



**UNESP**

**UNIVERSIDADE ESTADUAL PAULISTA  
“JÚLIO DE MESQUITA FILHO”  
Faculdade de Ciências e Letras  
Campus de Araraquara - SP**

LORENA ZARDO TRINDADE

**DISTRIBUIÇÃO POPULACIONAL E EVOLUÇÃO DO  
TAMANHO DOS MUNICÍPIOS BRASILEIROS: uma análise  
espacial do período 1920-2000**



ARARAQUARA – S.P.  
2010

LORENA ZARDO TRINDADE

**DISTRIBUIÇÃO POPULACIONAL E EVOLUÇÃO DO TAMANHO DOS  
MUNICÍPIOS BRASILEIROS: uma análise espacial do período 1920-2000**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Economia da Faculdade de Ciências e Letras de Araraquara da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como requisito para obtenção do título de Mestre em Ciências Econômicas

**Orientador:** Prof. Dr. Alexandre Sartoris Neto.

ARARAQUARA – S.P.  
2010

Zardo Trindade, Lorena

Distribuição populacional e evolução do tamanho dos  
Municípios Brasileiros: uma análise espacial do  
período 1920-2000 / Lorena Zardo Trindade – 2010  
64 f.

Dissertação (Mestrado em Economia) – Universidade  
Estadual Paulista "Júlio de Mesquita Filho",  
Faculdade de Ciências e Letras (Campus Araraquara)  
Orientador: Alexandre Sartoris Neto

1. distribuição populacional brasileira. 2.  
convergência. 3. autocorrelação espacial. 4. modelo  
SUR espacial. 5. probabilidade de transição espacial.  
I. Título.

Ficha catalográfica elaborada pelo sistema automatizado  
com os dados fornecidos pelo(a) autor(a).

LORENA ZARDO TRINDADE

**DISTRIBUIÇÃO POPULACIONAL E EVOLUÇÃO DO TAMANHO DOS  
MUNICÍPIOS BRASILEIROS: uma análise espacial do período 1920-2000**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Economia da Faculdade de Ciências e Letras de Araraquara da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como requisito para obtenção do título de Mestre em Ciências Econômicas

**Orientador:** Prof. Dr. Alexandre Sartoris Neto.

Data da defesa: 04/03/ 2010

**Membros componentes da Banca Examinadora**

---

**Presidente e Orientador: Prof. Dr. Alexandre Sartoris Neto**

Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP - Araraquara

---

**Membro Titular: Prof. Dr. André Luiz Correa**

Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP - Araraquara

---

**Membro Titular: Prof. Dr. André Luiz Squarize Chagas**

Fundação Escola de Comércio Álvares Penteado

**Local:** Universidade Estadual Paulista

Faculdade de Ciências e Letras

UNESP – Campus de Araraquara

## **AGRADECIMENTOS**

Embora uma dissertação seja, pela sua finalidade acadêmica, um trabalho individual, há contribuições que não devem deixar de ser realçadas. Por essa razão, desejo expressar os meus sinceros agradecimentos:

Ao meu Orientador Prof. Doutor Alexandre Sartoris Neto pela compreensão, paciência, incentivo e presteza revelados ao longo da orientação. E pelos ensinamentos essenciais para a realização dessa dissertação

Aos professores Dr. Marcelo Pinho, Dr. Elton Eustáquio Casagrande e Dr. Eduardo Haddad pelas oportunas e relevantes sugestões para a estrutura final desse trabalho.

A todo corpo docente e técnico do Programa de Pós-Graduação em Economia da Faculdade de Ciências e Letras de Araraquara da Universidade Estadual Paulista “Júlio de Mesquita Filho.

A Guilherme da Rocha Bezerra Costa, querido amigo, por todo incentivo e companheirismo diariamente demonstrados ao longo do curso.

Aos demais colegas de classe.

A Ana Paula Vitali Janes Vescovi e ao Instituto Jones dos Santos Neves pela compreensão e interesse manifestados por este trabalho. A todos os colegas das Coordenações de Estudos Econômicos e Estudos Sociais pelo auxílio na conciliação das minhas obrigações profissionais e acadêmicas.

Ao amigo Alexsandro Hoffman pela porta que me abriu em meu retorno ao Espírito Santo. A Henrique Tápias de Salles pelo apoio e carinho. Aos meus amigos, Silvia Varejão e Celso Bissoli Sessa pela grande amizade, estímulo e paciência com que sempre me ouviram desde os tempos de graduação. A Karoline Pereira Ferreira, notável companheira na reta final deste trabalho.

E, finalmente, a meus pais, Silas e Margarete, e minhas irmãs, Lívia e Flávia, pela oportunidade, privilégio e apoio incondicionais que sempre me foram dados.

## RESUMO

Este estudo objetiva um exame mais detalhado da evolução da relação entre o tamanho das cidades brasileiras e sua distribuição populacional no período de 1920 a 2000. Para isso, utilizaremos dois tipos de métodos, um estático – que envolve um indicador de convergência baseado na Lei de Zipf – e outro dinâmico, que mostra, através de uma cadeia de Markov, movimentos nas posições relativas das cidades dentro da distribuição por tamanhos. Nas duas análises, verificamos uma persistente concentração populacional em um número pequeno de áreas. Os efeitos espaciais, considerados em ambas as análises, mostraram ter importante influência nos resultados obtidos.

**Palavras chave:** Distribuição populacional brasileira, convergência, autocorrelação espacial, modelo SUR espacial, probabilidade de transição espacial.

## **ABSTRACT**

This study aims a more detailed examination on the evolution of the relationship between Brazilian cities size and their population distribution in the period from 1920 to 2000. For this purpose, two type of methods are performed. A static one - which involves the estimation of convergence indicator given by the Zipf's Law - and a dynamic one - which shows, by means of a Markov chain, movements in cities relative positions within the distribution of their sizes. In both analyses, there is a persistent concentration of people residing in a small number of areas, growing at a greater rate than smaller ones. Spatial effects, considered in the estimation of both analyses, is clearly an important influence on the results

**Keywords:** Brazilian population distribution, convergence, spatial autocorrelation, spatial SUR models, spatial transition probability.

## LISTA DE TABELAS

Tabela 1 - Número de municípios nos Censos e AMC nos períodos intercensitários: Brasil, 1920-2000.....	39
Tabela 2 - Valores médios da distribuição populacional: Brasil, AMC, 1920/2000 .....	42
Tabela 3 - Intervalos inter-quartis (IQ*) da distribuição populacional: Brasil, AMC, 1920/2000 .....	43
Tabela 4 - Coeficiente de Pareto por MQO .....	44
Tabela 5 - Coeficiente de Pareto estimado para diferentes tamanhos de AMC.....	45
Tabela 6 - Coeficiente de Pareto por MQO e teste de especificação.....	48
Tabela 7 - Coeficiente de Pareto por SUR erro espacial .....	49
Tabela 8 - Distribuição do tamanho das AMCs brasileiras por classe de tamanho relativo.....	51
Tabela 9 - Matriz de probabilidade de transição 1920-2000: Brasil, população relativa .....	52
Tabela 10 - Distribuição versus ergódica 1920-2000: Brasil, população relativa .....	53
Tabela 11 - Probabilidade de transição com efeitos espaciais .....	54

## LISTA DE GRÁFICOS

Gráfico 1 - Evolução do coeficiente de Pareto estimado para diferentes tamanhos de AMC ..	46
Gráfico 2 - Densidade log-relativa do tamanho das AMCs: Toda amostra, Brasil, 1920/1960/2000 .....	47
Gráfico 3 - Evolução do coeficiente de Pareto estimado .....	50

## LISTA DE FIGURAS

Figura 1 - Brasil: Áreas Mínimas Comparáveis, 1920-2000. ....	40
Figura 2 - Brasil: Áreas mínimas comparáveis, 1920-2000 e área dos municípios em 2000...	41

## SUMÁRIO

INTRODUÇÃO .....	11
1 REVISÃO DE LITERATURA.....	14
2 ASPECTOS METODOLÓGICOS: A ABORDAGEM DA ECONOMETRIA ESPACIAL .....	19
2.1 PESO E DEFASAGEM ESPACIAL.....	21
2.1.2 O problema da unidade de área modificável.....	23
2.2 MODELOS DE REGRESSÃO ESPACIAL .....	24
2.2.1 Defasagem espacial e modelo Erro-espacial.....	25
2.3 MODELOS ESPACIAIS DE REGRESSÃO LINEAR PARA DADOS NO TEMPO-ESPAÇO .....	27
3.3.1 Regressões espaciais aparentemente não relacionadas ( <i>Spatial seemingly unrelated regression – spatial SUR</i> ).....	30
2.4 TESTES DE ESPECIFICAÇÃO .....	34
2.4.1 I de Moran.....	34
2.4.2 Teste de Máxima-Verossimilhança.....	34
3.4.3 Teste Kelejian-Robinson.....	35
2.5 CADEIA DE MARKOV .....	36
3 RESULTADOS ECONOMÉTRICOS .....	39
3.1 DESCRIÇÃO DOS DADOS .....	39
Fonte: Extraído de Reis, Pimentel e Alvarenga (2008). .....	41
3.2 EVOLUÇÃO DA DISTRIBUIÇÃO POPULACIONAL BRASILEIRA .....	43
3.2.2 Análise não-paramétrica do expoente de Pareto.....	47
3.2.1 O efeito espacial na estimativa do expoente de Pareto .....	48
3.3 ANÁLISE DINÂMICA ESPACIAL DA DISTRIBUIÇÃO DO TAMANHO DOS MUNICÍPIOS BRASILEIROS .....	50
CONSIDERAÇÕES FINAIS .....	55
REFERÊNCIAS.....	57

## INTRODUÇÃO

As modificações do cenário econômico brasileiro no decorrer do século XX impulsionaram transformações significativas na dinâmica da população brasileira. A queda da importância relativa do setor agrícola na economia nacional na década de 1930; a ampliação dos investimentos estatais em atividades industriais na década de 1950; o “milagre econômico” dos anos 1970; a recessão da década de 1980 e a estagnação da década de 1990; todos foram responsáveis pela dinâmica populacional do país no decorrer do século e pela atual distribuição da população no território nacional.

Ao longo de todo o século, os movimentos populacionais entre os estados brasileiros foram pautados pela dinâmica e pela localização das atividades industriais. No período 1920-1950, quando a transição de um modelo econômico agroexportador para um urbano-industrial ainda estava em curso, apenas 8% de seu total viviam nas grandes cidades<sup>1</sup>. (MATA, 1973; BRITO, 2000; NETTO JR. E TARGINO, 2003). Na década de 1970, 41% da população total do país poderia ser encontrada em regiões com mais de 500 mil habitantes: São Paulo e Rio de Janeiro, áreas de grande concentração populacional, consolidaram-se como os principais centros econômico-financeiros do país, gerando, respectivamente, 20,4% e 19,6% do Produto Interno Bruto (PIB) brasileiro, ou seja, quase metade da produção nacional. No que se refere à produção industrial nacional, São Paulo participava com 27,1% do valor adicionado e Rio de Janeiro com 9,5%, uma soma de 36,6% da produção industrial.

Porém, a partir dos anos 80, os obstáculos gerados pela recessão econômica, concentrada principalmente nas grandes metrópoles, diminuíram consideravelmente o poder de atração exercido pela indústria nos estados do Rio de Janeiro e São Paulo (BRITO E CARVALHO, 2006). Nesse período ocorria uma relativa desconcentração espacial da atividade industrial. São Paulo e Rio de Janeiro, que juntos geravam, em 1980, 26,4% do produto industrial brasileiro diminuiu sua participação para 16,8% em 2000. Como resultado desse processo, uma tendência de concentração populacional de ritmo mais lento tornou-se predominante, e em 2000, 56% da população encontrava-se nos grandes centros populacionais.

---

<sup>1</sup> Conforme metodologia do Instituto Brasileiro de Geografia e Estatística (IBGE), as grandes cidades são aquelas que possuem mais de 500 mil habitantes.

Braga (2006) verifica que as últimas décadas do século XX foi palco de uma alteração substancial nos padrões de crescimento populacional no Brasil. Se anteriormente os investimentos industriais, a urbanização e as altas taxas de natalidade foram protagonistas na geração do excedente populacional que impulsionou os fortes movimentos migratórios em direção às grandes metrópoles e áreas de fronteiras de recursos, agora, esses mesmos atores - mudanças econômicas e demográficas - causaram um esgotamento de antigos padrões que acompanhavam a dinâmica populacional até então. Dessa forma, Brito (2000, p. 38) argumenta que o processo de urbanização da sociedade brasileira continuou, porém, menos acelerado do que nas décadas anteriores, quando o Estado implementava políticas de incentivo à ocupação demográfica das regiões Norte e Centro-Oeste, assim como promovia uma primeira onda de “expansões dos sistemas de transporte e de telecomunicações que reforçaram enormemente a cultura migratória”. Conforme o autor, a contínua ampliação sistemas de transporte e de telecomunicações, associada ao desenvolvimento das redes de interação social, tiveram um efeito arrefecedor na cultura migratória brasileira ao passo que “as externalidades positivas das grandes cidades, das regiões metropolitanas, em particular, que tanto atraíam os migrantes, foram superadas pelas externalidades negativas” - violência urbana, desemprego, dificuldades de acesso aos serviços públicos básicos e à moradia (BRITO 2000, p. 39).

Embora esta breve análise descritiva forneça algumas ideias gerais sobre as mudanças na distribuição da população brasileira, certas características especiais permanecem ocultas uma vez que por meio de estatísticas descritivas não é possível verificar se o crescimento da população concentra-se em alguns pontos do território nacional, ou se é distribuído de maneira mais uniforme em áreas de diferentes tamanhos. Neste caso, um exame mais detalhado do comportamento dos movimentos populacionais é extremamente importante para a melhor compreensão da dinâmica populacional brasileira no século XX, e das mudanças ocorridas nesses movimentos durante todo este período.

Nas últimas décadas, a literatura de economia regional tem realizado debates significantes em torno da distribuição dos tamanhos das cidades e de sua evolução ao longo do tempo. Trata-se, no entanto, de um difícil tema, que envolve teorias sujeitas às propriedades dinâmicas da distribuição, em que flutuações e interações são produzidas de forma contínua.

Na literatura nacional poucos são os estudos empíricos que analisam como a população brasileira distribui-se entre suas cidades e como essa distribuição tem evoluído ao longo do tempo. Assim, visando preencher esta lacuna, o presente estudo tem por objetivo realizar uma

análise sobre a evolução das relações entre a distribuição populacional e o tamanho das cidades brasileiras no período de 1920 a 2000. Por meio de diferentes técnicas avançadas de estatística e econometria, esse estudo ainda pretende agregar aos estudos realizados anteriormente sobre a distribuição do tamanho das cidades brasileiras – Oliveira (2004) e Ruiz (2005) – uma vez que pretende utilizar testes e especificações de dependência espacial, tanto na parte estática quanto na dinâmica, para captar a influência do espaço na convergência e nas probabilidades de transição de tamanho, realizando dessa forma, um exame mais completo da velocidade do movimento e da forma de convergência na distribuição do tamanho das cidades.

Para atingir os objetivos citados, o estudo foi dividido em três capítulos além dessa introdução e das considerações finais. No primeiro capítulo é feita uma revisão de literatura que apresenta o debate internacional sobre metodologias de descrição da distribuição do tamanho das cidades, considerando a integração de aspectos de localização geográfica com estimações e testes de modelos econométricos. O segundo capítulo apresenta os principais aspectos teóricos e metodológicos utilizados neste estudo, abordando desde conceitos básicos da Econometria Espacial - peso e defasagem espacial - a conceitos mais complexos - modelos espaciais de regressão linear para dados no tempo-espaço e análise dinâmica com uso de Cadeia de Markov. O capítulo 3 apresenta resultados empíricos, incluindo análises paramétricas e não-paramétricas sobre a evolução da distribuição populacional brasileira e o possível efeito espacial nesta.

## 1 REVISÃO DE LITERATURA

A literatura tem realizado debates significantes na descrição da distribuição do tamanho das cidades, pois, na busca por uma distribuição estatística que mais se assemelhasse à distribuição populacional foram encontradas duas regularidades empíricas robustas que remetem a duas abordagens distintas, a da lei de Gibrat e a da lei de Zipf.

A abordagem da lei de Gibrat, desenvolvida em 1931, se baseia, de acordo com Eeckhout (2004), na hipótese de que as cidades crescem ou se contraem por meio de um processo de crescimento proporcional, em que a taxa de crescimento populacional de uma determinada cidade não depende de seu tamanho inicial. Segundo o autor, esse processo de crescimento é válido uma vez que as taxas de crescimento entre diferentes cidades possuem a mesma variância. Contudo, mesmo que isso não ocorra, não é possível verificar algum tipo de padrão sistemático em relação ao tamanho inicial, indicando que o processo estocástico de crescimento é o mesmo para todas as cidades.

O processo de crescimento proporcional do tamanho das cidades pode ser verificado por meio da regressão de estimadores não-paramétricos<sup>2</sup>, que além de indicar o formato da distribuição do tamanho das cidades – que nesse caso trata-se de uma log-normal -, estabelece uma relação entre o crescimento das cidades e seus respectivos tamanhos iniciais, permitindo testar se há uma relação de dependência entre essas duas variáveis (ROSEN E RESNICK, 1980; IOANNIDES E OVERMAN, 2000; EECKHOUT, 2004).

Essa abordagem, apesar de ter sido umas das primeiras formulações a fornecer uma descrição mais detalhada da distribuição do tamanho das cidades, começou a ser aplicada pela literatura de economia urbana dominante apenas nos últimos anos. Anteriormente, o método mais utilizado para realizar esse tipo de análise baseava-se na suposição de que a distribuição estatística mais adequada para explicar o tamanho das cidades seria a distribuição de Pareto, descrita por uma lei de potência

$$R(S) = aS^{-b}, \quad (2.1)$$

em que,  $R$  é a classificação das cidades por população na distribuição;  $S$  é o tamanho dessa população; e  $a$  e  $b$  são parâmetros. O expoente de Pareto - parâmetro  $b$  que assume sempre valores positivos - pode servir como um indicador de convergência da distribuição do tamanho

---

<sup>2</sup> Geralmente estimados por meio do modelo  $g_i = m(S_i) + \varepsilon_i$ , em que  $g_i$  é a taxa de crescimento normalizada,  $m$  é estimador não paramétrico e  $S_i$  é o logaritmo do tamanho da população de cada cidade (Eeckhout, 2004).

das cidades. Assim, quando  $b < 1$  tem-se uma estrutura urbana concentrada – poucas grandes cidades – e com cidades que crescem a taxas menos convergentes no que diz respeito ao número de habitantes. Mas quando  $b > 1$ , o inverso ocorre, tendo-se uma estrutura urbana dispersa – vários grandes centros urbanos – e com cidades que crescem a taxas mais convergentes (ROSEN E RESNICK, 1980).

