Universidade Federal de Pernambuco
Centro de Ciências Sociais Aplicadas

Programa de Pós-Graduação em Economia
PIMES – UFPE

Distribution, Growth and Spatial Interactions: An Analysis of Brazilian Population Dynamics During the Period 1970-2010

Diego Firmino Costa da Silva

TESE DE DOUTORADO

Recife
31 de janeiro de 2014

Universidade Federal de Pernambuco
Centro de Ciências Sociais Aplicadas

Programa de Pós-Graduação em Economia

PIMES – UFPE

Diego Firmino Costa da Silva

Distribution, Growth and Spatial Interactions: An Analysis of Brazilian Population Dynamics During the Period 1970-2010

Trabalho apresentado ao Programa de Pós-graduação em Economia do Departamento de Economia da Universidade Federal de Pernambuco como requisito para obtenção do grau de Doutor em Economia.

Orientador: Raul da Mota Silveira Neto

Recife

31de janeiro de 2014

Catalogação na Fonte
Bibliotecária Ângela de Fátima Correia Simões, CRB4-773

S586d Silva, Diego Firmino Costa da
Distribution, growth and spatial interactions: an analysis of brazilian population dynamics during the period 1970-2010 / Diego Firmino Costa da Silva. - Recife : O Autor, 2014.
70 folhas : il. 30 cm.

Orientador: Prof. Dr. Raul da Mota Silveira Neto.
Tese (Doutorado em Economia) – Universidade Federal de Pernambuco, CCSA, 2014.
Inclui referências e apêndices.

1. Distribuição populacional. 2. Crescimento populacional. 3. Lei de Zipf. I. Silveira Neto, Raul da Mota (Orientador). II. Título.

330 CDD (22.ed.) UFPE (CSA 2014 – 046)

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE ECONOMIA
PIMES/PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

PARECER DA COMISSÃO EXAMINADORA DE DEFESA DE TESE DE DOUTORADO EM ECONOMIA DE:

DIEGO FIRMINO COSTA DA SILVA

A Comissão Examinadora composta pelos professors abaixo, sob a presidência do primeiro, considera o Candidato Diego Firmino Costa da Silva **APROVADO.**

Recife, 13/03/2014

Prof. Dr. Raul da Mota Silveira Neto
Orientador

Prof. Dr. Tatiane Almeida de Menezes
Examinador Interno

Prof. Dr. Carlos Roberto Azzoni
Examinador Externo/USP

Prof. Dr. Ciro Biderman
Examinador Externo/FGV-SP

Prof. Dr. André Luis Squarize Chagas
Examinador Externo/USP

Agradecimentos

Agradeço, primeiramente, aos meus pais, João e Angela, ao meu irmão, Bruno, juntamente à toda minha família, por terem me apoiado e incentivado incondicionalmente.

Agradeço à minha namorada, Ana Luiza, e à sua família pela demonstração de estarem sempre na torcida ao meu favor.

Agradeço ao meu orientador Professor Raul Silveira pela orientação e apoio neste trabalho e no doutorado como um todo.

Agradeço ao professor Paul Elhorst pela excelente oportunidade e experiência de estudar durante alguns meses na Universidade de Groningen, Holanda.

Agradeço aos meus amigos e colegas do PIMES-UFPE. Entre eles destaco Igor Ézio, Janaína Ratis, Lucas Varjão, Renata Caldas, Diogo Baerlocher, Henrique Veras, Rodrigo Arruda, Flavius Sodré, Luiza Campos, Sammara Cavalcanti, Thiago Bananeira, Leandro Coimbra, Poema Isis, André Melo e Deborah Seabra por terem sempre boa vontade na hora de ajudar e passar seus conhecimentos, assim como pelos momentos de descontração.

Agradeço a todos os meus amigos pelas amenidades que, sem dúvida, contribuiram de maneira bastante positiva nos últimos quatro anos.

Acknowledgment

Thank, first, to my parents, John and Angela, my brother, Bruno, and all my family for their unconditional encouragement.

Thanks also to my girlfriend, Ana Luiza, and his family.

I thank my advisor professor Raul Silveira for guidance and support in this work and the PhD as a whole.

I am grateful to Professor Paul Elhorst for the excellent opportunity and experience of studying at the University of Groningen, Netherlands.

I thank my friends and colleagues at the Federal University of Pernambuco / Department of Economics: Igor Ézio, Janaína Ratis, Lucas Varjão, Renata Caldas, Diogo Baerlocher, Henrique Veras, Rodrigo Arruda, Flavius Sodré, Luiza Campos, Sammara Cavalcanti, Thiago Bananeira, Leandro Coimbra, Poema Isis, André Melo e Deborah Seabra for having always willing to help and share their knowledge as well as the moments of casualness.

I thank all my friends for the amenities that undoubtedly contributed in a very positive way in the last four years.

Abstract

This thesis focuses on the Brazilian population dynamics between 1970 and 2010. In this sense, the first objective is to explore the behavior of the Brazilian population distribution, revisiting the traditional rank-size rule and Markov chain approaches. In order to bring up more accurate information on the dynamics and evolution of the population distribution, the spatial dependence is introduced through the analysis of LISA Markov and Spatial Markov Chains. The distribution shape may indicate that divergence in population size of Minimum Comparable Areas (MCAs) is decreasing. The Zipf's law estimation indicates that the population distribution is, every decade, moving away from Pareto law. Markov chain approach brings as main evidence the high-persistence of MCAs to stay in their own class size from one decade to another over the whole period, and different spatial contexts have different effects on transition for regions. The second and main objective of the present thesis is to model population growth dynamics of Brazilian MCAs in order to assess the determinants of population growth of these units between 1970 and 2010 and examine the existence and magnitude of spatial interaction and spatial spillover effects associated with these determinants. In this sense, the population growth model developed by Glaeser et al. (1995) and Glaeser (2008) is extended to include spatial interaction effects, and it is tested by estimating a dynamic spatial panel model with spatial and time period fixed effects and by comparing the performance of a wide range of potential neighborhood matrices using Bayesian posterior model probabilities. Six of the thirteen determinants of population growth considered in this thesis turn out to produce significant spatial interaction effects. This implies that a change of such a variable in one unit, also significantly affects population growth in other units, a phenomena that in most of the previous studies on population growth has been ignored.

Keywords: Population distribution, Zipf's Law, Markov Chains, Population growth, Spatial dependence, Dynamic Spatial Panel models, Spillover effects

Resumo

Esta tese tem como foco principal a dinâmica populacional brasileira entre 1970 e 2010. Neste sentido, o primeiro objetivo é explorar o comportamento da distribuição populacional, utilizando tanto a abordagem tradicional de *rank* quanto as cadeias de Markov. A fim de obter informações mais precisas sobre a dinâmica e a evolução da distribuição populacional, a dependência espacial é introduzida através da análise de LISA Markov e *Spatial Markov Chains.* O formato da distribuição indica que a divergência no tamanho populacional das Áreas Mínimas Comparáveis (AMC) é decrescente. A estimação da lei de Zipf traz evidências de que a distribuição populacional está, a cada década, de distanciando da distribuição de Pareto. A abordagem utilizando as cadeias de Markov traz como principais evidências a alta persistência das AMCs permanecerem nas suas classes iniciais com o passar das décadas e o fenômeno que diferentes contextos espaciais tem efeitos diferentes sobre a transição das localidades. O segundo e principal objetivo da tese é modelar a dinâmica do crescimento populacional das AMCs brasileiras a fim de avaliar os determinantes do crescimento populacional destas unidades entre 1970 e 2010, bem como examinar a existência e magnitude da interação espacial e dos efeitos de *spillovers* espaciais associados a estes determinantes. Neste sentido, o modelo de crescimento populacional desenvolvido por Glaeser et al (1995) e Glaeser (2008) é ampliado para incluir efeitos de interações espaciais. Este modelo é, então, testado empiricamente através da estimação de um modelo espacial dinâmico com dados em painel incluindo efeitos fixos e comparando a performance de uma ampla gama de matrizes de vizinhança através de modelos Bayesianos de probabilidade posterior. Seis dos treze determinantes do crescimento populacional considerados nesta tese apresentaram efeitos de interação espacial significantes. Isto implica que uma mudança em uma destas variáveis de uma unidade também afeta significantemente o crescimento populacional nas unidades vizinhas, um efeito que tem sido ignorado na maiorias dos estudos anteriores a este.

Palavras-Chave: Distribuição Populacional, lei de Zipf, Cadeias de Markov, Crescimento Populacional, Modelos espaciais dinâmicos em dados de painel, *Spillovers Effects*

Contents

1. Introduction 1
2. Distribution Dynamics of Population and Spatial Dependence 5
2.1 Literature Review 5
2.2 Zipf's Law 8
2.3 Markov Chains and Spatial Dependence...9
3. A Spatial Economic Model of Population Growth Dynamics in Brazilian MCAs.............17
3.1 Literature Review 17
3.2 A Spatial Extension of Glaeser's Population Growth Model..19
3.3 Econometric Methodology Spatial Panel Data Models..25
4. Results 31
4.1 The Distribution Dynamics of Brazilian MCAs: Results...31
4.1.1 Data Implementation 31
4.1.2 Evolution of Brazilian MCAs size distribution 32
4.1.3 The Rank Size Rule for Brazilian MCAs 36
4.1.4 Brazilian Population Distributional Dynamics..37
4.2 Spatial Evaluation of Population Growth Dynamics of Brazilian MCAs: Results........48
4.2.1 Data Implementation 48
4.2.2 Population Growth of Brazilian MCAs 51
5. Conclusions 61
References 64
Appendix 69

1. Introduction

The Brazilian urbanization process is surely one of the most significant and robust social phenomena in the country in the past four decades. Throughout all economic and social changes of the last decades, the percentage of people living in urban centers in Brazil has been steadily growing; according to the last Brazilian Demographic Census of the year 2010, it increased from 55.9% in 1970 to 84.4% in 2010 (IBGE, 2011). Several studies have shown that this process has been driven by improved economic and social prospects in cities (Yap, 1976; Henderson, 1988; Da Mata et al., 2007). In addition, Ramalho and Silveira-Neto (2012) have pointed out that the most important migration flux of people in Brazil since the 1990s occured between cities (urban-urban). In other words, apart from demographic factors, the most important sources of population growth dynamics of Brazilian cities in more recent decades are associated to specific urban characteristics.

Population growth of cities is also associated with the population movement between municipalities. To illustrate, according to data from the Brazilian Institute of Geography and Statistics (IBGE) in 1970 there were 66 cities in Brazil with more than 100 000 inhabitants, 105 cities in the early 1980s, 398 in 2000, and this number increased to 488 in 2010. Meanwhile, in the last decade the number of cities with over 1 million inhabitants went only from 13 to 14. In this sense, the concern is not only for the scale of urbanization, but also for the distribution of population across the urban hierarchy that becomes a challenge for policy makers to establish strategies for cities of different sizes. This observation raises some questions: How cities of different sizes grow during the process of development and transformation of a country? Is the degree of cities-size mobility slow or fast in the last 40 years? Are the movements within the distribution affected by spatial dependence?

Regarding this context, some authors investigated the behavior of cities's size distribution (Dobkins and Ioannides, 1999; Black and Henderson, 2003; Gabaix and Ioannides, 2004; Gallo and Chasco, 2007). About the Brazilian case, there are few studies on the behaviour of population distribution. Oliveira (2004a), when analyzing the evolution of city-size distribution in Brazil between 1936 and 2000, found evidence that smaller cities grew less than large ones until the 1990s. Trindade and Sartoris (2009) examined the evolution of size distribution of cities in Brazil between the 1920-2000 period and the results show evidence of divergence, similarly to Oliveira (2004a). Justo (2012) finds evidence of low interclass mobility and high persistence in the population distribution behavior of 431 minimum comparable areas between 1910 and 2010. Moro and Santos (2013) also found low mobility for the period of 1970-2010, but they only as sample the municipalities that existed in 1970, not covering all Brazilian territory.

For studies on the characterization and evolution of population distribution, the national literature available does not include information in their databases beyond the year 2000 (e.g.: Oliveira, 2004a; Trindade and Sartoris, 2009), do not go further in the investigation of spatial effects (e.g.: Justo, 2012) or do not cover all Brazilian territory (e.g.: Moro and Santos, 2013). In this sense, the first focus of this thesis is to assess the behavior of the population size distribution of Brazilian Minimum Comparable Areas (MCA) covering all Brazilian territory between 1970 and 2010. [1] Furthermore, the intention of the first part of the study is to advance in providing more accurate evidences, taking into account the possibility of spatial dependence in population size distribution by using current spatial techniques. In analyzing the evolution of Brazilian minimum comparable areas (MCA) population distribution, we begin by revisiting the traditional rank-size rule and Pareto distribution approaches. Therefore, in order to bring up information on the dynamics and evolution of the population distribution, we lead with the estimation of transition probability matrices associated with discrete Markov chains (Kemeny and Snell, 1976). And then, the spatial dependence is introduced through the analysis of LISA Markov and Spatial Markov Chains, both developed by Rey (2001).

Much less is known, however, about the specific factors conditioning population growth in Brazilian cities. Among the few studies available, Henderson (1988) has shown that population growth of Brazilian cities between 1960 and 1970 was positively associated with initial levels of schooling of the population. More recently, Da Mata et al. (2007) have analyzed the growth of cities between 1970 and 2000 and have provided evidence that favorable supply and demand conditions measured by market potential variables, better schooling and decreasing opportunities in the agricultural sector favored the growth of Brazilian cities. One obvious limitation of these studies is that they do not analyze the more recent period of 2000-2010. Firstly, after a decade of high inflation, this was a period of price stability. Secondly, it was also a period characterized by income convergence among Brazilian states (Silveira-Neto and Azzoni, 2012). Finally, the great increase in the production of commodities and agricultural goods during this period had a positive impact on the number of opportunities in towns that are distant from large urban centers.

[1] A MCA is a municipality or an aggregation of municipalities necessary to enable consistent spatial analyses over time. More details follow in section 4.2.

Besides this limitation, in their study, Da Mata et al. (2007) selected only those municipalities with over 75,000 inhabitants, which corresponds to 75% of Brazilian urban population. As has now been largely recognized (Boarnet et al., 2005), spatial dependence is much more severe for small spatial units such as municipalities. Recently, by analyzing income dynamics at different levels of spatial aggregation units, Resende et al. (2013) confirmed the importance of spatial dependence for the case of Brazilian MCAs.

These findings are not surprising since the Brazilian Constitution of 1988 states that the municipality is the third and lowest political administrative unit of the country, with autonomy for collecting taxes on services activities (ISS: Imposto sobre Serviços) and on urban real estate property (IPTU: Imposto Predial e Territorial Urbano) and for legislating land use. All these policies can potentially affect the location of firms and people and, hence, may generate reactions from or affect neighboring cities (Brueckner, 2008). There is a scarcity of studies related to population growth with a high level of geographical disaggregation, since most studies available in Brazilian literature are based on geographic units representing large aggregates, including Vergolino and Jatoba (2000) and Da Mata et al. (2007) using micro-regions and urban agglomerations, respectively. By using these large aggregates, the researcher runs the risk of being negligent with potential spatial effects, for example the effects of tax competition between municipalities to attract new ventures. Some of the studies that attempt to analyze population growth defend the use of smaller geographic units, mainly because these are the most open economies allowing greater movement of capital, labor and ideas (Glaeser et al., 1995), better reflecting the spatial heterogeneity unlike large aggregates such as countries, states and macro regions (Beenson et al., 2001) and greater relevance for planning and policy-making (Chi, 2009). Furthermore and more generally, spatial technological spillovers (Ertur and Koch, 2007) are potentially more prevalent among small urban neighboring centers than among large ones. By considering spatial urban centers at a lower and fuller scale, there is also the possibility that common or shared specific amenities will affect population welfare. All of these factors may cause spatial dependence on population growth dynamics of Brazilian cities and its determinants.

Therefore, in order to adequately analyze population growth dynamics of all Brazilian cities and identify the factors conditioning such dynamics, it appears fundamental to model spatial dependence among these spatial units explicitly. Thus, the second and central objective of the present thesis is to model population growth dynamics of Brazilian Minimum Comparable Areas (MCA) in order to assess the determinants of population growth of these units between 1970 and 2010 and to examine the existence and magnitude of spatial interaction and spatial spillover effects associated with these determinants. Specifically, we motivate the empirical investigation by presenting a spatial extension of the city population growth model developed by Glaeser (2008). This spatial extension accounts for spatial interaction effects among both productivity and city amenities and implies an empirical specification for population growth dynamics consisting of spatial interaction effects in the dependent and independent variables. The empirical strategy of modeling population growth dynamics of Brazilian cities (MCA) consists of a dynamic spatial panel model with controls for spatial and time specific effects. This strategy offers the opportunity to disentangle the magnitude and significance levels of spatial spillover effects both in the short and in the long term, and guarantees that any evidence in favor of these effects is not due to ignoring time-specific effects that areas have in common.

This dissertation is organized in correlated chapters and is not composed by independent papers. In addition to these introductory notes, in the chapter 2 we present the foundations for distributions dynamics of population and spatial dependence through the literature review and presentation of Zipf's law and Markov Chain approaches. Chapter 3 builds the foundations for the population growth dynamics. In this chapter, besides the literature review, we set out our theoretical spatial extension of Glaeser's population growth model as well as we present the econometric methodology behind our dynamic spatial panel data model. Chapter 4 describes the dataset and reports the results of the empirical analysis regarding distributional dynamics and population growth, respectively. The conclusions are presented in Chapter 5.

2. Distribution Dynamics of Population and Spatial Dependence

In this chapter, we present the literature review and methodologies used to analyze the behavior of the MCAs's population distribution covering the entire Brazilian territory in the period 1970-2010. The first methodology deals with the traditional estimation of Zipf's Law, which verifies that the distribution of city sizes follows the Pareto distribution. This technique only provides some information about static distribution at each point in time. As suggested by Duranton (2006), even though it may be only a rough first approximation, Zipf's law nonetheless remains a useful benchmark to think about the distribution of city sizes. Therefore, in order to bring up information on the dynamics and evolution of the population distribution, i.e., to access information on the movements of the localities within the distribution, techniques based on the Markov Chain will be explored in the remaining topics in this section.

2.1 Literature Review

Based on the integration of spatial statistic with modeling using Markov chains, Rey (2001) studies the evolution of the regional distribution of income taking into account the transitions of both the individual economies and those of their respective geographic neighbours within a distribution of income. Using data from 48 U.S. states for the period 1929-1994, the main result found by the author is that the rates of upward or downward mobility of the states within the distribution was sensitive to the position of its neighbour in the same distribution. And a possible implication in terms of policy is that, for example, a policy to reduce regional disparities could be more effective when the receiving state of the policy is surrounded by less disparate states.

Black and Henderson (2003) examine the evolution of the US city size distribution by applying Zipf 's Law to the 20th century US city size distribution. They then turn to a more general approach to analyze the evolution and trends of the size distribution of cities by modeling the transition process of cities directly. In relation to the Zipf's Law, the Pareto parameter estimated for the whole sample lending modest support to the view of increasing urban concentration in recent decades. For the top one-third of cities, the rise in the Pareto parameter would suggest decreasing urban concentration in the US over time. But the fact is that this difference between the estimated parameters suggests that the relationship between rank and city size is not log-linear. In relation to city size distribution, the authors conclude that existing cities tend to move up the size distribution 'fairly quickly', but to move down extremely slowly. Additionally, there is some tendency in the USA towards increasing urban concentration, with a greater proportion of cities in large relative size categories.

Gallo and Chasco (2007) analyzed the evolution of population growth for a group of 722 Spanish municipalities during the period 1900-2001. Attentive to the fact that the omission of spatial autocorrelation could cause a bias to the OLS estimator, the authors followed the strategy suggested by Anselin (1988) and estimated a spatial model SUR for Zipf's Law's (size distribution of cities follows a Pareto Law) and two main phases are found: divergence (1900-1980) and convergence (1980-2001). Furthermore, the authors estimated transition matrices associated with discrete Markov chains to obtain information concerning the movements of the urban groups within the population distribution. In this case, the results indicated that the municipalities located on the ends of the distribution would be more persistent in staying in those positions in the ranking, while medium-sized cities were more likely to move into smaller categories. The authors, however, do not explore the approach suggested by Rey (2001) who proposes the use of a spatial Markov matrix.

