although it could be somehow outperformed by another distribution of distributions (as it suggests the outcome of the KS and CM tests).

4.5.3 Results for Germany

For the case of Germany we carry out the same analysis for the period considered, 1996-2006. Table 4.5 show the estimation results. More important, in Table 4.6 we present the results of the KS and CM tests. In this case, the Student's-t is the only not rejected for either test, meaning that this distribution presents a better fit for the sample.

The results of the AIC and BIC are presented in Table 4.7. In this case we obtain that the Student's-t is the preferred distribution for both criteria. We can conclude that the city size log-growth of *Gemeinden* can be described by the Student's-t distribution.

4.5.4 Results for Italy

In Table 4.8 we show the estimated parameters of the three distributions for the Italian samples. These estimations are quite precise as well.

In Table 4.9 we show the results of the KS and CM tests. In this case, for the KS test, neither the α-stable nor the Student's-t are rejected in 9 out of 11 samples (81.8%). For the CM test the α-stable is not rejected in 10 out of 11 samples (90.9%), while the Student's-t is not rejected in 8 out of 11 (72.7%). The normal distribution is rejected in all cases.

In Table 4.10 we show the values of the AIC and BIC information criteria. For the AIC, the α-stable distribution is preferred in 6 out of 11 samples (54.5%), with the Student's-t being the most preferred in the other cases. The BIC present other results, the Student's-t is the one preferred in 7 out of 11 samples (63.6%) and the α-stable in the other four.

In short, both distributions (the α-stable and the Student's-t) present a good fit in different periods, but there is not a clear-cut distinction of what it is the best one.

4.5.5 Results for Spain

In Table 4.11 we show the estimation results. In Table 4.12 we present the results of the KS and CM tests. The α-stable distribution is not rejected in 8 out of 11 samples for the KS test and 9 times for the CM test. The Student's-t distribution is not rejected in 8 out of 11 cases in both tests. Again, the normal distribution is rejected in all cases.

Finally, in Table 4.13 we show that the AIC selects the α-stable for 5 out of 11 cases, and the Student's-t for the other 6 samples. For the BIC, the α-stable is preferred only in 3 out of 11 cases, the Student's-t being preferred in the other 8.

With these results, we can see that both the α-stable and the Student's-t present a good fit for the samples, althought the Student's-t is the preferred one in most cases. There is not, however, a clear sign that either one of the functions present a better result in explaining the distribution of our samples.

Table 4.2: Exact estimators for the French samples. Estimators and 95% confidence intervals of the parameters

France	normal		α-stable			
	μ	σ	α	β	γ	δ
1990-1999	0.05	0.13	1.75±0.01	0.36±0.05	0.07±0.00	0.04±0.00
1999-2009	0.1	0.15	1.79±0.01	0.64±0.05	0.09±0.00	0.08±0.00
	Student's-t					
	μ	σ	ν			
1990-1999	0.04±0.00	0.09±0.00	4.47±0.21			
1999-2009	0.09±0.00	0.11±0.00	5.24±0.26			

Table 4.3: p-values (statistics) of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for French samples and the density functions used. Non-rejections are marked in bold

	normal		α-stable	
France	KS	CM	KS	CM
1990-1999	0 (0.0513)	0 (30.9433)	0 (0.0138)	0 (1.4806)
1999-2009	0 (0.0533)	0 (31.9016)	0.0043 (0.0107)	0.0064 (0.9137)
	Student's-t			
	KS	CM		
1990-1999	0 (0.0143)	0 (2.1683)		
1999-2009	0 (0.0234)	0 (4.4126)		

Table 4.4: Maximum log-likelihoods, AIC and BIC for French samples. The lowest values of AIC and BIC for each sample are marked in bold

	normal			α-stable		
France	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
1990-1999	23616.5	-47228.9	-47211.9	25819	-51630	**-51915.7**
1999-2009	17514.6	-35025.3	-35008.3	20439	**-40870**	**-40836**
	Student's-t					
	log-likelihood	AIC	BIC			
1990-1999	25876.5	**-51747.1**	-51721.6			
1999-2009	20166.6	-40327.2	-40301.6			

Table 4.5: Exact estimators for the German samples. Estimators and 95% confidence intervals of the parameters

Germany	normal		α-stable			
	μ	σ	α	β	γ	δ
1996-2006	0.01	0.11	1.77±0.03	0.07±0.09	0.07±0.00	0.01±0.00
	Student's-t					
	μ	σ	ν			
1996-2006	0.01±0.00	0.08±0.00	4.54±0.35			

Table 4.6: *p*-values (statistics) of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for German samples and the density functions used. Non-rejections are marked in bold

	normal		α-stable	
Germany	KS	CM	KS	CM
1996-2006	0 (0.0511)	0 (9.9794)	0.0091 (0.0157)	0.0229 (0.5991)
	Student's-t			
	KS	CM		
1996-2006	**0.147 (0.0109)**	**0.0793 (0.3847)**		

