



## UM SÉCULO DE CRESCIMENTO DAS CIDADES BRASILEIRAS: UMA ABORDAGEM ESPACIAL (1910-2010)

Wellington Ribeiro Justo  
URCA/UFPE  
justowr@yahoo.com.br

### 1 Introdução

Toda atividade econômica tem uma localização, embora atividades dissimilares floresçam em diferentes áreas. A economia regional e urbana se preocupa em considerar o efeito do impacto da locação e a distância na atividade econômica. Economistas regionais buscam identificar onde determinadas atividades econômicas irão prosperar ou porque determinadas atividades se localizam de forma mais concentrada em certas áreas e porque determinadas famílias escolhem se localizar em certas áreas. Outro foco deste campo de pesquisa é compreender o que fazem as pessoas migrarem, pois este fenômeno afeta as taxas de crescimento das cidades. Assim, os economistas regionais combinam instrumentos da micro e macroeconomia e economia internacional para analisar padrões e outros componentes das taxas de crescimento das localidades. Nesse contexto conceitos tais como proximidade e custos de transporte, retornos crescentes de escala e externalidades são empregados. Estes componentes mudam a face da teoria econômica tradicional. Já os economistas urbanos, por sua vez, focam nas relações entre áreas urbanas e periféricas bem como no padrão de uso do solo dentro das cidades (EDWARDS, 2007).

A análise econômica da distribuição espacial das atividades econômicas e, por conseguinte da população segundo Edwards (2007) vem desde o tratado de Cantillon em 1755 ao analisar porque as cidades se desenvolviam próximas às áreas mais produtivas. Entre 1800 e 1950 a escola alemã esteve à frente ao enfatizar a questão espacial na atividade econômica.

O processo de desenvolvimento econômico está associado com a distribuição da população entre os municípios de um país. Neste contexto surge a indagação de como cidades de portes diferentes crescem durante o processo de desenvolvimento. A distribuição do tamanho das cidades pode ocorrer no sentido que as menores cidades crescem mais rapidamente que as maiores ou no extremo oposto, as cidades maiores poderiam crescer mais que as menores aumentando ainda mais a concentração populacional beneficiando poucas localidades.

Muitos são os fatores envolvidos na dinâmica de crescimento das cidades. Decisões de políticas públicas podem tornar determinados municípios mais ou menos atraentes para empreendedores. Firms trazem empregos e elevam a compra de bens e serviços locais influenciando na atração de mais firms e pessoas. Desta forma, atração de empresas é um dos fatores chaves no processo de desenvolvimento elevando o padrão de vida dos residentes.

As firms buscam minimizar custos de transportes ao obter ganhos de escala por estarem próximas dos grandes mercados colaborando para a aglomeração de firms as beneficiando com economias de aglomeração. Já, as amenidades também podem favorecer na decisão de empreendedores em aportarem seus investimentos em determinada localidade. Por outro lado, pode haver deseconomias de aglomeração limitando o crescimento indeterminado de certas localidades.

A existência de *clusters* de firms em uma área pode gerar economias de aglomeração ao criarem forças centrípetas que reforça o processo de aglomeração. Contudo, os *clusters* podem crescer muito e surgir deseconomias de aglomeração. Estas, por sua vez, criam forças centrífugas que repelem a concentração econômica (Edwards, 2007).

Uma dificuldade presente durante muito tempo para os pesquisadores de economia regional era a falta de formalização das questões regionais que dificultava testar empiricamente suas hipóteses. Por outro lado, havia uma regularidade observada e testada com vários conjuntos de dados, mas que ao contrário carecia de explicações teóricas que é a regra da ordem do tamanho das cidades (Lei de Zipf). Esta regularidade empírica mostra como acontece a distribuição espacial das cidades ao longo do tempo. Fujita, Krugman and Venables (2001) e Duranton (2006), contudo buscaram explicar teoricamente tal regularidade.

A análise do crescimento das cidades, por sua vez, permite entender de que forma este crescimento afeta as cidades de vários portes possibilitando identificar a

efetividade de políticas públicas locais na solução de problemas endêmicos nas áreas urbanas.