The international literature that refers to the behavior of population distribution is much wider than the one discussed in this thesis. Dobkins and Ioannides (1999) using the data from U.S. Census and cover metropolitan areas between 1900 and 1990, the authors found evidences of divergent growth, if spatial evolution is ignored, and convergent growth in the presence of very significant regional effects. Lalanne (2013) investigate the hierarchical structure of the Canadian urban system. Some papers discuss theoretical issues on city size distributions (Gabaix, 1999; Gabaix and Ioannides, 2003; Duranton, 2006; Gan et al., 2006). The diversity of international studies is very large, but this is not the case of the national literature. Below we list some work featured in national literature regarding distributional population behavior.

Oliveira (2004a), when studying the evolution of size distribution in Brazilian cities and testing the validity of Zipf's Law, estimates Pareto coefficient for Brazil between 1936 and 2000. The obtained results do not allow the conclusion that the rule of order and size applies to Brazil. Only in 1960 and 1970 this rule is true, but represents a transition period, since the coefficient decreased constantly over the period studied. This reduction represents an increase in inequality in the size of Brazilian cities. In this study, the author does not take into account the spatial factors that could influence the results.

Trindade and Sartoris (2009) examine the evolution of the relationship between the size of Brazilian cities and their population distribution for the period between 1920 and 2000. Using models based on Zipf's Law, Markov chain and taking spatial effects into account, the authors find in their results that there is a persistent population concentration in a small number of areas. As Gallo and Chasco (2007), the authors do not use the approach of spatial Markov matrix proposed by Rey (2001), which would have clearer information about the spatial relation existing in size distribution of cities.

Monastério (2009) analyzes the changes in the spatial distribution of population and manufacturing employment in Brazil between 1872 and 1920. To this end, the used tools, which combine the spatial analysis techniques Exploratory Spatial Data Analysis (ESDA) of the Markov chain, as, suggested by Rey (2001). The sample consists of minimum comparable areas in the Northeast, Southeast and South, and the state of Goiás. That is, the other states of the Midwest and the northern region are outside the sample used. The analysis revealed differences in the trajectories of the areas within states, the role of space in the dynamics and the tendency to increase in concentration during the studied period, especially with regard to manufacturing occupation. The analysis using the Markov matrix spatially conditioned indicated that the neighborhood was essential to the destinations of AMCs. Localities with little dense neighbors tended to approach the low-density profile of its neighbors.

Justo (2012) seeks to identify the dynamics of population growth for a group of 431 minimum comparable areas in Brazil between 1910 and 2010. For this, the author estimated spatial models for Zipf's law achieving results that point to the divergence, which has been losing strength in recent decades. Furthermore, through the estimation of functions of non-parametric densities, the author attempts to characterize the population distribution and through a process stationary first order Markov Chain shown the growth process of Brazilian cities. The results point to a low interclass mobility and high persistence. The probability of remaining cities on the class itself between a decade and another over the last hundred years is high. Despite being a very recent paper, as Trindade and Sartoris (2009), the author does not use the approach of spatial Markov matrix proposed by Rey (2001).

Moro and Santos (2013) test the Zipf’s Law in order to describe the spatial distribution of the Brazilian cities and Markov Chains analysis to examine the dynamics of the cities within the urban system. Additionally, the authors introduce spatial dependence in both Zipf’s law estimation and Markov. To estimate the Zipf’s law equation, they used the full sample of municipalities between 1970 and 2010. The results point that the Pareto coefficient is much smaller than 1, featuring a polarized and asymmetric urban structure. Regarding the spatial Markov approach, the results show strong evidence that the probability of urban growth of a municipality depends on the surrounding urban context, and there is a low mobility for the period 1970-2010. However, in the Markov chain analysis, they use as sampling only municipalities that existed in 1970, not covering all Brazilian territory in the following decades. In this way, territory and population of new municipalities (created from the subdivision of former municipalities) will be excluded from the sample, skewing the results with selection bias.

2.2 Zipf's Law

The evolution of the size distribution of cities is explored through the law of Zipf, or rank-size rule. Zipf (1949) stated that the size distribution of cities follows a Pareto law (Pareto, 1896) by claiming that:

$$R = a \cdot S^{-b} \qquad (1)$$

where R is the classification order of the city in the size distribution of population, S is the city's population, a and b are parameters, b is the Pareto exponent. Formally, the size distribution of cities depends on the value of b parameter. In the limit, if b tends to infinity, all cities will have the same size. The smaller the value of b, the greater the inequality in the size distribution of cities.

In terms of the Pareto distribution, this means that the probability of city size be greater than some S is proportional to $1/S$: $P(Size > S) = \alpha / S^b$, the statement of Zipf's Law implies a Pareto exponent of unity, $b=1$. According to this law, populations of cities within any group of cities at any point in time are inversely proportional to the ranking of their populations in this group. According to Gabaix (1999), one proposed explanation for Zipf's Law is if cities grow randomly, with the same expected growth rate and the same standart deviation, the limit distribution will converge to Zipf's law.

At this moment to is interesting to point the differences between Zipf's law and rank-size rule. Using Gabaix (1999) words, Zipf's law states that the probability that a city has a size greater than S decreases as *1/S*. The rank-size rule states that we should expect the size S_i of a city of rank *I* to follow a power law: the size of the city of rank *I* varies as *1/i* , and the ratio of the second largest city to the largest city should be ½, the ratio of city 3 and city 2, 2/3, and so on. These size ratios are often used to compare actual urban patterns with "ideal" (Zipf) patterns. In fact, even if Zipf's law is verified exactly, the rank-size rule will be verified only approximately, if our probabilistic interpretation of Zipf's law is correct.

The *b* parameter can be interpreted as an indication of inequatility. More precisely, the high value of b represents a greater possibility of mobility. That is, as the inequality in the localities's size is small, the possibility of mobility is higher in the rank. Greater dispersion of population among the cities implies increasing convergence of cities's sizes and a greater number of cities with a population close to average size (the smaller the size variance). Empirically, the logarithms are taken on both sides of equation (1) and the linear expression for each city each year is estimated:

$$\ln R_{it} = \ln a_t - b_t \cdot \ln S_{it} + \varepsilon_{it} \qquad (2)$$

According to Rosen and Resnick (1980), beyond the question of how closely city-size data obey the rank-size rule is the more fundamental question of how well these populations fit the general Pareto distribution. To test for non-Pareto behaviour, the authors suggest the addition of a non-linear term to basic logarithmic version (2), giving,

$$\ln R_{it} = \ln a_t - b_t \ln S_{it} + c_t (\ln S_{it})^2 + \varepsilon_{it} \qquad (3)$$

The authors state that more interesting than the significance of these extra terms is the direction of curvature which they indicate. Then, $c_t > 0$ indicating upward concavity and $c_t < 0$ downward concavity. The upward concavity means that the ranking variation rate increases with city size. According to Rosen and Resnick (1980), this may reflect scale economies. In many countries, the highest-ranking cities, taking advantage of scale economies, have grown most quickly.

Unfortunately, it is not possible to have information on the dynamics of the distribution only from the estimation of equations (2) and (3). According to Gallo and Chasco (2007), Zipf's law allows the characterization of the overall evolution of the size distribution of cities, but gives no information on the movements of the cities within the distribution. It is not to possible to answer, for example, why is it that some cities are present in certain positions of the distribution over time. Another limitation of Zipf's law to study the population distribution of cities is that in addition to not realize movements within the distribution, it does not take into account the possibility that these movements are affected by spatial dependence. To clarify these issues, in the next topics we lead with the estimation of transition probability matrices associated with discrete Markov chains (Kemeny and Snell, 1976), which will make it possible to follow the progress of each group of cities of a certain size in time. And then, the spatial dependence is introduced through the analysis of LISA Markov and Spatial Markov Chains, both developed by Rey (2001).

2.3 Markov Chains and Spatial Dependence

The study of distributive population dynamics according to the position of the cities and the trend configuration of population distribution over time is a method aimed at describing the law of motion driving the evolution of the distribution as a Markovian stochastic process. Once estimated the motion up or down probabilities in the population hierarchy during a transition period of a given length, the law is used to calculate a limiting population distribution characterizing a stochastic steady-state income distribution to which the system converges over time. Through modeling of the transition process of the minimum comparable areas directly, we can examine the evolution and trends in the MCAs size distribution. Compared to continuous stochastic kernels, for example, one of advantages of using this method listed by Gallo and Chasco (2007) is that discrete probability distribution and transition matrices are easier to interpret: various descriptive indices and the long-run or ergodic distribution are easier to compute. The Zipf's law, as density functions, allows the characterization of the evolution of the global distribution, but Gallo and Chasco do not provide information about the movements of the localities within the distribution. Specifically, they do not say if the locations that were in a region of the distribution at the beginning remain or not in the same region of the distribution at the end of period.

We denote F_t as the distribution of the cross-section population of the minimum comparable areas at time *t* related to an average in the country. Defining a set of *K* different size classes, we discretize the population distribution in K relevant classes. To proceed with the estimation, first we need to assume that the distribution frequency follows a first order stationary process of Markov. This assumption requires transition probabilities, p_{ij}, of order 1, which means independence of the classes at the beginning periods (*t-2, t-3, ...*). If the order is higher, the transition matrix will not be clearly specified. That is, we only have part of the necessary information to describe the true evolution of the population distribution. Following this assumption, the evolution of a size distribution is represented by a transition probability matrix, *M*, in which each element (*i, j*) indicates which is the probability that a city in class *i* at time *t* will be in the class *j* in the following period. The (*K, 1*) vector indicating the frequency of cities in each size class at time *t*, is described by:

$$F_{t+1} = MF_t \tag{4}$$

where M is the matrix of transition probability ($k \times k$) representing the transition between two distributions:

$$M = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{bmatrix} \tag{5}$$

where each element $p_{ij} \geq 0$ represents the probability that the cities of a particular size class *k* at time *t-1* will be in the class *j* at time *t* and $\sum_{j=1}^{K} p_{ij} = 1$.

The elements of the matrix *M* can be estimated by the frequency of changes from a size class to another. According to Amemiya (1985) or Hamilton (1994), the maximum likelihood estimator of p_{ij} is:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i} \tag{6}$$

where n_{ij} is the total number of cities moving from class *i* in the decade *t-1* to *j* in decade *t* and n_i is the total number of municipalities which remains in *i* for all *T-1* transitions.

If the transition probabilities are stationary, in other words, if the probabilities of two classes do not vary over time, then:

$$F_{t+s} = F_t M^s \tag{7}$$

Thus, we can define the steady state distribution (also known as ergodic distribution) of F_t, which is characterized when *s* tends to infinity in the equation (7), since the changes represented by *M* are repeated an arbitrary number of times. Such distribution of steady state exists if the Markov chain is regular, which means, if and only if, for an *m*, M^m has no inputs with a value equal to zero. In this case, the matrix of transition probabilities converges to M^* of rank 1. Then the existence of a steady state distribution, F^*, is characterized by:

$$MF^* = F^* \tag{8}$$

The vector F^* describes the future distribution of cities' sizes, if the movements observed in the sample period are repeated *ad infinitum*. Each row of M^t tends to the limit of the distribution when $t \to \infty$.

To get a sense of speed with which the urban localities move within the distribution, it is possible to calculate the matrix of mean first passage times M_P, where one element M_{Pij} indicates the expected time for a unit of observation to move from class *i* to class *j* for the first time. For a regular Markov chain, M_P is defined as

$$M_P = (I_K - Z + ee'Z_{dg})D \tag{9}$$

where I_K is the identity matrix of order K, Z is the fundamental matrix: $Z = (I_K - M + M^*)^{-1}$, M^* is the limiting matrix, *e* is the unit vector, Z_{dg} results from Z setting off-diagonal entries to 0, and D is the diagonal matrix with diagonal elements $d_{ii} = 1/\alpha_i$, given that $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_r\}$ is the limiting probability vector for M (Kemeny and Snell, 1976).

A consideration about limitations of traditional Markov chains to study the dynamics of cities is that they do not capture the spatial dependence that may exist between the studied observational units. This spatial dependence can arise from measurement errors, such as boundary mismatches between the administrative data and market processes (Rey, 2001), and may reflect amenities and knowledge spillovers or trade and migration flows. In this sense, in order to bring up information on the spatial dependence in the population distribution, techniques developed by Anselin (1995) and Rey (2001) will be explored in the remainder of this section.

Local Indicators of Spatial Association

Anselin (1995) suggest a class of Local Indicators if Spatial Association (LISA) for the analysis of spatial clustering and hot spots. These local statistics can provide more detailed insights on the location specific nature of spatial dependence. The local Moran statistic is given by

$$I_i = \frac{z_i}{\sum_i z_i^2} \times z_i^{\circ} \tag{10}$$

where z_i expresses the observation for region I on a variable as a deviation from the mean, and z_i° is the spatial lag for location i

$$z_i^{\circ} = \sum_{j=1}^{n} w_{ij} z_j \tag{11}$$

As pointed by Rey (2001), the local Moran can be used to map an observation's location in absolute space into a relative space that consider not only its point in an a-spatial variate distribution, but also the location of its neighbours in the same distribution. The position of neighbours is summarized by the spatial lag from z_i°. The LISA implementation, then, divides the observations in four classes according to their local Moran statistic, as summarized in table below. Additionally, several geographical aspects can be viewed in a Moran scatter plot, consisting of pairs of local Moran and its spatial lag local Moran for each location.

Table 1 – LISA Classifications

Class	Own Value (z_i)	Neighbours Value (z_i°)
HH	Above Average	Above Average
HL	Above Average	Below Average
LH	Below Average	Above Average
LL	Below Average	Below Average

Source: Rey (2001)

Rey (2001) suggest an extension of LISA approach that integrates the local indicators of spatial integration into a dynamic framework based on Markov chain. The main motivation for this extension is the fact that Moran scatter plot only brings spatial information about the locational distribution of a given variable at a point in time.

In each period, the local Moran statistic for each observation can be classified into four mutually exclusive classes. Thus, there are twelve possible transitions a local Moran may experience over two or more periods. Moreover, These twelve transitions can be divided in three groups: Type 0 – The region-neighbours pair remains at same level; Type I – Only the region moves, but its neighbours were in the same category; Type II – Involves a transition of only the neighbours in relative space, but the region in question remains in the previous state; and Type III – Involves a transition of both a locality and its neighbours. The Type III can be broken down into two subgroups: IIIA – which occurs when both state and neighbours move in the same direction in the distribution; and IIIB - occurs when locality and neighbours move in opposite directions. These transitions types are summarized in Table X.

Table 2 – Spatial Transitions

t	t+1			
	HH	HL	LH	LL
HH	0	II	I	IIIA
HL	II	0	IIIB	I
LH	I	IIIB	0	II
LL	IIIA	I	II	0

Source: Rey (2001)

Rey (2001) also suggests two interesting measures that can be obtained using the frequency of each type of transition between two periods. One is a flow measure, which can be understood as a measure of instability in the short-term spatial dynamics. A measure of instability or flux of the short run can be given by

$$\text{Flux}_t = 1 - \frac{F_{0,t}}{n} \tag{12}$$

Where $F_{I,t}$ is defined as the number of observation that experience a transition of type *I* in the period $t \rightarrow t+1$, and $n = F_{0,t} + F_{I,t} + F_{II,t} + F_{III,t}$. This flux measure varies between 0 and 1, where 1 indicates a high instability.

Since the relationship between the locality and its neighbours remains cohesive under Type 0 and Type IIIA, a measure of spatial cohesion is given by

$$\text{cohesion}_t = \frac{F_{IIIA,t} + F_{0,t}}{n} \qquad (13)$$

This cohesion measure varies between 0 and 1, where 1 indicates a high cohesion. It is the percentage of locations that move in the same direction of its spatial lag or locality-neighbours pair that remains in the same class from the previous period.

In the original implementation of LISA, developed by Anselin, a bifurcation of high and low value relative to the mean was used. This correspond to discretize the distribution in k=2 classes. According to Rey (2001), with respect to Markov chains, such classification may too aggregate and darken some of transitional dynamics in the income distribution.

Spatial Markov Chains

Rey (2001) suggests a modification in the traditional Markov matrix, conditioning the transition probability (p_{ij}) to the *j* initial class of the spatial lag of the variable in question. Here, this conditioning concerns the population size class of the spatial lag in the initial period. This combination of traditional Markov matrix with the spatial autocorrelation is called spatial Markov matrix. Conditioned on the class of spatial lag in the initial period, this matrix can be constructed by dividing the traditional matrix ($k \times k$) in *k* conditional matrices of dimension (*k,k*), this is, the traditional matrix ($k \times k$)is decomposed into a system ($k \times k \times k$). In other words, an explicit test of adherence or propulsive influence of neighbours of an economy can be based on the comparison between the different state transitions conditioned to the initial state of its spatial lag (Rey, 2001).

For the *k-th* matrix conditional, an element $p_{ij|k}$ is the probability of a region in class *i* at time *t* convert the class *j* in the next moment on the understanding that its spatial lag was in class *k* at time *t*. This matrix is shown in Table 1 below where *k = 4*.

Table 3 – Spatial Markov Transition Probability Matrix

Spatial Lag	t_i	t_i+1			
		1	2	3	4
1	1	$p_{11\|1}$	$p_{12\|1}$	$p_{13\|1}$	$p_{14\|1}$
	2	$p_{21\|1}$	$p_{22\|1}$	$p_{23\|1}$	$p_{24\|1}$
	3	$p_{31\|1}$	$p_{32\|1}$	$p_{33\|1}$	$p_{34\|1}$
	4	$p_{41\|1}$	$p_{41\|1}$	$p_{43\|1}$	$p_{44\|1}$
2	1	$p_{11\|2}$	$p_{12\|2}$	$p_{13\|2}$	$p_{14\|2}$
	2	$p_{21\|2}$	$p_{22\|2}$	$p_{23\|2}$	$p_{24\|2}$
	3	$p_{31\|2}$	$p_{32\|2}$	$p_{33\|2}$	$p_{34\|2}$
	4	$p_{41\|2}$	$p_{41\|2}$	$p_{43\|2}$	$p_{44\|2}$
3	1	$p_{11\|3}$	$p_{12\|3}$	$p_{13\|3}$	$p_{14\|3}$
	2	$p_{21\|3}$	$p_{22\|3}$	$p_{23\|3}$	$p_{24\|3}$
	3	$p_{31\|3}$	$p_{32\|3}$	$p_{33\|3}$	$p_{34\|3}$
	4	$p_{41\|3}$	$p_{41\|3}$	$p_{43\|3}$	$p_{44\|3}$
4	1	$p_{11\|4}$	$p_{12\|4}$	$p_{13\|4}$	$p_{14\|4}$
	2	$p_{21\|4}$	$p_{22\|4}$	$p_{23\|4}$	$p_{24\|4}$
	3	$p_{31\|4}$	$p_{32\|4}$	$p_{33\|4}$	$p_{34\|4}$
	4	$p_{41\|4}$	$p_{41\|4}$	$p_{43\|4}$	$p_{44\|4}$

Notes: Elaboration by the Author based on Rey (2001)

Table 3 can be used to test the negative or positive influence of geographic neighbours in a region. In this case, dividing the cities in four size classes (small, medium-small, medium-large and large), for example. If we want to know the effect of medium-large sized neighbours on the transition to move up or down of a city, we analyze the matrix elements in the third conditional, where the spatial lag is medium-large. Per instance, the $p_{34|3}$ element stands for the possibility of a region in the medium-large class to move upwards given that its neighbours are in medium-large class.

Furthermore, it is possible to know the influence of spatial dependence on the transition probability comparing the elements of the traditional transition matrix with the elements of the spatial Markov matrix. For example, if $p_{34} > p_{34|3}$, then the probability of an upward movement in the classification of a city in the medium-large class is higher than the probability of one in the medium-large class with neighbours in the same class. Generally speaking, if the neighbourhood has no effect on the probability of transition, then the conditional probability is equal to the probability of the traditional Markov matrix

$$p_{ij|1} = p_{ij|2} = \dots = p_{ij|k} = p_{ij} \quad \forall ij \tag{14}$$

The main gain in analyzing the dynamics of the spatial conditioning is capturing the influence of the location and thus the influence of the dimensions of the neighbors about the possibilities for mobility of minimum comparable areas within the populational hierarchy. Beyond to providing a more detailed view of the geographic dimension of population distribution, some interesting questions concerning the characteristics of population mobility can be formulated, in analogy with the questions that Rey(2001) has brought forward for the income distribution theme. Some of these: Is MCA's probability of moving up or down the distribution related to the current, or past movements of its neighbors? Is this form of spatial dependence of a similar magnitude for upward as opposed to downward moves in the distribution? These are some of the questions that can be answered by using of a spatial Markov matrix.