Table 4.7: Maximum log-likelihoods, AIC and BIC for Germany samples. The lowest values of AIC and BIC for each sample are marked in bold

	normal			α-stable		
Germany	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
1996-2006	9431.68	-18859.4	-18844.5	10186	-20363	-20333
	Student's-t					
	log-likelihood	AIC	BIC			
1996-2006	10202	**-20398**	**-20375.8**			

Table 4.8: Exact estimators for the Italian samples. Estimators and 95% confidence intervals of the parameters

Italy	normal		α-stable			
	μ	σ	α	β	γ	δ
1901-1911	0.05	0.11	1.85±0.03	0.00±0.17	0.07±0.00	0.05±0.00
1911-1921	0.02	0.09	1.78±0.03	-0.18±0.12	0.06±0.00	0.03±0.00
1921-1931	0.00	0.13	1.77±0.03	-0.45±0.11	0.08±0.00	0.01±0.00
1931-1936	0.00	0.09	1.65±0.03	-0.54±0.07	0.04±0.00	0.01±0.00
1936-1951	0.05	0.13	1.90±0.02	0.09±0.20	0.08±0.00	0.05±0.00
1951-1961	-0.05	0.16	1.89±0.02	0.75±0.14	0.10±0.00	-0.06±0.00
1961-1971	-0.06	0.20	1.83±0.03	0.60±0.12	0.12±0.00	-0.08±0.00
1971-1981	0.00	0.15	1.80±0.03	0.37±0.12	0.09±0.00	-0.01±0.00
1981-1991	0.00	0.13	1.68±0.03	0.07±0.09	0.06±0.00	0.00±0.00
1991-2001	0.01	0.11	1.76±0.03	-0.10±0.11	0.07±0.00	0.01±0.00
2001-2011	0.04	0.12	1.84±0.03	0.39±0.14	0.07±0.00	0.04±0.00
	Student's-t					
	μ	σ	ν			
1901-1911	0.05±0.00	0.09±0.00	5.90±0.71			
1911-1921	0.03±0.00	0.07±0.00	5.15±0.57			
1921-1931	0.01±0.00	0.10±0.00	5.01±0.52			
1931-1936	0.01±0.00	0.05±0.00	3.29±0.25			
1936-1951	0.05±0.00	0.10±0.00	6.82±0.90			
1951-1961	-0.05±0.00	0.13±0.00	6.65±0.82			
1961-1971	-0.07±0.00	0.16±0.00	5.46±0.58			
1971-1981	0.00±0.00	0.11±0.00	4.84±0.49			
1981-1991	0.00±0.00	0.08±0.00	3.53±0.29			
1991-2001	0.01±0.00	0.08±0.00	4.62±0.47			
2001-2011	0.04±0.00	0.09±0.00	5.71±0.67			

Table 4.9: p-values (statistics) of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for Italian samples and the density functions used. Non-rejections are marked in bold

	normal		α-stable	
Italy	KS	CM	KS	CM
1901-1911	0 (0.0355)	0 (4.166)	**0.2848 (0.0116)**	**0.3139 (0.1783)**
1911-1921	0 (0.0393)	0 (4.7116)	**0.3628 (0.0109)**	**0.5133 (0.1157)**
1921-1931	0 (0.0518)	0 (8.5967)	**0.1939 (0.0125)**	**0.2299 (0.2212)**
1931-1936	0 (0.097)	0 (32.7248)	**0.3155 (0.0111)**	**0.4832 (0.1231)**
1936-1951	0 (0.037)	0 (3.6067)	0.0137 (0.0182)	**0.0517 (0.456)**
1951-1961	0 (0.0381)	0 (3.3609)	**0.8117 (0.0073)**	**0.6819 (0.0817)**
1961-1971	0 (0.042)	0 (5.7044)	**0.3049 (0.0112)**	**0.5111 (0.1162)**
1971-1981	0 (0.0551)	0 (6.1134)	0.0074 (0.0193)	0.0147 (0.6848)
1981-1991	0 (0.089)	0 (24.5235)	**0.0903 (0.0144)**	**0.0761 (0.3919)**
1991-2001	0 (0.0501)	0 (7.2769)	**0.243 (0.0118)**	**0.181 (0.2559)**
2001-2011	0 (0.0829)	0 (20.0764)	**0.3251 (0.011)**	**0.3583 (0.1608)**
	Student's-t			
	KS	CM		
1901-1911	**0.6877 (0.0084)**	**0.7725 (0.0666)**		
1911-1921	**0.1145 (0.0141)**	**0.2356 (0.2178)**		
1921-1931	0.0177 (0.0177)	0.0098 (0.7513)		
1931-1936	0 (0.0273)	0 (1.4002)		
1936-1951	**0.4549 (0.0099)**	**0.4158 (0.1418)**		
1951-1961	**0.2335 (0.012)**	**0.1528 (0.2814)**		
1961-1971	**0.1295 (0.0135)**	0.0389 (0.5047)		
1971-1981	**0.2933 (0.0113)**	**0.1694 (0.2657)**		
1981-1991	**0.7262 (0.008)**	**0.5542 (0.1065)**		
1991-2001	**0.8177 (0.0073)**	**0.7576 (0.069)**		
2001-2011	**0.3204 (0.011)**	**0.2876 (0.1901)**		