Difundido por George Zipf, em seu livro *Human Behavior and the Principle of Least Effort* de 1949, esse método destacou-se devido à regularidade encontrada em sua aplicação. Segundo Brañas (2000) e Ruiz (2005), Zipf deparou-se com indícios de que, frequentemente,  $b$  assumia o valor 1; o que ocorria porque a partir do momento em que todas as cidades de uma determinada região possuíssem o mesmo potencial de crescimento, independente de seu tamanho ou posição nos seus espaços econômicos, o produto do tamanho da população e de sua classificação na distribuição (*rank*) seria aproximadamente constante. Dessa forma, a maior cidade dessa região teria uma população  $n$  vezes maior que a  $n$ -ésima, de modo que a segunda maior cidade em população teria a metade de habitantes da primeira, a terceira teria um terço, a quarta um quarto, e assim por diante. A essa regularidade foi dada o nome de lei de Zipf ou regra de ordem-tamanho.

Essas duas abordagens, a da lei de Gibrat e a da lei de Zipf, por muito tempo foram consideradas incompatíveis, contudo, em estudos recentes observou-se que, com algumas considerações, essas podem ser utilizadas como análises complementares. Dentre esses estudos cabe destacar Gabaix (1999a, 1999b), Ioannides e Overman (2000) e Eeckhout (2004).

O trabalho de Gabaix (1999a, 1999b), que segue uma linha mais teórica, sugerem que a lei de Zipf seria na verdade uma consequência da lei de Gibrat. Em termos formais os autores consideram que o processo de crescimento das cidades, ao menos no extremo superior da distribuição, segue a forma  $S_{t+1}^i = y_{t+1}^i S_t^i$ , em que  $y_{t+1}^i$  são independentes e variáveis aleatórias identicamente distribuídas. Como a taxa de crescimento da cidade  $i$  segue uma distribuição  $f(y) \cdot y_{t+1}^i - 1$ , o tamanho médio da cidade deve ser normalizado e permanecer constante ( $\sum_{j=1}^N S_t^j = 1$ ), o que requer que  $E[y] = 1$ , ou  $\int_0^\infty y f(y) dy = 1$ .

Ao denominar  $G_t(S) := P(S_t > S)$  como o extremo da distribuição do tamanho das cidades no tempo  $t$ , tem-se como equação de movimento para  $G_t$ :

$$G_{t+1}(S) = P(S_{t+1} > S) = P(y_{t+1} S_t) = E \left[ \mathbf{1}_{S_t > \frac{S}{y_{t+1}}} \right] = E \left[ E \left[ \frac{\mathbf{1}_{S_t > \frac{S}{y_{t+1}}}}{y_{t+1}} \right] \right] = E \left[ G_t \left( \frac{S}{y_{t+1}} \right) \right] = \int_0^\infty G_t \left( \frac{S}{y} \right) f(y) dy. \quad (2.2)$$

Supondo que exista um processo de estado estacionário que ateste  $G_t = G$ , logo,

$$G(S) = \int_0^{\infty} G \frac{S}{y} f(y) dy. \quad (2.3)$$

Comparada à equação  $\int_0^{\infty} y f(y) dy = 1$ , a distribuição do tipo  $G(S) = \frac{a}{S}$  satisfaz a equação do estado estacionário  $G(S) = \int_0^{\infty} G \frac{S}{y} f(y) dy$ , o que indica que a Lei de Zipf é um bom candidato para descrever distribuições de estado estacionário.

A partir desse raciocínio, a maior contribuição de Gabaix (1999b), foi a indicação de que quando o crescimento das cidades obedece a lei de Gibrat e, concomitantemente, é submetido a choques exógenos estocásticos de mesmo desvio-padrão, a distribuição limite do seu tamanho resultará na lei de Zipf. Esse resultado ainda seria alcançado mesmo que algumas cidades possuíssem potenciais de crescimento diferentes, pois nestas circunstâncias, a partir de sugestões do próprio Gabaix (1999b) e de Ioannides e Overman (2000), basta que se considere o processo de crescimento das cidades como um movimento browniano geométrico variante no tempo para que a distribuição limite do tamanho das cidades automaticamente tenda a uma lei de potência com um expoente compatível com lei de Zipf.

O trabalho de Eeckhout (2004) chega à mesma conclusão geral de que a lei de Zipf não passa de um caso especial da lei de Gibrat, no entanto, para chegar a esse resultado o autor usa uma análise empírica da distribuição do tamanho das cidades dos Estados Unidos. Nessa análise, em que a amostra utilizada compreende a população de todas as cidades americanas no ano de 2000, verifica-se que a distribuição da população no território americano segue uma distribuição log-normal e não de Pareto. Mas, como na parte superior da amostra essas duas distribuições possuem formatos semelhantes, a lei de Zipf ainda pôde ser verificada para parte da amostra.

Os estudos que pretendem analisar a distribuição do tamanho das cidades podem, então, considerar tanto a abordagem da lei de Gibrat quanto à da lei de Zipf. Alguns autores utilizam ambas as abordagens; Eeckhout (2004), Black e Henderson (2003) para os Estados Unidos e Anderson e Ge (2004) para a China. No que se refere à lei de Gibrat, com exceção de Black e Henderson (2003), todos encontram resultados que sugerem que essa se aplica ao crescimento populacional das cidades. Mas em relação à lei de Zipf, os resultados encontrados diferem entre os trabalhos, pois, ainda que certa regularidade seja verificada, observa-se uma forte sensibilidade das estimativas do parâmetro  $b$  com relação a variações na amostra e ao método de estimação empregado.

Essa sensibilidade do parâmetro  $b$  pode se vista de maneira mais clara na comparação entre as estimações feitas por Rosen e Resnick (1980), Soo (2002) e Oliveira (2004). Rosen e Resnick (1980), ao utilizar uma amostra que abrangia cidades com mais de 15 mil habitantes de diversos países, obteve um coeficiente de Pareto de 1,1341 para o Brasil, ao utilizar dados de 2000, enquanto Soo (2002), usando dados populacionais de todas as cidades do país, obteve para o mesmo ano 0,8690.

Oliveira (2004, p. 7), ao verificar variações do coeficiente de Zipf para diferentes tamanhos de amostras compostas por cidades brasileiras ressalta que a sensibilidade do parâmetro  $b$  sugere que “quando se estabelecem restrições sobre a amostra, o coeficiente de Pareto é maior, o que indica que a desigualdade é menor entre cidades médias e grandes do que quando se acrescenta as cidades menores”.

Até então, os métodos utilizados para examinar a relação entre distribuição populacional e o tamanho das cidades têm revelado características essenciais para um melhor entendimento da dinâmica populacional ao longo do tempo e do espaço. Contudo, como a verificação da lei de Zipf se trata de uma análise estática, não é possível identificar como são as mudanças nas posições relativas das cidades na distribuição de seus tamanhos. Lanaspá et al. (2003) afirma que essa é uma deficiência que dever ser corrigida, dada a dificuldade de se aceitar que a estrutura urbana não passou por mudanças profundas ao longo de um século. Em outras palavras, Lanaspá et al. (2003) explica que mesmo que a distribuição tenha permanecido essencialmente a mesma por um longo período de tempo, mudanças significantes podem ter ocorrido na posição relativa das cidades na distribuição.

Para um exame mais detalhado da dinâmica dos dados, aplicou-se mesma metodologia utilizada por Eaton e Eckstein (1997), Black e Henderson (2003), Lanaspá et al. (2003) e Le Gallo e Chasco (2008). Inicialmente elaborada por Qua (1993), essa metodologia baseia-se no pressuposto de que a distribuição interna da dinâmica populacional de determinada área entre dois pontos diferentes no tempo pode ser modelada como um processo estocástico com espaço paramétrico de tempo discreto e um espaço finito de estados. Formalmente, denota-se  $F_t$  como a distribuição transversal do tamanho populacional das AMCs no tempo  $t$  relativa à média brasileira. Define-se um conjunto de  $K$  classes de diferentes tamanhos, que propiciam uma aproximação discreta da distribuição populacional. Primeiro, assume-se que a frequência da distribuição segue um processo de Markov estacionário de primeira ordem.

Outra característica ignorada pelos estudos aplicados ao caso brasileiro é a possibilidade da distribuição do tamanho das cidades sofrer efeitos espaciais. Devido à natureza geográfica dos dados utilizados nesse tipo de análise, o uso de uma metodologia econométrica que considere essas características foi essencial para que uma inferência estatística adequada pudesse ser realizada. Le Gallo e Chasco (2008) introduziram em sua análise, testes e especificações de dependência espacial, tanto na parte estática quanto na dinâmica, para captar a influência do espaço na convergência e nas probabilidades de transição, realizando dessa forma, um exame mais completo da velocidade do movimento e da forma de convergência na distribuição do tamanho das cidades. Na análise da convergência, os resultados encontrados pelo método dos Mínimos Quadrados Ordinários (MQO) indicaram a existência de resíduos não-normais, que por sua vez, apresentaram heteroscedasticidade e autocorrelação espacial. E na análise da cadeia de Markov foram encontradas evidências de que a probabilidade de um movimento dentro da distribuição ocorrer difere em diferentes áreas urbanas. Ambos os resultados confirmam a suposição de que há influência espacial na distribuição do tamanho das cidades.

Diante dessas evidências, para a realização de uma análise completa sobre o comportamento da distribuição do tamanho das cidades, seus aspectos estáticos e dinâmicos devem ser considerados sob a ótica do tempo e do espaço.

## 2 ASPECTOS METODOLÓGICOS: A ABORDAGEM DA ECONOMETRIA ESPACIAL

Os trabalhos empíricos realizados na área de ciências regionais são, geralmente, baseados em amostras de dados recolhidos de acordo com sua localização no espaço. Na década de 1970, frente à necessidade de integrar esses aspectos de localização da amostra de dados multiregionais com estimações e testes de modelos econométricos, Jean Paelinck introduziu uma nova forma analítica denominada de Econometria Espacial, que associada a instrumentais da econometria tradicional, contemplava a importância da questão espacial nos estudos regionais.

Em Anselin (1988), o termo econometria espacial foi definido como uma coleção de técnicas delimitadas por cinco características distintas: (i) o papel da interdependência espacial nos modelos espaciais; (ii) a assimetria nas relações espaciais; (iii) a importância de fatores explicativos localizados em outros espaços; (iv) diferenciação entre interações *ex-post* e *ex-ante*; e (v) a modelagem explícita do espaço. Em outras palavras, a econometria espacial contempla questões metodológicas que surgem na consideração explícita dos efeitos espaciais, como autocorrelação espacial e heterogeneidade espacial nas relações estimadas pelos modelos (ANSELIN, 1988 & 1999; LE SAGE, 1998).

De acordo com Le Sage (1998), estas duas questões, ignoradas pela econometria tradicional, violam as propriedades básicas do teorema de Gauss-Markov, essenciais para modelos clássicos de regressão. No que diz respeito à violação pela existência de dependência espacial, as propriedades do teorema de Gauss-Markov exigem que as variáveis explicativas sejam fixas em amostragem repetidas. Do mesmo modo, a heterogeneidade espacial viola a propriedade referente à existência de uma única relação linear em toda a amostra de dados observados.

No contexto das regressões, os efeitos espaciais pertencem à dependência espacial, ou, segundo Anselin (1999), a suas expressões mais fracas, heterogeneidade espacial e autocorrelação espacial<sup>3</sup>. A heterogeneidade espacial refere-se à instabilidade estrutural dos erros ou dos coeficientes de modelos compostos por dados de cortes transversais de diferentes

---

<sup>3</sup> Anselin (1999) ressalta que, apesar de utilizar os termos dependência espacial e autocorrelação espacial de maneira indiferente, ambos não são idênticos. Assim, no sentido de manter uma padronização dos termos, a forma mais fraca foi utilizada.

unidades espaciais<sup>4</sup>. Mas, uma vez que este estudo não tem como objetivo verificar os efeitos da heterogeneidade espacial na distribuição do tamanho dos municípios brasileiros, não cabe aqui detalhar as especificidades de sua formalização.

A autocorrelação espacial é encontrada com frequência na maioria dos trabalhos aplicados de ciências regionais, em que os dados são obtidos para observações dispostas no espaço, ou no tempo e no espaço. Nessas situações, as observações podem ser caracterizadas por sua localização absoluta, utilizando um sistema de coordenadas, ou por sua localização relativa, baseada em distâncias métricas particulares, e os dados são, portanto, organizados em observações de unidades espaciais (ANSELIN, 1988).

Como define Anselin (1988), a autocorrelação espacial pode ser considerada como a ausência de independência estatística de observações obtidas ao longo do espaço geográfico;

$$y_i = f(y_1, y_2, \dots, y_N) \quad (3.1)$$

em que cada observação da variável  $y$  em  $i \in S$  (com  $S$  como conjunto de unidades espaciais da observação) é relacionada formalmente por meio da função  $f$  à magnitude das variáveis em outras unidades espaciais do sistema. Segundo o autor, sua verificação condiciona-se à hipótese de que essa relação de dependência entre as observações seja um subproduto das medidas de erros para cada unidade espacial contígua. Formalmente,

$$\text{Cov}[y_i, y_j] = E[y_i y_j] - E[y_i] \cdot E[y_j] \neq 0, \text{ para } i \neq j \quad (3.2)$$

em que  $i, j$  refere-se às observações individuais de determinado espaço geográfico, e  $y_{i(j)}$  é o valor da variável aleatória investigada em cada espaço. Em escala agregada, para que a matriz de covariância, composta pelos elementos resultantes de 3.2, se torne significativa no que diz respeito à existência de autocorrelação espacial, a configuração particular de pares não-nulos  $i, j$  deve apresentar interação ou combinação espacial de suas observações. Nessas situações, geralmente, há pouca correspondência entre o escopo espacial do fenômeno em estudo e o

---

<sup>4</sup> De maneira mais precisa, parâmetros e formas funcionais variam de acordo com o local e não são homogêneos ao longo do conjunto de dados. A heterogeneidade espacial pode ser abordada por meio de instrumentos econométricos padrões. Contudo, Anselin (1999) destaca três razões para tratar heterogeneidade espacial de maneira explícita. Primeiro, a estrutura por trás da instabilidade é espacial (ou geográfica) no sentido de que a localização das observações é crucial na determinação da forma da instabilidade. Considere um conjunto  $S$  de  $N$  unidades geográficas (i.e. estados, países, etc), divididos em  $R$  subconjuntos compactos não sobrepostos  $S_r$ , em que  $r = 1, 2, \dots, R$  e para cada  $r, s$  ( $r \neq s$ ),  $S_r \cap S_s = \emptyset$ . A heteroscedasticidade espacial ao longo do grupo seguiria então na forma de variâncias dos erros espacialmente aglomerados para a observação  $i$ .  $\text{Var}[\varepsilon_i] = \sigma_r^2$  quando  $i \in S_r$ . Similarmente, variabilidades nos coeficientes da regressão poderiam ser especificadas para corresponderem aos denominados regimes espaciais, ou subconjuntos  $S_r$  dos dados em que a inclinação do modelo é diferente,  $\beta_i = \beta_r$ , para uma observação  $i \in S_r$ . Em segundo lugar, como a estrutura dos dados é espacial, heterogeneidade ocorrerá frequentemente concomitantemente com autocorrelação espacial, portanto as técnicas econométricos padrões não são apropriadas. Em terceiro lugar, em dados de cortes transversais unitários, autocorrelação espacial e heterogeneidade espacial podem ser equivalentes.

delineamento das unidades espaciais das observações. Como consequência, se as variações apresentadas por uma variável aleatória  $y$  na unidade espacial  $i$  em função de variações na unidade contígua  $j$  não forem consideradas, haverá erros de medidas espacialmente correlacionados.

Tratando-se de dados em cortes transversais com  $N$  observações, não há informações suficientes para se estimar, diretamente dos dados, a matriz de covariância  $N$  por  $N$  dada por 3.2, o que impossibilita, por sua vez, a estimativa dos pesos atribuídos às unidades contíguas. Para isso, um método alternativo deve ser aplicado para sua estimação.

## 2.1 PESO E DEFASAGEM ESPACIAL

De maneira análoga à análise de séries temporais, processos estocásticos espaciais podem ser caracterizados como processos autoregressivos espaciais (SAR) ou médias móveis espaciais (SMA), entretanto existem diferenças importantes entre os contextos de cortes transversais e de séries temporais. Em contraste com a noção inequívoca de “deslocamento” ao longo de eixo temporal, Anselin (1999) destaca que não há um conceito correspondente no domínio espacial, especialmente quando as observações se localizam de maneira irregular no espaço. Ao invés da noção de deslocamento, um operador de defasagem espacial é utilizado, dado pela média ponderada de variáveis aleatórias em localizações contíguas (vizinhas). Nesse conceito, segundo o autor, a definição de um conjunto de unidades contíguas para cada unidade espacial é essencial. Para isso, associam-se a cada unidade contígua  $j$  (coluna) à unidade espacial  $i$  (linha), elementos não-nulos  $w_{ij}$ , organizados em uma matriz positiva  $N$  por  $N$  de pesos espaciais fixos  $\mathbf{W}$ .

Formalmente, a defasagem espacial para  $y$  em  $i$  é expressa como

$$[\mathbf{W}\mathbf{y}]_i = \sum_{j=1, \dots, N} w_{ij} \cdot y_j \quad (3.3)$$

ou em notação matricial, como

$$\mathbf{W}\mathbf{y} \quad (3.4)$$

em que  $\mathbf{y}$  é um vetor  $N$  por 1 de observações de variáveis aleatórias. Como para cada  $i$  os elementos  $w_{ij}$  da matriz são não-nulos para aqueles em que  $j \in S_i$  (sendo  $S_i$  o conjunto de unidades contíguas), somente os pesos correspondentes de  $y_j$  são incluídos na defasagem.