3. A Spatial Economic Model of Population Growth Dynamics in Brazilian MCAs

In this third chapter, we build the foundations for the population growth dynamics. Firstly, we present the literature review that discusses studies related to our work. Secondly, we set out our theoretical spatial extension of Glaeser's population growth model. And in the remaining topic, we present the econometric methodology behind our dynamic spatial panel data model.

3.1 Literature Review

Glaeser et al (1995) studied how urban growth is related to several characteristics of the initial period, providing a theoretical framework and empirical analysis. Conceptually, in the model developed by Glaeser and more sophisticatedly in Glaeser (2008), including housing market, the cities are treated as separated economies that share common pools of labour and capital[2]. Because of assumption of shared labour and capital, cities differ only in the level of productivity and their quality of life. Additionally, under the assumption of spatial equilibrium, which assumes that, with free migration of workers, welfare is equalized across space. Population growth thus reflects both productivity growth and improvement in quality of life of a locality. The empirical analysis is based on cross-section data of industrial cities in the United States between 1960 and 1990. The results show that a city's income and population grow together and they are positively related to the education of the initial period, negatively related to initial employment, negatively related to the initial share of employment in manufacturing. Moreover, the results also suggest the importance of a well-educated work force, suggesting that higher education levels influence later growth not through saving rates but through influencing the growth of technology.

In 2001, Beeson, DeJong and Troesken, examined the location and growth of urban population of North American cities using data from the U.S. Census between 1840 and 1900. The authors investigated how natural and produced characteristics of cities in 1840 influenced the growth of the subsequent 150 years. The first results indicated that the natural characteristics strongly influenced where the population was located in 1840. Then the authors found that produced characteristics, many of which are viewed as proxies for agglomeration economies (i.e., industry mix, educational infrastructure, and human capital), have had a significant effect on population growth. For example, there is evidence suggesting

---

[2] See also Glaeser and Gottlieb (2009).

that growth is influenced not only by the stock of human capital, but also factors that facilitate the production of more human capital (educational infrastructure).

Oliveira (2004b) intends to identify the determinants of economic and population growth of Brazilian northeastern cities in the 1990s. The variables used represent the initial characteristics of these cities coming from the IBGE censuses for the years 1991 and 2000. Based on Glaeser (1995) theoretical framework, the author found that cities that had the highest levels of human capital in 1991 were the fastest growing ones. Additionally, cities with higher income and quality of life in 1991 grew fastest. This result, according to Oliveira, can indicate migratory movement in search of better living conditions.

Da Mata et al (2007) consider a model for population growth composed by supply and demand factors, similar to that found in Henderson (1988). To measure these factors, they use market potential variables, which are partly dependent on the road transport network. From a spatial econometric viewpoint, market potential variables can be interpreted as spatially lagged explanatory variables or exogenous interaction effects, since they measure the impact of X variables in one spatial unit on the dependent variable in another spatial unit. The disadvantage of using market potential variables is that a certain neighborhood matrix structure is imposed on these *X* variables without first testing this structure. Following this theoretical approach, the authors estimated equations that describe the size and growth of a group of Brazilian urban agglomerations between 1970 and 2000. For these estimates econometric techniques including Generalized Methods of Moments (GMM) and spatial GMM were used in order to correct endogeneity in the presence of spatially autocorrelated errors. The results showed that the increase of rural population supply, improved connectivity inter-regional transport and qualification of the labor force all have strong impacts on the growth of cities. On the other hand, crime and violence variables and public industrial capital had negative impacts on the growth of urban areas.

Vieira (2009), using specifications for spatial models in cross section, intended to compare the growth of municipalities in São Paulo state between 1980 and 2000 through the influence of spatial externalities in their growth trajectory variables and of correlated variables with their growth rate. The employed model to obtain these growth ratios was the one proposed by Glaeser et al. (1995), with added variables that seek to capture the agglomeration effects. Among the findings, the average schooling variable did not present a positive correlation with population growth, on the contrary; it was negative and statistically significant. One hypothesis to explain such a result would be that a higher educational level of the municipality’s population boosted the expulsion of the inhabitants of those cities less favored with skilled jobs opportunities comparing to the capital and regional centers. Likewise, some of his results indicate the presence of negative spatial externality caused by the higher education of the neighbour. The municipal infrastructure also proved relevant to the growth rate as well as the composition of the municipality's economy. According to the author, the results show that this municipality that had greater involvement of industry in total production has tended to grow more. In relation to these interpretations of results, Vieira did not perform the calculation of the partial derivatives (direct and indirect effects). He interprets the estimated coefficients directly as marginal effects. But the estimated coefficients on spatial models do not represent the marginal effects of changes in the explanatory variables on the dependent variable. According to Elhorst (2010), no observation of this characteristic of spatial models leads some empirical studies to erroneous conclusions.

Recently, Duranton and Puga (2013) reviewed the key theories with implications for urban growth followed by a discussion of their respective empirical evidences. According to the aurhors, a large literature based in Alonso (1964), Mills (1967), and Muth (1969), has focused on the importance of location within the city and its impact on commuting costs as a key determinant of land use and housing development in cities, factors that drive the population size of cities. Another theoretical segment has been devoted to modelling the productive advantages of cities or agglomeration economies (Fujita, 1988; Helsley and Strange, 1990; Glaeser, 1999; Duranton and Puga, 2001), these models follow Rosen (1979) and Roback (1982) works. Finally, Duranton and Puga attempt to explain the random urban growth models, witch are motivated by existence of regularities in the size distribution of cities and in the patterns of urban growth (e.g.: Gabbaix, 1999). However, as the authors themselves warn, the empirical evidences presented in their paper are mostly focused on cities in developed economies, most of them originated from the United States in particular.

3.2 A Spatial Extension of Glaeser's Population Growth Model

In this section, we provide a theoretical framework for the empirical analysis of population growth across Brazil. As remembered by Glaeser and Gottilieb (2009), due to the seminal works of Mills (1967), Rosen (1979) and Roback (1982) on population changes within a country, the spatial equilibrium condition became one of the main theoretical tools of urban economists. This assumption states that welfare is equalized across space, provided that labor is mobile; higher wages in urban areas are offset by negative urban attributes such as higher prices and negative amenities.

In the urban growth model developed by Glaeser et al. (1995),[3] cities are treated as independent economies that share common pools of labor and capital and that differ in their level of productivity (A_{it}) and quality of life (Q_{it}). Total output of an economy is determined by this productivity level and modeled by a Cobb-Douglas production function dependent on population size. Total welfare of a potential migrant to this economy equals wages multiplied by the quality of life, which decreases with population size. The net result is a population growth regression containing factors that determine quality of life, such as crime, housing prices, traffic congestion and population size of an economy. Glaeser concludes that population growth is the most useful indicator for urban prosperity or welfare. As explained by Cheshire and Magrini (2007), people vote with their feet, and if the combination of real wage and quality of life they could receive in some other city is higher, then they will move to it.

One objection to Glaeser's theoretical framework is that it ignores spatial interaction effects among economies, especially between a locality and its surroundings. One way to deal with this problem is to increase the scale of the geographical units that are used in the empirical analysis, thereby assuming that these interaction effects at this higher scale no longer exist (Glaeser et al., 1995). The other way, more prevalent and fundamental, is to model these spatial interaction effects explicitly.

Let total output of an economy be given by

$$Y_{it} = A_{it} N_{it}^{\beta} K_{it}^{\gamma} \overline{Z}_i^{1-\beta-\gamma} \tag{15}$$

where A_{it} represents the level of productivity in economy i at time t, N_{it} denotes population size, K_{it} is traded capital, and $\overline{Z}_i$ is fixed non-traded capital.

The first extension includes productivity interaction effects among economies. In a seminal paper, Ertur and Koch (2007) state that there is no clear reason to constrain knowledge externalities (that improve technology) within the borders of the economy. They argue that knowledge accumulated in one economy depends on knowledge accumulated in other economies, although with diminished intensity due to frictions caused by socio-economic and institutional dissimilarities which can be captured by geographical distance or border effects. More formally,

$$A_{it} = a_{it} \prod_{j \neq i}^{N} a_{jt}^{\rho w_{ij}} \tag{16}$$

[3] A more sophisticated approach, including the housing market, can be found in Glaeser (2008).

where the productivity level of an economy, A_{it}, is taken to depend on urban differences in the productivity of labor related to social, technological, and political sources in the own economy i, a_{it}, and those in neighboring economies denoted by j, a_{jt}. The parameter ρ reflects the degree of interdependence among economies, with $0 < \rho < 1$. Although this parameter is assumed to be identical for all economies, the net effect of these interaction effects on economy i depends on its relative location, the effect of being located closer or further away from other economies. This relative location of different economies is represented by the exogenous terms, w_{ij}, which are assumed to be non-negative, non-stochastic and finite numbers setting up an N by N row-normalized neighborhood matrix W with $0 \le w_{ij} \le 1$ and $w_{ij} = 0$ if $i = j$.

On substituting (16) into (15), total output of an economy is given by

$$Y_{it} = \left(a_{it} \prod_{j \neq i}^{N} a_{jt}^{\rho w_{ij}} \right) N_{it}^{\beta} K_{it}^{\gamma} \overline{Z}_{i}^{1-\beta-\gamma} \tag{17}$$

The first order conditions for capital and labor, i.e., capital income (with a normalized price of one) and labor income of worker S_{it} are equal to their marginal products, result in the following labor demand equation, provided that the optimal solution for capital is substituted in the condition for labor

$$S_{it} = \beta \gamma^{\frac{\gamma}{1-\gamma}} \left(a_{it} \prod_{j \neq i}^{N} a_{jt}^{\rho w_{ij}} \right)^{\frac{1}{1-\gamma}} N_{it}^{\frac{\beta+\gamma-1}{1-\gamma}} \overline{Z}_{i}^{\frac{1-\beta-\gamma}{1-\gamma}} \tag{18}$$

This labor demand equation shows that higher wages reflect higher productivity and fewer workers.

Consumers have Cobb-Douglas utility functions defined over tradable goods and housing, respectively denoted by C_{it} and H_{it}. In addition, it is assumed that welfare is partly due to the (dis)amenities of the local economy, Θ_{it}. In other words, the locality's (dis)amenities will directly affect the welfare of people who live there. These (dis)amenities may interfere negatively or positively with somebody's utility and, moreover, can be natural (e.g. climate, beaches, vegetation) or generated by humans (e.g. violence, entertainment, traffic, pollution)

$$U_{it} = C_{it}^{1-\alpha} H_{it}^{\alpha} \Theta_{it} \tag{19}$$

where α is a constant. The price of tradable goods is normalized to 1, while the housing price is p_H. Consumers maximize their utility subject to the budget constraint

$$C_{it} + p_H H_{it} = S_{it} \tag{20}$$

by choosing C_{it} and H_{it}.

The second extension includes amenity interaction effects among economies. Some (dis)amenities may also (dis)benefit individuals living in other economies (Brueckner, 2003). For example, people may use facilities in other localities; the violence in a particular neighborhood can also generate feelings of insecurity in adjacent neighborhoods; and water pollution can cause health damage downstream. In mathematical terms

$$\Theta_{it} = \left(\theta_{it} \prod_{j \neq i}^{N} \theta_{jt}^{\eta w_{ij}} \right) \tag{21}$$

where the total amenities of an economy, Θ_{it}, are taken to depend on local amenities θ_{it} and those in neighboring economies denoted by θ_{jt}, whose impact decrease with geographical distance. The parameter η measures the degree of interdependence among economies, with $0 < \eta < 1$. According to Glaeser et al. (1995), many of the potential (dis)amenities are reflected by the level of population and the population growth rate; the greater the size of a city, the lower the quality of life. One reason is that the costs of migration are rising in the number of in-migrants. In addition, if the population size of a city grows rapidly, it takes time to build up certain public goods, infrastructure, or housing. Consequently, the residents of quickly growing cities may suffer in terms of quality of life. Together this yields the utility function

$$U_{it} = C_{it}^{1-\alpha} H_{it}^{\alpha} \left(\theta_{it} \prod_{j \neq i}^{N} \theta_{jt}^{\eta w_{ij}} \right) N_{it}^{-\varphi} \left(\frac{N_{it}}{N_{it-1}} \right)^{-\tau} \tag{22}$$

where $\varphi > 0$ and $\tau > 0$. Total city demand for housing is given by

$$H_D = N_{it} \frac{\alpha S_{it}}{p_H} \tag{23}$$

The spatial equilibrium condition implies that welfare is equalized across space, which we denote by $\overline{V_t}$ at a particular point in time t. Substituting the demand equation for housing derived in (23) into (22), yields the indirect utility function in (24), which equals the common utility level $\overline{V_t}$

$$V(S_{it}, p_H) = \alpha (1-\alpha)^{1-\alpha} \left(\theta_{it} \prod_{j \neq i}^{N} \theta_{jt}^{\eta w_{ij}} \right) S_{it} p_H^{-\alpha} N_{it}^{-\varphi} \left(\frac{N_{it}}{N_{it-1}} \right)^{-\tau} = \overline{V_t} \tag{24}$$

Following Glaeser (2008), housing floor space is produced competitively either by land (L) or by height (h). A fixed quantity of land at a particular location ($\overline{L}$) will determine an endogenous price for land (p_L) and housing (p_H), while the cost of producing hL units of structure on top of L units of land is given by $c_0 h^{\delta} L$, where $\delta > 1$. The developer then maximizes profits given by

$$\pi = p_H hL - c_0 h^{\delta} L - p_L L \tag{25}$$

The first order condition of this maximization problem for height, $h = (p_H / \delta c_0)^{\frac{1}{\delta-1}}$, implies the total housing supply equation

$$h\overline{L} = (p_H / \delta c_0)^{\frac{1}{\delta-1}} \overline{L} \tag{26}$$

By setting housing demand in equation (23) to housing supply in equation (26), the housing price equation is obtained

$$p_H = \left(\frac{N_{it} \alpha S_{it}}{\overline{L}} \right)^{\frac{\delta-1}{\delta}} (\delta c_0)^{\frac{1}{\delta}} \tag{27}$$

The labor demand equation in (18), the indirect utility equation in (24), and the housing price equation in (27) form a system of three equations with three unknown variables (N_{it}, S_{it}, and p_H). Solving this system for population, N_{it}, yields

$$\begin{aligned} \log N_{it} = D_N + \psi \Bigg\{ & \log \theta_{it} + \left(\eta \sum_{j \neq i}^{N} w_{ij} \log \theta_{jt} \right) + \left(\frac{\delta - \alpha\delta - \alpha}{\delta} \right) \left(\log a_{it} + \rho \sum_{j \neq i}^{N} w_{ij} \log a_{jt} \right) \\ & + \tau \log N_{it-1} + \log \overline{V}_t \Bigg\} \end{aligned} \tag{28a}$$

where

$$\psi = \frac{\delta(1-\gamma)}{(1-\gamma)\big(\alpha(\delta-1) + \delta\varphi + \delta\tau\big) - (\delta - \alpha\delta + \alpha)(\beta + \gamma - 1)} \tag{28b}$$

$$\begin{aligned} D_N = \psi \Bigg(& \alpha \log \alpha + (1-\alpha)\log(1-\alpha) + \left(\frac{\delta - \alpha\delta - \alpha}{\delta} \right) \left(\log \beta + \left(\frac{\gamma}{1-\gamma} \right) \log \gamma + \left(\frac{1-\beta-\gamma}{1-\gamma} \right) \log \overline{Z} \right) \\ & - \frac{\alpha(\delta-1)}{\delta} - \frac{\alpha}{\delta} \log(\delta c_0) + \left(\frac{\alpha(\delta-1)}{\delta} \right) \log \overline{L} \Bigg) \end{aligned} \tag{28c}$$

According to Glaeser and Gottlieb (2009), the spatial equilibrium condition formally means that in a dynamic model only lifetime utility levels will be equalized across space. However, as long as housing prices or rents can change quickly, or to a reasonable extent within the observation periods being considered, which in this study is ten years,[4] the price adjustment suffices to maintain the spatial equilibrium. Under this condition, the change in utility level between time *t* and *t +1* is the same across space, $\overline{V}_{t+1} / \overline{V}_t$, as a result of which equation (28a) can be rewritten as

$$\begin{aligned} \log\left(\frac{N_{it+1}}{N_{it}} \right) = \psi \Bigg(& \log\left(\frac{\theta_{it+1}}{\theta_{it}} \right) + \left(\eta \sum_{j \neq i}^{N} w_{ij} \log\left(\frac{\theta_{jt+1}}{\theta_{jt}} \right) \right) + \left(\frac{\delta - \alpha\delta - \alpha}{\delta} \right) \left(\log\left(\frac{a_{it+1}}{a_{it}} \right) + \rho \sum_{j \neq i}^{N} w_{ij} \log\left(\frac{a_{jt+1}}{a_{jt}} \right) \right) \\ & + \tau \log\left(\frac{N_{it}}{N_{it-1}} \right) + \log\left(\frac{\overline{V}_{t+1}}{\overline{V}_t} \right) \Bigg) \end{aligned} \tag{29}$$

[4] According to Duranton and Puga (2013, p.18), cyclical behavior and sluggish adjustment also suggest measuring population growth over periods of five or ten years.

Following Glaeser et al. (1995), we assume that X_{it} is a vector of city characteristics at time t that determine the growth of both city specific amenities and of city specific productivity

$$\log\left(\frac{\theta_{it+1}}{\theta_{it}}\right) = X'_{it}\lambda_{\theta} \tag{30a}$$

$$\log\left(\frac{a_{it+1}}{a_{it}}\right) = X'_{it}\lambda_{a} \tag{30b}$$

Combining the sets of equations (28) and (30) yields the dynamic spatial population growth equation

$$\log\left(\frac{N_{it+1}}{N_{it}}\right) = \psi\left[\begin{array}{l}\tau\log\left(\frac{N_{it}}{N_{it-1}}\right)+\left(1+\frac{\delta-\alpha\delta+\alpha}{\delta}\right)X'_{it}\left(\lambda_{\theta}+\lambda_{a}\right)\\ +\left(\eta+\frac{\delta-\alpha\delta+\alpha}{\delta}\rho\right)\sum_{j\neq i}^{N}w_{ij}X'_{jt}\left(\lambda_{\theta}+\lambda_{a}\right)+\log\left(\frac{\bar{V}_{t+1}}{\bar{V}_{t}}\right)\end{array}\right] \tag{31a}$$

$$\log\left(\frac{N_{it+1}}{N_{it}}\right) = \psi\left[\begin{array}{l}\tau\log\left(\frac{N_{it}}{N_{it-1}}\right)+\left(1+\frac{\delta-\alpha\delta+\alpha}{\delta}\right)X'_{it}\left(\lambda_{\theta}+\lambda_{a}\right)\\ +\left(\eta+\frac{\delta-\alpha\delta+\alpha}{\delta}\rho\right)\sum_{j\neq i}^{N}w_{ij}X'_{jt}\left(\lambda_{\theta}+\lambda_{a}\right)+\log\left(\frac{\bar{V}_{t+1}}{\bar{V}_{t}}\right)\end{array}\right]+\mu\sum_{j\neq i}^{N}w_{ij}\varepsilon_{jt+1}+\varepsilon_{it+1} \tag{31b}$$

The equation (31a) contains spatial interaction effects among the explanatory variables Empirically, there is the possibility of spatial interaction among the error terms, as showed in equation (31b). In the spatial econometrics literature, such a model specification in (31b) is known as the spatial Durbin error model (SDEM, see LeSage and Pace, 2009). Since the right-hand side of this model also contains the dependent variable lagged one period time, this model may also be labeled as a dynamic SDEM model.