Nesse sentido este trabalho busca identificar a dinâmica de crescimento das cidades brasileiras no último século melhorando o conhecimento do sistema urbano no Brasil. Mais especificamente procura responder as seguintes indagações: Como tem sido a distribuição espacial das cidades brasileiras no último século? Há alguma mobilidade das cidades dentro da distribuição?

Este trabalho avança na literatura ao utilizar um banco de dados que abrange um maior período de tempo disponível para a economia brasileira desagregada ao nível de cidades. Outra contribuição é incorporar na análise da estimação do coeficiente de Pareto os efeitos espaciais de vizinhança. E finalmente, ao analisar a dinâmica de crescimento das cidades com as matrizes markovianas para captar possíveis movimentos dentro da distribuição.

Este artigo está organizado da seguinte forma. Além desta introdução, na seção seguinte apresentam-se os modelos a serem estimados. Na terceira seção apresentam-se e discutem-se os resultados. Finalmente a última seção traz as conclusões.

## 2 Metodologia

Os dados utilizados foram obtidos junto ao IPEA (Instituto de Pesquisa Econômica Aplicada). Utilizaram-se as AMC (Áreas Mínimas Comparáveis) compatibilizada para o último século (1910-2010). Esta amostra compreende 432 observações incluindo o território de Fernando de Noronha. Neste trabalho, desconsiderou-se esta observação por não ter dados disponíveis para toda a série e por ser uma ilha causa alguns problemas na construção das matrizes de vizinhança.. A vantagem desta amostra para a análise da lei de Zipf é porque se deixa o aspecto arbitrário do pesquisador ao fazer uma escolha *ad hoc* do tamanho das cidades a ser considerado o que no caso brasileiro torna-se um problema pela criação de municípios entre uma década e outra. No caso brasileiro Justo (2007) e Oliveira (2005) estimaram o coeficiente de Pareto para vários tamanhos de cidades e os valores obtidos variavam a depender desta escolha. Similarmente no caso internacional, este problema também foi levantado por Lanaspá *et al* (2003) que considerou para o caso espanhol cidades a partir de 50 mil habitantes e Mella e Chasco (2006).

### 2.1 A regra da ordem do tamanho ou a lei de Zipf

Na estimação da do expoente de Pareto ou regra da ordem do tamanho seguiu-se o modelo original proposto por Zipf (1949). O autor sugeriu que a distribuição das cidades seguia a regra de (Pareto, 1897) de acordo com o seguinte modelo:

$$R = a \cdot S^{-\beta} \quad (1)$$

Onde R é o ranking da ordem da distribuição da população; S é a população da cidade; e a e  $\beta$  são parâmetros, sendo este último o expoente de Pareto sendo positivo por construção.

De acordo com Le Gallo and Chasco (2009) a princípio a regra da ordem do tamanho surgiu de constatação de uma regularidade nos dados observados sem nenhuma fundamentação da teoria econômica. Krugman (1996) Eaton and Eckstein (1997), Overman and Yoannides (2001), Gabaix and Ibragimov (2006) entre outros estimaram o referido modelo constatando esta regularidade para a Lei de Zipf para as cidades. Duranton (2006), contudo propiciou uma fundamentação teórica baseado em modelo de crescimento endógeno sugerido por Grossman and Helpman (1991) em uma estrutura urbana e analisou o efeito de investimentos em P&D nas taxas de crescimento das cidades.

Formalmente nesta estrutura, o tamanho da distribuição das cidades depende do valor do expoente do Pareto ( $\beta$ ). No limite, se  $\beta$  tende a infinito, então todas as cidades terão a mesma população. Quando  $\beta$  é igual a 1, tem-se o que é conhecido como a regra da ordem do tamanho ou lei de Zipf. De acordo com esta regra, a população entre quaisquer grupos de cidades no tempo é inversamente proporcional ao ranking da sua população no grupo. O Expoente de Pareto pode ser interpretado como um indicador de convergência. Quando o valor do coeficiente cai ao longo do tempo indica uma importância relativa maior para as grandes cidades. Além disso, isto causa uma tendência de divergência no grupo das cidades ou maior concentração nas maiores cidades. Por outro lado, um aumento de 1% no tamanho da cidade produz uma menor queda (em %) no rank quando  $\beta$  aumenta. Assim, elevação no valor de  $\beta$  representa uma dinâmica de convergência, ou em outras palavras, uma maior dispersão da população fora dos grandes centros urbanos e maior equilíbrio na distribuição da população entre os centros urbanos de diferentes tamanhos (Le GALLO AND CHASCO, 2009).