Table 4.10: Maximum log-likelihoods, AIC and BIC for Italian samples. The lowest values of AIC and BIC for each sample are marked in bold

	normal			α-stable		
Italy	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
1901-1911	5959.71	-11915.4	-11901.5	6387	-12766	-12738
1911-1921	7235.13	-14466.3	-14452.4	7579.5	-15151	-15123
1921-1931	4814.06	-9624.13	-9610.13	5535	**-11062**	**-11034**
1931-1936	7915.57	-15827.1	-15813.1	9833	**-19658**	**-19630**
1936-1951	5275.53	-10547.1	-10533.1	5658.6	**-11309**	-11281
1951-1961	3321.99	-6639.99	-6625.99	3694.1	**-7380.2**	**-7352.2**
1961-1971	1523.75	-3043.5	-3029.5	2021	**-4033.9**	**-4005.9**
1971-1981	4123.95	-8243.9	-8229.9	4561	**-9114**	-9086
1981-1991	4935.42	-9866.83	-9852.83	6506.5	-13005	-12977
1991-2001	6358.46	-12712.9	-12698.9	6798.8	-13590	-13562
2001-2011	5854.65	-11705.3	-11691.3	6513.57	-13019.14	-12991.15
	Student's-t					
	log-likelihood	AIC	BIC			
1901-1911	6387.92	**-12769.8**	**-12749**			
1911-1921	7594.49	**-15183**	**-15162.1**			
1921-1931	5514.67	-11023.3	-11002.3			
1931-1936	9755.6	-19505.2	-19484.2			
1936-1951	5656.89	-11307.8	**-11286.8**			
1951-1961	3640.46	-7274.92	-7253.92			
1961-1971	1968.16	-3930.32	-3909.32			
1971-1981	4559.47	-9112.95	**-9091.95**			
1981-1991	6545.97	**-13085.9**	**-13064.9**			
1991-2001	6837.22	**-13668.4**	**-13647.4**			
2001-2011	6512.85	**-13019.7**	**-12998.7**			

Table 4.11: Exact estimators for the Spanish samples. Estimators and 95% confidence intervals of the parameters

Spain	normal		α-stable			
	μ	σ	α	β	γ	δ
1900-1910	0.05	0.12	1.71±0.03	0.18±0.1	0.07±0.00	0.04±0.00
1910-1920	0.01	0.13	1.78±0.03	0.44±0.11	0.07±0.00	0.00±0.00
1920-1930	0.03	0.14	1.72±0.03	0.18±0.1	0.08±0.00	0.03±0.00
1930-1940	0.01	0.14	1.73±0.03	-0.22±0.1	0.08±0.00	0.01±0.00
1940-1950	0.02	0.13	1.71±0.03	0.12±0.1	0.07±0.00	0.02±0.00
1950-1960	-0.05	0.18	1.65±0.03	0.49±0.08	0.09±0.00	-0.08±0.00
1960-1970	-0.2	0.31	1.72±0.03	0.18±0.1	0.18±0.00	-0.22±0.01
1970-1981	-0.2	0.31	1.74±0.03	-0.1±0.1	0.18±0.00	-0.19±0.01
1981-1991	-0.1	0.24	1.67±0.03	0.00±0.09	0.12±0.00	-0.11±0.00
1991-2001	0.01	0.24	1.53±0.03	0.46±0.06	0.11±0.00	-0.03±0.00
2001-2010	0.04	0.24	1.67±0.03	0.78±0.06	0.13±0.00	-0.01±0.01
	Student's-t					
	μ	σ	ν			
1900-1910	0.04±0.00	0.08±0.00	4.03±0.38			
1910-1920	0.01±0.00	0.09±0.00	4.48±0.45			
1920-1930	0.03±0.00	0.1±0.00	3.79±0.32			
1930-1940	0.01±0.00	0.1±0.00	3.93±0.34			
1940-1950	0.02±0.00	0.08±0.00	3.58±0.29			
1950-1960	-0.07±0.00	0.11±0.00	3.25±0.25			
1960-1970	-0.21±0.01	0.22±0.01	4.05±0.37			
1970-1981	-0.2±0.01	0.22±0.01	4.13±0.38			
1981-1991	-0.1±0.00	0.16±0.00	3.52±0.28			
1991-2001	-0.02±0.00	0.13±0.00	2.46±0.16			
2001-2010	0.01±0.00	0.17±0.00	3.71±0.31			