The utility function specified in (22) assumes that the quality of life for potential migrants declines both in the level of population and in the growth rate of population. Just as knowledge and amenities in one economy are assumed to interact with knowledge and amenities in other economies, so might the level of population and the growth rate of population depend on those in neighboring economies. If residents of quickly growing cities may suffer in terms of quality of life, they might move to neighboring areas. In view of this, if the utility of individuals is also negatively correlated with the level of population (population size) and with the population growth rate of their neighbors, the utility function takes the form

$$U_{it} = C_{it}^{1-\alpha}H_{it}^{\alpha}\left(\theta_{it}\prod_{j\neq i}^{N}\theta_{jt}^{\eta w_{ij}}\right)N_{it}^{-\varphi}\left(\frac{N_{it}}{N_{it-1}}\right)^{-\tau}\left(\prod_{j\neq i}^{N}N_{jt}^{-\nu w_{ij}}\right)\left(\prod_{j\neq i}^{N}\left(\frac{N_{it}}{N_{it-1}}\right)^{-\sigma w_{ij}}\right) \tag{32}$$

where $\nu > 0$ and $\sigma > 0$. Solving the system for population, N_{it}, with this alternative specification of the utility function yields the population growth equation

$$\log\left(\frac{N_{it+1}}{N_{it}}\right) = \psi \begin{bmatrix} \tau \log\left(\frac{N_{it}}{N_{it-1}}\right) + \sigma \sum_{j \neq i}^{N} w_{ij} \log\left(\frac{N_{jt}}{N_{jt-1}}\right) - (\nu + \sigma) \sum_{j \neq i}^{N} w_{ij} \log\left(\frac{N_{jt+1}}{N_{jt}}\right) + \\ \left(1 + \frac{\delta - \alpha\delta + \alpha}{\delta}\right) X'_{it} \left(\lambda_\theta + \lambda_a\right) \\ + \left(\eta + \frac{\delta - \alpha\delta + \alpha}{\delta} \rho\right) \sum_{j \neq i}^{N} w_{ij} X'_{jt} \left(\lambda_\theta + \lambda_a\right) + \log\left(\frac{\bar{V}_{t+1}}{\bar{V}_t}\right) \end{bmatrix} \tag{33a}$$

$$\log\left(\frac{N_{it+1}}{N_{it}}\right) = \psi \begin{bmatrix} \tau \log\left(\frac{N_{it}}{N_{it-1}}\right) + \sigma \sum_{j \neq i}^{N} w_{ij} \log\left(\frac{N_{jt}}{N_{jt-1}}\right) - (\nu + \sigma) \sum_{j \neq i}^{N} w_{ij} \log\left(\frac{N_{jt+1}}{N_{jt}}\right) + \\ \left(1 + \frac{\delta - \alpha\delta + \alpha}{\delta}\right) X'_{it} \left(\lambda_\theta + \lambda_a\right) \\ + \left(\eta + \frac{\delta - \alpha\delta + \alpha}{\delta} \rho\right) \sum_{j \neq i}^{N} w_{ij} X'_{jt} \left(\lambda_\theta + \lambda_a\right) + \log\left(\frac{\bar{V}_{t+1}}{\bar{V}_t}\right) \end{bmatrix} + \mu \sum_{j \neq i}^{N} w_{ij} \varepsilon_{jt+1} + \varepsilon_{it+1} \tag{33b}$$

where ψ is defined as in (28b). In addition to spatial interaction effects among the explanatory variables, this model specification also contains spatial interaction effects among the dependent variable. Empirically, there is the possibility of spatial interaction among the error terms, as showed in equation (33b). In the spatial econometrics literature, such a model specification in (33b) is known as the general nesting spatial model (GNS, see Elhorst, 2014), or, if we also account for the dependent variable lagged one period in time, as a dynamic GNS model.

3.3 Econometric Methodology: Spatial Panel Data Models

In spatial models research, panel data refer to observations made on a number of spatial units over time. Recently, after contributions from Elhorst (2003), Baltagi et al (2006), Elhorst (2005), LeSage and Pace (2009), a number of studies on models of spatial panels increased and the quality of information is increasingly improving.

According to Elhorst (2011), the central focus of spatial econometrics was originally a type of interaction in a single equation configured for data in cross-section. Thus, the punctual estimation of the equation's coefficient was used to test the hypothesis that the spatial effect existed or not. Recently, however, such focus has changed to more than one type of interaction effect, panel data and the marginal effects of explanatory variables in the model. The main advantages of using panel data are that they are more informative (including the possibility of controlling for fixed effects) and contain more variation and less collinearity among the variables. The use of panel data results in a better availability of degrees of freedom, increasing the efficiency of estimation. Panel data also allows the specification of more complicated tests of hypotheses, including effects that can not be addressed using data on purely cross-section.

The econometric spatial model to cross-section can be expanded to a model in panel of N observations and T time periods and can be written as:

$$\boldsymbol{Y_t} = \boldsymbol{\delta W Y_t + X_t \beta + W X_t \theta + \mu + \lambda_t \iota_N + v_t}, \quad \boldsymbol{v_t = \lambda W v_t + \varepsilon_t}, \tag{34}$$

Where $\boldsymbol{Y}_t$ denotes an $N\times1$ vector consisting of one observation of the dependent variable for every economy (i=1,...,N) in the sample at time t (t=1,...,T), in this study the population growth rate $\log(N_{it+1}/N_{it})$, and $\boldsymbol{X}_t$ is an $N\times K$ matrix of exogenous or predetermined explanatory variables observed at the start of each observation period (t-1,t). A vector or a matrix premultiplied by $\boldsymbol{W}$ denotes its spatially lagged value. W is a weight matrix $(N \times N)$ describing the spatial distribution of spatial units and being w_{ij} the (i, j) element of W. It is assumed that this matrix is composed of known constants that the diagonal elements are equal to zero and that the characteristic of the matrix, ω_i , is known. According to Elhorst (2003) the first assumption excludes the possibility that the matrix is parametric, the second assumption implies that a spatial unit which can not be adjacent to itself and the third assumption infers that the matrix's characteristics can be precisely computed, allowing the use in empirical research. The parameter δ represents the response parameter of the dependent variable lagged in space, $\boldsymbol{WY}_t$. Additionally, the autoregressive spatial coefficient δ is assumed to be restricted to the interval $\left(1/\omega_{\min}, 1/\omega_{\min}\right)$ where $\omega_{\min}$ is the smallest and $\omega_{\max}$ is the largest characteristic root of W, if this matrix is symmetric. If W is a symmetric matrix similar to a row-stochastic matrix, where $\omega_{\max} = 1$, the interval for δ becomes $\left(1/\omega_{\min}, 1\right)$(LeSage and Pace, 2009). The symbols $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ represent $K\times1$ vectors of response parameters of the exogenous explanatory variables. The error term specification consists of three components. The vector $\boldsymbol{v}_t$ reflects the error term specification of the model, which is assumed to be spatially correlated with autocorrelation coefficient λ. The $N\times1$ vector $\boldsymbol{\varepsilon}_t=(\varepsilon_{1t},\ldots,\varepsilon_{Nt})^T$ consists of i.i.d. disturbance terms, which have zero mean and finite variance σ^2. The $N\times1$ vector $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_N)^T$ contains spatial specific effects, μ_i, and are meant to control for all spatial specific, time-invariant variables that are difficult to be measured or obtained, but if omitted, could bias the estimates in a typical cross-sectional study (Baltagi, 2005). Similarly, λ_t (t=1,…,T) denote time-period specific effects, where ι_N is a $N\times1$ vector of ones, meant to control for all time-specific, unitinvariant variables, which, if omitted, could bias the estimates in a typical time-series study. More details about the fixed effects are discussed in the next topic.

Dynamic Model

The econometric counterpart of the dynamic spatial GNS model, the final equation implied by the theoretical model presented in the previous section 3.2, reads as (in vector form)

$$\boldsymbol{Y_t = \tau Y_{t-1} + \delta W Y_t + \eta W Y_{t-1} + X_t \beta + W X_t \theta + \mu + \lambda_t \iota_N + v_t}, \quad \boldsymbol{v_t = \lambda W v_t + \varepsilon_t}, \tag{35}$$

A vector or a matrix with subscript t-1 in (35) denotes its time-lagged value. The parameters τ, δ and η are the response parameters of successively the dependent variable lagged in time, $\boldsymbol{Y_{t-1}}$, the dependent variable lagged in space, $\boldsymbol{WY_t}$, and the dependent variable lagged in both space and time, $\boldsymbol{WY_{t-1}}$. In order to avoid possible confusion with the parameters, Table 4 below brings the linkage between theoretical model equation and econometric model equation.

Table 4 – Linkage Between Theoretical Model Equation and Econometric Model Equation

Econometric Model	Theoretical Model
$\boldsymbol{\tau Y_{t-1}}$	$\psi\tau \log\left(\frac{N_{it}}{N_{it-1}}\right)$
$\boldsymbol{\delta W Y_t}$	$\psi(-(\nu+\sigma))\left(\sum_{j\neq i}^{N} w_{ij} \log\frac{N_{jt+1}}{N_{jt}}\right)$
$\boldsymbol{\eta W Y_{t-1}}$	$\psi\sigma\left(\sum_{j\neq i}^{N} w_{ij} \log\frac{N_{jt}}{N_{jt-1}}\right)$
$\boldsymbol{X_t \beta}$	$\left(1+\frac{(\delta-\alpha\delta+\alpha)}{\delta}\right)X'_{it}(\lambda_\theta+\lambda_a)$
$\boldsymbol{W X_t \theta}$	$\left(\eta+\rho\frac{(\delta-\alpha\delta+\alpha)}{\delta}\right)\sum_{j\neq i}^{N} w_{ij}X'_{it}(\lambda_\theta+\lambda_a)$
$\boldsymbol{v_t}$	$\log\left(\frac{\bar{V}_{t+1}}{\bar{V}_t}\right)+\mu\sum_{j\neq i}^{N} w_{ij}\varepsilon_{it+1}+\varepsilon_{it+1}$

Note: Elaboration by the Author.

Generally, the spatial and time-period specific effects may be treated as fixed or as random effects. In the fixed effects model, a dummy variable is introduced for each spatial unit and for each time period (except one to avoid multicollinearity). In the random effects model, $\boldsymbol{\mu}$ and λ_t are treated as random variables that are independently and identically distributed with zero average and variance of σ_μ^2 e σ_λ^2, respectively. It is also assumed that the random variables, $\boldsymbol{\mu}$, λ_t and ε_i are independent from each other. As Elhorst (2014) has pointed out, the random effects model might not be an appropriate specification when observations of adjacent units in an unbroken study area are used and the whole population is sampled, i.e., each spatial units represents itself and has not be sampled randomly. The random effects model would make sense if a limited number of MCAs would be drawn from Brazil, but then the elements of the neighborhood matrix cannot be defined and the impact of spatial interaction effects not be consistently estimated. Only when neighboring units are also part of the sample, it is possible to measure the impact of these neighboring units. In this respect, this study distinguishes itself from many urban studies trying to explain economic growth in cities, such as those of Glaeser et al. (1995) and Glaeser (2008). Whereas we include both urban and rural regions so as to cover a whole country and to model the interactions between them, these previous studies ignore potential interaction effects with their surroundings and treat cities as independent entities.

Unlike the non-spatial models, the estimated coefficients on spatial models do not represent the marginal effects of changes in the explanatory variables on the dependent variable. Direct interpretation of the coefficients in the dynamic GNS model is difficult because they do not represent true partial derivatives. The interpretation of the partial derivative of the impact of changes in a variable represents a more valid ground to test the hypothesis whether there are spatial spillovers or not (LeSage and Pace, 2009). Debarsy et al. (2012) and Elhorst (2012) show that the matrix of (true) partial derivatives of the expected value of the dependent variable with respect to the k^{th} independent variable for i=1,…,N in year t in the short term is given by

$$\left[\frac{\partial E(\boldsymbol{Y})}{\partial x_{1k}} \quad \cdots \quad \frac{\partial E(\boldsymbol{Y})}{\partial x_{Nk}}\right]_t = (\boldsymbol{I} - \delta\boldsymbol{W})^{-1}[\beta_k \boldsymbol{I}_N + \theta_k \boldsymbol{W}] \tag{36}$$

In addition, LeSage and Pace (2009) advocate for the decomposition of the marginal effects into direct (own-economy) and indirect (spillover) effects to other economies. Direct effects are given by the own-partial derivatives along the diagonals of (36). They constitute the effect on the dependent variable resulting from a change in the k^{th} regressor, x_k, in economy i in the short term, i.e., the direct effect arises from the effect of the independent variable on the dependent of the individual. The off-diagonal elements represent short-term indirect effects, the impact of this neighbour's independent variable on the dependent variable of the individual. Since the direct and the indirect effects are different for different units in the sample, LeSage and Pace (2009) propose to report one summary indicator for the direct effects measured by the average of the diagonal elements, and one summary indicator for the indirect effects measured by the average of the column sums of the non-diagonal elements of that matrix. Since the matrix on the right-hand side of equation (36) is independent of the time index *t*, these calculations are equivalent to those presented in LeSage and Pace (2009) for a cross-sectional setting, i.e., these calculations are the short term effect. From Elhorst (2012) it further follows that the long-term marginal effects are given by

$$\begin{bmatrix} \frac{\partial E(\boldsymbol{Y})}{\partial x_{1k}} & \cdots & \frac{\partial E(\boldsymbol{Y})}{\partial x_{Nk}} \end{bmatrix} = [(1-\tau)\boldsymbol{I} - (\delta+\eta)\boldsymbol{W}]^{-1}[\beta_k \boldsymbol{I}_N + \theta_k \boldsymbol{W}] \quad (37)$$

By examining the relationship between population growth and its determinants using the decompositions in both (36) and (37), we are able to explore the direct and indirect effects into short-term and long-term effects. The insight of these two types of effects is quite interesting for the case of municipalities, considering that it will become possible to get to know specifically how a city is affected by variables related to it and its neighboring cities. For example, it is possible to estimate not only the effect of education on population growth, but also the effect of education of the neighbours on city growth.

One problem of the dynamic GNS model is that its parameters are not identified. Recently, Anselin et al. (2008), Gibbons and Overman (2012), and Halleck Vega and Elhorst (2012) paid attention to this issue. Interaction effects among the dependent variable and among the error terms cannot formally distinguished from each other, provided that interaction effects among the explanatory variables are also included. This implies that one of these two spatial interaction effects should be left aside to obtain consistent parameter estimates. If the spatial interaction effects among the dependent variable are left aside, the dynamic SDEM specification results, which is fully consistent with the utility function specified in equation (22). This implies that $\delta=\eta=0$ in (35), that the spatial multiplier matrix $(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}$ in (36) reduces to the identity matrix and the spatial multiplier matrix $[(1-\tau)\boldsymbol{I}-(\delta+\eta)\boldsymbol{W}]^{-1}$ in (37) to the matrix $1/(1-\tau)\boldsymbol{I}$.

The dynamic Spatial Durbin model (SDM) results if the spatial interaction effects among the error terms are left aside. This model specification is consistent with the utility function specified in equation (32). Although interaction effects among the error terms are not accounted for in this specification, thereby reducing the efficiency of the parameter estimates, this does not affect the consistency of the parameter estimates. This also follows from the fact that imposing the restriction $\lambda=0$ in (35) does not affect the direct and indirect effects derived in equations (36) and (37). In the words of LeSage and Pace (pp. 155-158), the cost of ignoring spatial dependence in the dependent variable and/or in the independent variables is relatively high since the econometrics literature has pointed out that if one or more relevant explanatory variable are omitted from a regression equation, the estimator of the coefficients for the remaining variables is biased and inconsistent (Greene, 2005, pp. 133-134). In contrast, ignoring spatial dependence in the disturbances, if present, will only cause a loss of efficiency.

Another important difference between the SDEM and SDM specifications is that the indirect or spatial spillover effects in the first model are local, while in the second model they are global in nature. Anselin (2003) describes the difference. In the SDM specification, a change in X at any location will be transmitted to all other locations following the matrix inverse in equation (36), also if two locations according to $\boldsymbol{W}$ are unconnected. In contrast, local spillovers are those that occur at other locations without involving an inverse matrix, i.e., only those locations that according to $\boldsymbol{W}$ are connected to each other.

To be able to choose between the SDM and SDEM specifications, and related to that, between the utility functions specified in equation (22) or (32) and between a global or local spillover model, we apply a Bayesian approach recently developed by LeSage (2013). If the probability of one model is higher than that of the other model, we may conclude that the former model describes the data better. Another advantage of taking a Bayesian perspective is that different specifications of the neighborhood matrix $\boldsymbol{W}$ can be tested against each other too. Bayesian methods do not require nested models to carry out these comparisons. By repeating the previous analysis for different specifications of the neighborhood matrix and by comparing their performance based on the value of the log marginal likelihood, one can select the neighborhood matrix that outperforms other potential specifications.

Depending on the outcome of these tests, one can finally estimate either the SDM or the SDEM specification by ML. If the previous test points to the dynamic SDM, this model can be estimated by the bias corrected ML estimator developed by Yu et al. (2008) or Lee and Yu (2010), dependent on whether time period fixed are included. If it points to the dynamic SDEM, this model can be estimated by the bias corrected ML estimator developed by Elhorst (2005).

4. Results

In this chapter we present the empirical analysis referent to the population distribution dynamics and the population growth dynamics of Brazilian minimum comparable areas between 1970 and 2010. Therefore, in the next sections, besides information about the dataset implementation, we report and interpret the estimated results, based in the theoretical and methodological foundations presented in chapters 2 and 3.

4.1 The Distribution Dynamics of Brazilian MCAs: Results

In order to examine the behavior of the population distribution between the minimum comparable areas covering the entire Brazilian territory, a serie of empirical evidences are presented according to the characteristics of the methodologies applied. Besides the data implementation, the following two subsections deal with the estimation of density functions and Zipf's law. Then, techniques based on Markov Chain will be explored, in order to bring up information on the dynamics and evolution, as well as the possibility of spatial dependence on the behavior of population distribution.

4.1.1 Data Implementation

The main source of data is the Brazilian Demographic Census for the years 1970, 1980, 1991, 2000 and 2010 conducted by the Brazilian Institute of Geography and Statistics (IBGE). Although the municipality constitute the smallest unit of observation in political and administrative terms to which is possible to obtain economic and demographic data with coverage of entire Brazilian territory for various periods of time, the intertemporal comparisons in a strictly municipal geographic level become inconsistent with changes in the number, area, and border of municipalities that occurred over the decades. Specifically, over the period 1970-2010, the number of municipalities increases from 3952 to 5565. Therefore, to allow consistent comparisons over time, it is necessary to aggregate these municipalities into broader geographical areas, called Minimum Comparable Areas (MCA). Based on the aggregation of municipalities developed by IPEA (Reis et al., 2010), this study has 3659 MCAs relating to the aggregation of all Brazilian municipalities for each census from 1970 to 2010, covering all territory and avoiding selection bias problem. Figure 1 shows a map of the geographical delineation of all areas taken up in the sample.

Figure 1 – Brazil: Minimum Comparable Areas (1970 – 2010)

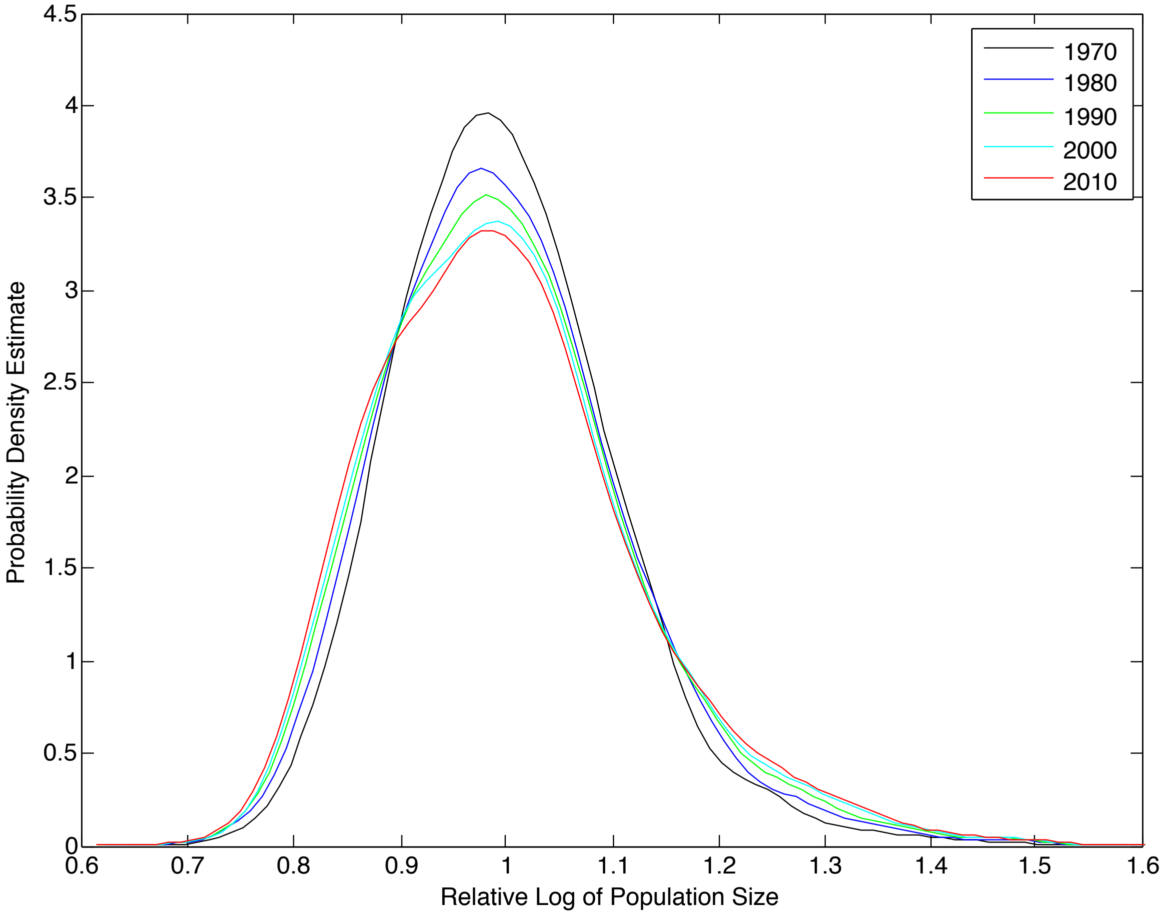
Notes: Elaboration by the Author based on IPEA's shape file

4.1.2 Evolution of Brazilian MCAs size distribution

To investigate the evolution of the Brazilian population distribution shape for 1970-2010 period, a non-parametric normal kernel density with bandwidth value of 0.0245 was estimated for the urban population distribution for each decade[5]. Following Gallo and Chasco (2007), relative population size are considered and the Figure 2, below, shows the distributions of the relative log of population size in 1970, 1980, 1991, 2000 and 2010. The Kernel density plot may be interpreted as the continuous equivalent of a histogram in which the number of intervals has been set to infinity. Adopting a similar strategy to interpretation of Gallo and Chasco, 1 on the horizontal axis indicates Brazilian average MCA size, 1.5 indicates 50% higher than the average, and so on.