O modelo empírico derivado da equação (1) pode ser estimado na seguinte forma funcional:

$$\ln R_{it} = \ln \alpha_t - \beta_t \cdot \ln S_{it} + \varepsilon_{it} \quad (2)$$

De acordo com Gabaix and Ibragimov (2006) a estimação por Mínimos Quadrados Ordinários (MQO) da equação (2) é viesada para pequenas amostras. Desta forma seguiram-se os autores Le Gallo e Chasco (2009) e fez-se a correção do viés subtraindo-se  $\frac{1}{2}$  do rank e estimou-se o modelo:  $\ln(\text{Rank} - \frac{1}{2}) = \alpha - \beta \ln(\text{tamanho})$ . Testou-se a estimação sem esta correção e comprovou-se a robustez da correção sugerida, mantendo assim esta correção.

## 2.2 O efeito espacial

Os estimadores de MQO podem ser afetados pela omissão de autocorrelação espacial. Se o processo gerador da autocorrelação espacial dos resíduos for autorregressivo, os estimadores de MQO são não-viesados, mas não eficientes. A inferência estatística, então é viesada neste caso. Já se eles são devido à omissão da autocorrelação espacial das variáveis, os estimadores de MQO são viesados.

Desta forma, este trabalho procura identificar e incorporar modelos que apreendam o efeito de vizinhança na estimação do expoente de Pareto, tornando assim, um avanço na literatura nacional sobre o tema.

A autocorrelação espacial, também conhecida como dependência espacial, interação espacial ou interação local é definida como uma medida de similaridade entre dois valores de um atributo que estão próximos espacialmente. De acordo com Pacheco and Tirrel (2002) a autocorrelação espacial pode ser mensurada por vários índices entre os quais o mais conhecido é o I de Moran que mede o grau de associação linear entre um atributo (y) em um dado local e a média ponderada dos atributos nas localidades vizinhas (Wy) e pode ser interpretada como a inclinação da regressão de (y) em (Wy). A autocorrelação espacial pode ser visualmente ilustrada em um gráfico em que (Wy) é plotado no eixo vertical e (y) no eixo horizontal.

Esta estatística segue a seguinte expressão segundo Battisti and Vaio (2009):

$$I = \frac{n}{q} \frac{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j} \quad (3)$$

Onde,  $\omega_{ij}$  é um elemento da matriz de peso W,  $x_i$  é uma variável específica para a observação  $i$ ,  $n$  é o número de observações,  $q$  é um fator de escala igualando a soma de todos os elementos da matriz. Neste artigo, seguiu-se Justo *et. al.* (2010) e utilizou-se a matriz binária de linha padronizada baseada na estrutura de vizinhança com um número fixo de vizinhos próximos (k-nearest), nos quais os elementos são:

$$\left\{ \begin{array}{l} \omega_{ij}(k) = 0 \text{ se } i=j \\ \omega_{ij}(k) = 1 \text{ se } d_{ij} \leq d_i(k) \\ \omega_{ij}(k) = 0 \text{ se } d_{ij} > d_i(k) \end{array} \right\}$$

Onde  $d_i$  é um valor crítico, definido para cada observação  $i$ , assegurando que cada município tenha o mesmo número de vizinhos<sup>1</sup>.

A heterogeneidade espacial (também conhecida como uma estrutura espacial, não-estacionaridade, tendência global de larga escala dos dados) refere-se às diferenças na média e/ou variância, e/ou covariância incluindo autocorrelação dentro de uma região espacial. Diferentemente da autocorrelação espacial requer que a média e a variância de um atributo seja constante no espaço, e a autocorrelação espacial de um atributo em quaisquer duas locações dependa de um *lag* da distância entre duas locações, mas não da própria locação (JUSTO, *et al.*, 2010).