Table 4.12: p-values (statistics) of the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests for the Spanish samples and the density functions used. Non-rejections are marked in bold

	normal		α-stable	
Spain	KS	CM	KS	CM
1900-1910	0 (0.0581)	0 (9.6953)	0.0183 (0.018)	**0.0942 (0.3571)**
1910-1920	0 (0.0605)	0 (9.0233)	**0.0756 (0.015)**	**0.0581 (0.4365)**
1920-1930	0 (0.0705)	0 (13.1563)	**0.7758 (0.0077)**	**0.7111 (0.0767)**
1930-1940	0 (0.0653)	0 (15.2822)	**0.4971 (0.0097)**	**0.5132 (0.1157)**
1940-1950	0 (0.0761)	0 (18.9552)	**0.0642 (0.0153)**	**0.2498 (0.2095)**
1950-1960	0 (0.0858)	0 (19.342)	**0.2287 (0.0121)**	**0.2388 (0.2159)**
1960-1970	0 (0.0586)	0 (8.9549)	**0.4562 (0.01)**	**0.419 (0.1408)**
1970-1981	0 (0.0523)	0 (8.4837)	**0.484 (0.0097)**	**0.4694 (0.1266)**
1981-1991	0 (0.0762)	0 (14.9052)	0.0227 (0.0173)	0.0335 (0.5314)
1991-2001	0 (0.1198)	0 (35.2587)	0.0208 (0.0174)	0.0173 (0.6534)
2001-2010	0 (0.0829)	0 (20.0764)	**0.1317 (0.0135)**	**0.2085 (0.2352)**

	Student's-t	
	KS	CM
1900-1910	**0.0903 (0.0146)**	**0.1003 (0.3469)**
1910-1920	**0.1833 (0.0128)**	**0.2567 (0.2057)**
1920-1930	**0.3543 (0.0109)**	**0.4175 (0.1412)**
1930-1940	**0.1824 (0.0128)**	**0.1722 (0.2633)**
1940-1950	**0.6361 (0.0087)**	**0.641 (0.0891)**
1950-1960	0.0015 (0.0221)	0.0063 (0.9195)
1960-1970	**0.2147 (0.0123)**	**0.3708 (0.1564)**
1970-1981	**0.7003 (0.0082)**	**0.7489 (0.0704)**
1981-1991	**0.0561 (0.0155)**	**0.0654 (0.4171)**
1991-2001	0 (0.0287)	0 (1.7459)
2001-2010	0 (0.0336)	0 (2.3608)

Table 4.13: Maximum log-likelihoods, AIC and BIC for the Spanish samples. The lowest values of AIC and BIC for each sample are marked in bold

	normal			α-stable		
Spain	log-likelihood	AIC	BIC	log-likelihood	AIC	BIC
1900-1910	5637.68	-11271.4	-11257.4	6141.6	-12275	-12247
1910-1920	5085.31	-10166.6	-10152.7	5719.4	**-11431**	-11403
1920-1930	4129	-8254	-8240.08	5013.1	-10018	-9990.4
1930-1940	4114.54	-8225.07	-8211.13	5239.1	**-10470**	-10442
1940-1950	5068.87	-10133.7	-10119.8	6345.7	-12683	-12655
1950-1960	2517.51	-5031.01	-5017.06	3481	**-6954**	**-6926.1**
1960-1970	-1995.35	3994.69	4008.64	-1478.5	2964.9	2992.8
1970-1981	-1864.72	3733.43	3747.4	-1317	2642	2669.9
1981-1991	232.092	-460.184	-446.201	1102.6	-2197.2	-2169.2
1991-2001	143.037	-282.073	-268.079	1581.5	**-3154.9**	**-3126.9**
2001-2010	-80.254	164.508	178.501	913.081	**-1818.16**	**-1790.18**
	Student's-t					
	log-likelihood	AIC	BIC			
1900-1910	6178.04	**-12350.1**	**-12329.2**			
1910-1920	5715.74	-11425.5	**-11404.6**			
1920-1930	5017.05	**-10028.1**	**-10007.2**			
1930-1940	5237	-10468	**-10447.1**			
1940-1950	6347.02	**-12688**	**-12667.1**			
1950-1960	3439.6	-6873.2	-6852.28			
1960-1970	-1451.87	**2909.74**	**2930.67**			
1970-1981	-1292.45	**2590.89**	**2611.84**			
1981-1991	1128.86	**-2251.71**	**-2230.74**			
1991-2001	1534.93	-3063.87	-3042.88			
2001-2010	749	-1492	-1471.01			

4.5.6 An informal graphical approximation

For the purpose of analyzing the goodness of fit of the distributions, the graphical representation of the estimated density function along with the data has its own shortcomings. In short, the graphical fit can not take into account the number of observations, while the KS and CM tests are strongly dependent on this aspect of the data. Also, (see previous chapter), there is an amplification effect on the diferences between the estimated cdf's and the empirical data for both tails.