[5] The largest bandwidth among the optimum values calculated for each decade was chosen. The optimum values for each decade were calculated using the Matlab function ksdensity.

Figure 2 – Normal Kernel Density Functions for the Population Distribution of MCAs, 1970-2010



Notes: Elaboration by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

Observing the figure, it is remarkable that over the decades there is a loss of concentration of AMCs around the mean. However, this presents a deconcentration rhythm regressive each decade, i.e., the distance between the lines is decreasing. The behavior in the distribution shape may indicate that divergence in population size of MCAs is decreasing, or polarized sizes distribution. In other words, the size of the localities are not converging to the same level, but diverging at a diminishing rate.

Table 5 below shows a statistical summary information, clarifying the Figure 1. Through observation of the columns, obviously the mean is equal to 1 since the values are normalized, the median value decreases over the decades while the standard deviation increases. Clearly, the 2010 distribution is more dispersed around the mean, and this seems to be the trend between 1970 and 2010. Specifically, the distribution became more dispersed in approximately 20% between 1970 and 2010. However, as already pointed out, this divergence is decreasing, between 1970 and 1980 and between 2000 and 2010 the standard deviation growth was 9% to 2.1%, respectively. Still observing the values of the table, with the reduction of the median over the decades we note that the greater dispersion arises mainly from the increased presence of cities above the mean.

Table 5 – Summary Statistics – Relative Population of Brazilian MCAs, 1970 – 2010

Year	Mean	Median	Standart Deviation
1970	1	0.9914	0.1044
1980	1	0.9905	0.1137
1991	1	0.9902	0.1194
2000	1	0.9885	0.1231
2010	1	0.9880	0.1258

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

The evidence of decreasing divergence is similar to others found in the literature, as in Justo (2012) for 431 brazilians MCAs between 1910 and 2010. On the other hand, Trindade and Sartoris (2009) found evidence that the behavior of the Brazilian population distribution already shows a trend of increasing number of municipalities with population below average between 1920 and 2000.

Tables 6a and 6b (below) allow us to contextualize such regional changes in population distribution emphasized in the preceding paragraphs. While the numbers remain fairly stable over the decades, the North region had the largest growth in the number of MCAs above the national median from 2.27% in 1970 to 3.03% in 2010. In this period, as can be seen in Table 6 the North Region's share in the total population increased from 4.43% to 8.32%. On the other hand, the southern region had the highest percentage reduction of MCAs above the median from 9.51% in 1970 to 7.76% in 2010. The participation of the southern region in the total population decreased from 17.71% to 14.36% between 1970 and 2010.

Table 6a – Relative Population Above the Median per Region, 1970 – 2010

Region	Above Median (%)				
	1970	1980	1991	2000	2010
North	2.27	2.71	2.84	2.92	3.03
Northeast	19.24	19.90	19.98	19.98	19.70
Southeast	16.64	16.40	16.67	16.64	16.81
South	9.51	8.72	8.03	7.82	7.76
Midwest	2.32	2.27	2.46	2.62	2.71

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

Table 6b – Participation on the Total Population per Region, 1970 – 2010

Region	Percentage of Population (%)				
	1970	1980	1991	2000	2010
North	4.43	5.56	6.83	7.60	8.32
Northeast	30.18	29.25	28.94	28.12	27.82
Southeast	42.79	43.47	42.73	42.65	42.13
South	17.71	15.99	15.07	14.79	14.36
Midwest	4.89	5.72	6.42	6.85	7.37
Total	93,134,846	119,011,052	146,825,475	169,799,170	190,747,731

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

Table 7 provides information about what percentage of MCAs within each region was higher than the median of the country, which allows us to observe how these changes occurred in the population distribution within each region. As can be seen, the biggest changes in population distribution occurred within the North and Midwest regions. For example, from 143 minimum comparable areas of the North, 58% were above the median in 1970 and this percentage increased to 77.62% in 2010. In the South, there is a reduction of 58.6% to 47.81% in the percentage of MCAs with population above the national median.

Table 7 – MCAs with Population Above the Median per Region, 1970 – 2010

Region	Above Median (%)					
	1970	1980	1991	2000	2010	Total
North	58.04	69.23	72.73	74.83	77.62	143
Northeast	54.24	56.09	56.32	56.32	55.55	1298
Southeast	43.47	42.83	43.54	43.47	43.90	1401
South	58.59	53.70	49.49	48.15	47.81	594
Midwest	38.12	37.22	40.36	43.05	44.39	223

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

The non-parametric normal kernel density functions estimates, as well as the descriptive tables foregoing, have as main role to illustrate the city size distribution patterns. One of the limitations of this information is that it does not permit us to make more precise statements about the size distribution of cities. In this sense, the next topics will bring evidences obtained through the Zipf's law approach, which allows the characterization of the overall evolution of the size distribution of cities, and, in order to bring up information on the dynamics and evolution of the population distribution, techniques based on Markov Chain will also be explored.

4.1.3 The Rank Size Rule for Brazilian MCAs

Table 8 presents the estimation of rank-size equation (2) for all Brazilian MCAs in each decade using an OLS estimator. In the 1970s, the estimated Pareto coefficient approaches the Zipf's law, with an estimated value of 0.95. According to this rule, city populations among any group of cities at any time are proportional to the inverse of the ranking of their populations in that group (Gallo and Chasco, 2007). In the following decades, this coefficient deviates increasingly from the unit value, reaching 0.77 in 2010. This parametric analysis is consistent with the previously obtained evidence of the decreasing distances between the size population distributions of MCAs seen in Figure 1, acquired through the non-parametric normal kernel density analysis. What seems natural, for a more equitable distribution of population between locations (with less variance; 1970 in relation to 2010, for example), the change in position between the cities become easier due to the fact that most localities have closer sizes.

Table 8 – Classic Rank-Size Equation for log(rank) as dependent variable, 1970 – 2010

Explanatory Variables	OLS									
	1970		1980		1991		2000		2010	
Variable										
Intercept	16.1272	**	15.4354	**	15.1158	**	14.8880	**	14.7113	**
ln Population	-0.9517	**	-0.8698	**	-0.8251	**	-0.7950	**	-0.7713	**
No. Obs.	3659		3659		3659		3659		3659	
R-squared	0.913		0.920		0.925		0.929		0.928	
Log Likelihood	-2458		-2311		-2220		-2110		-2150	
JB stat	3302	**	4373	**	6572	**	9942	**	12381	**

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

This result, from the classical equation of rank-size, is qualitatively consistent with Trindade and Sartoris (2009), Justo (2012) and Moro and Santos (2013). Of these, what the first two have in common with our work is the use of the entire Brazilian territory and both rural and urban populations of the observed units, which allows a comparison with our analysis. Moreover, unlike our analysis, both two studies use a very high level of aggregation, 920 MCAs between 1920 and 2000, and 431 observational units between 1910 and 2010, respectively. In 1970 their estimated coefficients were 0.794 and 0.77, respectively. The only difference between these two estimates and our work is the level of aggregation, but 3659 observational units is much closer to reality for 1970. Therefore, we can see that the high aggregation level used by these two studies lead to results that indicate a higher population concentration than the reality, making their evidence inaccurate[6]. Although Moro and Santos (2013) use municipality as observational unit (more disaggregated than MCAs), they only take the urban population into account, which makes our results quantitatively incomparable.

Following the suggestion from Rosen and Resnick (1980), we test for non-Pareto beahavior using the quadratic form for Rank-size equation (3). If the coefficient of quadratic log of population is positive, $c_t > 0$, there is a positive correlation between ranking variation and size. On the other hand, if $c_t < 0$ there is a negative correlation between ranking variation and size. The results of quadratic form are presented in Table 9.

Table 9 – Quadratic Rank-Size Equation for log(rank) as dependent variable, 1970 – 2010

Explanatory Variables	OLS quadratic									
	1970		1980		1991		2000		2010	
Variable										
Intercept	2.4910	**	3.2885	**	3.5169	**	3.7033	**	3.5493	**
ln Population	1.8708	**	1.6153	**	1.5170	**	1.4424	**	1.4470	**
(Ln Population)^2	-0.1444	**	-0.1254	**	-0.1165	**	-0.1101	**	-0.1084	**
No. Obs.	3659		3659		3659		3659		3659	
R-squared	0.978		0.985		0.990		0.993		0.995	
Log Likelihood	34		769		1477		2198		2604	
JB stat	3.8E+06	**	2.7E+06	**	1.4E+06	**	9.0E+05	**	7.2E+05	**

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

The quadratic term has a negative coefficient and significant at 1% for all analysed decades. The curvature presents downward concavity, there is a negative correlation between ranking variation and size, thus producing more cities in the intermediate size classes than would be predicted by a Pareto distribution. This result, a significant value for the quadratic term representing a deviation from the Pareto for the Brazilian case, is concurring with Oliveira (2004a) and Moro and Santos (2013), and contrary to the results found by Rosen and Resnick (1980).

4.1.4 Brazilian Population Distributional Dynamics

Unfortunately, it is not possible to have information on the dynamics of the distribution estimating the Zipf's law equations. The approach of the last topic gives no information on the movements of the cities within the distribution. Apart from this, it does not take into account the possibility that these movements are affected by spatial dependence. To

[6] In 1970 the Brazilian territory was divided into 3952 municipalities.

assess these empirical issues on the size distribution of Brazilian minimum comparable areas, in the next topics we lead with the estimation of transition probability matrices associated with discrete Markov chains (Kemeny and Snell, 1976), which will make it possible to follow the progress of each group of Brazilian MCAs in time. And then, the spatial dependence is introduced through the analysis of LISA Markov and Spatial Markov Chains, both developed by Rey (2001).

Traditional Markov Chains

In order to observe the behaviour of transition from the relative population levels over time, Table 10 shows the traditional Markov transition probability matrix for four classes of relative population according to quartiles for each decade between 1970 and 2010. The class with the MCAs with smaller populations relative is represented by the first quartile. Therefore, if a MCA is in the first quartile class means that it is among the 25% smaller in terms of relative population, and if the MCA is inserted in the fourth quartile means that 25% is among the largest in terms of relative population.

Table 10 – Markov Transition Probability Matrix for Brazilian MCAs population, 1970-2010

t_i	t_i+1			
	1	2	3	4
1	0.9208	0.0781	0.0005	0.0005
2	0.0787	0.8249	0.0951	0.0014
3	0.0005	0.0971	0.8356	0.0667
4	0.0000	0.0000	0.0686	0.9314

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

From Table 10, several points can be observed. Firstly, the transitions probabilities on the main diagonal are relatively high. If the MCA is in the ith class, the probability of being in the same class the decade after is at least 82.49% and at most up 93.14%. Specifically, the probability that the MCA in the second quartile remain in this class in the next period is 82.49%. The high probabilities on the main diagonal show a low interclass mobility, a high persistence of MCAs to stay in their own class from one decade to another over the whole period. However, since these probabilities are not exactly equal to 1, we have the possibility to analyse how the MCAs in each cell move to other cells. Secondly, the probability to continue in the initial state, given by the diagonal elements, is higher in the two extreme classes. In particular, the largest and smallest MCAs have less probability of moving to another categories, i.e., these localities have less interclass mobility than the medium-size cities. Since the elements of main diagonal do not assume the value 1, so there is no possibility of parallel or uniform growth between MCAs. This result is an evidence that population distribution structure suffered changes during the period 1970-2010.

Continuing with the reading of Table 10, we realize that the non-diagonal elements are extremely smaller than elements in the main diagonal. Nevertheless, during 1970 to 2010, the medium classes (2 and 3) have more probability of inter-class mobility than extreme classes (1 and 4), the biggest transition probability among different classes is 9.71% which occurs from third to second quartile; next is the second to third class moving, 9.51%. That is, the largest flows occur between the MCAs that are in the second and third classes. This evidence, together with high persistence of both largest and smallest MCAs to stay in the initial class, highlights the major role of medium-size localities in the processes of urban agglomeration that occurred in Brazil during the last 40 years. This evidence is in agreement with Andrade and Serra (2001), as they assert that Brazilian population is undergoing a process of polarization reversal, in which the medium-sized cities play a decisive role in an automatic decentralization of economic activities. In addition, the probabilities of MCAs move up or down more than two steps are extremely small.

The first mean passage time indicates the expected time for a locality to move from class *i* to class *j* for the first time. To determine the speed with which the urban municipalities move within the distribution, Table 11 displays the mean first passage time matrix for relative population based on equation (9)[7].

Table 11 – First Mean Passage Time Matrix in Decades for Brazilian MCAs population, 1970-2010

t_i	t_i+1			
	1	2	3	4
1	4.00	13.00	33.16	75.36
2	37.47	4.00	20.72	63.31
3	57.47	20.28	4.00	43.73
4	72.05	34.86	14.58	4.00

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

On average, the number of years to reach any class other than the original is relatively high: the shortest time passage is 13 decades and the longest is 75.36 decades. As expected, more distant classes take longer to be reached. For example, for a MCA that was originally in class 1 achieve the class 3, it takes on average 33.6 decades. The faster declines in the 3 and 4 classes (20.28 and 14.58 decades, respectively) may indicate that localities in these classes are

[7] $M_P = (I_K - Z + ee'Z_{dg})D$

more likely to lose relative population. This evidence suggests a general progressive suburbanization process in which big cities stop to grow, favouring the progressive appearance of smaller population cores (Gallo and Chasco, 2007).

The ergodic distribution can be interpreted as the long-run equilibrium in the distribution of relative population of the MCAs. As stated by Gallo and Chasco (2007), given a regular transition matrix, with the passage of many periods, there will be a time when the distribution of urban municipalities will not change any more: that is the ergodic or limit distribution. As the population relative discretization was made from the quartiles, the ergodic distribution naturally will be similar to the initial distribution of classes (25% of MCAs in each class) and does not bring interesting results to be interpreted.

A consideration about limitations of traditional Markov chains to study the dynamics of cities is that they do not capture the spatial dependence that may exist between the studied observational units. In order to take into account the possibility of spatial dependence in population distribution dynamics of Brazilian minimum comparable areas, we introduce in the following topics the spatial dependence through the analysis of LISA Markov, that integrates the local indicators of spatial association into a dynamic framework based on Markov chains, and Spatial Markov Chains, that extends the transition probabilities from traditional Markov chains to be conditioned on the initial relative population class of its spatial lag. Both approaches were developed by Rey (2001).

LISA Markov

Table 12 below summarizes the spatial transitions using the same classification system proposed by Rey (2001) shown in section 2, Table 2. The LISA Markov matrix was estimated for four different time intervals: between 1970 and 1980, 1970 and 1991, 1970 and 2000, and between 1970 and 2010. This way, various types of evidence can be seen, among them the role of time on the behaviour of MCA's population distribution in relation to their respective spatial lags[8]. The standard contiguity neighbours matrix (W) was used for estimation of following LISA Markov matrices. In this spatial weight matrix, the w_{ij} of the contiguity neighbours are equal to 1 and zero otherwise, the diagonal elements also have values set to zero since no spatial unit can be viewed as its own neighbour.

The first result that emerges is a high probability of minimum comparable areas and their spatial lag to remain in the same classification (Type 0). Specifically, between 1970 and

[8] The probabilities presented should be seen under an exploratory perspective, it would be prudent to interpret these probabilities with a certain caution (Rey, 2001).

1980 this probability was 89.86%, indicating a low mobility between classes in this decade. However, it is possible to observe an increase in mobility according to an increase in the time interval. For example, in a forty-years interval the probability of remaining in the same size for a MCA and its neighbours drops to 80.13%. Moro and Santos (2013) found that these probabilities are 96.2% and 88.5%, respectively, for urban population of Brazilian municipalities that already existed in 1970, not covering the entire Brazilian territory between 1980 and 2010, since new municipalities were created in this period. This way of obtaining the sample may bring harm to the results, considering that territories and, consequently, population of new municipalities (created from the subdivision of former municipalities) will be excluded from the sample. That is, the results of these authors could be affected by a selection bias in the choice of geographical units.

Table 12 – LISA Spatial Transitions for Brazilian MCAs population, 1970-2010

Interval	Type of Transition					Cohesion	Flux
	Type 0	I	II	IIIA	IIIB		
	0.8986	0.0443	0.0533	0.0038	0.0000	0.9024	0.1014
1 decade	UP	0.0139	0.0128	0.0003	-		
	DOWN	0.0303	0.0404	0.0036	-		
	0.8500	0.0596	0.0798	0.0098	0.0008	0.8598	0.1500
2 decades	UP	0.0186	0.0221	0.0011	-		
	DOWN	0.0410	0.0577	0.0087	-		
	0.8163	0.0711	0.0981	0.0139	0.0005	0.8303	0.1837
3 decades	UP	0.0251	0.0271	0.0022	-		
	DOWN	0.0459	0.0711	0.0118	-		
	0.8013	0.0749	0.1071	0.0161	0.0005	0.8174	0.1987
4 decades	UP	0.0257	0.0284	0.0033	-		
	DOWN	0.0492	0.0787	0.0128	-		

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

A second result we can see through the table 12, besides the transition from Type 0, the most common type of transition is Type II, which involves a transition of only the neighbours in relative space, but the locality in question remains in the previous state. Furthermore, the probability of this transition type also increases when the time interval increases. Another result concerning this type of transition is that there is a greater probability of downward movement, i.e. a higher probability of most populous neighbours to become less populated.

Also regarding the Type II transition, we can obtain a subgroup of MCAs that were populated above average and had less populous neighbours, while in the following period the neighbours became highly populated. We can investigate further highly populated communities that propelled the neighbours. Thus, in the interval between 1970 and 1980, we identified 14 MCAs that played this role. These minimum comparable areas are equivalent to the current territory of 48 municipalities, three of these are located in Pará state (North), 3 in Pernambuco and 3 in Bahia (Northeast), 6 in São Paulo state (Southeast), 3 in Paraná state (South), and finally, 27 municipalities in state of Mato Grosso and one in the Federal District (Midwest)[9]. Only 3 (Curitiba, Brasília and Cuiabá) of these 48 municipalities are capitals of their respective states. To get an idea of the important role played by these localities, 45 of these municipalities had these features for all time intervals. Probably, these municipalities boosted population growth in theirs neighbourhoods. To investigate more deeply these municipalities is an interesting suggestion for future research.

The second most frequent transition type is Type I, which occurs when only the locality moves, but its neighbours remained in the same category. Similar to our results, Moro and Santos (2013) found in their study that transition Type I is the second most frequent, while less frequent is the Type IIIB.

Additionally, the cohesion measure is decreasing with the increase of the time interval. That is, over longer time intervals the probability of the MCAs move in the same direction of theirs spatial lag between different classes decreases. The flux measurement indicates that there is an increased instability in behaviour of MCAs relative to its neighbours in the population distribution when the time interval increases.