Os elementos da matriz de pesos espaciais são tipicamente normalizados, assim, para cada  $i$ ,  $\sum_j w_{ij} = 1$ . Consequentemente a defasagem espacial pode ser interpretada como uma média

ponderada, de peso  $w_{ij}$ , pelas unidades contíguas. Entretanto, essa noção é perfeitamente geral e especificações alternativas de peso espacial podem ser construídas tendo como base a distância geográfica, seu inverso ou o quadrado do seu inverso, a estrutura de redes sociais, a distância econômica, os  $k$  vizinhos mais próximos, ou medidas de interações de troca.

De acordo com Abreu, De Groot e Florax (2004), a especificação da matriz de pesos espaciais é um grande ponto de contenção na literatura de econometria espacial, pois diferentes escolhas podem determinar diferentes resultados. A relação intrincada entre medidas de associação espacial e a escolha de uma matriz de conectividade tem várias implicações para o desempenho dos estimadores e das estatísticas de testes. De acordo com os autores, do ponto de vista metodológico, relatar as propriedades das diferentes técnicas de estabelecimento de conectividades ou dependência espacial é importante para a estruturação de experimentos simulatórios para avaliar o desempenho de estimadores produzidos por modelos que reproduzem contextos empíricos realísticos.

Contudo, Abreu, De Groot e Florax (2004) corroboram com a conclusão de Anselin (1988) de que a estrutura da dependência espacial incorporada à matriz de peso espacial deve ser escolhida judiciosamente e relacionada a conceitos gerais da teoria da interação espacial, como as noções de acessibilidade e potencial. Na linha dos modelos direcionados para econometria espacial, a matriz de pesos deve suportar uma relação direta para uma conceitualização teórica da estrutura de dependência, ao invés de refletir uma descrição *ad hoc* dos padrões espaciais.

Nos trabalhos empíricos, a matriz simples de contigüidade é a escolha mais comum, seguida pela matriz de distância e pela combinação de contigüidade e distância. Para os autores, é razoável esperar que efeitos transbordamentos devido a guerras ou instabilidade afetam primariamente os países limítrofes, assim, uma matriz de contigüidade é mais apropriada. Por outro lado, efeitos transbordamentos devido à difusão tecnológica devem ter um alcance mais abrangente, assim, uma matriz baseada em distância é mais apropriada.

Dada à importância da escolha das propriedades de pesos para a interpretação de modelos espaciais, Anselin (1988) destaca que algumas características inerentes das matrizes de pesos ainda devem ser consideradas, pois como os elementos da matriz espacial referem-se a unidades espaciais com limites irregulares e arbitrários, não é possível definir uma estrutura espacial única e identificável com propriedades estatísticas claras e independentes do modo como os dados são organizados em unidades espaciais.

### 2.1.2 O problema da unidade de área modificável

O problema da unidade de área modificável pertence ao fato de que medidas estatísticas para dados em cortes transversais são sensíveis ao modo como as unidades espaciais estão organizadas. De maneira específica, o nível de agregação e a disposição espacial em conjuntos de unidades contíguas afetam a magnitude de medidas do modelo de regressão, como coeficientes de correlação espacial e parâmetros. Portanto, o problema da unidade de área modificável pode ser considerado como uma combinação de dois problemas familiares da econometria: agregação e identificação (ANSELIN, 1988).

Segundo Anselin (1988), o primeiro aspecto do problema pertence à agregação de unidades espaciais. Como é notório, o processo de agregação é significativo apenas se realizado de modo homogêneo ao longo de todas as unidades observadas. Caso contrário, a heterogeneidade inerente aos dados e sua instabilidade estrutural devem ser consideradas para qualquer agregação. Esse aspecto do problema da unidade de área modificável é, na realidade, uma questão de especificação relacionada à heterogeneidade espacial, e não somente uma questão determinada pela organização espacial dos dados.

O segundo aspecto, pertence à identificação apropriada da estrutura da dependência espacial. Como discutido anteriormente, uma análise de associação espacial é realizada por meio da associação de uma variável a uma contraparte espacialmente defasada, construída como uma combinação linear de observações. Em geral, essa associação é realizada por meio de um coeficiente de correlação ou regressão; uma variável  $y$  deve ser relacionada a  $\mathbf{W}y$ , por meio de  $\rho$ , que é um coeficiente espacial autoregressivo. Assim, como diferentes escolhas de  $\mathbf{W}$  resultam em diferentes valores de  $\rho$ , a medida da associação espacial é indeterminada.

Do ponto de vista econométrico, esse problema só deve ser encarado como um problema de identificação se houver informações insuficientes para a construção do modelo, de modo a impedir uma especificação completa das interações espaciais simultâneas. Nesse sentido, uma formulação de autocorrelação espacial linear pode ser considerada como um caso especial de um sistema de equações simultâneas lineares, com uma observação para cada equação

$$y_1 = \sum_j \beta_{ij} y_j, \forall i \quad (3.5)$$

no sistema.

Para que pelo menos alguns parâmetros do modelo sejam identificáveis a abordagem usual da econometria espacial introduz uma variável espacialmente defasada, que resulta em

$$y_1 = \rho \sum_j w_{ij} y_j \quad (3.6)$$

Como ressaltado, diferentes escolhas de pesos  $w_{ij}$  podem resultar em diferentes estimativas para  $\rho$ . Contudo, razões teóricas *a priori* determinam a forma particular para as restrições de identificação, similares à abordagem utilizada em um sistema de equações simultâneas. Especificações competitivas podem subseqüentemente ser comparadas por meio dos testes de especificação e procedimentos de seleção de modelos (ANSELIN, 1988).

## 2.2 MODELOS DE REGRESSÃO ESPACIAL

A abordagem mais utilizada para expressar autocorrelação espacial é por meio da especificação de processos espaciais estocásticos, i.e., uma relação funcional entre uma variável aleatória em dada localização e a mesma variável em outras localizações. A estrutura de covariância surge então da natureza do processo. Por exemplo, dada uma matriz de pesos  $\mathbf{W}$  ( $N$  por  $N$ ), um vetor  $N$  por 1 de variáveis aleatórias  $\mathbf{y}$ , e um vetor  $N$  por 1 de erros aleatórios  $\boldsymbol{\varepsilon}$  independentemente e identicamente distribuídos (i.i.d.), um processo espacial autoregressivo (SAR) é definido como

$$(\mathbf{y} - \mu\mathbf{i}) = \rho\mathbf{W}(\mathbf{y} - \mu\mathbf{i}) + \boldsymbol{\varepsilon}, \text{ ou } (\mathbf{y} - \mu\mathbf{i}) = (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon} \quad (3.7)$$

em que  $\mathbf{I}$  é uma matriz identidade  $N$  por  $N$ ,  $\mathbf{i}$  é um vetor  $N$  por 1,  $\mu$  é a média comum das variáveis aleatórias  $y_i$ , os termos erros i.i.d. de média zero possuem variância  $\sigma^2$ , e  $\rho$  e  $\lambda$  são respectivamente os parâmetros autorregressivos e de média móvel (ANSELIN, 1999).

As restrições impostas pela estrutura de pesos, juntamente com a forma específica do processo espacial determina a matriz de variância-covariância para  $\mathbf{y}$  como uma função de dois parâmetros, variância  $\sigma^2$  e coeficiente espacial  $\rho$  ou  $\lambda$ . Para a estrutura SAR em (3.7), isso produz, a partir  $E[\mathbf{y} - \mu\mathbf{i}] = 0$ ,

$$\text{Cov}[(\mathbf{y} - \mu\mathbf{i}), (\mathbf{y} - \mu\mathbf{i})] = E[(\mathbf{y} - \mu\mathbf{i})(\mathbf{y} - \mu\mathbf{i})'] = \sigma^2[(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})]^{-1} \quad (3.8)$$

Essa é uma matriz completa que determina que choques em qualquer uma das unidades espaciais afetam todas as outras por meio de efeitos multiplicadores espaciais, ou interações globais. Como a heteroscedasticidade depende da estrutura da vizinhança incorporada na matriz de pesos espaciais  $\mathbf{W}$ , o processo em  $\mathbf{y}$  não possui covariância estacionária. A estacionariedade é obtida apenas em casos muito raros, por exemplo, em estruturas regularmente entrelaçadas quando cada observação possui estrutura de peso idêntica às demais. Essa falta de estacionariedade possui importante implicação para o Teorema do

Limite Central e para a Lei dos Grandes Números que precisam ser invocados para a obtenção das propriedades assintóticas dos estimadores e dos testes de especificação.

Segundo Anselin (1999), diversas especificações de modelos para processos espaciais foram sugeridos na literatura e aplicados empiricamente. Contudo, alguma estrutura pode ser imposta, guiada pelo princípio de que técnicas econométricas podem ser essencialmente aplicadas para modelos agrupados da mesma maneira.

### 2.2.1 Defasagem espacial e modelo Erro-espacial

Nos modelos padrões de regressão linear, a dependência espacial pode ser incorporada de duas maneiras distintas: como um regressor adicional na forma de variáveis dependentes espacialmente defasadas ( $\mathbf{W}\mathbf{y}$ ), ou em estruturas de erro ( $E[\varepsilon_i \varepsilon_j] \neq 0$ ). Este último é referido como um modelo espacialmente defasado e é apropriado quando o foco de interesse é a avaliação da existência e da força da interação espacial. Isto é interpretado como dependência espacial substantiva no sentido de ser diretamente relacionada ao modelo espacial. Dependência espacial no termo de perturbação da regressão, ou um modelo erro-espacial pode assumir qualquer uma das formas descritas no início do capítulo e são referidos como *nuisance dependence*. Isto é apropriado quando se pretende corrigir o modelo econométrico pelas influências potencialmente viesantes da autocorrelação espacial, que ocorrem em função do uso de dados espaciais.

Formalmente, um modelo espacialmente defasado, ou um modelo espacial autorregressivo é expresso como

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.9)$$

em que  $\rho$  é um coeficiente espacial autorregressivo,  $\boldsymbol{\varepsilon}$  é um vetor de termos erros, e as demais variáveis permanecem como o descrito anteriormente. De modo contrário ao que assegura a contrapartida de séries temporais desse modelo, o termo de defasagem espacial  $\mathbf{W}\mathbf{y}$  é correlacionado com as perturbações, mesmo quando este último é i.i.d. . Isso pode ser visto a partir da forma reduzida de (3.9)

$$\mathbf{y} = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon} \quad (3.10)$$

em que cada inverso pode ser expandido em séries infinitas, incluindo ambas as variáveis explicativas e os termos erros em todas as unidades espaciais, por meio de multiplicadores espaciais. Consequentemente, o termo de defasagem espacial deve ser tratado como uma variável endógena e métodos apropriados de estimação devem ser considerados para essa

endogeneidade, uma vez que os parâmetros estimados pelo Método dos Mínimos Quadrados (MQO) serão viesados e inconsistentes devido ao viés de simultaneidade.

Um modelo erro-espacial é um caso especial de regressão com termo erro não esférico, em que os elementos da diagonal da matriz de covariância expressam a estrutura da dependência espacial. Conseqüentemente, os parâmetros estimados por MQO permanecem não viesados, mas não eficientes e os estimadores clássicos para erros-padrões são viesados. Como descrito na seção 3.1, a estrutura espacial pode ser especificada em diferentes maneiras, e, exceto para a abordagem não-paramétrica, resulta em uma matriz de erros de variância-covariância na forma de

$$E[\varepsilon\varepsilon'] = \Omega(\theta). \quad (3.11)$$

em que  $\theta$  é um vetor de parâmetros, como os coeficientes em um processo SAR ou SMA. Quando o processo de erro é SAR, o modelo resultante também pode ser expresso como uma especificação espacialmente defasada, na forma de Durbin espacial ou modelo espacial de fator comum. O modelo de erro SAR é

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \text{ e } \varepsilon = \lambda\mathbf{W}\varepsilon + \mathbf{u}. \quad (3.12)$$

Como  $\varepsilon = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u}$ , logo  $\mathbf{y} = \mathbf{X}\beta + \varepsilon = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u}$ , (3.12) é equivalente a

$$\mathbf{y} = \lambda\mathbf{W}\mathbf{y} + \mathbf{X}\beta - \lambda\mathbf{W}\mathbf{X}\beta + \varepsilon, \quad (3.13)$$

que é um modelo espacialmente defasado com um conjunto adicional de variáveis exógenas espacialmente defasadas ( $\mathbf{W}\mathbf{X}$ ) e um conjunto de  $k$  restrições não-lineares aos coeficientes. Isso indica que o produto do coeficiente autorregressivo com os coeficientes da regressão  $\beta$  deve igualar a negatividade dos coeficientes  $\mathbf{W}\mathbf{X}$ . A similaridade entre o modelo erro (3.13) e o modelo espacialmente defasado “puro” (3.9), na prática, traz complicações para os testes de especificação, uma vez que os testes formulados para modelos alternativos de defasagem espacial também terão poder contra modelos alternativos de erros-espaciais e vice-versa.

A maioria dos modelos de regressão espacial, utilizados na prática, é baseada em uma única matriz de pesos espaciais. Entretanto, em princípio, modelos de ordem superior também são possíveis, como modelos SAR de ordem superior, espaciais autorregressivos, especificações de média móvel ou SARMA, e modelos que incluem tanto variáveis dependentes espacialmente defasadas quanto processos SAR de erro. Na especificação de modelos dessa ordem, deve-se cuidar para que os pesos sejam únicos, ortogonais, e que todos os coeficientes sejam identificados.

Existem duas formas de autocorrelação espacial: defasagem espacial (*spatial lag*) e erro-espacial (*spatial error*). O modelo que incorpora a defasagem espacial, SAR (*Spatial Autoregressive model*), é apresentado por

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.14)$$

onde  $\mathbf{y}$  é o vetor de observações da variável endógena,  $\mathbf{W}$  é a matriz das vizinhanças,  $\mathbf{X}$  é a matriz das observações das variáveis exógenas,  $\boldsymbol{\beta}$  é o vetor dos coeficientes,  $\rho$  é o coeficiente espacial autoregressivo e  $\boldsymbol{\varepsilon}$  é o vetor dos erros. O coeficiente  $\rho$  é uma medida que explica como as observações vizinhas afetam a variável dependente. O modelo de erro espacial, SEM (*Spatial Error model*), é expresso por

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (3.15)$$

em que a dependência espacial está considerada no termo erro

$$\boldsymbol{\varepsilon} = \rho \mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu}, \quad (3.16)$$

### 2.3 MODELOS ESPACIAIS DE REGRESSÃO LINEAR PARA DADOS NO TEMPO-ESPAÇO

Nos trabalhos empíricos das ciências regionais, os modelos compostos por observações dispostas em duas dimensões, tempo e espaço, tornam-se cada vez mais relevantes. A introdução da dimensão temporal aumenta consideravelmente a complexidade de questões que devem ser levadas em consideração na especificação dos modelos de econometria espacial. Em termos econométricos, os modelos espaciais de regressão para dados no tempo-espaço são compostos por dados empilhados em cortes transversais e em séries temporais (*pooled cross-section e time serie data*), e consideram padrões de dependência transversal (*cross-sectional dependence*) e heterogeneidade (ANSELIN, 1988).

A formalização dessa categoria de modelo foi desenvolvida por Anselin (1988), que como ponto de partida, considerou a especificação a seguir para expressar um conjunto completo de potenciais dependências tempo-espaço e formas de heterogeneidade

$$y_{it} = x_{it}\beta_{it} + \varepsilon_{it} \quad (3.17)$$

em que  $x_{it}$  é uma coluna de vetores de observações para a unidade espacial  $i$  no tempo  $t$ ,  $\beta_{it}$  é um vetor de parâmetros específicos de tempo-espaço, e  $\varepsilon_{it}$  é um termo erro. O termo erro é caracterizado pelas seguintes condições

$$E[\varepsilon_{it}\varepsilon_{it}] = \sigma, \quad \text{variância constante;}$$

$$E[\varepsilon_{it}\varepsilon_{it}] = \sigma_i, \text{ heterogeneidade espacial;}$$

$$E[\varepsilon_{it}\varepsilon_{it}] = \sigma_t, \text{ heterogeneidade temporal;}$$

$$E[\varepsilon_{it}\varepsilon_{it}] = \sigma_{it}, \text{ variância tempo-espaço específica.}$$

Com  $i \neq j$  e  $t = s$ , a dependência é uma correlação espacial contemporânea, que pode ser a mesma para todos os períodos, ou específica a cada  $t$ . Formalmente

$$E[\varepsilon_{it}\varepsilon_{jt}] = \sigma_{ij}(t), \text{ correlação contemporânea.}$$

Com  $i = j$  e  $t \neq s$ , a dependência está no domínio do tempo, novamente, ou constante ao longo de todas as unidades espaciais, ou variando com o local  $i$ . Formalmente

$$E[\varepsilon_{it}\varepsilon_{is}] = \sigma_{ts}(i), \text{ correlação temporal.}$$

Quando  $i \neq j$  e  $t \neq s$ , o padrão de dependência atinge ao longo do tempo e do espaço simultaneamente

$$E[\varepsilon_{it}\varepsilon_{js}] = \sigma_{ij}(ts), \text{ correlação tempo-espaço.}$$

Devido à falta de graus de liberdade para estimar  $\beta_{it}$  para cada observação a expressão (3.17) não é operacional. A forma específica em que cada restrição é imposta à forma geral permite várias modelagens interessantes nas dimensões tempo-espaço.

O oposto exato de (3.17) em termos de variabilidades dos coeficientes é

$$y_{it} = x_{it}\beta + \varepsilon_{it} \quad (3.18)$$

em que os parâmetros são fixos ao longo de todas as observações no tempo e no espaço.

Quando o parâmetro  $\beta_{it}$  é fixado em uma das duas dimensões e os termos erros são correlacionados por meio da outra dimensão, tem-se como resultado uma regressão aparentemente não-relacionada (*seemingly unrelated regression* - SUR), desenvolvida por Zellner (1962).