Although the graphical tools are, then, less suited for the purpose of the selection of a distribution function, we offer, for completeness, a representative case for every country, using the distribution that has a better fit with the data in that period: The α-stable for the case of France (1999-2009) and Spain (2001-2010). And the Student's-t for Germany (1996-2006) and Italy (1991-2001).

For the case of France and the Student's-t, the fit can be improved for both tails, in which the discrepancies are more important. The case of Germany is similar, and although the fit is somewhat better, there appear remarkable deviations at the ends of the tails.

For the case of Spain and Italy we have again something similar. The fit of the densities are nearly perfect visually, but the tails present important differences between the estimated function and the data.

This graphical approximation, in which the tails of the distributions are poorly fitted, shows that an improvement may be possible.

4.6 Conclusions

In the preceding subsections we have seen that a very appropriate parametric models for the log-growth rate distribution of France, Germany, Italy and Spain are the α-stable and Student's-t distributions. The normal, instead, offers a quite poor description of the empirical log-growth rates.

In any case, of the distributions studied, there is none which is clear-cut preferable, but instead the best ones are close competitors. Furthermore, none of them is preferable at all instances.

This leads us to think that there might exist one or more distributions which would outperform the ones analyzed in this chapter, based if possible on theoretical arguments.

We hope to address this issue in future research.

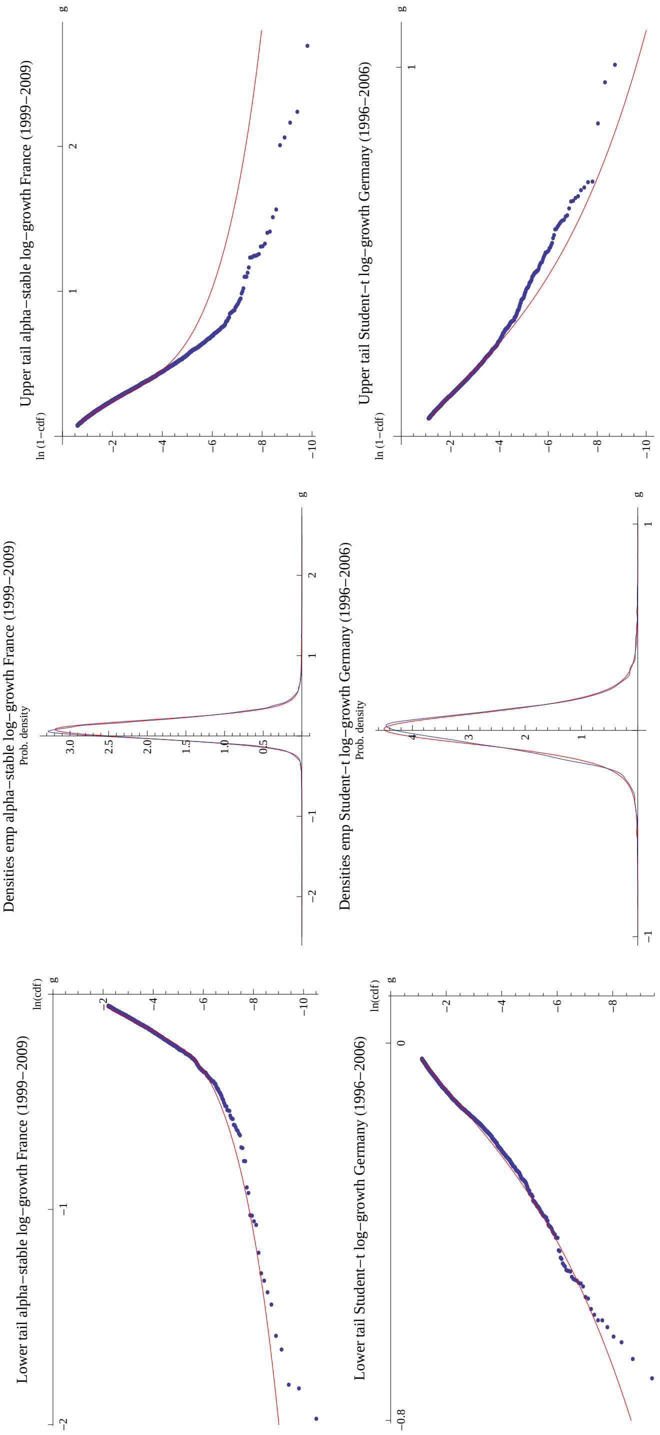

Figure 4.1: For the French (1999-2009) and German (1996-2006) log-growth of city size. Left-hand column: Empirical and estimated ln(cdf) of α-stable and Student's-t for the lower tails (empirical in blue, estimated in red). Center column: Empirical (Gaussian adaptive kernel density) and estimated α-stable and Student's-t density functions (empirical in blue, estimated in red). Right-hand column: Empirical and estimated $\ln(1-\text{cdf})$ of α-stable and Student's-t for the upper tails (empirical in blue, estimated in red)