Nem sempre é fácil distinguir a heterogeneidade espacial da autocorrelação espacial. A presença de *clusters*, por exemplo, pode induzir a autocorrelação espacial entre vizinhos, mas também pode ser sinal de diferentes possibilidades de regimes espaciais (Anselin, 2001). Testes para determinar autocorrelação espacial ou heteroscedasticidade podem gerar resultados inconclusivos.

Segundo LeSage e Pace (2009) ao considerar a autocorrelação espacial em um conjunto de dados é necessário estabelecer a estrutura da vizinhança para cada localidade especificando àquelas locações que são consideradas vizinhas. Em particular, faz-se necessário especificar uma matriz de pesos correspondentes à estrutura de

<sup>1</sup> Esta é uma das razões de deixar de fora o território de Fernando de Noronha.

vizinhança tal que a matriz de variância–covariância possa ser expressa como uma função de um pequeno número de parâmetros estimáveis compatíveis com o tamanho da amostra (Anselin, 2002). Os tipos de matrizes de pesos utilizadas na econometria espacial incluem entre outros os seguintes tipos: torre, rainha, matriz de contigüidade, matriz de pesos espaciais por meio de uma distância limite, mas com um número fixo de vizinhos próximos (*k*-nearest), matriz de peso das distâncias, e a matriz do inverso das distâncias. Normalmente a matriz de pesos é definida exogenamente e após a comparação entre vários tipos de matrizes. Segundo Voss and Chi (2006) criam-se vários tipos de matrizes de pesos e escolhe-se aquela que apresenta a mais alta significância estatística<sup>2</sup>.

Segundo Chi and Zhu (2008) há dois problemas associados com a especificação dos pesos espaciais na prática. Um problema é que a estrutura de pesos pode ser afetada pela qualidade dos dados geo-referenciados. O outro problema é que o uso de alguma matriz de peso de distância pode requerer um valor limiar, que pode ser difícil de ser determinado especialmente quando há forte heterogeneidade espacial. Um pequeno limiar pode produzir muitas ilhas, enquanto um grande limiar cria uma quantidade excessiva de vizinhos. Uma solução para este caso proposto por Anselin (2002) é a estrutura de matriz de pesos espaciais por meio de uma distância limite, mas com um número fixo de vizinhos próximos (*k*-nearest).

Segundo Chi e Zhu (2008) em relação ao modelo de regressão linear, o modelo de regressão linear espacial comumente usado inclui além dos usuais coeficientes das variáveis explicativas ( $\beta$ ) e a variância do termo de erro ( $\sigma^2$ ) um coeficiente autorregressivo espacial ( $\rho$ ), que mede a força da autocorrelação espacial. Inclui também uma matriz de pesos ( $W$ ) correspondente à estrutura de vizinhança e a matriz de pesos ( $D$ ) que são pré-especificadas<sup>3</sup>.

Especificar-se-á a partir de agora um modelo de regressão linear espacial quando os termos de erros são especificados. Dois modelos mais comumente utilizados serão apresentados: o Modelo de Lag espacial cuja estrutura é assim modelada:

$$Y = X\beta + \rho WY + \varepsilon_i \quad (4)$$

Onde  $Y$  é o vetor das variáveis dependentes,  $X$  a matriz de variáveis explicativas,  $W$  a matriz de pesos espacial, e  $\varepsilon$  o vetor dos termos de erros que são independentes, mas

<sup>2</sup> Neste trabalho foram testados vários tipos de matrizes e optou-se pelo *k*-nearest com ( $k = 6$ ) por apresentar os melhores resultados.

<sup>3</sup> Para aprofundar nestes modelos veja LeSage e Pace (2009).

não necessariamente identicamente distribuídos. O outro modelo é o modelo de erro espacial que é especificado da seguinte forma:

$$Y = X\beta + u, u = \rho Wu + \varepsilon_i \quad (5)$$

Onde os termos são definidos como no modelo anterior.