Segundo Anselin (1988), quando o coeficiente varia ao longo do espaço, mas é constante ao longo do tempo, o termo erro é contemporaneamente correlacionado

$$y_{it} = x_{it}\beta_i + \varepsilon_{it} \quad (3.19)$$

com

$$E[\varepsilon_{it}\varepsilon_{jt}] = \sigma_{ij} \quad (3.20)$$

A autocorrelação dos erros na dimensão tempo para cada localização  $i$  é também introduzida se

$$E[\varepsilon_{it}\varepsilon_{is}] = \sigma_{ts}. \quad (3.21)$$

Por outro lado, na situação em que o coeficiente é específico a cada período de tempo e constante ao longo do espaço, um tipo diferente de modelo é obtido quando há correlação espacial dos erros ao longo do tempo. Nessa situação tem-se, portanto, o modelo SUR espacial

$$y_{it} = x_{it}\beta_t + \varepsilon_{it} \quad (3.22)$$

com

$$E[\varepsilon_{it}\varepsilon_{is}] = \sigma_{ts}. \quad (3.23)$$

e

$$E[\varepsilon_{it}\varepsilon_{jt}] = \sigma_{ij}. \quad (3.24)$$

Nessa especificação, uma variável espacialmente defasada, ou uma variável temporalmente defasada pode ser incluída. De maneira análoga, a heteroscedasticidade pode ser introduzida ao tornar a variância do erro diferente ao longo dos períodos de tempo no modelo (3.19), e diferente ao longo das unidades espaciais no modelo (3.22).

Uma questão particularmente interessante, de acordo com Anselin (1988), a respeito da especificação, consiste da escolha entre uma expressão geral ou uma parametrização específica para a dependência espacial ou temporal nos termos erros. Para ilustrar essa situação, o autor utiliza como exemplo um contexto espacial econométrico usando o modelo (3.22), em que a correlação contemporânea pode ser expressa na forma de covariância geral  $\sigma_{ij}$ , ou pode ser parametrizada como

$$\boldsymbol{\varepsilon} = \rho \mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu} \quad (3.25)$$

sendo que para cada período de tempo, o  $\boldsymbol{\varepsilon}$  é um vetor de termos erros ao longo do espaço.

Em contraste à situação pura de cortes transversais, quando essa parametrização é uma necessidade, nas dimensões tempo-espaço a escolha entre as formas estruturadas e não-estruturadas para dependência pode ser baseada em outras considerações (ANSELIN, 1988). Assim, formas especiais para a covariância de erros são resultantes quando a heterogeneidade é expressa em termos de coeficientes aleatórios de variação. A variação do coeficiente  $\beta_{it}$  é formalmente expressa como

$$\beta_{it} = \beta + \mu_{it} \quad (3.26)$$

em que  $\beta$  é um valor médio para os coeficientes e o termo erro  $\mu_{it}$  permite variações ao longo do tempo  $t$ , espaço  $i$ , ou tempo-espaço. Consequentemente, em função das suposições sobre  $\mu_{it}$ , vários padrões de dependência dos erros e heterogeneidades dos erros podem ser englobados dentro da estrutura geral delineada acima.

### 3.3.1 Regressões espaciais aparentemente não relacionadas (*Spatial seemingly unrelated regression – spatial SUR*)

Com base em Anselin (1988), os modelos do tipo Regressões Aparentemente Não Relacionadas (SUR, *Seemingly Unrelated Regression*) é um daqueles que possibilitam contemplar a dependência espacial. Originalmente sugerido por Arnold Zellner em 1962, foi designado para situações empíricas onde existe um limitado grau de simultaneidade na forma de dependência entre os erros de diferentes equações. Se as equações pertencem a uma série temporal de diferentes regiões a dependência resultante pode ser considerada como uma forma de autocorrelação espacial.

Segundo o autor, o SUR e o SUR espacial são casos especiais da taxonomia geral de modelos de espaço-tempo, em que os dados das variáveis dependentes,  $y_{it}$ , e dos vetores ( $I$  por  $K$ ) das variáveis explicativas,  $x_{it}$ , estão organizados em unidades espaciais  $i$  ( $i = 1, \dots, N$ ) e nos períodos de tempo  $t$  ( $t = 1, \dots, T$ ).

Retomando a discussão da seção anterior, em seu formato tradicional, o SUR apresenta coeficientes  $\beta_i$  que variam de acordo com a unidade espacial, mas são constantes ao longo do tempo. Os erros, espacialmente correlacionados, exibem covariância constante entre si para diferentes unidades espaciais, medidas no mesmo ponto do tempo. De maneira formal, o modelo é expresso pela equação

$$y_{it} = x_{it}\beta_i + \varepsilon_{it}, \quad (3.27)$$

com

$$E[\varepsilon_{it} \cdot \varepsilon_{jt}] = \sigma_{ij}. \quad (3.28)$$

Na forma matricial, a equação para cada período de tempo  $t$  torna-se

$$\mathbf{y}_t = \mathbf{X}_t\beta_t + \boldsymbol{\varepsilon}_t \quad (3.29)$$

em que  $\mathbf{y}_t$  e  $\boldsymbol{\varepsilon}_t$  são vetores  $T$  por 1 e  $\mathbf{X}_t$  é uma matriz  $T$  por  $K_t$  de variáveis explicativas. Em geral, o número de variáveis explicativas,  $K_t$ , pode ser diferente para cada equação (unidade espacial).

No desenvolvimento econométrico realizado por Anselin (1988), o SUR, em seu formato espacial, exhibe coeficientes  $\beta_i$  constantes no espaço, mas variáveis ao longo do tempo. Os erros  $\varepsilon$  são temporalmente correlacionados, i.e., existe uma covariância constante entre resíduos de mesma unidade espacial, mas de diferentes períodos. O modelo pode ser expresso pela equação

$$y_{it} = x_{it}\beta_t + \varepsilon_{it}, \quad (3.30)$$

com

$$E[\varepsilon_{it} \cdot \varepsilon_{is}] = \sigma_{is}. \quad (3.31)$$

Na forma matricial, a equação para cada período de tempo  $t$  torna-se

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t \quad (3.32)$$

em que  $\mathbf{y}_t$  e  $\boldsymbol{\varepsilon}_t$  são vetores  $N$  por  $1$  e  $\mathbf{X}_t$  é uma matriz  $N$  por  $K$  de variáveis explicativas. Em geral, o número de variáveis explicativas,  $K_t$ , pode ser diferente para cada equação, que se refere a cada período de tempo.

De acordo com Anselin (1988), o modelo SUR espacial pode ser operacionalizado somente quando mais observações estão disponíveis na dimensão espacial do que na dimensão temporal ( $N > T$ ). Essa modelagem é particularmente adequada para situações em que dados transversais são obtidos para um pequeno número de períodos de tempo, como no caso de dados censitários decenais.

A estimação e o teste de hipóteses nos modelos SUR e SUR espacial podem ser tratados como casos especiais da estrutura com uma matriz geral de variância do erro não esférica. Isto é ilustrado mais facilmente quando as equações da regressão são combinadas em formato empilhado (ANSELIN, 1988). Para o modelo SUR espacial, as equações para períodos de tempo  $1$  a  $T$  são combinadas como

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & X_T \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix} \quad (3.33)$$

ou, agrupados,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.34)$$

em que  $\mathbf{Y}$  é um vetor  $NT$  por  $1$  de variáveis dependentes,  $K$  é o número total de coeficientes ( $=\sum_t K_t$ ),  $\mathbf{X}$  é uma matriz diagonal de dimensões  $NT$  por  $K$ ,  $\boldsymbol{\beta}$  é o vetor de coeficientes de dimensão  $K$  por  $1$ , e  $\boldsymbol{\varepsilon}$  é um vetor erro  $NT$  por  $1$ .

A dependência entre os vetores de erro é tanta que para cada par de períodos de tempo  $t,s$

$$E[\varepsilon_t \cdot \varepsilon_s'] = \sigma_{ts} \cdot \mathbf{I}. \quad (3.35)$$

Isso produz uma matriz de variância do erro  $\Omega$  de formato

$$E[\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}'] = \Omega = \Sigma \otimes \mathbf{I} \quad (3.36)$$

em que  $\Sigma$  é uma matriz  $T$  por  $T$  com  $\sigma_{ts}$  como seus elementos, e  $\otimes$  é o produto de Kronecker.

Quando os elementos de  $\Sigma$  são considerados como conhecidos, o método dos Mínimos Quadrados Generalizados (MQG) pode ser aplicado ao sistema completo, como

$$\mathbf{b}_{\text{MQG}} = [\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{X}]^{-1} \mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y} \quad (3.37)$$

com matriz de covariância:

$$\text{var}(b_{\text{MQG}}) = [\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{X}]^{-1}. \quad (3.38)$$

Para Anselin (1988), a estrutura da matriz de variância dos erros  $\Omega$  e o uso do produto de Kronecker opõem-se à inversão de matrizes de dimensão completa  $NT$  requerida pela estimação. Assim, somente a inversão das matrizes de ordem  $K$  (para  $\mathbf{X}'\Omega^{-1}\mathbf{X}$ ) e  $T$  (para  $\boldsymbol{\Sigma}$ ) é necessária.

Geralmente, elementos de  $\boldsymbol{\Sigma}$  não são exatamente conhecidos, mas precisam ser estimados junto aos outros coeficientes do modelo. Como consequência, a inferência deve basear-se somente em considerações assintóticas. A forma de estimação mais apropriada dos coeficientes do SUR-espacial é por meio do MQG ou por Máxima Verossimilhança (MV).

Ambos os modelos de defasagem espacial e erro-espacial são casos especiais de uma especificação mais geral que também podem incluir formas de heteroscedasticidade. Isso também propicia bases para a estimação de modelos SUR espaciais com defasagem espacial ou erro-espacial por meio de Máxima-Verossimilhança.

*a SUR espacial com variáveis dependentes espacialmente defasadas*

A presença de uma variável espacialmente defasada por si só é suficiente para excluir MQO de produzir coeficientes eficientes. Como consequência, quando variáveis espacialmente defasadas estão presentes, o passo inicial em um procedimento SUR espacial não pode ser baseado em uma estimação por MQO.

Essa situação é similar à dos modelos de equações simultâneas de erros relacionados e pode ser abordada pelos métodos MV ou Variáveis Instrumentais (VI). Formalmente cada equação no sistema é expressa por

$$y_t = \gamma_t W y_t + X_t \beta_t + \varepsilon_t \quad (3.39)$$

ou,

$$A_t y_t = X_t \beta_t + \varepsilon_t \quad (3.40)$$

com,

$$A_t = I - \gamma_t W \quad (3.41)$$

e, como antes,

$$E[\varepsilon_t \varepsilon_s'] = \sigma_{ts}. \quad (3.42)$$

O sistema completo pode ser representado em forma empilhada como

$$\mathbf{AY} = \mathbf{X}\beta = \boldsymbol{\varepsilon} \quad (3.45)$$

em que,

$$A = \mathbf{I} - (\mathbf{\Gamma} \otimes \mathbf{W}) \quad (3.46)$$

com  $\mathbf{\Gamma}$  como uma matriz diagonal  $T$  por  $T$  com  $\gamma_t$  na diagonal, e  $\mathbf{I}$  como uma matriz identidade de dimensão  $NT$ .

*b SUR espacial com autocorrelação espacial dos erros*

Como o SUR espacial consiste em equações para cada período de tempo estimadas para dados de cortes transversais de unidades espaciais, a autocorrelação espacial dos erros, inerentes às equações, é um problema em potencial. Portanto, de maneira análoga à autocorrelação serial dos erros, tem-se que os erros seguem um processo autoregressivo em cada equação, com parâmetros diferentes para cada  $t$  e correlacionados entre si.

$$\mathbf{y}_t = \mathbf{x}_t \boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \quad (3.47)$$

com

$$\boldsymbol{\varepsilon}_t = \lambda_t \mathbf{W} \boldsymbol{\varepsilon}_t + \boldsymbol{\mu}_t, \quad (3.48)$$

e

$$E[\boldsymbol{\mu}_t \boldsymbol{\mu}'_s] = \sigma_{ts} \cdot \mathbf{I}. \quad (3.49)$$

O vetor de erros espacialmente dependentes  $\boldsymbol{\varepsilon}_t$  pode ser considerado como uma transformação do termo independente  $\boldsymbol{\mu}_t$ , como

$$\boldsymbol{\varepsilon}_t = (\mathbf{I} - \lambda_t \mathbf{W})^{-1} \boldsymbol{\mu}_t. \quad (3.50)$$

Consequentemente tem-se que

$$\begin{aligned} E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_s] &= E[(\mathbf{I} - \lambda_t \mathbf{W})^{-1} \boldsymbol{\mu}_t \cdot \boldsymbol{\mu}'_s (\mathbf{I} - \lambda_s \mathbf{W})^{-1}] \\ &= \sigma_{ts} \cdot [(\mathbf{I} - \lambda_t \mathbf{W})' (\mathbf{I} - \lambda_s \mathbf{W})]^{-1} \end{aligned} \quad (3.51)$$

ou

$$E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_s] = \sigma_{ts} \cdot \mathbf{B}_t \cdot \mathbf{B}_s' \quad (3.52)$$

em que, para uma simplificação notacional,

$$\mathbf{B}_t = (\mathbf{I} - \lambda_t \mathbf{W})^{-1}. \quad (3.53)$$

A covariância do erro para o sistema completo,  $\Omega$ , toma o seguinte formato

$$\Omega = E[\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}'] = \mathbf{B} (\boldsymbol{\Sigma} \otimes \mathbf{I}) \mathbf{B}' \quad (3.54)$$

em que  $\boldsymbol{\varepsilon}$  é um vetor de erros empilhados  $NT$  por  $1$ ,  $\boldsymbol{\Sigma}$  é a matriz de covariância  $T$  por  $T$  das equações, e  $\mathbf{B}$  é uma matriz diagonal  $NT$  por  $NT$ :

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & 0 & 0 \\ 0 & \mathbf{B}_2 & 0 \\ 0 & 0 & \mathbf{B}_T \end{bmatrix} \quad (3.55)$$

Alternativamente, usando a suposição de uma  $\mathbf{W}$  constante em todas as equações,  $B$  pode ser expresso como:

$$B = [\mathbf{I} - (\mathbf{\Lambda} \otimes \mathbf{W})]^{-1} \quad (3.56)$$

em que  $\mathbf{\Lambda}$  é uma matriz diagonal  $T$  por  $T$  contendo  $\lambda_t$ , e  $\mathbf{I}$  é uma matriz identidade  $NT$  por  $NT$ .

## 2.4 TESTES DE ESPECIFICAÇÃO

### 2.4.1 I de Moran

Teste de especificação utilizado com maior frequência para identificar autocorrelação espacial, é derivado da estatística desenvolvida por Moran (1948) surgiu como uma analogia bidimensional ao teste de Durbin-Watson para séries de tempo. Em notação matricial, a estatística I de Moran é

$$I = (N/S_0)(\mathbf{e}'\mathbf{W}\mathbf{e}/\mathbf{e}'\mathbf{e}) \quad (3.57)$$

com  $\mathbf{e}$  como um vetor de resíduos MQO e  $S_0 = \sum_i \sum_j w_{ij}$ , um fator de padronização que corresponde à soma dos pesos para os produtos não-nulos. A inferência do I de Moran é baseada em uma aproximação normal, utilizando um valor-z normalizado obtido a partir das expressões para a média e variância da estatística.

### 2.4.2 Teste de Máxima-Verossimilhança

Quando modelos de regressões espaciais são estimados por máxima-verossimilhança, inferências para os coeficientes autorregressivos espaciais podem ser baseadas em Wald ou no teste  $t$  assintótico (da matriz de variância assintótica) ou em testes de razão de verossimilhança (LR). Ambas as abordagens requerem que o modelo (hipótese) alternativo, i.e. o modelo espacial, seja estimado. Em contraste, uma série de estatísticas de teste baseadas no princípio do multiplicador de Lagrange (LM) somente requer a estimação do modelo sob a hipótese nula. O teste LM também permite distinções entre erro-espacial e defasagem espacial.

O teste LM de uma alternativa erro-espacial foi originalmente sugerido por Burridge (1980) e assume o seguinte formato

$$LM_{err} = [\mathbf{e}'\mathbf{W}\mathbf{e}/(\mathbf{e}'\mathbf{e}/N)]^2 / [\text{tr}(\mathbf{W}^2 + \mathbf{W}'\mathbf{W})] \quad (3.58)$$

Essa estatística possui uma distribuição assintótica  $\chi^2(1)$  e, dissociada do fator de escala, corresponde ao quadrado do I de Moran. A partir de vários experimentos simulatórios (ANSELIN & REY, 1991; ANSELIN & FLORAX, 1995b) tem-se que o I de Moran possui

desempenho levemente superior ao teste  $LM_{err}$  em amostras pequenas, mas o desempenho de ambos os testes torna-se indistinguível em amostras de tamanho médio ou grande.

O teste LM de uma alternativa defasagem espacial foi descrito em Anselin (1988c) e toma a forma de

$$LM_{lag} = [\mathbf{e}'\mathbf{W}\mathbf{e}/(\mathbf{e}'\mathbf{e}/N)]/D \quad (3.59)$$

em que  $D = [(\mathbf{W}\mathbf{X}\boldsymbol{\beta})'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X})(\mathbf{W}\mathbf{X}\boldsymbol{\beta})/\sigma^2] + \text{tr}(\mathbf{W}^2 + \mathbf{W}'\mathbf{W})$ . Essa estatística também possui uma distribuição assintótica  $\chi^2(1)$ . O princípio de LM também pode ser estendido para alternativas espaciais mais complexas, como processos de ordem superior, erro componente espacial e de modelos de representação direta a dados em painel.

### 3.4.3 Teste Kelejian-Robinson

Segundo Anselin e Moreno (2003), o teste Kelejian-Robinson (KR) é o teste robusto para identificar dependência espacial dos erros, que além de não exigir uma especificação completa do processo gerador do termo erro, foi desenvolvido para ajustes gerais e também não requer normalidade ou linearidade. Em vez disso, a alternativa é que a covariância entre dois erros contíguos é não-nula, ou que

$$\text{Cov}[\varepsilon_i, \varepsilon_j] = \sigma_{ij} = z_{ij}\gamma \quad (3.60)$$

em que  $z_{ij}$  é um vetor de covariadas 1 por  $q$ , tipicamente tomado como função de variáveis explicativas originais em  $i$  e  $j$  como localizações contíguas em um ordenamento espacial geral das observações. Por exemplo, o  $z_{ij}$  pode ser construído a partir dos produtos cruzados de  $x_i$  e  $x_j$ . O vetor de coeficientes  $\gamma$  ( $q$  por 1), indica o grau em que as covariadas em  $z$  podem explicar a covariância não-nula em (3.60). Intuitivamente, a ausência de autocorrelação espacial não dever produzir uma relação significativa entre  $\text{Cov}[\varepsilon_i, \varepsilon_j]$  e  $z_{ij}$ , ou, as estimativas para os coeficientes  $\gamma$  podem não ser significantes. Formalmente, o teste KR resume-se em um teste sob a hipótese nula  $H_0: \gamma=0$  em (3.60).