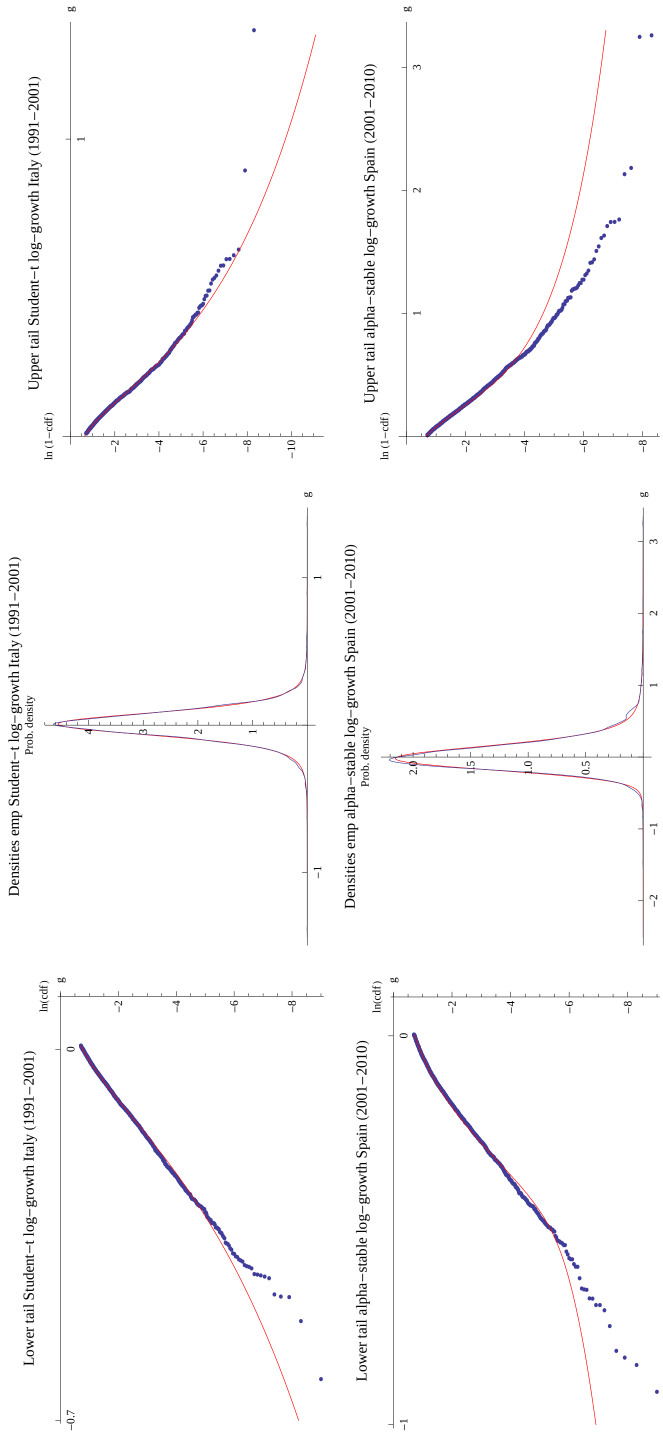

Figure 4.2: For the Italian (1991-2001) and Spanish (2001-2010) log-growth of city size. Left-hand column: Empirical and estimated ln(cdf) of Student's-t and α-stable for the lower tails (empirical in blue, estimated in red). Center column: Empirical (Gaussian adaptive kernel density) and estimated Student's-t and α-stable density functions (empirical in blue, estimated in red). Right-hand column: Empirical and estimated $\ln(1-\text{cdf})$ of Student's-t and α-stable for the upper tails (empirical in blue, estimated in red)

References

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 353(1674):401–419.

Barndorff-Nielsen, O. and Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 38:309–311.

Barndorff-Nielsen, O. and Stelzer, R. (2005). Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse Gaussian Lévy processes. *Scandinavian Journal of Statistics*, 32:617–637.

Bickel, P. and Doksum, K. (2001). *Mathematical statistics (second edition). Volume I.* Prentice Hall.

Bottazzi, G. and Secchi, A. (2011). A new class of asymmetric exponential power densities with applications to economics and finance. *Industrial and Corporate Change*, 20(4):991–1030.

Delli Gatti, D., Di Guilmi, C., Gaffeo, E., Giulioni, G., Gallegati, M., and Palestrini, A. (2005). A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic Behavior & Organization*, 56:489–512.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 94(5):1429–1451.

González-Val, R., Lanaspa, L., and Sanz-Gracia, F. (2013). Gibrat's law for cities, growth regressions and sample size. *Economics Letters*, 118:367–369.