No modelo de Lag espacial, a autocorrelação espacial é modelada por uma relação linear entre a variável dependente ( $y$ ) e associada a uma variável espacialmente defasada ( $Wy$ ). No caso no modelo de erro espacial, a autocorrelação espacial é modelada por um termo ( $u$ ) e o termo de erro espacialmente defasado ( $Wu$ ). Em qualquer um dos modelos, a interpretação de um coeficiente auto-regressivo espacial significativo nem sempre é direto. Um significativo termo de erro espacial indica autocorrelação espacial nos erros que podem ser devido às variáveis explicativas importantes que não foram incluídas no modelo (ANSELIN, 1995).

De acordo Chi and Zhu (2008) para um dado conjunto de dados vários modelos de regressões podem ser especificados. Se os modelos são aninhados pode-se utilizar um teste de razão de verossimilhança (LR)<sup>4</sup> para comparar os modelos. Se os modelos não são aninhados, pode-se utilizar o AIC (Akaike's Information Criterion) e BIC (Schwartz's Bayesian Information Criterion). Modelos com menores BIC e AIC são considerados melhores.

### 2.3 A Mobilidade Dentro do Sistema Urbano Brasileiro

A função de densidade e a Lei de Zipf permitem a caracterização da evolução da distribuição global, mas não propiciam quaisquer informações sobre o movimento das cidades dentro da distribuição. Por exemplo, não se sabe se os municípios que estão na cauda inferior da distribuição em 1910 são os mesmos que estão em 2010. Le Gallo and Chasco (2009) basearam-se no trabalho de Kemeny and Snell (1976) e sugerem que um caminho para tratar esta questão é verificar a evolução da posição relativa de cada cidade ao longo do período analisado por meio da estimação da matriz de probabilidades de transição com as cadeias de Markov. Black and Hendersson (2003) utilizaram deste procedimento para o sistema urbano americano.

A análise da evolução ao longo do tempo da distribuição *cross-section*, ou dito de outra forma, a análise da dinâmica, é uma metodologia que tem como objetivo descrever o processo estocástico Markoviano. Neste sentido, trabalhar com estado-espaço tem várias vantagens de acordo com Bulli (2001). A distribuição das

---

<sup>4</sup> Likelihood Ratio Tests.

probabilidades discretas e as matrizes de transição são mais fáceis de interpretar que um processo estocástico de Kernel. Outra vantagem é que índices descritivos e a distribuição ergótica de longo prazo são mais fáceis de ser calculadas. Porém, nesta metodologia surge um problema que é a estratificação, ou seja, os resultados podem ser condicionados à escolha das classes que dividem a distribuição.

Seguindo Le Gallo and Chasco (2009), a formulação da distribuição *cross-section* do tamanho populacional das cidades brasileiras no tempo  $t$  será denotado por  $F_t$ . Define-se um conjunto de  $K$  diferentes tamanhos de classes, que propiciam uma discreta aproximação da distribuição da população. Assume-se que a frequência da distribuição segue um processo Markoviano de primeira ordem. Neste caso, a evolução da distribuição do tamanho das cidades é representada pela matriz de transição de probabilidades,  $M$ , na qual cada elemento  $(i, j)$  indica a probabilidade que uma cidade que está em uma classe no período  $t$  passe para uma classe superior no período  $t+1$ . O vetor  $F_t (K, 1)$  indica a frequência das cidades em cada classe no tempo  $t$ . É descrita então pela seguinte equação:

$$F_{t+1} = M F_t \quad (6)$$

Onde  $M$  é a matriz de transição de probabilidades  $(K,K)$  representando as duas distribuições da seguinte forma:

$$M = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1K} \\ p_{21} & p_{22} & \dots & p_{2K} \\ \dots & \dots & \dots & \dots \\ p_{K1} & p_{K2} & \dots & p_{KK} \end{bmatrix} \quad (7)$$

Onde cada elemento  $p_{ij} \geq 0$ ,  $\sum_{j=1}^K p_{ij} = 1$ . A estacionária matriz transição de probabilidades  $p_{ij}$  capta a probabilidade que cada cidade na classe  $t-1$  passe para a classe  $j$  em  $t$ .