O teste é aplicado ao regredir os produtos cruzados  $h_n$  dos erros obtidos por MQO para localizações contíguas,  $\hat{C}_{ij} = e_i e_j$  pelos correspondentes produtos cruzados das variáveis explicativas,  $z_{ij} = x_i \cdot x_j$ . Com  $\hat{\gamma} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{C}}$  como as estimativas de MQO nessa regressão, em que  $\mathbf{Z}$  e  $\hat{\mathbf{C}}$  são, respectivamente, uma matriz  $h_n$  por  $q$  e um vetor  $h_n$  por 1. O teste é construído como

$$KR = \frac{\hat{\gamma}'\mathbf{Z}'\mathbf{Z}\hat{\gamma}}{\hat{\sigma}^4} \quad (3.61)$$

Sob a hipótese nula, a estatística KR possui distribuição assintótica  $\chi^2(q)$ , em que  $q$  corresponde ao número de colunas da matriz  $\mathbf{Z}$ .

Mesmo que o teste KR não se refira explicitamente a uma matriz de pesos espaciais  $\mathbf{W}$ , há uma noção subjacente de ordenamento espacial. Esse ordenamento é dado como equivalente à noção de contigüidade de primeira ordem, que corresponde a elementos não-nulos da parte triangular superior ou inferior da matriz de pesos espaciais baseada em contigüidade. A seleção dos pares  $i, j$  tem o objetivo de identificar os pares que correspondem a covariâncias não-nulas (3.60).

## 2.5 CADEIA DE MARKOV

Para um exame mais detalhado da dinâmica dos dados, aplicou-se mesma metodologia utilizada por Eaton e Eckstein (1997), Black e Henderson (2003), Lanaspá et al. (2003) e Le Gallo e Chasco (2008). Inicialmente elaborada por Qua (1993), essa metodologia baseia-se no pressuposto de que a distribuição interna da dinâmica populacional de determinada área entre dois pontos diferentes no tempo pode ser modelada por um processo estocástico com espaço paramétrico de tempo discreto e um espaço finito de estados. Formalmente,  $\mathbf{F}_t$  representa a distribuição transversal do tamanho populacional das AMCs no tempo  $t$  relativa à média brasileira. Define-se um conjunto de  $K$  classes de diferentes tamanhos, que propiciam uma aproximação discreta da distribuição populacional. Primeiro, assume-se que a frequência da distribuição segue um processo de Markov estacionário de primeira ordem.

A evolução da distribuição do tamanho das AMCs é representada por uma matriz de probabilidade de transição  $\mathbf{M}$ , em que cada elemento  $(i, j)$  indica a probabilidade de uma AMC se deslocar do estado  $i$  no momento  $t$  para o estado  $j$  no período seguinte  $(t + 1)$ . Formalmente, o vetor  $\mathbf{F}_t$  ( $K, 1$ ), que indica a frequência das AMCs em cada classe no tempo  $t$ , é descrita pela seguinte equação:

$$\mathbf{F}_{t+1} = \mathbf{M}\mathbf{F}_t, \quad (2.4)$$

em que  $\mathbf{M}$  é uma matriz de probabilidade de transição ( $K, K$ ) que representa a transição entre as duas distribuições como segue:

$$\mathbf{M} = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{pmatrix}, \quad (2.5)$$

em que cada elemento  $p_{ij} \geq 0$  e  $\sum p_{ij} = 1$ . A probabilidade de transição estacionária  $p_{ij}$  captura a probabilidade de uma AMC na classe  $i$  no período  $t-1$  passar para a classe  $j$  no período  $t$ .

Os elementos de  $\mathbf{M}$  podem ser estimados das frequências observadas nas mudanças de classes de um período para o outro. Portanto, com base em Lanaspa, Pueyo e Sanz (2003) e Le Gallo e Chasco (2008), o estimador de  $p_{ij}$  por máxima verossimilhança é

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i}, \quad (2.6)$$

em que  $n_{ij}$  é o total do número de AMCs que se deslocaram da classe  $i$  na década  $t-1$  para a classe  $j$  na década  $t$  imediatamente seguinte ao longo de todas as transições e  $n_i$  é a soma total de AMCs que estiveram em  $i$  ao menos uma vez ao longo de todas as transições.

Se as probabilidades de transição são estacionárias, isto é, se as probabilidades entre duas classes não variam no tempo, então

$$\mathbf{F}_{t+s} = \mathbf{M}^s \mathbf{F}_t \quad (2.7)$$

Nessa estrutura, é possível determinar a distribuição ergódica (também denominada de distribuição de longo prazo, equilíbrio, estado estacionário) de  $F_t$ , caracterizada quando  $s$  tende ao infinito na equação 2.7, o que significa que uma vez que as mudanças representadas pela matriz  $\mathbf{M}$  são repetidas um número arbitrário de vezes. Tal distribuição existe se a cadeia de Markov é regular, ou seja, se e somente se, para um dado  $m$ ,  $\mathbf{M}^m$  não possuir entradas nulas. Nesse caso, a matriz de transição converge para uma matriz limite  $\mathbf{M}^*$  de posto 1. A existência de uma distribuição ergódica  $F^*$ , e então caracterizada por:

$$\mathbf{F} * \mathbf{M} = \mathbf{F}^* \quad (2.8)$$

Esse vetor descreve a distribuição futura das AMCs se os movimentos observados na amostra são repetidos infinitamente. Cada linha de  $\mathbf{M}^t$  tende à distribuição limite quando  $t \rightarrow \infty$ . Essa distribuição limite é dada pelo autovetor associado com o autovalor unitário de  $\mathbf{M}$ .

A suposição de um processo de Markov estacionário de primeira-ordem requer que as probabilidades de transição  $p_{ij}$  sejam de ordem 1, ou seja, sejam independentes das classes iniciais nos períodos anteriores ( $t-2$ ,  $t-3$ , ...). Se a cadeia é de ordem superior, a informação necessária para descrever a evolução verdadeira da distribuição populacional. Além disso, a propriedade de Markov implicitamente assume que as probabilidades de transição,  $p_{ij}$  dependem de  $i$  (isto é, que o processo não é de ordem 0).

Para testar essa propriedade, Bickenbach e Bode (2003) enfatizam o papel dos teste de independência temporal. Na determinação da ordem da cadeia de Markov, Tan e Yilmaz (2002) sugerem testar, primeiramente, a ordem 0 contra a ordem 1; e em seguida, testar ordem

1 contra ordem 2; e assim por diante. Se o teste de ordem 0 contra ordem 1 é rejeitado, e o teste de ordem, 1 contra ordem 2 não é rejeitado, o processo pode ser determinado como de ordem 1.

Para testar para ordem 0, a hipótese nula  $H_0: \forall i: p_{ij} = p_j (i = 1, \dots, K)$  testado contra a seguinte alternativa  $H_a: \exists i \setminus p_{ij} \neq p_j$ . O teste estatístico de razão verossimilhança (LR) apropriado é

$$LR^{(0(0))} = 2 \sum_{i=1}^K \sum_{j \in A_i} n_{ij}(t) \ln \frac{\hat{p}_{ij}}{\hat{p}_i} \sim asy\chi^2 [(K-1)^2], \quad (2.9)$$

supondo que  $\hat{p} > 0, \forall j (= 1, \dots, K)$ .  $A_i = \{j: \hat{p}_{ij} > 0\}$  é o conjunto de probabilidades de transição sob  $H_a$ . Para testar ordem 1 contra 2, uma cadeia de Markov de segunda-ordem é definida ao considerar as classes de tamanho populacional  $k (k=1, \dots, K)$  nas quais estavam as AMCs no tempo  $t-2$  e supondo que os pares sucessivos de classes  $k$  e  $i$  formam uma classe composta. Então, a probabilidade de uma AMC que deslocar-se da classe  $j$  no tempo  $t$ , dado que estava em  $k$  em  $t-2$  e em  $i$  em  $t-1$ , é  $p_{ij}$ . O número absoluto de transições é  $n_{kij}(t)$ , com frequência marginal sendo  $n_{ki}(t-1) = \sum_j n_{kij}(t-1)$ .

Para  $H_0: \forall k: p_{kij} = p_{ij} (k = 1, \dots, K)$  testado contra a seguinte alternativa  $H_a: \exists k: p_{kij} \neq p_{ij}$ ,  $p_{kij}$  são estimados como  $\hat{p}_{kij} = n_{kij}/n_{ki}$ , em que  $n_{ki} = \sum_{t=2}^T n_{kij}(t)$  e  $n_{kij} = \sum_{t=2}^T n_{ki}(t-1)$ . O  $p_{ij}$  são estimados do conjunto de dados completo como  $\hat{p}_{ij} = n_{ij}/n_i$ . O teste LR apropriado segue:

$$LR^{(0(1))} = 2 \sum_{k=1}^K \sum_{i=1}^K \sum_{j \in A_i} n_{kij}(t) \ln \frac{\hat{p}_{kij}}{\hat{p}_{ki}} \sim asy\chi^2 [\sum_{i=1}^K (c_i - 1)(d_i - 1)]^2 \quad (2.10)$$

Similarmente à notação acima,  $C_i = \{j: \hat{p}_{ij} > 0\}$ ,  $c_i = \#C_i$ ,  $C_{ki} = \{j: \hat{p}_{kij} > 0\}$  e  $d_i = D_i = \#\{k: n_{ki} > 0\}$ .

Se ambas as ordens 0 e 1 são rejeitadas, os testes podem ser aplicados em ordens superiores ao introduzir dimensões adicionais para o tamanho da população no período  $t-3$ ,  $t-4$ , e assim por diante. Entretanto, como o número de parâmetros para serem estimados aumenta exponencialmente com o número de defasagens temporais, enquanto o número de observações diminui linearmente para um dado conjunto de dados, a confiabilidade das estimativas e o poder do teste diminuem rapidamente. Portanto, um limite máximo de ordem ao qual a cadeia de Markov pode ser testada deve ser estabelecido.

### 3 RESULTADOS ECONOMÉTRICOS

#### 3.1 DESCRIÇÃO DOS DADOS

Para alcançar o objetivo proposto por esse trabalho e, portanto, analisar a distribuição do tamanho dos municípios brasileiros ao longo do século XX, foram adotadas como unidade espacial básica, as Áreas Mínimas Comparáveis (AMC). Construídas pelo Instituto de Pesquisa Econômica Aplicada (IPEA), as AMCs são áreas geográficas geradas por meio da união ou junção das áreas dos municípios alterados. De acordo com Reis, Pimentel e Alvarenga (2008), tal construção teve o intuito de compatibilizar as divisões político-administrativas apresentadas nos vários censos e com isso, tornar possível a construção de painéis de dados econômicos estaduais e municipais no período que se estende desde 1872 a 2000.

A Tabela 1 apresenta o número de municípios brasileiros recenseados de 1872 a 2000 e também de AMCs para os períodos intercensitários que se estendem dos respectivos anos censitários até 2000. A comparação das duas cifras mostra que, em termos percentuais, há uma perda de aproximadamente 30 por cento no número de observações quando passamos de municípios para AMC no caso do Censo de 1872, sendo progressivamente menor para os demais anos censitários.

Tabela 1 - Número de municípios nos Censos e AMC nos períodos intercensitários: Brasil, 1920-2000

Anos censitários	Número de municípios	Período intercensitário	Número de AMC
1920	1.305	1920-2000	953
1940	1.575	1940-2000	1.275
1950	1.891	1950-2000	-
1960	2.768	1960-2000	2.407
1970	3.974	1970-2000	3.659
1980	3.991	1980-2000	3.692
1991	4.491	1991-2000	4.267
2000	5.507	-	-

Fonte: Extraído de Reis, Pimentel e Alvarenga (2008).

Segundo os dados da tabela 1, no período de interesse do presente estudo, o número de municípios recenseados no Brasil aumentou de 1.305, no Censo de 1920, para 5.507, no

Censo de 2000, revelando inúmeras alterações no número, área e fronteira dos municípios. Ao longo do século XX, essas alterações causaram mudanças significativas na delimitação geográfica das Unidades político-administrativas do território brasileiro<sup>5</sup>.

Para ilustrar os resultados da agregação de municípios realizada pelo IPEA, a figura 1 apresenta o mapa das AMCs no período 1920-2000, distinguindo-se por meio de cores as Unidades da Federação atuais.

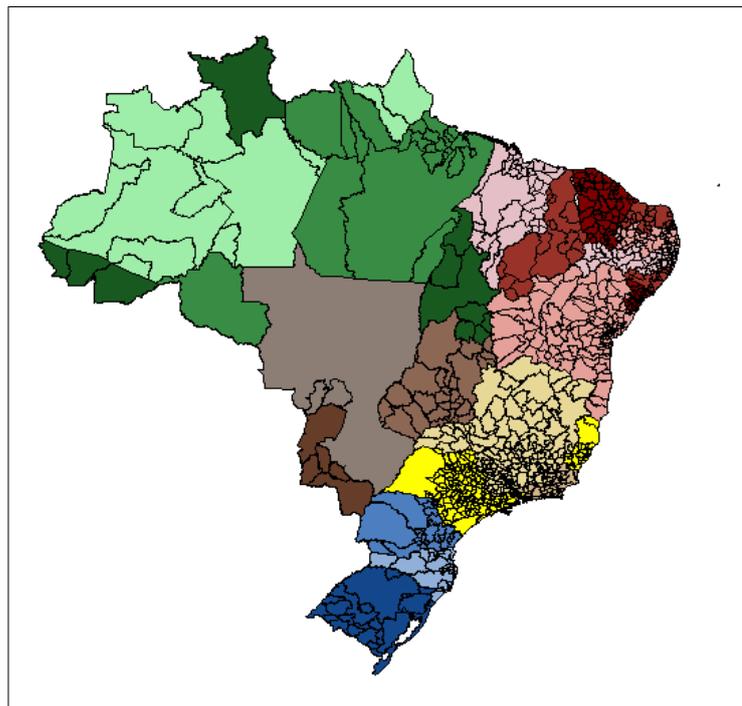


Figura 1 - Brasil: Áreas Mínimas Comparáveis, 1920-2000.

Fonte: Extraído de Reis, Pimentel e Alvarenga (2008).

O elevado número de variações na quantidade de unidades municipais brasileiras, torna inconsistente qualquer comparação intertemporal em nível geográfico estritamente municipal. Assim, a utilização de AMCs como unidade espacial básica para o estudo da distribuição do tamanho dos municípios brasileiros é justificável. Contudo, essa utilização

<sup>5</sup>De acordo com Reis, Pimentel e Alvarenga (2008) entre 1920 e 1940, foram criados os territórios de: Fernando Noronha, a partir de Pernambuco; Guaporé, posteriormente Rondônia e desmembrado de Mato Grosso e Amazonas; Amapá, a partir do Pará; Roraima do Amazonas; Ponta-Porã, do Mato Grosso; e Iguaçu, desmembrado de Paraná. Em 1946, Ponta-Porã e Iguaçu foram reincorporados aos seus respectivos estados de origem e Mato Grosso e Paraná. Em 1960, devido à transferência do Distrito Federal para Brasília, foi criado o Estado da Guanabara, extinto em 1975 e incorporado ao Estado do Rio de Janeiro. Em 1962, o Acre foi elevado à condição de estado da Federação. Em 1975, foi criado o estado do Mato Grosso do Sul, a partir do Mato Grosso. Em 1988, Fernando de Noronha foi reincorporado a Pernambuco; Rondônia e Amapá foram elevados à condição de estado da Federação; e, por fim, foi criado o estado de Tocantins, desmembrado de Goiás.

deve ser realizada com ressalvas uma vez que nos estados das regiões Centro-Oeste e Norte, onde ocorreu grande número de criações e desmembramentos de municípios no período de 1920 a 2000, a agregação para AMC implica perda de um número significativo de observações municipais. Especificamente no caso do estado de Rondônia, todos os municípios foram agregados em uma única AMC e o mesmo ocorreu com grande parte dos municípios que compunham o antigo estado do Mato Grosso (REIS, PIMENTEL E ALVARENGA, 2008).

A Figura 2 apresenta em linhas fortes o contorno das AMCs para o período 1920-2000, comparando-as com a malha municipal utilizada no Censo de 2000 - em linhas mais claras.



Figura 2 - Brasil: Áreas mínimas comparáveis, 1920-2000 e área dos municípios em 2000.

Fonte: Extraído de Reis, Pimentel e Alvarenga (2008).

Em suma, a base de dados utilizada para analisar a distribuição do tamanho dos municípios brasileiros ao longo do século XX foi composta por dados populacionais censitários para as AMCs brasileiras nos anos de 1920, 1940, 1960, 1970, 1980, 1991 e 2000.

Com base no método de análise utilizado por Arribas-Bell (2008), a identificação de padrões gerais da distribuição do tamanho das AMCs pode ser realizada por meio de

diferentes medidas: (i) tamanho absoluto dos municípios em termos populacionais ( $S_i$ ); (ii) tamanho dos municípios em relação à média populacional (População relativa:  $S_{rel} = \frac{S_i}{\bar{S}_i}$ , em que  $\bar{S}_i = \sum_{i=1}^n S_i$  e  $n$  o número total de municípios) e; (iii) tamanho dos municípios em relação ao total populacional (Participação relativa:  $P_{part} = \frac{P_i}{\sum_{i=1}^n P_i}$ );

Segundo o autor, intuitivamente, a medida (i) parece ser a mais adequada. Contudo, ela pode resultar em análises problemáticas uma vez que não associa o tamanho de determinado município em relação ao demais, desconsiderando as mudanças apresentadas pela distribuição. Por outro lado, as medidas relativas (ii) e (iii), além de permitirem análises comparativas, são normalizações necessárias para a estimação de distribuições de equilíbrio (*steady-state*).

A tabela 2 apresenta os valores médios das distribuições do tamanho das AMCs brasileiras referente aos dois períodos extremos da amostra, 1920 e 2000.