In Table 13, below, the LISA Markov matrix was estimated for decennial time interval: between 1970 and 1980, 1980 and 1991, 1991 and 2000, and between 2000 and 2010. The main motivation is the distinction between the decades over the last 40 years of changes in the population configuration of the country. Performing the estimation with LISA decennial intervals give us an idea of how the changes on the behavior of MCA's population distribution in relation to their respective spatial lags have occurred in every period. As can be seen, the transition probability of Type 0 is high; there is a higher probability of minimum comparable areas and their spatial lag to remain in the same classification. Additionally, this probability increases each decade indicating a low mobility between the classes. Specifically, between 1970 and 1980 the probability of transition Type 0 was 86.89%, and between 2000

[9] The listing of these 48 municipalities is in Table A1 in the Appendix.

and 2010 this probability increases to 96.91%. This evidence of stability in the population distribution behavior over time corroborate with the normal non-parametric kernel density functions estimates (Figure 1) and with the summary statistics (Table 4).

Table 13 – LISA Spatial Transitions (Decennial) for Brazilian MCAs population, 1970 – 2010

Interval	Type of Transition					Cohesion	Flux
	Type 0	I	II	IIIA	IIIB		
1970 to 1980	0.8986	0.0443	0.0533	0.0038	0.0000	0.9024	0.1014
	UP	0.0139	0.0128	0.0003	-		
	DOWN	0.0303	0.0404	0.0036	-		
1980 to 1991	0.9322	0.0295	0.0358	0.0019	0.0005	0.9341	0.0678
	UP	0.0098	0.0123	0.0005	-		
	DOWN	0.0197	0.0235	0.0014	-		
1991 to 2000	0.9516	0.0191	0.0292	0.0000	0.0000	0.9516	0.0484
	UP	0.0096	0.0093	0.0000	-		
	DOWN	0.0096	0.0200	0.0000	-		
2000 to 2010	0.9691	0.0120	0.0189	0.0000	0.0000	0.9691	0.0309
	UP	0.0046	0.0063	0.0000	-		
	DOWN	0.0074	0.0126	0.0000	-		

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

Other interesting results can be obtained by Type II transitions, which involves a transition of only the neighbours in relative space, but the locality in question remains in the previous state. The probability of this transition type, as expected, decreases with each decade. As in the previous analysis, there is a greater probability of downward movement, i.e. higher probabilities of most populous neighbours become less populated.

Similarly to the exercise for different time intervals, we can investigate further highly populated communities that propelled the neighbours, i.e., MCAs that were populated above average and had less populous neighbours, while in the following period the neighbours became highly populated. Of course, in the interval between 1970 and 1980, we identified the same 14 MCAs (48 municipalities) that played this role. Between 1980 and 1991, 13 MCAs (25 municipalities) played this role, ten of these municipalities are located in Pará state (North); 1 in Rio Grande do Norte, 1 in Paraíba and 6 in Bahia (Northeast); 2 in Minas Gerais and 3 in São Paulo state (Southeast); and 2 in Santa Catarina state (South). Of these 25 municipalities, only 4 are capitals (Belém, Natal, João Pessoa and Salvador). Between 1991 and 2000, only 4 MCAs (5 municipalities) played this role, 1 in Sergipe (Northeast); 2 in Minas Gerais and 1 in São Paulo state (Southeast); and 1 in Santa Catarina (South). At this period, Sergipe's state capital (Aracaju) played this role of highly populated community that propelled the neighbours. Finally, between 2000 and 2010, also only 4 MCAs (8 municipalities) played this role, all of them in Northeast, 4 in Maranhão state (including the capital São Luís) and 4 in Ceara state[10]. Again, investigate more deeply the particularities of these municipalities is an interesting suggestion for future research.

In relation to Type I transition, which occurs when only the locality moves, but its neighbours remain in the same category, this is less likely than Type II and is decreasing over the decades. Regarding the probability of Type IIIa, which occurs when both MCA and neighbors move in the same direction in the distribution, the results indicate that the probability of such transition became null between the decades of 1991 and 2000 and between 2000 and 2010. That is, in the last 20 years there were no transitions of minimum comparable areas together with their neighborhood within the urban hierarchy. Finally, the cohesion and flow measurements are decreasing over time. In other words, over the time the probability of the MCAs move in the same direction of its spatial lag between different classes decreases, and the flux measurement indicates that there is a decreasing instability in behaviour of MCAs relative to its neighbours in the population distribution, as already evidenced.

Spatial Markov Chains

From the traditional Markov matrix, Rey (2001) suggests an extending modification, so that the transition probabilities are conditioned on the initial relative population class of its spatial lag. The spatial Markov matrix, as called by Rey, speaks to the question of whether a locality's transition in the relative population distribution is related to the relative population of its neighbours. As explained by Rey, the spatial provides a great deal of information regarding the transitions of regions and the possible association between the direction and rate of the transitions and the regional context faced by each economy.

A spatial Markov transition probability matrix was constructed to analyse the spatial-temporal dynamics of relative population distribution, i.e., considering the possible influence from neighbours on the transition of regions. The standard contiguity neighbours matrix (W) is used for estimation of following spatial Markov matrices, reported in Table 14.

---

[10] The listings of municipalities for each decennial interval are in Table A1 in the Appendix.

Table 14 – Spatial Markov Transitions Probabilities Matrix for Brazilian MCAs population, 1970-2010

Spatial Lag	t_i	t_i+1			
		1	2	3	4
1	1	0.9546	0.0448	0.0006	0.0000
	2	0.1088	0.8391	0.0522	0.0000
	3	0.0000	0.1088	0.8407	0.0505
	4	0.0000	0.0000	0.0507	0.9493
2	1	0.9096	0.0894	0.0010	0.0000
	2	0.0850	0.8470	0.0670	0.0009
	3	0.0011	0.1105	0.8475	0.0409
	4	0.0000	0.0000	0.0700	0.9300
3	1	0.8866	0.1134	0.0000	0.0000
	2	0.0559	0.8410	0.1030	0.0000
	3	0.0009	0.0986	0.8489	0.0516
	4	0.0000	0.0000	0.0811	0.9189
4	1	0.8628	0.1326	0.0000	0.0047
	2	0.0631	0.7474	0.1836	0.0059
	3	0.0000	0.0761	0.8073	0.1166
	4	0.0000	0.0000	0.0657	0.9343

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

Initially, some evidences can be seen from Table 14. Firstly, spatial background appears to play an important role in the dynamics of relative population distribution. In other words, the neighbours of a MCA have an impact on its transition probabilities over time. If the spatial context did not work, then the four conditional matrices should be the same and are equal to the traditional Markov matrix (Table 10). But in fact it is the opposite. Specifically, a chi-squared test of the difference between each of the spatial conditioned transition submatrices against the overall (a-spatial) transition matrix rejects the null hypothesis that these matrices are equal at 1%[11]. Secondly, different spatial contexts have different effects on transition for regions. Specifically, the probability of upward transitions will increase for MCAs with neighbours in high classes. For example, for a MCA in the first quartile with neighbours in the same class, the probability of moving upward is 4.48%, while if it is adjacent to localities in fourth quartile this probability increases to 13.26%. A similar phenomenon to this occurs also for MCAs originally in classes 2 and 3.

In table 14, as in Table 10, the medium classes (2 and 3) have more probability of inter-class mobility than extreme classes (1 and 4). However, considering the spatial dimension, we can observe that the MCAs grouped into medium classes have a higher probability of a downward transition if your neighbours are in a less populated class (class 1). The opposite happens if the neighbours are the most populous class (class 4). This evidence

[11] Table A2 in appendix.

that the MCAs in the third quartile is more likely to move within the distribution to a larger class when they are near the most populated places, certainly means that there is an overflow of the population of large cities to medium-sized cities. The latter evidence highlights again the major role of medium-size localities in the processes of urban agglomeration that occurred in Brazil during the last 40 years, even considering the spatial context.

Furthermore, as suggested by Rey (2001), it is possible to know the influence of spatial dependence on the transition probability comparing the elements of a traditional transition matrix with the elements of the spatial Markov matrix. For example, ignoring spatial context (Table 10), the probability of a MCA in the third quartile to move down to the second quartile is 9.71%, this probability rises to 10.88% if the neighbours are in the first quartile (less populated class). We can also observe, by comparing the traditional and spatial matrix, that the less populous MCAs that have highly populated neighbours decreases the probability of persistence in the same class distribution. Specifically, ignoring the spatial context, the probabilities of MCAs to in the first and second quartiles are 92.08% and 82.49%, respectively. These probabilities are reduced to 86.28% and 74.74% when high-populated neighbours surround these locations. Additionally, we can explore steady state distribution implied by each estimated conditional transition probability matrix from Table 14, calculated as steady state distribution that was defined in section 2.3. The steady state distribution spatially conditioned is presented in Table 15.

Table 15 – Steady State Distribution for Brazilian MCAs population, 1970 – 2010

Spatial Lag	(%)			
	1	2	3	4
1	0.5472	0.2284	0.1125	0.1119
2	0.3225	0.3401	0.2101	0.1273
3	0.1570	0.3130	0.3239	0.2062
4	0.0537	0.1167	0.2939	0.5357

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

The long run distribution for MCAs with neighbouring relatively less populated (class 1) has 54.72% of localities in the first quartile and 11,19% in the fourth quartile, for example. On the other hand, the long run distribution for MCAs with neighbouring relatively high populated (class 4) has just 5.37% of localities in the first quartile and 53.57% in the fourth quartile. According to Gallo and Chasco (2007), concentration of the frequencies in some of the classes, that is, a multimodal limit distribution, may be interpreted as a tendency towards stratification into different convergence clubs. As can be noted in main diagonal, there would be a higher concentration of frequency on a particular class according to the spatial lag that can be an evidence of different convergence clubs according to spatial lag.

Finally, we can determine the speed with which the urban municipalities move within the relative population distribution, conditioned to spatial lag. Table 16 displays the expected time for a locality to move from class *i* to class *j* for the first time based on equation (9) for each submatrices in Table 14, i.e., conditional on quartile in population distribution of its neighbours.

Table 16– Spatial Markov First Mean Passage Time in Decade for Brazilian MCAs population, 1970 - 2010

Spatial Lag	t_i	t_i+1			
		1	2	3	4
1	1	1.83	22.27	84.52	240.90
	2	17.98	4.38	63.34	219.72
	3	36.31	18.33	8.89	156.38
	4	56.02	38.04	19.71	8.93
2	1	3.10	11.23	39.10	135.51
	2	23.10	2.94	28.35	124.74
	3	37.06	14.30	4.76	97.93
	4	51.35	28.59	14.29	7.86
3	1	6.37	8.82	23.31	70.82
	2	47.37	3.20	14.49	62.00
	3	63.37	16.52	3.09	47.51
	4	75.71	28.86	12.34	4.85
4	1	18.62	9.04	15.78	26.86
	2	126.64	8.57	8.26	20.26
	3	163.11	36.46	3.40	13.19
	4	178.33	51.68	15.22	1.87

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

As can be seen in Table 16, MCAs with relative population in the first quartile with neighbours in the first quartile return to the first quartile after 1.83 decades, after leaving the first quartile. This time is 18.62 decades for localities in the first quartile with neighbours high populated. Furthermore, MCAs in first class with neighbours also in the first quartile enter the third class 84.52 decades after leaving the first quartile, on average. On the other hand, this time frame falls to 23.31 decades if the spatial lag is in third quartile.

In this section, we presented an overview of the population distribution of Brazilian minimum comparable areas between 1970 and 2010. The Zipf's law estimation indicates that the population distribution is, every decade, moving away from Pareto law. The traditional Markov chain approach brings as main evidence the high probabilities on the main diagonal indicating a low interclass mobility. Finally, the mobility rates for MCAs in the population distribution were found to be sensitive to the relative positions of their neighbors in the same distribution.

4.2 Spatial Evaluation of Population Growth Dynamics of Brazilian MCAs: Results

In this chapter we present the data implementation followed by the results for population growth equations developed in a spatial economic model of population growth dynamics in Brazilian MCAs section. The objective of this empirical analysis is to assess the determinants of population growth of Brazilian MCAs between 1970 and 2010 and to examine the existence and magnitude of spatial interaction and spatial spillover effects associated with these determinants.

4.2.1 Data Implementation

In order to estimate the population growth equation developed in chapter 3, the main source of data from the Brazilian Demographic Census is complemented with data collected by the Brazilian Institute for Applied Economic Research (IPEA). Due to an ongoing process of changes in the number, area, and borders of municipalities, this dataset is also based on the aggregation through minimum comparable areas (MCAs). Thus, this study is able to cover a spatial panel of 3659 MCAs over the 1970-2010 period. Therefore, if we talk about cities or municipalities below, we are actually referring to MCAs.

Based on the theoretical model set out in chapter 2, literature review and data availability, population growth rate is further taken to depend on the population growth rate in the previous decade and thirteen explanatory variables, which will be further explained below. Although Brazil is an emerging economy rather than a developed country, and we consider population growth in both urban and rural areas so as to be able to model spatial interaction effects, the explanations put forward based in literature review remain helpful in selecting the explanatory variables for this study. The main difference is that sometimes the variables selected need to be placed in a different context, in relation to studies for developed countries. For example, a variety of studies includes variables representing the human capital in an attempt to explain population growth, among them, Glaeser and Mare (1994), Nardinelli and Simon (1996), Simon (1998), Da Mata et al (2007), Chi (2009), Chi and Vos (2011), Duranton and Puga (2013). Educational attainment can be related with local economic dynamism, ability to absorb and generate innovation, and adaptability. Whereas Duranton and Puga (2013, p.39) observe a tendency in the U.S. literature to measure human capital by the share of university graduates, this study focuses on the share of people aged 25 years and over that is literate, which increased from 48% in 1970 to 82% in 2010.

Regarding demographic characteristics, we consider lagged population growth rate, rural population, density, mean age, and birth rate. As explained in the literature review, in Da Mata et al. (2007), the market potential variables are inversely related to transport costs, which in turn are linearly related to distance, without further testing whether this degree of distance decay is faster or slower. In this sense, just as in Da Mata et al. (2007), the population growth rate is therefore taken to depend on GDP per capita, rural GDP per capita, rural population size, and their spatially lagged values, but then as separate variables in order to test for agglomeration effects. Population density is used to control for housing supply and (dis)amenities, two of key drivers of city growth identified by Duranton and Puga (2013). Many cities have the capacity to receive more people. However, since it takes time to build up certain public goods, infrastructure, or housing, the cost of living increases, the residents of quickly growing cities may suffer in terms of quality of life and might also deter prospective migrants. Since each age group in a population behaves differently, and the distribution across age groups changes over time, economic opportunities may be boosted or slowed down temporarily. This leads us to believe that it is also important to control for variables such as mean of age. In addition, if the population of a particular region is relatively immobile, differences in population growth among areas within that region are mainly due to difference in fertility (Glaeser et al., 1995), which is an important reason to consider the birth rate.

Percentage of population working on agricultural sector and relationship between the number of employees in the manufacturing industry and the service sector were measures for industrial composition. Moreover, the percentage of the economically active population that was occupied was also used in this analysis. These variables provide information about types of jobs available, and productive structure. According to Glaeser et al (1995), employment variables can reflect how workers respond to business cycle shocks; can be a proxy for human capital, in which the percentage of employed may reflect that there are skills needed in the workforce; and low percentage of workforce occupied may generate negative amenities. We expect that the share of employment in agriculture will have a negative effect on population growth due to less economic opportunities, especially for women.

According to Duranton and Puga (2013), apart from housing supply, infrastructure and (dis)amenities are key drivers of city growth. Cities infrastructures are present in this analysis because they can capture factors of attractiveness for prospective migrants and housing qualities that might lead prices up. The variables used are the following: percentage of households supplied by water company and percentage of households supplied by sewer company. Considering the hypothesis that crime rates may be disamenity that negatively influences the decision to settle in a particular locality. Just as in Da Mata et al (2007), we use the homicide rate, the number of homicides per 100,000 inhabitants, to represent crime rate.

The following Table 17 describe the means and standard deviation for these variables. Note that dependent variable is the rate of population growth in the decade and we have no information for the decade of 1960, so there is no information for the population growth rate between 1970 and 1980. Also, there is no information available about the homicide rates for the years 1970 and 2010, but as we shall see, the estimated equation does not need the explanatory variables for 1970 and 2010.

Among the information generalized to the country as a whole, some stand out: (i) the reduction of the rural population over the 4 decades, (ii) starting in 1991, an increase in density, i.e. more people sharing the same area, (iii) in this period there was also an expressive increase in the literacy rate, which whent from 48% in 1970 to 82% in 2010, (iv) reduction in the percentage of people employed in agriculture, (v) increase in the logarithm of GDP per capita between 1970 and 1980 and between 2000 and 2010, (vi) increase in the percentage of households with piped water from 14% in 1970 to 71% in 2010, while, the percentage of homes with sewer system goes out of 5 % to 37% in the same period - these two variables reflects improvements in infrastructure and basic health conditions that occurred in Brazil in the last 40 years.

Table 17 - Means and Standard Deviations of Variables for Brazilian MCAs, 1970 - 2010

Explanatory Variables	1970		1980		1991		2000		2010	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
population growth rate	-	-	0.09	0.27	0.12	0.21	0.08	0.15	0.07	0.12
ln rural population	8.87	1.04	8.69	1.25	8.5	1.38	8.32	1.43	8.19	1.5
ln density	1.8	1.2	1.8	1.3	1.4	1.2	1.6	1.2	1.8	1.2
mean age	22.9	1.7	24.2	2	25.6	2.4	28.1	2.7	32	2.9
birth rate	0.67	0.72	0.76	0.97	0.9	1.23	0.39	0.4	0.33	0.35
literacy rate	0.48	0.17	0.58	0.19	0.66	0.19	0.76	0.15	0.82	0.13
agriculture	0.32	0.11	0.29	0.12	0.24	0.11	0.2	0.11	0.17	0.1
manufacte/service	1.02	1.85	1.15	1.63	0.8	0.98	1.74	1.99	0.95	1.02
workforce ocuppied	0.99	0.02	0.98	0.03	0.96	0.03	0.88	0.06		
ln of GDP per capita	0.12	0.81	0.9	0.81	0.82	0.84	1.12	0.71	1.38	0.67
ln of rural GDP per capita	-0.89	0.98	-0.16	1.1	-0.43	1.29	-0.89	1.25	-0.71	1.28
homicide rate	-	-	3.8	44.1	8.3	92.7	12.4	135.4	14.25	78.75
water company	0.14	0.18	0.29	0.24	0.48	0.23	0.61	0.2	0.71	0.17
sewer company	0.05	0.11	0.11	0.19	0.18	0.26	0.29	0.3	0.37	0.32

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.

4.2.2 Population Growth of Brazilian MCAs

In this section, we examine how the characteristics of MCAs were correlated with subsequent population growth rates. Starting with equations (31), (33), and (35), the dependent variable of our empirical analysis, Y_{it}, is measured by the rate of population growth in one particular MCA over one decade (t-1,t), where i runs from 1 to 3659, t in equation (35) runs from 1980 to 2010 and the number "1" represents a decade. Population growth rate is taken to depend on the population growth rate in the previous decade, and when the dynamic spatial Durbin model is adopted also on the population growth rate in neighboring units in the contemporaneous and previous decades.

The estimation results are presented in Table 18. The first column reports the estimation results of a standard linear panel data model extended to include spatial and time-period fixed effects, but without any spatial interaction effects. The estimation results of the dynamic SDM specification are recorded in the second column of Table 18. We first discuss the results of several specification tests before we turn to the estimation results reported in the second column of Table 18.

Table 18 – Population Growth of Brazilian MCAs: Non-Spatial and Dynamic Spatial Models

Explanatory Variables	OLS		Dynamic SDM + Fixed Effects (bias correction)			
	Coeff	t	Coeff	t	Spatial	t
lagged population growth rate	-0.0271	**	0.0755	**	0.0681	**
ln rural population	-0.0433	**	-0.0391	**	0.0068	
density	-0.1248	**	-0.1256	**	-0.0221	**
mean age	0.0135	**	0.0089	**	-0.002	
birth rate	0.0172	**	0.015	**	0.0072	**
literacy rate	0.1361	**	0.0681	**	0.0395	
agriculture	-0.2612	**	-0.2315	**	0.1063	**
manufacturing/service	0.0045	**	0.0021	**	0.0016	
occupied workforce	0.4911	**	0.3535	**	-0.0681	
ln GDP per capita	0.0513	**	0.0527	**	-0.0248	**
ln rural GDP per capita	0.0088	**	0.0135	**	-0.0095	**
homicide rate	-0.003	**	0.0006		-0.0042	*
water company	0.0081		0.0274		-0.0255	
sewer company	-0.0123		-0.0365	**	-0.0058	
WY (delta)					0.3439	**
No. Obs.	10977		10977			
R-squared	0.711		0.743			
Log Likelihood	11144		5580.37			
Spatial lag, OLS model:						
LM	909.32	**	Spatial lag, SDM model:			
LM(robust)	114.89	**	*Wald*		54.39	**
Spatial error, OLS model:						
LM	796.34	**	Spatial error, SDM model:			
LM(robust)	1.91		*Wald*		134.23	**
Joint significance						
LR(spatial fe=0)	8674.6	**				
LR(time fe=0)	789.06	**				

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010. ** and * significant at 1% and 5%, respectively.