González-Val, R., Lanaspa, L., and Sanz-Gracia, F. (2014). New evidence on Gibrat's Law for cities. *Urban Studies*, 51(1):93–115.

Johnson, N., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions. Volume 2.* John Wiley & Sons.

Nolan, J. (2015). *Stable distributions-models for heavy tailed data.* Forthcoming. Birkhäuser.

Schluter, C. and Trede, M. (2013). Gibrat, Zipf, Fisher and Tippet: City size and growth distributions reconsidered. Working Paper 27/2013, Center for Quantitative Economics WWU Münster.

Stanley, M., Amaral, L., Buldyrev, S., Havlin, S., Leschhorn, H., Maass, P., Salinger, M., and Stanley, H. (1996). Scaling behaviour in the growth of companies. *Nature*, 379:804–806.

Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature*, 35:40–59.

Uchaikin, V. and Zolotarev, V. (1999). *Chance and stability. Stable distributions and their applications.* VSP.

Zolotarev, V. (1986). *One-dimensional stable distributions.* American Mathematical Society.

Conclusions

The first two chapters of the present PhD Thesis define two new models within the field of the New Economic Geography.

The first chapter analyzes the spatial distribution of heterogeneous firms in a theoretical model based on the Footloose Entrepreneur Model, that allows us to study not only the localization of economic activity but also the different dynamics derived by the competition between firms. From the model we can obtain several conclusions: First, the most productive firms will be concentrated in one region, the core. Second, an increase in the freeness of trade will decrease the market crowding effect and will increase the home market effect, fostering the creation of small firms in the core. Third, the presence of a higher heterogeneity among firms increases the concentration of small firms, although this process will take place in the core (periphery) if the value of the freeness of trade parameter is higher (lower) than a certain threshold. Fourth, a greater heterogeneity also implies a more concentrated industrial economic activity in the region where the most productive firms are located, measured by employment, revenues or profits. Fifth, with a sufficiently high freeness of trade, either the core is fully specialized in the industrial sector or the periphery is completely specialized in the agricultural sector. Sixth, the reduction of transportation costs increases the welfare of all agents in both regions. And seventh, the most productive workers will have a superior utility (because of their income level) and the workers of the core will have a higher utility than the workers of the periphery (thanks to a greater access to cheap commodities), except in the case of free trade, in which they will have the same level of welfare.

The second chapter develops an alternative framework to the Iceberg approach for the analysis of the transportation costs in which the explanation of the frictions imposed by the space is not due to the loss of a certain percentage of the quantity of the goods being traded. In our case, we introduce an explicit cost that is assumed by the firm, who will demand a certain number of workers to carry out the transportation process. This allows us to model the transportation costs in a more suitable correspondence to the empirical evidence: Increasing in a concave manner with respect to the distance, and with transport costs per ton-kilometer decreasing with respect to the total amount of goods being shipped. Also, our theoretical approach does not maintain the unitary elasticity between the domestic and foreign price, a very unrealistic outcome, and is more convincing in the consideration of non homogeneous space. Finally, the theoretical determination of the freeness of trade parameter changes too, and will no longer depend only on the total cost of transportation, but on the ratio between the cost of transportation and the cost of production; in this way, even an industry with a more costly process of transportation can have a higher freeness of trade if its production costs are relatively higher.

The next two chapters analyze the city size distribution and the city growth rate distribution for four European countries.

In the third chapter, we check the correspondence of the city size distribution with the lognormal, the double Pareto lognormal and the normal-Box-Cox, estimating them, by maximum likelihood (ML), for France, Germany, Italy and Spain. Having computed the Kolmogorov–Smirnov (KS) and Cramér–Von Mises (CM) tests and the performance of the distributions according to the Akaike and Bayesian information criteria, we can deduce that there is mixed evidence as to which function better fits the data. The double Pareto lognormal describes the French city size distribution accurately. In the case of Germany, this function is outperformed by the normal-Box-Cox, although the latter could be improved, since the statistical tests reject it. For the Italian case, the double Pareto lognormal can be used as an excellent theoretical density function, as in France. Spain shows that the normal-Box-Cox is preferable, but is rejected by the statistical tests, as in the German case.

In the fourth chapter, we carry out a similar analysis but with respect to the population growth of the urban units for the same four countries. In this case we use the normal, the α-stable and the Student's-t functions, to fit the distribution of the growth rates of city sizes. We conclude that the case of France can be explained with the α-stable, and Germany with the Student's-t. For Italy and Spain, there is not, however, a dominant function. In all cases, the normal is the least preferable of the three.

With these results we can conclude that there is no clear theoretical explanation of the distribution of population across all cities, at least, for the four countries analyzed. Future research should be done in this regard for a complete understanding of the city size distribution and the corresponding log-growth.

Conclusiones

Los dos primeros capítulos de la presente Tesis definen dos nuevos modelos pertenecientes al campo de la Nueva Geografía Económica.