Tabela 2 - Valores médios da distribuição populacional: Brasil, AMC, 1920/2000

<b>Medidas</b>	<b>1920</b>	<b>2000</b>	<b>Crescimento</b>
População absoluta	32.116	174.080	4,13
População relativa (média)	1,00	1,00	-1,00
Participação relativa (total)	1,05E-03	1,05E-03	-0,05

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

De acordo com os dados da tabela 2, na média, as AMCs brasileiras aumentaram seu tamanho, em termos populacionais, de 32 mil em 1920, para 174 mil habitantes em 2000. Essa variação significa um crescimento médio de 4,13% para cada AMC. Mas quando analisadas as medidas relativas, média ou total, verifica-se que as taxas médias de crescimento são negativas, -1,0% e -0,05% respectivamente. Isso implica que as AMCs brasileiras diminuíram em termos populacionais relativos. Em outras palavras, existem muitas AMCs com crescimento populacional menos acelerado do que o crescimento médio (4,13%).

De maneira análoga, a tabela 3 apresenta os intervalos interquartis (IQ) da distribuição. Por definição, o maior IQ refere-se à população absoluta, seguida pela população relativa e participação relativa.

Tabela 3 - Intervalos interquartis (IQ\*) da distribuição populacional: Brasil, AMC, 1920/2000

Medidas	Q1	Q3	IQ=Q3-Q1	ΔIQ
População absoluta - 1920	11.346	33.635	22.288	298%
População absoluta - 2000	28.876	117.518	88.641	
População relativa (média) - 1920	0,35	1,05	0,69	-27%
População relativa (média) - 2000	0,17	0,68	0,51	
Participação relativa (total) - 1920	0,24	0,81	0,57	-100%
Participação relativa (total) - 2000	3,00 E-04	9,47E-04	6,47E-04	

Nota(\*):Diferença entre o terceiro quartil e o primeiro quartil; informa sobre a amplitude do intervalo que inclui 50% das observações centrais. Ver também Arribas-Bel (2008, p.4).

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

A partir dos dados acima, observa-se que o IQ aumentou para a medida absoluta e diminuiu para as relativas. Isso indica que as AMCs que se encontravam nos extremos da distribuição de seu tamanho populacional tornaram-se, entre 1920 e 2000, ainda maiores ou ainda menores.

### 3.2 EVOLUÇÃO DA DISTRIBUIÇÃO POPULACIONAL BRASILEIRA

Uma vez identificada a existência de padrões gerais na distribuição do tamanho das AMCs brasileiras, métodos estatísticos mais apurados foram aplicados para descrever seu comportamento ao longo do período de análise.

Em termos paramétricos, a relação entre a distribuição da população e a dimensão das 953 AMCs (i) em cada década (t) é dada pela equação (1) linearizada

$$\ln R_{it} = b_t \ln S_{it} + \varepsilon_{it}, \quad (5.1)$$

estimada para os dados de cada uma das 8 décadas consideradas pelo estudo; 1920, 1940, 1950, 1960, 1970, 1980, 1991 e 2000. Os parâmetros de cada modelo são estimados pelo método dos Mínimos Quadrados Ordinários (MQO).

De acordo com a tabela 4, o expoente de Pareto  $b$ , estimado por MQO, assume valores decrescentes durante o período 1920-2000, sugerindo que em 1920 a população crescia de maneira mais uniforme ao longo das AMCs brasileiras do que em 2000. Até 1940, os valores estimados para  $b$  são superiores a 1. Isso indica que entre 1920 e 1940, a estrutura municipal brasileira, descrita por meio de AMCs, apresenta número considerável de municípios relativamente simétricos. A partir de 1950, seguindo tendência decrescente,  $b$  assume valores inferiores a 1. Por sua vez, isso revela um aumento do número de municípios de tamanhos assimétricos.

Tabela 4 - Coeficiente de Pareto por MQO

Variável dependente: ln R				
	Variável	Coefficiente	Desvio-padrão	P-valor
<b>1920</b>	Constante	17,099	0,108	0,000
	ln S	-1,131	0,011	0,000
			R <sup>2</sup>	0,920
<b>1940</b>	Constante	16,648	0,099	0,000
	ln S	-1,066	0,010	0,000
			R <sup>2</sup>	0,926
<b>1950</b>	Constante	15,874	0,104	0,000
	ln S	-0,976	0,010	0,000
			R <sup>2</sup>	0,907
<b>1960</b>	Constante	15,594	0,090	0,000
	ln S	-0,931	0,009	0,000
			R <sup>2</sup>	0,926
<b>1970</b>	Constante	15,021	0,088	0,000
	ln S	-0,863	0,008	0,000
			R <sup>2</sup>	0,920
<b>1980</b>	Constante	14,639	0,088	0,000
	ln S	-0,815	0,008	0,000
			R <sup>2</sup>	0,914
<b>1991</b>	Constante	14,508	0,087	0,000
	ln S	-0,790	0,008	0,000
			R <sup>2</sup>	0,914
<b>2000</b>	Constante	14,476	0,085	0,000
	ln S	-0,779	0,008	0,000
			R <sup>2</sup>	0,917

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

Ao analisar quantitativamente a queda nos valores de  $b$ , percebe-se uma diminuição nas variações apresentadas nas duas últimas décadas, 1980-1991 e 1991-2000, o que sugere um arrefecimento da tendência predominante ao longo de todo o século.

Diversos trabalhos da literatura, como Rosen e Resnick (1980), Guérin-Pace (1995), Soo (2002), ressaltam a sensibilidade das estimações de  $b$  em relação ao tamanho da amostra utilizada. Como o presente trabalho utiliza dados referentes a todos os municípios brasileiros, é interessante verificar o comportamento da distribuição do tamanho dos municípios diante diferentes tamanhos de amostra para assim, verificar a sensibilidade do expoente  $b$  e permitir

conclusões mais robustas. A tabela 5 apresenta as novas estimações para  $b$ , considerando dois tamanhos adicionais de amostras (N); as 100 e as 50 maiores AMCs brasileiras.

Tabela 5 - Coeficiente de Pareto estimado para diferentes tamanhos de AMC

Variável dependente: ln R							
	100 maiores AMCs			50 maiores AMCs			
	Coefficiente	Desvio-padrão	P-valor	Coefficiente	Desvio-padrão	P-valor	
<b>1920</b>	Constante	21,2110	0,143	0,000	20,358	0,230	0,000
	ln S	-1,525	0,012	0,000	-1,456	0,019	0,000
			R <sup>2</sup>	0,993		R <sup>2</sup>	0,992
<b>1940</b>	Constante	20,057	0,186	0,000	19,730	0,383	0,000
	ln S	-1,384	0,015	0,000	-1,359	0,031	0,000
			R <sup>2</sup>	0,987		R <sup>2</sup>	0,976
<b>1950</b>	Constante	18,917	0,144	0,000	19,196	0,312	0,000
	ln S	-1,262	0,011	0,000	-1,283	0,025	0,000
			R <sup>2</sup>	0,991		R <sup>2</sup>	0,983
<b>1960</b>	Constante	18,356	0,121	0,000	18,697	0,252	0,000
	ln S	-1,180	0,009	0,000	-1,206	0,019	0,000
			0,010	0,993		R <sup>2</sup>	0,988
<b>1970</b>	Constante	18,076	0,137	0,000	19,107	0,247	0,000
	ln S	-1,129	0,010	0,000	-1,204	0,018	0,000
			R <sup>2</sup>	0,991		R <sup>2</sup>	0,989
<b>1980</b>	Constante	18,475	0,208	0,000	19,694	0,446	0,000
	ln S	-1,135	0,015	0,000	-1,222	0,033	0,000
			R <sup>2</sup>	0,981		R <sup>2</sup>	0,967
<b>1991</b>	Constante	18,072	0,298	0,000	20,134	0,680	0,000
	ln S	-1,131	0,022	0,000	-1,231	0,049	0,000
			R <sup>2</sup>	0,963		R <sup>2</sup>	0,930
<b>2000</b>	Constante	18,772	0,337	0,000	20,607	0,786	0,000
	ln S	-1,122	0,024	0,000	-1,250	0,056	0,000
			R <sup>2</sup>	0,953		R <sup>2</sup>	0,913

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

De acordo com os dados da tabela 5, há mudanças significativas nos valores estimados para  $b$  quando utilizadas amostras de diferentes tamanhos. Essas mudanças podem ser ilustradas no gráfico 1.

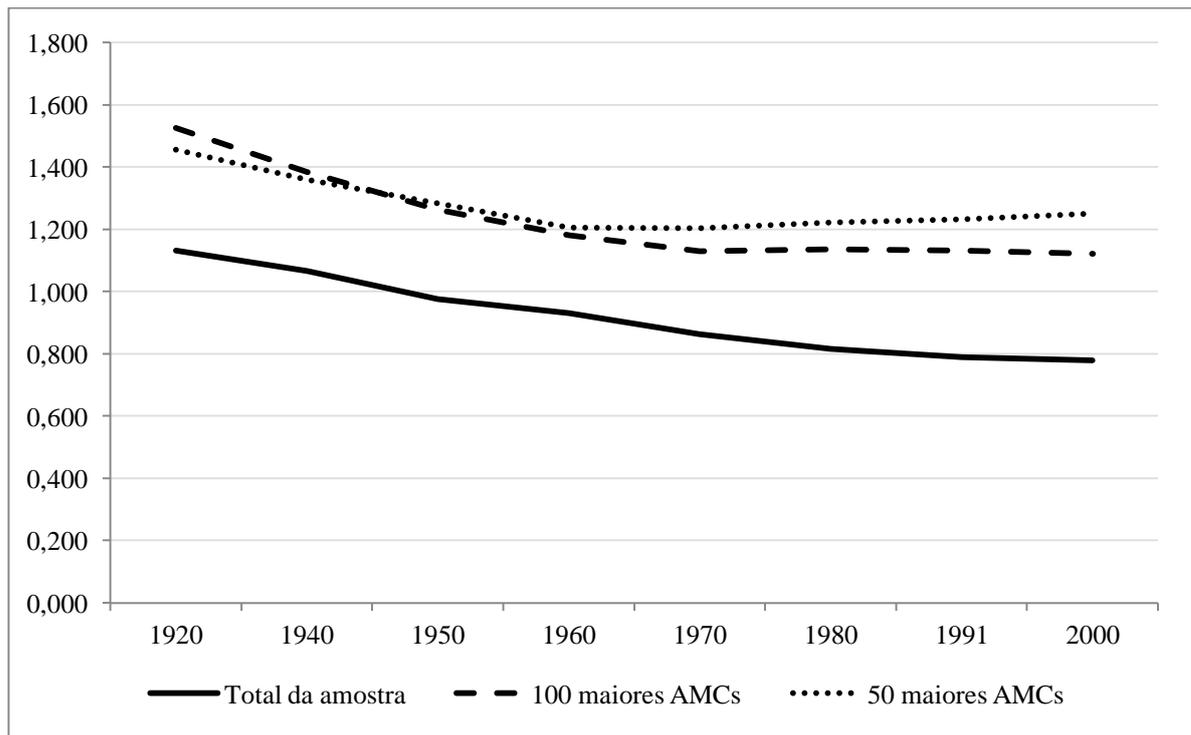


Gráfico 1 - Evolução do coeficiente de Pareto estimado para diferentes tamanhos de AMC

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

De maneira geral, verifica-se que os três tamanhos de amostra apresentam tendências diferentes. No que se refere à amostra das 100 maiores AMCs, o valor estimado para  $b$  decresce até 1970, quando parece estabilizar em torno de 1,13, valor próximo ao estabelecido pela lei de Zipf.

Assim como a amostra anterior, os 50 maiores AMCs, também apresentam mudança na declividade da estimativa de  $b$ . Contudo, os dados da tabela 5 e o gráfico 1 indicam que nessa amostra, essa mudança ocorre em 1960, uma década antes da amostra para as 100 maiores AMCs. Durante a década de 1960 e 1970,  $b$  apresenta estabilidade em torno de 1,20, mas a partir de 1980 assume tendência crescente, o que sugere que, considerando os limites das 50 maiores AMCs, houve um aumento da importância das AMCs de tamanho inferior na composição da distribuição. Tal fato pode ser explicado pelo notável processo de reversão de polarização pelo qual a população urbana brasileira passou entre 1970 e 1990, no qual as cidades médias assumiram papel decisivo. Conforme Andrade e Serra (1998), a distribuição concentrada da população urbana brasileira é caracterizada por uma tendência bem marcada do papel dos centros intermediários no crescimento demográfico do país.

### 3.2.2 Análise não-paramétrica do expoente de Pareto

A análise não paramétrica da distribuição do expoente de Pareto é uma maneira adicional de descrever a distribuição do tamanho das AMCs brasileiras por meio da construção de uma função de densidade de kernel. Uma das vantagens dessa construção é a obtenção de uma descrição mais completa do formato da função e da evolução do expoente de Pareto ao longo do tempo.

O gráfico 2 ilustra o logaritmo da distribuição relativa da população dos municípios ao longo 1920, 1960 e 2000. Tem-se por definição que o 0 (zero) do eixo horizontal indica o tamanho médio das AMCs brasileiras.

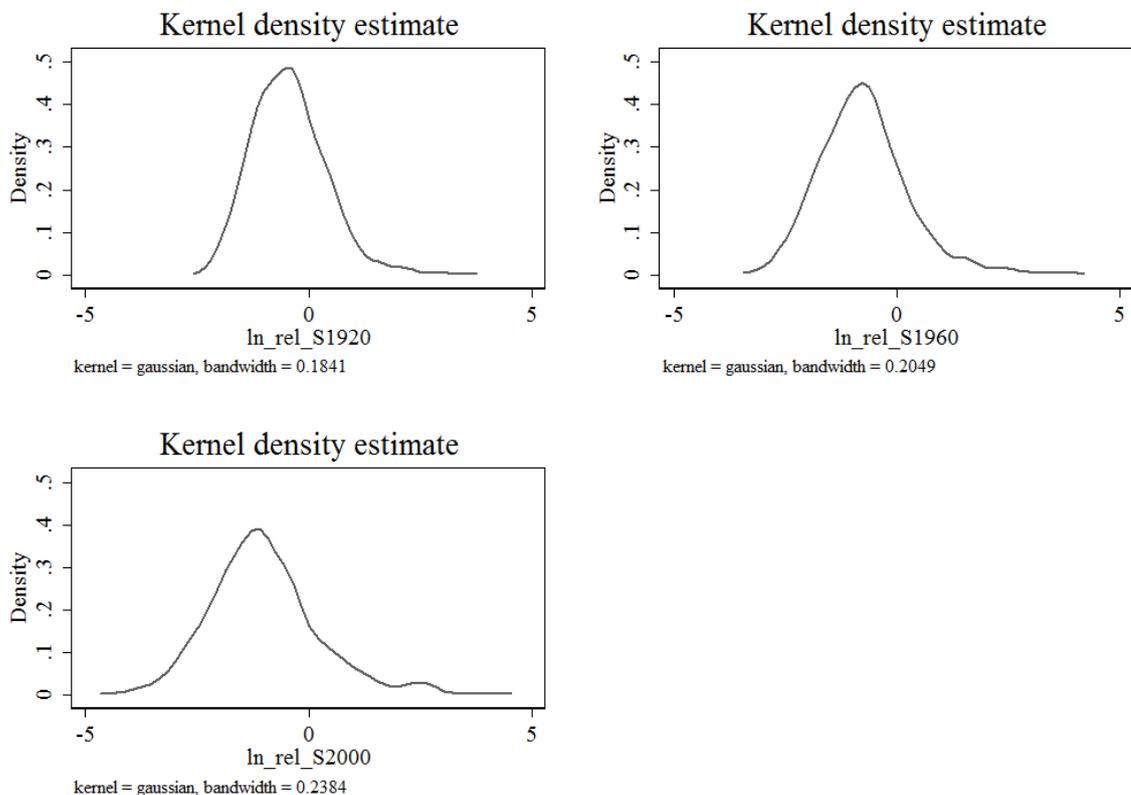


Gráfico 2 - Densidade log-relativa do tamanho das AMCs: Toda amostra, Brasil, 1920/1960/2000

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

Tradicionalmente, na análise das funções de densidade de kernel, dois aspectos devem ser tratados; seu formato e sua evolução. De maneira geral, as três distribuições apresentadas se assemelham, pois possuem formato unimodal concentrado em valor inferior à média. Contudo apresentam diferenças no que se refere à localização da massa de densidade

no eixo horizontal. Partindo de 1920, a massa de densidade desloca-se gradativamente para a esquerda. Este movimento pode ter sido causado por um grupo considerável de AMCs que inicialmente apresentavam maior importância relativa, mas ao longo das décadas perderam grande número de residentes.

Portanto, a distribuição da população brasileira parece assumir uma tendência crescente da quantidade de municípios que apresentam uma quantidade de população abaixo da média e que convergem para um nível populacional mais baixo do que o resto das AMCs. Portanto, estes resultados reafirmam o cenário descrito anteriormente, onde a população brasileira tende a se concentrar em um pequeno número de grandes AMCs e com crescimento superior àquelas de menor tamanho populacional.

### 3.2.1 O efeito espacial na estimativa do expoente de Pareto

Uma vez que as estimativas por meio de MQO podem ser afetadas pela omissão de autocorrelações espaciais, o modelo estimado anteriormente foi testado para sua presença.

A Tabela 6 mostra as estatísticas dos testes Jarque-Bera, Koenker-Basset e Kelejian-Robinson para cada modelo. Os resultados verificados sugerem, respectivamente, a existência de não-normalidade dos erros; heteroscedasticidade - mas apenas para os modelos de 1920 e de 1960 - e autocorrelação espacial - exceto para a distribuição de 1940. Conseqüentemente, os parâmetros obtidos por MQO podem ser considerados tendenciosos.

Tabela 6 - Coeficiente de Pareto por MQO e teste de especificação

Ano	$\alpha^*$	$\beta^*$	JB	KB	KR
1920	17,101	-1,132	620,573	13,456	9,535
1940	16,635	-1,064	658,713	0,347	2,864
1950	16,005	-0,988	2021,666	3,040	6,729
1960	15,579	-0,930	983,936	4,579	5,105
1970	15,061	-0,867	1285,525	3,672	8,283
1980	14,670	-0,817	2075,140	1,931	13,982
1991	14,535	-0,792	2836,998	0,257	12,908
2000	14,502	-0,781	3330,278	0,045	16,080

Notas:  $\alpha = \ln a$ . \* Hipótese nula rejeitada a 1%. A coluna JB refere-se aos resultados da aplicação do teste Jarque-Bera de não normalidade dos resíduos. A KB indica os resultados do teste Koenker-Basset para heteroscedasticidade. A KR apresenta os resultados do teste Kelejian-Robinson para autocorrelação espacial dos erros. Para KR e KB, o valor crítico de significância a 5% é de 3,8415.