To investigate the (null) hypothesis that the spatial fixed effects are jointly insignificant, we performed a likelihood ratio (LR) test. The results (8674.34, with 3658 degrees of freedom [df], $p < 0.01$) indicate that this hypothesis must be rejected. Similarly, the hypothesis that the time-period fixed effects are jointly insignificant must be rejected (789.06, 3 df, $p < 0.01$). These test results justify the extension of the model with spatial and time-period fixed effects.[12]

---

[12] Table A3 in appendix reports the correlation coefficients between the explanatory variables to show that multicollinearity in this empirical study is not a problem.

Additionally, to test whether this non-spatial model with spatial and time-period fixed effects should be extended with spatial interaction effects among the dependent variable (SAR specification) or spatial interaction effects among the error terms (SEM specification), we use LM tests applied to a first-order binary contiguity neighborhood matrix. In this spatial weight matrix, the w_{ij} are equal to 1 to contiguity neighbors and the diagonal elements have values set to zero, since no spatial unit can be viewed as its own neighbor. Furthermore, the matrix is normalized such that the sum of the elements of each row is equal to 1. These LM tests follow a chi-squared distribution with one degree of freedom and have a critical value of 3.84 at 5% significance or 2.71 at 10% significance. When using the classic LM tests, both the hypothesis of no spatially lagged dependent variable and the hypothesis of no spatially autocorrelated error term must be rejected. When using the robust tests, the hypothesis of no spatially lagged dependent variable must still be rejected. However, the hypothesis of no spatially autocorrelated error term cannot longer be rejected, also at 10% significance. These test results point to an extension of the non-spatial model with a spatially lagged dependent. In other words, the tests indicate that spatial dependence should be considered when estimating the minimum comparable areas' population growth; therefore to use standard OLS would prejudice the results. However, if a non-spatial model on the basis of (robust) LM tests is rejected in favor of the spatial lag model or the spatial error model, one should be careful to endorse one of these two models. LeSage and Pace (2009, Ch. 6) and Elhorst (2012) recommend to also consider the spatial Durbin model and then test whether or not this model can be simplified to the spatial lag or spatial error model. In this study, however, we take a broader view by applying the Bayesian approach set out in chapter 3. Firstly, we calculate the Bayesian posterior model probabilities of the SDM and SDEM specifications, as well as the simpler SAR and SEM specifications, to find out which model specification best describes the data. Secondly, we repeat this analysis for several specifications of the neighborhood matrix to find out which specification of $\boldsymbol{W}$ best describes the data. The results are reported in Table 19.

Table 19 – Comparison of model specifications and neighborhood matrices

W Matrix	Statistics	SAR	SDM	SEM	SDEM
Binary Contiguity	log marginal	3566.85	3616.03	3548.42	3611.80
	model probabilities	0.0000	0.9855	0.0000	0.0145
First and Second Order	log marginal	3562.21	3574.79	3558.60	3579.41
	model probabilities	0.0000	0.0097	0.0000	0.9903
First, Second and Third Order	log marginal	3527.98	3528.75	3535.86	3536.28
	model probabilities	0.0001	0.0003	0.3974	0.6022
Inverse distance	log marginal	3368.78	3444.87	3363.32	3455.44
	model probabilities	0.0000	0.0000	0.0000	1.0000
5 nearest neighbors	log marginal	3539.69	3601.04	3521.72	3597.88
	model probabilities	0.0000	0.9594	0.0000	0.0406
6 nearest neighbors	log marginal	3551.02	3613.06	3539.41	3613.60
	model probabilities	0.0000	0.3676	0.0000	0.6324
7 nearest neighbors	log marginal	3548.94	3606.39	3537.52	3606.54
	model probabilities	0.0000	0.4622	0.0000	0.5378
8 nearest neighbors	log marginal	3551.30	3607.94	3541.97	3610.07
	model probabilities	0.0000	0.1054	0.0000	0.8946
9 nearest neighbors	log marginal	3561.30	3610.94	3553.84	3613.93
	model probabilities	0.0000	0.0474	0.0000	0.9526
10 nearest neighbors	log marginal	3560.11	3607.68	3556.60	3609.52
	model probabilities	0.0000	0.1373	0.0000	0.8627
20 nearest neighbors	log marginal	3526.87	3552.07	3534.30	3552.99
	model probabilities	0.0000	0.2853	0.0000	0.7147

Notes: Calculations by the Author based on LeSage (2013).

The results show that the SAR and SEM models are always outperformed by either the SDM or the SDEM specifications. This indicates that spatially lagged explanatory variables (*WX)* are important and should be included in the model. By also considering the log-marginals of the different specifications of the spatial weight matrix, the worst performing spatial weight matrix appears to be the inverse distance matrix, corroborating with our observation made in the previous section that to decompose market potential variables into their underlying components, as well as to consider spatially lagged values of these components lead to a much greater degree of empirical flexibility. One can see that the first-order binary contiguity matrix and the SDM specification give the best performance of all 44 combinations. For this reason, we decided to estimate the dynamic SDM specification using the bias corrected ML estimator developed by Lee and Yu (2010).[13] The estimation results are recorded in the second column of Table 18. The results of this model are finally used to test

[13] This bias correction is needed since the dependent variables lagged in time (Y_{t-1}) and in both space and time (WY_{t-1}) at the right-hand side of (35) is correlated with the spatial fixed effects μ. This is the spatial counterpart of the Nickell (1981) bias, as shown by Yu et al. (2008) and Lee and Yu (2010) respectively for a dynamic spatial panel data model without and with time-period fixed effects.

whether the dynamic spatial Durbin can perhaps be simplified to the dynamic spatial lag model or the dynamic spatial error model. Both tests follow a chi-squared distribution with $K+1$ degrees of freedom (the number of spatially lagged explanatory variables and the spatially lagged dependent variable) and take the form of a Wald test, since these simplified models themselves have not been estimated. The results point out that both hypotheses need to be rejected. In conclusion, we can say that the empirical results point to the utility function specified in equation (32), which posits that the utility of individuals is also negatively correlated with the level of population (population size) and with the population growth rate of their neighbors, and the global spillover model, which posits that $\delta \neq 0$.

The results reported in second column of Table 18 show that six of the thirteen spatially lagged explanatory variables in the dynamic SDM specification appear to be statistically significant at the 5% significance level. Moreover, the coefficients of the spatially lagged dependent variable at time t and t-1, $\boldsymbol{WY}_t$ and the $\boldsymbol{WY}_{t-1}$, are also significant. Besides, a necessary and sufficient condition for stationarity is $\tau+\delta+\eta$=0.0755+0.3439+0.0681=0.4875 < 1, which is satisfied.

The spatial models explore a sophisticated dependence structure among spatial units. As consequence, the parameter estimates contain a wealth of information on relationships among the observations. This rich set of information also increases the difficulty of interpreting the resulting estimates (LeSage and Pace, 2009). Unlike the non-spatial models, the estimated coefficients on spatial models do not represent the marginal effects of changes in the explanatory variables on the dependent variable. According to Elhorst (2010), the denial of this characteristic of spatial models lead some empirical studies to erroneous conclusions, while the interpretation of the partial derivative of the impacts of a change in the explanatory variables represents a more valid basis for testing the hypothesis of spatial spillovers. Table 20 reports the short-term and long-term estimates of the direct, indirect and total effects, derived from the parameter estimates of this model using equations (36) and (37). The direct effects represent a change in the dependent variable associated with a change in the explanatory variable of the observational unit itself. Additionally, the indirect effects represent a change in the dependent variable associated with a spatially lagged explanatory variable. The indirect effects are properly the effects of spatial spillovers and, as has been emphasized during the text, its study is one of the main interests in this thesis. In order to draw inferences regarding the statistical significance of these effects, we used the variation of 100 simulated parameters combinations drawn from the variance-covariance matrix implied by the maximum likelihood estimates. The number of explanatory variables producing significant spatial spillover effects appears to be 3 in the short term and 6 in the long term. In other words, we find evidence of spatial spillovers in the determinants of population growth in Brazilian cities between 1970 and 2010. The reason that these numbers are lower than the number of significant spatial interaction effects is because they depend on more than just one single parameter, namely three parameters in the short term and five parameters in the long term, see equations (36) and (37).

Table 20 – Dynamic Spatial Model – Direct, Indirect and Total Effects

Explanatory Variables	Short Term Effects			Long Term Effect		
	Direct	Indirect	Total	Direct	Indirect	Total
lagged population growth rate				-0.918	0.137	-0.781
				(-113.146)	(7.030)	(-40.167)
ln rural population	-0.04	-0.011	-0.05	-0.043	-0.021	-0.064
	(-14.351)	(-1.253)	(-5.487)	(-13.992)	(-1.967)	(-5.548)
Density	-0.131	-0.092	-0.223	-0.145	-0.141	-0.286
	(-24.052)	(-6.158)	(-14.465)	(-22.473)	(-6.720)	(-12.907)
mean age	0.009	0.002	0.01	0.01	0.004	0.013
	(7.083)	(0.593)	(4.448)	(7.142)	(1.157)	(4.531)
birth rate	0.016	0.019	0.035	0.018	0.027	0.044
	(9.274)	(3.022)	(5.479)	(9.518)	(3.373)	(5.475)
literacy rate	0.074	0.103	0.177	0.083	0.143	0.226
	(2.325)	(1.558)	(2.464)	(2.384)	(1.733)	(2.480)
Agriculture	-0.227	0.037	-0.189	-0.247	0.004	-0.242
	(-8.076)	(0.612)	(-2.967)	(-8.093)	(0.058)	(-2.967)
manufacturing/services	0.003	0.003	0.005	0.003	0.004	0.007
	(2.200)	(0.944)	(1.621)	(2.240)	(1.061)	(1.625)
occupied workforce	0.359	0.08	0.439	0.394	0.167	0.561
	(10.297)	(0.905)	(4.423)	(10.353)	(1.506)	(4.470)
ln of GDP per capita	0.052	-0.009	0.043	0.057	-0.001	0.055
	(12.769)	(-0.875)	(3.986)	(12.649)	(-0.110)	(3.955)
ln of rural GDP per capita	0.013	-0.008	0.005	0.014	-0.008	0.006
	(5.647)	(-1.489)	(0.884)	(5.607)	(-1.139)	(0.885)
homicide rate	0.001	-0.006	-0.005	0.000	-0.007	-0.007
	(0.447)	(-1.842)	(-1.576)	(0.356)	(-1.804)	(-1.570)
water company	0.026	-0.026	0.000	0.028	-0.028	0.001
	(1.938)	(-0.846)	(0.010)	(1.927)	(-0.724)	(0.017)
sewer company	-0.036	-0.028	-0.064	-0.04	-0.043	-0.082
	(-2.660)	(-1.179)	(-2.594)	(-2.711)	(-1.403)	(-2.560)

Notes: Estimates by the Author. t-values in parentheses.

The long-term direct, indirect and total effect estimates of the growth rate represent significant convergence and deconcentration effects. Specifically, a 1% increase in the population growth rate from one location reduces the rate of population growth in the long run in 0.918 percentage point. In other words, the greater population growth in the own MCA has been in the previous decade, the smaller it will be in the next decade, and vice versa. This evidence points to conditional convergence in population growth rate of Brazilian MCAs. The localities that grow at higher rates in the past have reduced population growth rate in the future. The indirect effect of 0.137 is statistically significant at 1%. Increases in population growth in the neighborhood in the previous decade stimulate population growth in the future. This movement or deconcentration of people to neighboring areas, perhaps to escape the congestion of the city, is a convergence effect. However, the feedback effect of this behavior is that the city starts growing again, as a result of which the total convergence effect becomes smaller. This explains the reduction of the convergence effect from -0.918 to -0.781. These results are absolutely unknown in Brazilian literature about population growth; they indicate a potential force for deconcentration in the country's population in the past 40 years.

The direct effect of the natural logarithm of rural population is negative and highly significant in both short and long term. If a locality has a large rural population at the beginning of a decade, this will lead to a reduction in population growth after 10 years. Specifically, an increasing of 1 per cent in rural population in a particular MCA reduces the population growth rate at 0.04 percentage point in the short term and 0.043 percentage point in the long term. This result is similar to that found by Da Mata et al. (2007) and is consistent with the country urbanization process for the period, as mentioned in the introduction. The indirect effect is also negative and statistically significant at 5% in long term, suggesting that municipalities that have a large rural population in the neighbourhood will have a lower response in terms of population growth.

The direct effect of log of density on population growth is negative and significant, with magnitude greater in the long run than in the short-term. If a municipality has a population density increased, this will reduce the growth in population in the next decades, corroborating with the hypothesis that densely populated cities deter prospective migrants due to deteriorating living conditions. Specifically, if population density increases by 1%, population growth will be reduced by 0.13 percentage point in long term. Moreover, this adverse effect also spills over to neighboring MCAs. The indirect effect is also negative and highly significant. If there is an increase in population density in the neighbourhood of a particular MCA, there will be a decrease in population growth in this locality. If the spatially lagged population density increases by 1%, there will be a decrease of 0.09 percentage point in population growth in short term and 0.14 percentage point in long term. This evidence reinforces the analysis of spatial spillovers for studying population growth of cities.

The direct effect of the mean age of the population is positive and significant. If the mean age of the population increases by one year, the population growth rate increases by 0.01 percentage points. Since the mean age over the observation period increased from 23 in

1970 to 32 in 2010, this finding corroborates with the view that economic opportunities may boost when the number of working-age adults grows large relative to the dependent population. The birth rate variable has a direct effect on the population growth rate positive and significant, in both short and long term. If the birth rate increased by 1 child for every 1000 inhabitants in a given area, in the short term there will be an increase in the population growth rate of 0.016 percentage point and in the long run this rate will be increased by 0.018 percentage point. The indirect effect of birth rate is positive and significant in both short and long terms. That is, the population growth rate of Brazilian MCAs is positively affected by the amount of children born in the neighbourhood.

The direct effect of literacy rate is positive and significant. Areas that have a higher percentage of literate people in the initial period have a higher population growth rate in both the short and long term. Specifically, a 1% increase in literacy rate in a particular locality would raise the population growth rate of unit itself at approximately 0.072 and 0.079 percentage points in short and long run, respectively. The positive relationship between educational attainment and population growth is in line with Glaeser and Saiz (2003) and Da Mata et al (2007) when they argue that economies with more educated people generate positive amenities and are more adaptable to technological change due to spillover effect of knowledge, becoming more attractive places. The indirect effect is positive and statistically significant at 10% significance. The positive relationship between educational attainment and population growth in its surroundings is in line with the proposition introduced in equation (16) that knowledge accumulated in one economy depends on knowledge accumulated in other economies.

The percentage of people employed in activities related to agriculture has a negative and significant direct effect. The minimum comparable areas that had in the initial period a high percentage of people employed in agriculture experience a reduction in the rate of population growth in both short and long term due to less economic opportunities, especially for women. The indirect effect for this variable is positive and not significant. This result suggests that the percentage of people working in agriculture in a particular geographical area does not affect the population growth of their neighbourhood. The relationship between the number of people employed in the manufacturing sector and the number of people employed in the service sector have direct effects positive and statistically significant in both short and long terms. This result is in line with Da Mata et al. (2007) and suggests that areas with a high percentage of manufacturing in base period experienced a population growth faster. The indirect effect for manufacture/service variable is negative and not significant. The direct effect of percentage of occupied workforce is positive and significant. An increase in the number of workers employed in the initial period of a particular MCA exerts a positive effect on population growth rate of the locality. Specifically, a 1% increase in the percentage of employed in a particular locality in the long term its population growth rate will increase by 0.39 percentage points. The spillover effect is also positive but not statistically significant.

The direct effect of logarithm of per capita GDP is positive and significant. An increase in GDP per capita in a given locality results in an increase in its growth rate, in the short and long term. Specifically, a 1% increase in GDP per capita increases the population growth rate by 0.05 and 0.06 percentage point in short and long term, respectively. The spillover effect is not significant. The rural GDP per capita has a direct effect positive and statistically significant, indicating that municipalities that offer income opportunities remain attractive areas to live in. A reduction in rural GDP per capita of a location would reduce the population growth rate in the short and long term. The indirect effect has a negative sign and is not statistically significant. Da Mata et al. (2007, p. 266) reported their rural variables to perform poorly due to limited variation and multicollinearity. By decomposing the market potential variables, these problems do not occur in this study.

The homicide rate has statistically insignificant direct effects on the population growth rate. The spillover effect is negative and statistically significant at 10%, indicating that especially the surroundings of a city pay the price for this disamenity. This result is an evidence that neighbours amenities influences the utility of individuals and, consequently, the population growth of localities. In other words, MCAs with neighbouring areas presenting high crime ratio have a reduction in population growth rate, both in the short and long term. Regarding infrastructure, the share of households supplied by Water Company presents a positive and statistically significant direct effect. Specifically, the percentage of households supplied by water company of a particular MCA increase by 1% in base period, this area will have 0.026 and 0.028 percentage point increase in the rate of population growth, in short and long term, respectively. Moreover, the spillover effect is negative and not statistically significant. The share of households supplied by adequate public regular sewer system has negative and statistically significant direct effect. An increase in the share of households supplied by Sewer Company in a given MCA reduces its own population growth rate in the short and long term. One possible reason for the negative sign of direct effect is that the variable could be a proxy for the price of urban space, which affects the location of people and firms, especially at municipalities level. If the supply of housing with access to public sewer is relatively inelastic, the prices of this type of housing might increase so much that prospective migrants might be discouraged and the population growth rate falls. According to research from Economic Benefits Expansion of Sanitation (FGV, 2010), sanitation enables constructions with higher added value and appreciation of existing buildings. The indirect effect (spillover) for this variable also appears as negative, but is statistically insignificant.

Due to these significant indirect effects, it is interesting to compare the long-term total effects reported in Table 20 that have been derived from the dynamic SDM specification with those from the non-spatial model reported in the first column of Table 18. The long-term total effect of the non-spatial model can be obtained by $\beta/(1-\tau)$, where β is the coefficient estimate of a particular explanatory variable and τ the coefficient estimate of the dependent variable lagged one decade in time. The long-term total effect of the rural population according to the spatial model amounts to -0.064 and according to the non-spatial model to -0.0433/(1-(-0.0271)=-0.0422. This means that the latter effect in the non-spatial model is underestimated by 34.1%. For the other variables that produce significant spatial spillover effects, we find 57.5% for population density, 61.9% for the birth rate, 41.4% for the literacy rate, and 58.3% for the homicide rate. These results are reported in Table 21. On average, the degree of underestimation amounts to 27% taken over all explanatory variables, indicating that a non-spatial modeling approach does not reflect the full impact of policy measures that act on these variables.

Table 21 – Long-term Effects Comparison

Explanatory Variables	Non-spatial Coefficients	Long-term effect in Non-Spatial model $\beta/(1-\tau)$	Total Effect in Spatial model	Underestimation long-term effect in Non-Spatial model (%)
lagged population growth rate	-0.0271	-	-0.781	-
ln rural population	-0.0433	-0.0422	-0.064	34.1
Density	-0.1248	-0.1215	-0.286	57.5
mean age	0.0135	0.0131	0.013	-1.1
birth rate	0.0172	0.0167	0.044	61.9
literacy rate	0.1361	0.1325	0.226	41.4
Agriculture	-0.2612	-0.2543	-0.242	-5.1
manufacturing/services	0.0045	0.0044	0.007	37.4
occupied workforce	0.4911	0.4781	0.561	14.8
ln of GDP per capita	0.0513	0.0499	0.055	9.2
ln of rural GDP per capita	0.0088	0.0086	0.006	-42.8
homicide rate	-0.003	-0.0029	-0.007	58.3
water company	0.0081	0.0079	0.001	0
sewer company	-0.0123	-0.0120	-0.082	85.4

Notes: Elaboration by the Author.

5. Conclusions

The first objective of this thesis was to explore the behavior of the population size distribution of Brazilian Minimum Comparable Areas (MCAs) covering all Brazilian territory between 1970 and 2010, revisiting the traditional rank-size rule and Markov chain approaches. In order to bring up more accurate information on the dynamics and evolution of the population distribution, the spatial dependence is introduced through the analysis of LISA Markov and Spatial Markov Chains, both developed by Rey (2001). The second and main objective of the present thesis was to model population growth dynamics of Brazilian Minimum Comparable Areas (MCAs) in order to assess the determinants of population growth of these units between 1970 and 2010 and examine the existence and magnitude of spatial interaction and spatial spillover effects associated with these determinants.