El primero de ellos analiza la distribución espacial de empresas heterogéneas en un modelo teórico basado en el *Footloose Entrepreneur Model*, que nos permite estudiar no solo la localización de la actividad económica, sino también los diferentes efectos derivados de la competición entre empresas. A partir de este modelo podemos obtener varias conclusiones: Primera, la empresas más productivas están concentradas en una región, el centro (*core*). Segunda, un incremento en el parámetro de libertad de comercio disminuye el *market crowding effect* y aumenta el *home market effect*, fomentando la concentración de empresas pequeñas en el centro. Tercera, la presencia de una mayor heterogeneidad entre las empresas favorece la concentración de empresas pequeñas, aunque este proceso tendrá lugar en el centro (periferia) si el valor del parámetro de libertad de comercio es mayor (menor) que un cierto umbral. Cuarta, una mayor heterogeneidad también implica una mayor concentración de la actividad económica industrial en la región donde las empresas más productivas se han localizado, medida tanto por empleo, ingresos o beneficios. Quinta, con una libertad de comercio lo suficientemente alta, o bien el centro se especializa totalmente en el sector industrial o bien la periferia se especializa completamente en el sector agrario. Sexta, la reducción de los costes de transporte aumenta el bienestar de todos los agentes económicos, tanto en el centro como en la periferia. Y séptima, los trabajadores más productivos disfrutan de un bienestar mayor (derivado de su mayor nivel de renta) y los trabajadores del centro tienen un mayor nivel de bienestar que los trabajadores de la periferia (gracias al mejor acceso a bienes más baratos), excepto en el caso en el que haya total libertad de comercio, en el que el nivel de felicidad es el mismo para todos.

El segundo capítulo desarrolla una alternativa al marco teórico conocido como Iceberg para el análisis de los costes de transporte, en donde la explicación de la fricción impuesta por el espacio no se debe a la pérdida en un cierto porcentaje de la cantidad de bienes enviados. En nuestro caso, introducimos un coste explícito que es asumido por la empresa, que demanda una cierta cantidad de trabajadores para que lleven a cabo las tareas del proceso de transporte. Esto nos permite modelizar los costes de transporte con una mayor correspondencia con la evidencia empírica: Los costes aumentan con respecto a la distancia de forma cóncava y los costes por kilo y kilómetro con respecto a la cantidad total de bienes enviados son decrecientes. Además, nuestro marco teórico no mantiene la elasticidad unitaria entre el precio doméstico y externo, un resultado muy poco realista y, a su vez, es más convincente a la hora de considerar un espacio no homogéneo. Finalmente, la determinación teórica del parámetro de libertad de comercio cambia también, y no depende únicamente del coste total de transporte, sino del ratio entre el coste de transporte y el coste de producción; de esta forma, incluso una industria con un proceso de transporte más costoso puede tener una mayor libertad de comercio si sus costes de producción son relativamente más altos.

Los dos capítulos siguientes analizan la distribución del tamaño de las ciudades y la distribución de sus tasas de crecimiento para cuatro países europeos. En el tercero, estudiamos la correspondencia de la distribución urbana con la función lognormal, la doble Pareto lognormal y la normal-Box-Cox, estimándolas por el procedimiento de máxima verosimilitud para Francia, Alemania, Italia y España. Tras computar los contrastes de Kolmogorov-Smirnov y Cramér-Von Mises y de acuerdo con el criterio de Akaike y el criterio Bayesiano, podemos deducir que hay una evidencia mixta sobre qué función es la que mejor se ajusta a los datos. La doble Pareto lognormal describe la distribución del tamaño de las ciudades francesas con cierta exactitud. Para el caso de Alemania, esta función es superada por la normal-Box-Cox, aunque esta podría ser mejorada, ya que los contrastes estadísticos la rechazan. Para Italia, la doble Pareto lognormal puede usarse como una excelente función teórica, como en el caso de Francia. En España es preferible la normal-Box-Cox, pero también es rechazada por los contrastes estadísticos.

En el cuarto capítulo llevamos a cabo un análisis similar, esta vez con respecto al crecimiento poblacional de las unidades urbanas, para los mismos cuatro países. En este caso, usamos la función normal, la *alpha-stable* y la t de Student para ajustar la distribución de las tasas de crecimiento de los tamaños de las ciudades. Concluimos que el caso de Francia puede ser explicado con la función *alpha-stable* y Alemania con la t de Student. Para Italia y España no hay, sin embargo, ninguna función dominante. Para todos los países la función normal es la menos preferida de las tres. Con estos resultados podemos concluir que no hay una clara explicación teórica sobre la distribución de la población urbana sobre todas las ciudades, al menos para los cuatro países analizados. En el futuro hay que profundizar en esta línea de investigación, de forma que nos permita una mayor comprensión de la distribución urbana y su correspondiente crecimiento.