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

A exemplo de Le Gallo e Chasco (2008), uma especificação econométrica mais adequada foi utilizada com o intuito de obter estimativas mais robustas. Assim, uma vez que os modelos são compostos por dados de cortes transversais de diferentes pontos no tempo, o SUR – erro espacial (SSUERR) foi utilizado.

Os resultados estatisticamente significantes das estimativas por SUR – erro espacial, apresentados na tabela 7, confirmam a influência do espaço na distribuição populacional das AMCs brasileiras.

Tabela 7 - Coeficiente de Pareto por SUR erro espacial

<b>Ano</b>	<b><math>\alpha^*</math></b>	<b><math>\beta^*</math></b>	<b><math>\lambda^*</math></b>
1920	16,230	-1,045	-0,014
1940	15,639	-0,966	0,001
1950	15,064	-0,896	0,034
1960	14,722	-0,847	0,126
1970	14,291	-0,794	0,153
1980	13,980	-0,753	0,182
1991	13,890	-0,733	0,179
2000	13,883	-0,725	0,178

Nota: \* Hipótese nula rejeitada a 5%.

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

Para comparar a magnitude da diferença entre os expoentes de Pareto estimados por MQO e pelo SUR erro espacial, a evolução de seus valores ao longo do tempo é apresentada pelo gráfico 4.

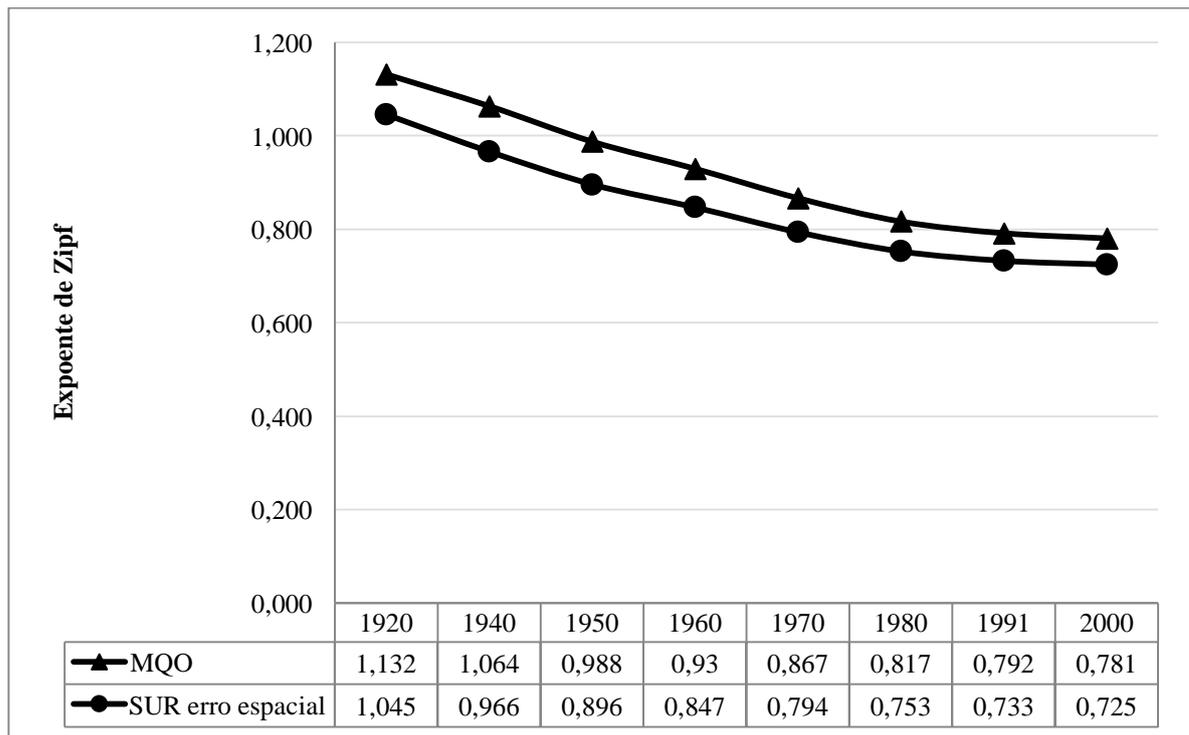


Gráfico 3 - Evolução do coeficiente de Pareto estimado

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

Os estimadores encontrados por MQO são sempre superiores aos do SSURERR, mas ambos seguem mesma evolução decrescente, permitindo mais uma vez verificar que distribuição da população brasileira em suas 953 AMCs tornou-se cada vez mais divergentes, nos últimos 80 anos, indica uma contínua tendência à concentração populacional no país. Contudo, nos resultados obtidos em ambos os modelos, verifica-se que a variação do coeficiente de Pareto tende a diminuir a partir da década de 1980, o que corrobora a hipótese levantada por Brito (2000) de que o processo de urbanização da sociedade brasileira continuou ao final do século XX, porém, menos acelerado do que nas décadas anteriores.

### 3.3 ANÁLISE DINÂMICA ESPACIAL DA DISTRIBUIÇÃO DO TAMANHO DOS MUNICÍPIOS BRASILEIROS

As estruturas locais de divisão territorial dos países são frequentemente expostas a mudanças ao longo de um século. No Brasil, em um período de 80 anos, essas mudanças resultaram em constantes redistribuições da população entre as AMCs. Os métodos utilizados para examinar a relação entre distribuição populacional e o tamanho dos municípios

brasileiros revelaram importantes informações para um melhor entendimento da dinâmica populacional ao longo do tempo e do espaço. Mas, como discutido anteriormente, uma vez que a verificação da lei de Zipf se trata de uma análise estática, ela não permite identificar como são as mudanças nas posições relativas das AMCs na distribuição.

Com base nos trabalhos de Eaton e Eckstein (1997), Black e Henderson (2003), Lanaspá et al. (2003), Le Gallo e Chasco (2008) aplicaram, - uma metodologia que permite a descrição do comportamento das cidades na distribuição populacional ao longo do tempo de maneira dinâmica ao descrever as leis de movimento da distribuição como um processo não-paramétrico de cadeia de Markov.

Para isso, a distribuição do tamanho das cidades foi dividida em seis classes de tamanho distintas<sup>6</sup>; (1) AMCs cuja população está abaixo dos 30 por cento da média; (2) entre 30 e 45 por cento da média; (3) entre 45 e 60 por cento da média; (4) entre 60 e 85 por cento da média; (5) entre 85 e 135 por cento da média e; (6) mais de 1,35 vezes a média. A tabela 8 apresenta o número absoluto de município presentes em cada classe ao longo do período estudado.

Conforme esse dados, as classes 2, 3 e 6 não apresentaram variações elevadas no número de AMCs que as compunha. De maneira contrária, as classes 1, 4 e 5, apresentaram variações significantes; as classes 4 e 5 tiveram redução do número de AMCs, enquanto a classe 1, que representa as AMCs de tamanho populacional de 30% média brasileira, tiveram aumento superior a 2,5 vezes no número de AMCs registradas no período inicial.

Tabela 8 - Distribuição do tamanho das AMCs brasileiras por classe de tamanho relativo

Ano	Total	1	2	3	4	5	6
		≤ 30%	≤ 45%	≤ 60%	≤ 85 %	≤ 135%	> 135%
<b>1920</b>	953	168	175	128	169	149	164
<b>1940</b>	953	214	175	143	145	133	143
<b>1950</b>	953	257	176	125	130	137	128
<b>1960</b>	953	313	169	134	114	101	122
<b>1970</b>	953	359	178	112	97	90	117
<b>1980</b>	953	397	164	100	108	67	117
<b>1991</b>	953	422	157	90	98	65	121
<b>2000</b>	953	437	157	82	98	57	122

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

<sup>6</sup> Inicialmente, três diferente discretizações (6, 7 e 8 classes) foram realizadas. Mas diante do melhor desempenho da discretização de 6 classes no teste de primeira ordem da cadeia de Markov, apenas os resultados dessa estimativa são apresentados.

A matriz de transição apresentada pela tabela 9 foi testada para a presença de um processo de Markov de ordem 0 ao comparar cada coluna da matriz de transição à distribuição populacional no tempo  $t$ , utilizando a estatística de teste descrita no capítulo 2 pela equação 2.9. Os resultados (LR = 13825.53; prob = 0; df = 25) atestam que o processo apresentado é dependente de sua condição inicial em  $t-1$  e pode portanto, ser classificado como um processo de Markov de ordem 1.

Tabela 9 - Matriz de probabilidade de transição 1920-2000: Brasil, população relativa

Classes (k)	1	2	3	4	5	6	Número de observações
	≤ 30%	≤ 45%	≤ 60%	≤ 85 %	≤ 135%	> 135%	
1	0,954	0,041	0,003	0,002	0,000	0,000	2130
2	0,292	0,616	0,070	0,016	0,003	0,002	1194
3	0,020	0,379	0,502	0,085	0,013	0,000	832
4	0,001	0,043	0,303	0,552	0,093	0,008	861
5	0,000	0,001	0,020	0,290	0,592	0,097	742
6	0,000	0,000	0,002	0,007	0,126	0,865	912

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

Portanto, uma vez que a tabela 9 contém a matriz de probabilidade de transição de primeira ordem para os anos 1920, 1940, 1950, 1960, 1970, 1980, 1991 e 2000 em que cada elemento  $(i, j)$  indica a probabilidade de uma AMC se deslocar do estado  $i$  no momento  $t$  para o estado  $j$  no período seguinte  $(t + I)$ , cabe destacar as seguintes conclusões:

- i. a probabilidade de ficar no estado inicial, dada pelos elementos da coluna diagonal central, são muito mais elevados nos dois estados extremos. A probabilidade de uma AMC presente na classe 1 em  $t$ , permanecer na classe 1 em  $t+1$  é de 95,4%, enquanto a probabilidade de uma AMC pertencente à classe 6 em  $t$  continuar nessa mesma classe em  $t+1$  é de 86,5%.
- ii. entre os períodos  $t$  e  $t+1$ , a probabilidade de uma AMC perder população é superior é superior à probabilidade de ganho populacional;
- iii. nos intervalos 2 e 3, as chances de passar para o estado 1 são mais elevadas (29,2% e 2,0% respectivamente) do que chances de mover-se ao estado 6 (respectivamente 0,2% e 0,0%), enquanto nos intervalos 4 e 5 as chances de passar para o estado 6 são mais elevadas (0,8% e 9,7% respectivamente) do que chances de mover-se ao estado 6 (0,1% e 0,0% respectivamente);

- iv. a probabilidade de subir um intervalo tendo com estado inicial 1 (4,1%) é inferior aos demais estados 2 (7%), 3 (8,5%), 4 (9,3%) e 5 (9,7%);
- v. a possibilidade de crescimento paralelo ou uniforme entre as AMCs não pode ser aceita pois vários elementos da coluna diagonal central da matriz não assumem valor 1, o que indica que a distribuição do tamanho da população brasileira sofreu mudanças estruturais durante ao longo de todo o período.

Além da matriz de transição, também foi considerada a distribuição ergódica do tamanho populacional dos municípios brasileiros. Essa distribuição é calculada a partir de uma distribuição inicial, que após sofrer infinitas transições atingirá um estado em que não mais sofrerá alterações. Assim, essa distribuição interpretada como uma distribuição de equilíbrio de longo-prazo.

Tabela 1 - Distribuição versus ergódica 1920-2000: Brasil, população relativa

	<b>1</b> ≤ 30%	<b>2</b> ≤ 45%	<b>3</b> ≤ 60%	<b>4</b> ≤ 85 %	<b>5</b> ≤ 135%	<b>6</b> > 135%
<b>Distribuição inicial</b>	0,319	0,179	0,125	0,129	0,111	0,137
<b>Distribuição ergódica</b>	0,799	0,123	0,036	0,022	0,010	0,010

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

A partir da tabela 10, verifica-se uma concentração de frequência na classe para AMCs de tamanho inferior a 30 % da média nacional. Essa concentração revela que, no equilíbrio, a distribuição do tamanho dos municípios brasileiros apresenta tendência de convergência para o tamanho representado pela classe 1. Portanto, mais uma vez conclui-se que a população brasileira tende a se concentrar nos municípios se tamanho populacional mais elevado.

Por último, a fim de captar a influência do espaço nos resultados da cadeia de Markov, a probabilidade de movimentos na distribuição populacional das AMCs brasileiras foi estimada considerando a distribuição média da população de AMCs vizinhas, como em Le Gallo e Chasco (2008). A correlação envolvendo a direção do movimento das AMCs na distribuição populacional e a média populacional de seus vizinhos é medida na tabela 11.

Tabela 2 - Probabilidade de transição com efeitos espaciais

<b>Defasagem espacial</b>	<b>Movimento</b>		
	<i>Desce</i>	<i>Permanece</i>	<i>Sobe</i>
<b>População Menor</b>	0,326	0,597	0,078
<b>Igual</b>	0,429	0,482	0,089
<b>População Maior</b>	0,286	0,653	0,061

Fonte: Estimativas elaboradas pela autora, produzidas com base nos Censos Demográficos, 1920 a 2000.

Pelos elementos de probabilidade de movimento relatados na tabela 11 (sobe, permanece e desce), é possível notar que a probabilidade de subir, permanecer ou descer difere de acordo com o contexto local da área. Por exemplo, considerando a população média das AMCs contíguas (vizinhas), a probabilidade de uma AMC subir na hierarquia de tamanho, é de 7,8% quando suas AMCs contíguas contêm, em média, população menor; 8,9% quando apresentam população igual e; 6,1% quando possuem população maior.

No caminho oposto, a probabilidade de uma AMC se descolar para uma posição inferior na hierarquia, considerando a população média das AMCs contíguas, é de 32,6% quando seus vizinhos apresentam, em média, população menor; 42,9% quando apresentam população igual e; 28,6% quando apresentam população maior. Portanto, a influência do espaço na probabilidade de transição das AMCs entre os intervalos da distribuição populacional parece mais relevante para movimentos descendentes.

## CONSIDERAÇÕES FINAIS

O principal objetivo deste estudo foi realizar uma análise empírica sobre a evolução das relações entre a distribuição populacional e o tamanho das cidades brasileiras no período de 1920 a 2000. Métodos paramétricos e não-paramétricos foram utilizados para descrever a evolução da distribuição populacional brasileira. Uma vez que a realização de uma análise completa sobre o comportamento da distribuição do tamanho das cidades exige a observação de aspectos estáticos e dinâmicos devem ser considerados sob a ótica do tempo e do espaço, diferentes técnicas avançadas de estatística, de econometria espacial e de análise dinâmica de previsão da futura distribuição do tamanho das cidades foram utilizadas na obtenção dos resultados.

O elevado número de variações na quantidade de unidades municipais brasileiras, tornaria inconsistente qualquer comparação intertemporal em nível geográfico estritamente municipal. Portanto, a utilização de AMCs como unidade espacial básica para o estudo da distribuição do tamanho dos municípios brasileiros se fez necessária.

Com base em análises descritivas, foi possível verificar que existem muitas AMCs com crescimento populacional menos acelerado do que o crescimento médio. Além disso, as AMCs que se encontravam nos extremos da distribuição de seu tamanho populacional tornaram-se, entre 1920 e 2000, ainda maiores ou ainda menores.

Supondo que a distribuição populacional brasileira segue uma distribuição tipo Pareto, estimou-se o expoente de Pareto para servir como indicador de convergência da distribuição do tamanho das cidades. Foram introduzidos à análise, testes e especificações de dependência espacial para captar a influência do espaço na convergência. Na análise da convergência, os resultados encontrados pelo método dos Mínimos Quadrados Ordinários (MQO) sugerem, respectivamente, a existência de não-normalidade dos erros; heteroscedasticidade - mas apenas para os modelos de 1920 e de 1960 - e autocorrelação espacial - exceto para a distribuição de 1940.

Ao analisar a evolução dos expoentes de Pareto no período de 1920 a 2000 verificou-se que, embora haja uma ligeira mudança de declividade nas últimas duas décadas, a tendência geral, durante todo o período é de divergência, ou seja, há uma contínua tendência à concentração populacional. Este resultado é enfatizado quando o comportamento das cidades na distribuição populacional ao longo do tempo de maneira dinâmica é descrito como um processo não-paramétrico de cadeia de Markov: AMCs menores tem maiores probabilidades de

perda de população relativa, enquanto AMCs com o dobro ou mais da média populacional nacional tendem a atrair relativamente mais população. Além disso, considerando efeitos espaciais, se uma AMC está cercada por vizinhos com número populacional superior, maior é a probabilidade de perder população.

Contudo, deve-se ressaltar que a mudança de declividade verificada da análise de convergência do expoente de Pareto corrobora a hipótese levantada por Brito (2000) de que o processo de urbanização da sociedade brasileira continuou ao final do século XX, porém, menos acelerado do que nas décadas anteriores. Além disso, ela pode sugerir que o século XXI, será caracterizado por uma nova fase de convergência, ou seja, dispersão, com pessoas que se deslocam de AMCs maiores para AMCs menores, em um processo intensificado de reversão de polarização, já vivenciado pela população urbana brasileira entre 1970 e 1990. Porém, visto que as dimensões continentais fazem do Brasil um país com diferenças significativas em termos de concentração populacional em seu território, sugestões de tendências gerais nacionais serão sempre contraditórias. Por isso, análises regionais específicas sobre o comportamento da distribuição do tamanho das cidades teriam o potencial de enriquecer a discussão do presente estudo.