Initially, the non-parametric normal kernel density functions estimates brought evidence that behavior in the distribution shape may indicate that divergence in population size of MCAs is decreasing. The Zipf's law estimation indicates that the population distribution is, decade by decade, moving away from Pareto law. In other words, this result shows that in the Brazilian case, over time, less and less the ranking of cities is influenced by its size. In the estimation of quadratic rank-size equation, the curvature presents downward concavity; there is a negative correlation between ranking variation and size.

The traditional Markov chain approach brings as main evidence the high probabilities on the main diagonal indicating a low interclass mobility, high-persistence of MCAs to stay in their own class size from one decade to another over the whole period. As suggested by Rey (2001), a spatial Markov transition probability matrix was constructed to analyse the spatial-temporal dynamics of relative population distribution, i.e., considering the possible influence from neighbours on the transition of regions. The results bring evidence that different spatial contexts have different effects on transition for regions. Specifically, the probability of upward transitions will increase for MCAs with neighbours in high classes. Another interesting results is that the MCAs grouped into medium classes have a higher probability of a downward transition if their neighbours are in a less populated class (class 1). The opposite happens if the neighbours are the most populous class (class 4). This evidence highlights again the major role of medium-size localities in the processes of urban agglomeration that occurred in Brazil during the last 40 years, even considering the spatial context. In relation to the LISA Markov approach, we found evidence of stability in the population distribution behavior over time that corroborate with the normal non-parametric kernel density functions estimates. Additionally, we investigated further highly populated communities that propelled the neighbours (MCAs that were populated above average and had less populous neighbours, while in the following period the neighbours became highly populated). It was identified that some municipalities, mainly in the north and northeast have played this role in the past 40 years, including some capitals of their respective states. Investigate more deeply the municipalities with this feature is an interesting suggestion for future research.

To achieve the second objective, we extended the population growth model developed by Glaeser et al. (1995) and Glaeser (2008) to include spatial interaction effects, both theoretically and empirically. Instead of treating cities as independent entities, we included interaction effects in the production function by assuming that knowledge accumulated in one economy may depend on knowledge accumulated in other economies, as well as interaction effects in the utility function by assuming that (dis)amenities in one economy may also (dis)benefit individuals living in other economies. Depending on whether or not utility of individuals is also assumed to be negatively correlated with population size and population growth of neighboring economies, we show that the population growth rate that can derived from this extended urban growth model eventually results in an econometric model that in the spatial econometrics literature is known as either a dynamic spatial Durbin error model (SDEM) or a dynamic general spatial nesting (GNS) model. To discriminate between these two models empirically and because the parameters of a dynamic GNS model are not identified, we test the dynamic SDEM and the dynamic spatial Durbin model (SDM) against each other. To carry out this test, we take a Bayesian perspective since it not only offers the opportunity to calculate Bayesian posterior model probabilities of both specifications, but also to compare the performance of these models using the log-marginals of different potential specifications of the neighborhood matrix. We find that the SDM specification in combination with the first-order binary contiguity matrix gives the best performance of all 44 possibilities being considered. Based on this finding, we conclude that the empirical results point to the utility function accounting for population size and population growth of neighboring economies and to the corresponding dynamic SDM specification.

According to the results, some spatially lagged variables, including the endogenous variable, showed highly significant coefficients. This result reinforces that these variables should not be excluded from the model. After considering a spatial panel data structure for Brazilian municipalities, we calculate the direct and indirect effects of short-term and long-term derived from population growth equation when the specification is the dynamic SDM model including time period fixed effects and space-specific fixed effects.

To make inferences regarding the statistical significance of the direct, indirect and total effects, were performed 100 simulated parameter combinations drawn from the multivariate Normal distribution implied by the ML Estimates. Some of the main results of the direct effects are: conditional convergence in population growth rate of Brazilian municipalities; a large rural population at the beginning of the decade will lead to a reduction in population growth after 10 years; areas that have a higher percentage of literate people in the initial period have a higher population growth rate in both short and long terms, an increase in GDP per capita in a given locality results in an increase in its growth rate, in short and long term.

Moreover, based on statistical significance verified through the t-statistics calculated, we found that some variables have statistically significant indirect effects. In other words, we found evidence of spatial spillovers in the determinants of population growth in Brazilian minimum comparable areas between the 1970s and 2010. Among these results, we find that five determinants of population produce significant indirect effects in the long term: rural population size, population density, the birth rate, the literacy rate and the homicide rate. This implies that a change of such a variable in one unit, also significantly affects population growth in other units, a phenomena that in most previous studies on population growth has been ignored. We find that such a non-spatial approach for Brazil underestimates the long-term total effects of the explanatory variables by 27% on average. Regarding the last four determinants we also find that the magnitude of the cumulative effect over all neighbors is as large as the magnitude of the impact on the city itself.

References

ALONSO, W. Location and Land Use; Toward a General Theory of Land Rent. Cambridge, ma: Harvard University Press, 1964.

AMEMIYA, T. Advanced econometrics. Harvard University Press, Cambridge, 1985.

ANDRADE, T.A.; SERRA, R.V. O Desempenho das cidades medias no crescimento populacional brasileiro no period 1970/2000. Cidades Médias Brasileiras, IPEA, 2001.

ANSELIN, L. Local Indicators of Spatial Association – LISA. Geographical Analysis, 27, 1995.

ANSELIN, L. Spatial externalities, spatial multipliers, and spatial econometrics. International Regional Science Review, v.26, pp.153-166, 2003.

ANSELIN, L., GALLO, J., JAYET, H. Spatial panel econometrics. In: MATYAS, L., SEVESTRE, P. (Eds.), The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice. Kluwer, Dordrecht, pp. 901-969, 3rd edition, 2008.

BALTAGI, B.H. Econometric Analysis of Panel Data. Wiley, Chichester, 3rd edition, 2005.

BALTAGI B.H.; SEUCK, H.S.;BYOUNG, C.J.;WON, K. Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data. Working Paper, Texa A&M University, College Station, 2006.

BENSSON, P.E.; DEJONG, D.N.; TROESKEN, W. Population Growth in U.S. Counties, 1840-1990. Regional Science and Urban Economics, v. 31, 2001

BLACK, D.; HENDERSON, V. Urban evolution in the USA. Journal of Economic Geography, v.3, pp. 343-372, 2003.

BOARNET, M.G.; CHARLERMPONG, S.; GEHO, E. Specification issues in models of population and employment growth. Papers in Regional Science, v.84, pp.21-46, 2008.

BRUECKNER, J.K. Strategic Interaction Among Governments: An Overview of Empirical Studies. International Regional Science Review, v. 26, pp. 175-188, 2003.

BRUECKNER, J.K. Strategic interaction among governments. In: ARNOTT, R.J., MCMILLEN, D.P. (Eds.), A Companion to Urban Economics. Malden, MA: Blackwell, pp. 332-347, 2008.

CHESHIRE, P.C.; MAGRINI, S., Population Growth in European Cities: Weather Matters – But only Nationally, Regional Studies, Vol. 40, 2006.

CHI, G. Can Knowledge Improve Population Forecast at Subcounty Levels? Demography, v. 46, 2009.

CHI, G.; VOSS, P.R. Small-area population forecasting: borrowing strength across space and time. Population, Space and Place v.17, pp. 505-520, 2011.

DA MATA, D.; DEICHMANN, U.; HENDERSON, J.V.; LALL, S.V.; WANG, H.G. Determinants of city growth in Brazil. Journal of Urban Economics, v.62, pp. 252-272, 2007.

DEBARSY, N.; ERTUR, C.; LESAGE, J.P. Interpreting dynamic space-time panel data models. Statistical Methodology 9, 158-171, 2012.

DOBKINS, L.H.; IOANNIDES, Y.M. "Dynamic Evolution of the U.S. City Size Distribution," Discussion Papers Series, Department of Economics, Tufts University 9916, Department of Economics, Tufts University, 1999.

DURANTON, G. Some foundations for Zipf's law: product proliferation and local spillovers. Reg Science and Urban Economics, v.36, pp. 542–563, 2006.

DURANTON, G.; PUGA, D. Nursery cities: Urban diversity, process innovation, and the life cycle of products. American Economic Review, v.91, pp. 1454–1477, 2001.

DURANTON, G.; PUGA, D. The Growth of Cities. CEPR Discussion Papers 9590, 2013.

ELHORST,J.P. Specification and Estimation of Spatial Panel Data Models. International Regional Science Review, v. 26, 2003.

ELHORST, J.P. Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. Geographical Analysis, v.37, pp. 85-106, 2005.

ELHORST, J.P. Applied Spatial Econometrics: Raising the Bar. Spatial Economic Analysis, Taylor and Francis Journals, v. 5, 2010.

ELHORST, J.P. Spatial panel models. Seminars, The University of York, 2011.

ELHORST, J.P. Dynamic spatial panels: models, methods and inferences. Journal of Geographical Systems, v.14, pp. 5-28, 2012.

ELHORST, J.P. Spatial Econometrics: From Cross-sectional Data to Spatial Panels. Springer, Heidelberg New York Dordrecht London, 2014.

ERTUR, C., KOCH, W. Growth, technological interdependence and spatial externalities: theory and evidence. Journal of Applied Econometrics, v.22, pp. 1033-1062, 2007.

FGV, Benefícios Econômicos da Expansão do Saneamento Básico. Instituto Trata Brasil - Fundação Getúlio Vargas, 2011.

FUJITA, M. A monopolistic competition model of spatial agglomeration: A differentiated product approach. Regional Science and Urban Economics, v.18, pp. 87–124, 1988.

GABAIX, X. Zipf's Law for Cities: An Explanation. The Quarterly Journal of Economics, vol. 114, n. 3, 1999.

GABAIX, X.; IOANNIDES, Y. M. The evolution of city size distributions, Handbook of Regional and Urban Economics, in: J. V. Henderson & J. F. Thisse (ed.), Handbook of

Regional and Urban Economics, edition 1, volume 4, chapter 53, pages 2341-2378, Elsevier, 2004.

GALLO, J.L.; CHASCO, C. Spatial Analysis of Urban in Spain, 1900-2001. Empirical Economics – Springer, 2007.

GIBBONS, S.; OVERMAN, H.G. Mostly pointless spatial econometrics? Journal of Regional Science, v.52, pp. 172-191, 2012.

GLAESER, E.L. Learning in cities. Journal of Urban Economics, v. 46, pp. 254–277, 1999

GLAESER, E.L. Cities, Agglomeration and Spatial Equilibrium. Oxford University Press, 2008.

GLAESER, E.L.; GOTTLIEB, J.D. The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States. NBER Working Paper Series, 2009.

GLAESER, E.L.; MARÉ, D.C. Cities and Skills. NBER Working Paper Series, 1994.

GLAESER, E.L.; SAIZ, A. The Rise of the Skilled City. NBER working Paper Series, 2003.

GLAESER, E.L.; SCHEINKMAN, J.A.; SHLEIFER, A. Economic growth in a cross-section of cities. Journal of Monetary Economics, v.36, pp. 252-272, 1995.

GREENE, W.H. Econometric Analysis. Pearson, New Jersey,5th edition, 2005.

HALLECK VEGA, S.; ELHORST, J.P. On spatial econometric models, spillover effects, and W. University of Groningen, Working paper, 2012.

HAMILTON, J.D. Time series analysis. Princeton University Press, Princeton, 1994.

HELSLEY, R.W.; STRANGE, W.C. Matching and agglomeration economies in a system of cities. Regional Science and Urban Economics, v.20, pp. 189–212, 1990.

HENDERSON, J.V. Urban Development, Theory, Fact, and Illusion. Oxford University Press, 1998.

IBGE. Sinópse do Censo Demográfico 2010 – Brasil. Available at: http://www.censo2010.ibge.gov.br/sinopse/, 2011.

JUSTO, W.R. Um Século de Crescimento das Cidades Brasileiras: Uma Abordagem Espacial. Anais da Anpec, 2012.

KEMENY, J.; SNELL, L. Finite Markov Chains. Springer, 1976.

LEE, L.F.; YU, J. A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory, v.26, pp. 564-597, 2010.

LESAGE, J.P. Spatial econometric panel data model specification: a Bayesian approach, available at ssrn: http://ssrn.com/abstract=2339124 or http://dx.doi.org/10.2139/ssrn.2339124, 2013.

LESAGE, J.P.; PACE, R.K. Introduction to Spatial Econometrics. CRC Press Taylor & Francis Group, Boca Raton, 2009.

MILLS, E.S. An aggregative model of resource allocation in a metropolitan area. American Economic Review, 1967.

MORO, S.; SANTOS, R.J. The Characteristics and Evolution of the Brazilian Spatial Urban System: Empirical Evidences for the Long-Run, 1970-2010. Texto Para Discussão 474, Cedeplar – UFMG, 2013.

MUTH, R. F. Cities and Housing. Chicago: University of Chicago Press, 1969.

NARDINELLI, C.; SIMON, C.J. The talk of the Town: Human Capital, Information, and the Growth of English Cities, 1861 to 1961. Explorations in Economic History. v. 33, 1996.

NICKELL, S. Biases in dynamic models with fixed effects. Econometrica, v.49, pp. 1417-1426, 1981.

OLIVEIRA, C.A. Evolução da Distribuição do Tamanho das Cidades Brasileiras: 1936-2000. Anais da ETENE, Banco do Nordeste, 2004a.

OLIVEIRA, C.A. Crescimento econômico das cidades nordestinas: um enfoque da nova geografia econômica. Encontro Regional De Economia, Anais da Anpec, 2004b.

PARETO, V. Cours d'Economie Politique. Geneva, Switzerland, 1896.

RAMALHO, H.M.B.; SILVEIRA NETO, R.M.A. A Inserção do migrante rural no mercado de trabalho urbano no Brasil: uma análise empírica da importância dos setores informal e formal. Estudos Econômicos 42, http://dx.doi.org/10.1590/S0101-41612012000400004, 2012.

REIS, E.; PIMENTEL, M.; ALVARENGA, A.I. Áreas Mínimas Comparáveis Para os Períodos Intercensitários de 1872 a 2000, IPEA, 2010.

REY, S.J. Spatial Empirics for Economic Growth and Convergence. Geographical Analisys, vol. 33, n. 3, 2001.

RESENDE, G.M., Spatial Dimensions of Economic Growth in Brazil. ISRN Economics 2013, 1-19. Hindawi Publishing Corperation, http://dx.doi.org/10.1155/2013/398021, 2013.

ROBACK, J. Wages, rents, and the quality of life. Journal of Political Economy, v.90, pp. 1257-1278, 1982.

ROSEN, S. Wage-based indexes of urban quality of life. In MIEZKOWSKI, P.N., STRASZHEIM, M.R. (Eds.), Current Issues in Urban Economics. John Hopkins University Press, Baltimore, pp. 74-104, 1979.

ROSEN, K.; RESNICK, M. The Size Distribution of Cities: An Examination of the Pareto Law and Primacy. Journal of Urban Economics, VIII, 1980.

SILVEIRA NETO, R.M., AZZONI, C.R. Social policy as regional policy: market and non-market factors determining regional inequality. Journal of Regional Science, v.51, pp. 1-18, 2012.

SIMON, C.J. Human Capital and Metropolitan Employment Growth. Journal of Urban Economics, v. 43, 1998.

TRINDADE, L.; SARTORIS, A. Distribuição populacional e evolução do tamanho das cidades brasileiras: uma análise espacial do período 1920-2000,Working Papers 09_11, Universidade de São Paulo, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, 2009.

VERGOLINO, J.R.; JATOBÁ, A.D.S. Os Determinantes do Crescimento da População e do Emprego das Microrregiões do Nordeste do Brasil – 1970/1996: Uma Abordagem Econométrica. Revista Econômica do Nordeste, v. 31, 2000

YAP, L.Y.L. Rural-urban migration and urban underemployment in Brazil. Journal of Development Economics, v.3, pp. 227-243, 1976.

YU, J., JONG, R. de, LEE, L. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. Journal of Econometrics, v. 146, pp. 118-134, 2008.

ZIPF, G.K. Human behavior and the principle of least effort. Addison-Wesley, Cambridge, 1949.

Appendix

Table A1 – High-Populated Municipalities that Propelled the Neighbours

1970-1980		1980-1991		1991-2000		2000-2010	
Municipality	FU	Municipality	FU	Municipality	FU	Municipality	FU
Terra Santa	PA	Belém	PA	Aracaju	SE	Presidente Sarney	MA
Faro	PA	Redenção	PA	Marataizes	MG	Pinheiro	MA
Oriximiná	PA	São Geraldo do Araguaia	PA	Itapemirim	MG	Pedro do Rosário	MA
Santa Filomena	PE	Conceição do Araguaia	PA	Jundiaí	SP	São Luís	MA
Ouricuri	PE	Xinguara	PA	Itajaí	SC	Itapagé	CE
Santa Cruz	PE	Rio Maria	PA			Tejuçuoca	CE
Mata de São João	BA	Piçarra	PA			Catunda	CE
Camaçari	BA	Floresta do Araguaia	PA			Santa Quitéria	CE
Dias d'Ávila	BA	Sapucaia	PA				
Araguari	ES	Pau d'Arco	PA				
Coronel Fabriciano	ES	Natal	RN				
Campinas	SP	João Pessoa	PB				
Itapevi	SP	Conceição do Coité	BA				
Saltinho	SP	Monte Santo	BA				
Piracicaba	SP	Madre de Deus	BA				
São Roque	SP	Salvador	BA				
Araçariguama	SP	Senhor do Bonfim	BA				
Curitiba	PR	Andorinha	BA				
Foz do Iguaçu	PR	Uberaba	ES				
Santa Terezinha de Itaipu	PR	Delta	ES				
Poxoréo	MT	Atibaia	SP				
Cana Brava do Norte	MT	Itapira	SP				
Campo Verde	MT	Sorocaba	SP				
Campinápolis	MT	Balneário Arroio do Silva	SC				
Cuiabá	MT	Araranguá	SC				
Jaciara	MT						
Novo São Joaquim	MT						
Juscimeira	MT						
Ribeirão Cascalheira	MT						
Alto Boa Vista	MT						
São José do Xingu	MT						
Canarana	MT						
Confresa	MT						
Santa Terezinha	MT						
Querência	MT						
Dom Aquino	MT						
Primavera do Leste	MT						
Luciara	MT						
Barra do Garças	MT						
Nova Xavantina	MT						
São Félix do Araguaia	MT						
São Pedro da Cipa	MT						
Cocalinho	MT						
Porto Alegre do Norte	MT						
Araguaiana	MT						
Água Boa	MT						
Vila Rica	MT						
Brasília	DF						

Notes: Elaboration by the Author. FU = Federal Unit (State).

Table A2 - Test of difference between the conditional transition matrix against the overall transition matrix

Spatial Conditioned Transitions Submatrices	Chi2	P-value	Degrees of freedon
1	61.8955	0.0000	9
2	24.1559	0.0041	9
3	26.3733	0.0018	9
4	148.6507	0.0000	9

Notes: Estimates by the Author.

Table A3 – Correlation coefficients between explanatory variables

ln rural population	1.00												
ln density	-0.25	1.00											
mean age	-0.32	0.13	1.00										
birth rate	-0.29	-0.16	-0.05	1.00									
literacy rate	-0.30	0.14	0.60	-0.15	1.00								
agriculture	0.23	-0.32	-0.31	0.21	-0.48	1.00							
manufacturing/services	-0.09	0.11	0.14	-0.07	0.20	-0.19	1.00						
workforce occupied	0.09	-0.04	-0.29	0.14	-0.20	0.48	-0.08	1.00					
ln GDP per capita	-0.26	0.19	0.40	-0.13	0.76	-0.42	0.19	-0.07	1.00				
ln rural GDP per capita	0.10	-0.44	0.02	0.19	0.05	0.46	-0.07	0.38	0.20	1.00			
homicide rate	-0.05	0.17	0.02	-0.13	0.14	-0.25	-0.03	-0.19	0.18	-0.19	1.00		
water company	-0.36	0.26	0.56	-0.18	0.65	-0.64	0.11	-0.40	0.55	-0.20	0.18	1.00	
sewer company	-0.30	0.26	0.51	-0.13	0.53	-0.45	0.11	-0.22	0.49	-0.10	0.07	0.64	1.00

Notes: Estimates by the Author. Database from Demographic Census for 1970, 1980, 1991, 2000 and 2010.