## REFERÊNCIAS

- ABREU, M.; GROOT, H. L. F. & FLORAX, R. J. G. M. “Space and growth: a survey of empirical evidence and methods”. **Tinbergen Institute Discussion Paper**, 129/3, 2004.
- ANDERSON, G.; GE Y. “The size distribution of Chinese cities”. **Regional Science and Urban Economics**, v. 35, n. 6, p. 756-776, 2005.
- ANDRADE, T. A.; SERRA, R. V. “O recente desempenho das cidades médias no crescimento populacional urbano brasileiro”. **Texto para Discussão**, 554. Rio de Janeiro: IPEA.1998.
- ANSELIN, L. **Spatial econometrics: methods and models**. Kluwer Academic Publishers: Dordrecht, 1988.
- ANSELIN, L. “Local Indicators of Spatial Association – LISA”. **Geographical Analysis** 27: 93-115, 1995.
- ANSELIN, L. **Spatial econometrics**. Bruton Center School of Social Sciences University of Texas at Dallas. 1999.
- ANSELIN, L.; FLORAX, J. G. M. “Directions in spatial econometrics: introduction”. In: ANSELIN, L.; FLORAX, R. J. G. M. (eds) **New Direction in Spatial Econometrics**, Springer, New York.1995.
- ANSELIN, L.; FLORAX, J. G. M. “Small sample of tests for spatial dependence in regression models: some further results”. In: Anselin, L. and Florax, R. J. G. M. (eds) **New Direction in Spatial Econometrics**, Springer, New York. 1995.
- ANSELIN, L.; MORENO, R. “Properties of tests for spatial error components”. **Regional Science and Urban Economics**, 33(5), 595-618. 2003.
- ARRIBAS-BEL, D. KANGAROOS (2008), “Cities and space: a first approach to the Australian Urban System”. **GeoDa Center Working Papers n. 2008-15**. Disponível em [http://geodacenter.asu.edu/drupal\\_files/200815\\_ArribasBel.pdf](http://geodacenter.asu.edu/drupal_files/200815_ArribasBel.pdf)
- BECKER, R.; HENDERSON, V. “Political economy of city sizes and formation”. **Journal of Urban Economics**, v. 48, p. 453-484. 2000.
- BECKMANN, M. J. “City hierarchies and the distribution of city size”. **Economic Development and Cultural Change**, v. 6, n. 3, Apr., p. 243-248. 1958.
- BERRY, B. “Urbanization and counter urbanization in the United States”. **Annals of the American Academy of Political and Social Science**, v. 451, Changing Cities: A Challenge to Planning, September, p. 13-20. 1980.
- BERRY, B. “Transnational urban migration, 1830-1980”. **Annals of the Association of American Geographers**, v. 83, n. 3, set., p. 389-405. 1993.

BICKENBACH, F.; BODE, E. “Evaluating the Markov property in studies of economic convergence”. **International Regional Science Review** 2003; v. 26, n. 363. 2003.

BLACK, D.; HENDERSON, V. “A theory of urban growth” **Journal of Political Economy**, v. 107 n. 2, p. 252-284. 1999a.

BLACK, D.; HENDERSON, V. “Spatial evolution of population and industry in the United States”. **The American Economic Review**, v. 89 n. 2, p. 321-327. 1999b.

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

BRAGA, F. G. (2006). **Migração interna e urbanização no Brasil contemporâneo: um estudo da rede de localidades centrais do Brasil (1980/2000)**. [artigo científico].

Disponível em:

<[http://www.abep.nepo.unicamp.br/encontro2006/docspdf/ABEP2006\\_573.pdf](http://www.abep.nepo.unicamp.br/encontro2006/docspdf/ABEP2006_573.pdf)> Acesso em: 7 maio 2008.

BRAÑAS, P.; ALCALÁ, F. “Aglomeración urbana española (1900-2000): estimaciones “rank-size” vs. tests LM”. **Working Paper 0101/Nº 19**, Departamento de Economía Aplicada, Universidad de Jaén. (s. d.).

BRAÑAS, P.; ALCALÁ, F. “Entropía, aglomeración urbana y la ley del ‘1’: evidencia para las regiones españolas”. **III Encuentro de Economía Aplicada**, Valença. 2000.

BRITO, F. **Brasil, final de século: a transição para um novo padrão migratório?** Anais da ABEP, Caxambu, Minas Gerais. 2000.

BRITO, F.; GARCIA, R.; SOUZA, R. **As tendências recentes das migrações Interestaduais e o padrão migratório**. Trabalho apresentado no XIV Encontro Nacional de Estudos Populacionais, ABEP, Caxambu, Minas Gerais, setembro. 2004.

CAMARANO, A. A.; BELTRÃO, K. I. “Distribuição espacial da população brasileira: mudanças na segunda metade deste século”. **Texto para discussão nº 766**. Rio de Janeiro: IPEA. 2000.

CAMPOLINA, C. C. “Repensando a questão regional brasileira: tendências desafios e caminhos”. In: CASTRO, A. C.(org.) **Desenvolvimento em Debate: novos rumos do desenvolvimento no mundo**. Rio de Janeiro: BNDES, vol. 2, p. 275-307, 2002.

CANO, W. Questão regional e política econômica nacional. In: CASTRO, A. C.(org.) **Desenvolvimento em debate: novos rumos do desenvolvimento no mundo**. Rio de Janeiro: BNDES, vol. 2, p. 275-307, 2002.

CARVALHO, A. Y. *et al.* **Ensaio sobre economia regional**. Instituto de Pesquisa Econômica Aplicada, Brasília, 2007.

COELHO, R. L. **Dois ensaios sobre a desigualdade de renda dos municípios brasileiros**. Universidade Federal de Minas Gerais, 2006.

COELHO, R. L.; FIGUEIREDO, L. “Uma análise da hipótese da convergência para os municípios brasileiros”. **Revista Brasileira de Economia**, v. 61, n. 3, Julho-Setembro, p. 331–352, 2007.

CONLEY, T.G. & LIGON, E. “Economic distance and cross-country spillovers”. **Journal of Economic Growth**, n. 7, p. 157-87, 2002.

CORDOBA J. C. **A generalized Gibrat’s law for cities**. Economics Department, Rice University, 2004a.

CORDOBA J. C. **On the distribution of city sizes**. Economics Department, Rice University, 2004b.

DAVIS, D.R.; WEINSTEIN, D.E. “Bones, bombs and break points: the geography of economic activity”. **The American Economic Review**, v. 92, n. 5, Dec, p. 1269-2002. 2002.

DAVIS, D.R.; WEINSTEIN, D.E. “A search for multiple equilibria in urban industrial structure”. **Working Paper 10252**, NBER Working Paper Series, 2004, Jan. National Bureau of Economic Research. 2004.

DOBKINS, L. H.; IOANNIDES, Y. M. “Dynamic evolution of the size distribution of U.S. cities”. In: J. M. Huriot and J. F. Thisse (ed.) **Economics of Cities**, p. 217–260. New York: Cambridge University Press. 2000.

DURANTON, G. **City size distributions as a consequence of the growth process**. Centre for Economic Performance. London School of Economics and Political Science. 2002.

DURANTON, F. “Some foundations for Zipf 's law: Product proliferation and local spillovers”. **Regional Science and Urban Economics**, v. 36, p. 542–563, 2006.

DURANTON, G. “Urban evolutions: the fast, the slow, and the still”. **The American Economic Review**, v. 97, n. 1, p. 197-221. 2007.

EATON, J. and ECKSTEIN, Z. “Cities and growth: theory and evidence from France and Japan”. **Regional Science and Urban Economics**, v. 27, p. 443–474. 1997.

EECKHOUT, J. “Gibrat's law for (all) cities”. **The American Economic Review**, v. 94, n. 5, December, p. 1429-1451. 2004.

FAN, C. C.; CASETTI, E. “The spatial and temporal dynamics of US regional income inequality, 1950-1989”. **The Annals of Regional Science**, n. 28, p. 177-196. 1994.

FINGLETON, B. “Estimates of time to economic convergence: an analysis of regions of the European Union”. **International Regional Science Review**, v. 22, n. 1, p. 3-34, 1999.

FINGLETON, B. “Theoretical economic geography and spatial econometrics: dynamic perspectives”. **Journal of Economic Geography** 1 (2001) pp. 201-225. 2001.

FUGUITT, G. V. "The nonmetropolitan population turnaround". **Annual Review of Sociology**, 11, p. 259-280. 1958.

GABAIX, X. "Zipf's law and the growth of cities". **The American Economic Review**, v. 89, n. 2, May, p. 129-132, 1999a.

GABAIX, X. "Zipf's law for cities: an explanation," **Quarterly Journal of Economics**, CXIV, August, p.739-767, 1999b.

GABAIX, X.; IOANNIDES, Y.M. "The evolution of city size distributions". **Handbook of Urban and Regional Economics, Volume IV: Cities and Geography**. 2003.

GABAIX, X; IBRAGINOV, R. "Rank-1/2: a simple way to improve the OLS estimation of tail exponents". **NBER Technical Working Paper Series** n. 342. Setembro, 2007.

GIBRAT, R. **Les inegalites economiques**. Paris: Librairie du Recueil Sirey, 1931.

GLAESER, E. L. **Cities, information, and economic growth**. Harvard University. 1994.

GLAESER, E. L. *et al.* "Growth in cities". **The Journal of Political Economy**, v. 100, n. 6, p. 1126-52.1992.

GLAESER, E. L.; SCHEINKMAN, J. A.; SHLEIFER, A. "Economic growth in a cross-section of cities". **Journal of Monetary Economics**, v. 36, p. 117-143. 1995.

GLAESER, E. L. "The economics approach to cities". **Nber Working Paper Series** N°13696. Dezembro, 2007.

GUÉRIN-PACE, F. "Rank-size distribution and the process of urban growth". **Urban Studies**, v. 32, n. 3, p. 551-562.1995.

HADDAD, E. A. **Regional inequality and structural changes: lessons from the brazilian economy**. Ashgate, Aldershot. 1999.

HAESSEL, W. "Macroeconomic policy, investment, and urban unemployment in less developed countries". **American Journal of Agricultural Economics**, v. 60, n. 1, Feb., p. 29-36. 1978.

HENDERSON, J. V. "The sizes and types of cities". **The American Economic Review**, v. 64, n. 4, September, p. 640-656. 1974.

HENDERSON, J. V. "The urbanization process and economic growth: the so-what question". **Journal of Economic Growth**, v. 8, p. 47-71. 2003.

HENDERSON, J. V. **Urbanization and growth**. Brown University. 2004.

HENDERSON, J. V.; THISSE, J. F. "On strategic community development". **The Journal of Political Economy**, v. 109, n. 3, Jun., p. 546-569. 2001.

HENDERSON, J. V.; VENABLES, A. J. **The dynamics of city formation: finance and governance**. 2005.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA - IPEADATA. Dados macroeconômicos e regionais. Disponível em: <<http://www.ipeadata.gov.br>>. Acesso em 6 jul. 2008.

IOANNIDES, Y.; OVERMAN, H. “Zipf’s law for cities: an empirical examination”, **Regional Science and Urban Economics**, v. 33, p. 127-137. 2003.

JANNUZZI, P. M. “Redistribuição regional da população no interior paulista nos anos 80:em busca dos determinantes estruturais do fenômeno”. **Textos NEPO 34**. Campinas, julho, 1998.

KALECKI, M. “On the Gibrat distribution”. **Econometrica**, v. 13, p.161– 170. 1945.

KELEJIAN, H.; PRUCHA, I. “A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances”. **The Journal of Real Estate Finance and Economics**, Volume 17, Number 1 / July 1998. 99-121. 1998.

KELEJIAN, H.; PRUCHA, I. “A generalized moments estimator for the autoregressive parameter in a spatial model”. **International Economic Review**, Volume 40 Issue 2, Pages 509 – 533, 1999.

KRUGMAN, P. “Confronting the mystery of urban hierarchy”. **Journal of the Japanese and International Economies** 10, 1120–1171. 1996.

KRUGMAN, P. “The final frontier”. **The Journal of Economic Perspectives**, Vol. 12, No. 2 (Spring, 1998), p. 161-174. 1998.

LANASPA L, PUEYO F, SANZ F. “The evolution of Spanish urban structure during the Twentieth Century”. **Urban Studies** v. 40, n. 3, p. 567–580. 2003.

LANASPA L; PERDIGUERO A.M.; SANZ F. **La distribución del tamaño de las ciudades en España, 1900-1999**. Mimeo, Departamento de Análisis Económico, Universidad de Zaragoza. 2004.

LE GALLO, J.; CHASCO, C. “Spatial analysis of urban growth in Spain (1900 – 2001)”. **Empirical Economics**, v. 34, n. 1, p. 59 – 80. 2008.

LESAGE, J. P. (1999) **Spatial econometrics**. Manuscrito não publicado disponível em <<http://rri.wvu.edu/regscweb.htm>>.

LEMOS, M. B. et al. **A dinâmica urbana das regiões metropolitanas brasileiras**. In: Encontro Nacional de Economia, 29., 2001, Salvador. Anais... Salvador: ANPEC, 2001. CD-ROM.

LEMOS, M. B. et al. “A nova configuração regional brasileira e sua geografia econômica”. **Estudos Econômicos**, v. 33, n. 4, p. 665-700, out./dez. 2003.

LEMOS, M. B.; GUERRA, L. P.; MORO, S. **A nova configuração regional brasileira: sua geografia econômica e seus determinantes locais da indústria**. In: Encontro Nacional de Economia, 28, 2000, Campinas. Anais... Campinas: ANPEC, 2000. CD-ROM.

LEVER, W. F. "Reurbanisation: the policy implications", **Urban Studies**, n. 30, p. 267– 284. 1993.

MARTINE, G., CAMARGO, L. "Crescimento e distribuição da população brasileira". **Revista de Estudos Populacionais**, Abep, v. 1, n. 1/2, p. 99-144, jan./dez. 1984.

MATA, M.; CARVALHO, E.W.; Castro e Silva, M. T. **Migrações internas no Brasil: aspectos econômicos e demográficos**. IPEA, coleção Relatórios de Pesquisa, n. 19, Rio. Janeiro. 1973.

MIRANDA, R. A. DE; BADIA, B. D. **A evolução da distribuição do tamanho das cidades de Minas Gerais: 1920 – 2000**. In: Anais do XII Seminário sobre a Economia Mineira, Cedeplar, Universidade Federal de Minas Gerais. 2006.

MORAN, P. "The interpretation of statistical maps". **Journal of the Royal Statistical Society B**, 10, 243-51. 1948.

NETTO JUNIOR, J. L. DA S.; TARGINO, I. **Migrações e diferenciais de renda estaduais: uma análise por dados em painel no período de 1950 - 2000**. Trabalho apresentado no Encontro Transdisciplinar Espaço e População, Campinas. 2003.

OLIVEIRA, C. A. DE. **A evolução da distribuição do tamanho das cidades brasileiras: 1936-2000**. 2004. [artigo científico]. Disponível em: <<http://www.bnb.gov.br/content/aplicacao/ETENE/Anais/docs/2004-aevolucao.pdf>>. Acesso em: 20 fev. 2008.

OLIVEIRA, C.A e GUIMARÃES NETO, L. "Emprego organizado e regiões nos anos 90: quem perdeu mais?" **Estudos Econômicos**, v. 27, n. especial, p. 37-64, São Paulo. 1997.

OVERMAN, H. and IOANNIDES Y. M. "Cross sectional evolution of the U.S. city size distribution". **Journal of Urban Economics**, 49, p. 543-566, 2001.

PATARRA, N. L. "Movimentos migratórios no Brasil: tempos e espaços". **Textos para Discussão da Escola Nacional de Ciências Estatísticas**, n. 7, 2003.

QUAH, D. "Empirical cross-section dynamics in economic growth". **European Economic Review**, v. 37, p. 426–434, 1993.

REIS, E.; PIMENTEL, M.; ALVARENGA, A.I. **Áreas mínimas comparáveis para os períodos intercensitários de 1872 a 2000**. Mimeo, IPEA/DIMAC, Rio de Janeiro. 2008.

RESENDE, M. "Gibrat's law and the growth of cities in Brazil: a panel data investigation". **Urban Studies**, v. 41, n. 8, p. 1537-1549, 2004.

- RESENDE, M. “O crescimento econômico dos municípios mineiros: externalidades importam?”. **Cadernos BDMG**, n. 11, p. 5-33, 2005.
- RIBEIRO, L.C. “Dinâmica socioterritorial das metrópoles brasileiras: dispersão e concentração.” **IPEA políticas sociais – acompanhamento e análise**, n. 12, fev. 2006
- RICHARDSON, H. W. “Theory of the Distribution of City Sizes: Review and Prospects”, **Regional Studies**, v. 7, p. 239-251, 1973.
- ROMER, P.M. “Endogenous technical change”. **Journal of Political Economy** 98, S71–S102, 1990.
- ROSEN, K. T.; RESNICK, M. “The size distribution of cities: an examination of the Pareto law and primacy”, **Journal of Urban Economics**, v. 8, p. 165-186, 1980.
- ROSSI-HANSBERG E, WRIGHT M. “Urban structure and growth”. **National Bureau of Economic Research**, Working Paper 11262, April. 2005.
- RUIZ, R. M. “Estruturas urbanas comparadas: Estados Unidos e Brasil”. **Pesquisa e Planejamento Econômico**, julho, p. 715-737, 2005.
- SABOIA, J. “A Dinâmica da descentralização industrial no Brasil”. **Texto para discussão Nº 452**. Universidade Federal do Rio de Janeiro, Instituto de Economia, julho, 2001.
- SHARMA, S. “Persistence and stability in city growth”. **Journal of Urban Economics**, v. 53, pp. 300– 320, 2003.
- SIMON, H. A. “On a class of skew distribution functions”. **Biometrika**, v. 42, n. 3/4, Dec., pp. 425-440, 1955.
- SOO, K. “Zipf’s law for cities: a cross country investigation”, **Working Paper, Centre for Economic Performance**, London School of Economics. 2002.
- SUÁREZ-VILLA, L. “Metropolitan evolution, sectoral economic change, and the city size distribution”, **Urban Studies**, v. 25, p. 1-20, 1988.
- SUZIGAN. **Aglomerções industriais como foco de políticas regionais**. Texto da aula magna do XXVIII Encontro Nacional de Economia da ANPEC. Campinas, SP, Brasil, dezembro, 2000.
- TAN, B; YILMAZ, K. “Markov chain test for time dependence and homogeneity: An analytical and empirical evaluation”. **European Journal of Operational Research** 137 (2002) 524-543, 2002.
- WILLIAMSON, J. “Regional inequality and the process of national development” **Economic Development and Cultural Change**, June, p. 3-45, 1965.

ZELLNER. "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias". **Journal of the American Statistical Association**, Vol. 57, No. 298 (Jun. 1962), p. 348- 368, 1962.

ZIPF, G. K. **Human Behaviour and the Principle of Least Effort**. Addison-Wesley, Reading, MA, 1949.