Os elementos da matriz  $M$  podem ser estimados da frequência observada nas mudanças de classes de um período para outro. De acordo com Amemiya (1985) e Hamilton (1994), o estimador de máxima verossimilhança de  $p_{ij}$  é:

$$p_{ij} = n_{ij} / n_i \quad (8)$$

Onde  $n_{ij}$  é o número total de cidades que se movem da classe  $i$  na década  $t-1$  para a classe  $j$  na imediata década seguinte  $t$  nas dez matrizes de transição e  $n_i$  é a soma total de cidades em  $i$  nas dez matrizes de transição.

Se as probabilidades de transição são estacionárias, ou seja, se a probabilidade entre duas classes não variam no tempo, então:

$$F_{t+s} = M^s F_t \quad (9)$$

Nesta estrutura, a distribuição ergótica (também chamada de distribuição de *steady state*) de  $F_t$ , caracterizada quando  $s$  tende para infinito na equação (9). Uma vez que as mudanças representadas pela Matriz  $M$  são repetidas em um número arbitrário de vezes. A distribuição existe se o Markov Chain for regular, ou seja, se somente se, para um  $m$ ,  $M^m$  não tem entradas com valores zero. Neste caso, segundo Le Gallo and Chasco (2009) a matriz de probabilidade de transição converge no limite para uma matriz  $M^*$  de rank igual a 1. A existência da distribuição ergótica,  $F^*$ , é denotada por:

$$F^* M = F^* \quad (10)$$

Este vetor  $F^*$  descreve a futura distribuição das cidades se os movimentos observados no período da amostra são repetidos infinitamente. De acordo com a equação (10), o limite da distribuição é dado pelo autovetor associado ao autovalor de  $M$ .

A hipótese do processo Markov Chain estacionário requer que a matriz de probabilidades,  $p_{ij}$ , seja de primeira ordem. Se as matrizes markovianas são de ordem superior ela irá conter apenas parte da informação necessária para descrever a verdadeira evolução da distribuição da população. Além disso, a propriedade da matriz markoviana implicitamente assume que as probabilidades de transição,  $p_{ij}$ , não são de ordem zero (Le GALLO AND CHASCO, 2009).

Neste caso, Bickenbach and Bode (2003) sugerem testar esta propriedade, ou seja, da independência temporal. Primeiro testa a ordem 0 versus a ordem 1; depois a ordem 1 versus a ordem 2 e assim sucessivamente. Se a hipótese de ordem 0 contra a hipótese de ordem 1 é rejeitada e se a hipótese de ordem 1 contra a ordem 2 não for rejeitada, então o processo é de ordem 1.

Para testar a a ordem 0, seguindo Le Gallo and Chasco (2009), a hipótese nula  $H_o : \forall_i : p_{ij} = p_j (i = 1, \dots, K)$  é testada novamente seguindo a alternativa  $H_a : \exists_i \setminus p_{ij} \neq p_j$ . O teste de razão de verossimilhança (LR) segue então:

$$LR^{(O(0))} = 2 \sum_{t=1}^K \sum_{j \in A_i} n_{ij}(t) \ln(\hat{p}_{ij} / \hat{p}_i) : asy \chi^2 [(K-1)^2] \quad (11)$$

Assumindo que  $\hat{p} > 0, \forall_j (j=1, \dots, K)$ .  $A_i = \{j : \hat{p}_{ij} > 0\}$  é o conjunto de probabilidades de transição não-zeros sob a hipótese nula.

Para testar a ordem 1 versus a ordem 2, o Markov Chain de segunda ordem é definido levando em consideração o tamanho da população na classe  $k$  ( $k=1, \dots, K$ ) na qual as cidades estavam no tempo  $t-2$  e assumindo que o par de sucessivas classes  $k$  formam uma classe. Então a probabilidade de uma cidade mover para uma classe  $j$  no tempo  $t$ , dado que estava em  $k$  no período  $t-2$  e em  $i$  no período  $t-1$ , é  $p_{kij}$ . O correspondente número absoluto de transições é  $n_{ij}(t)$ , com a frequência marginal sendo

$$n_{ki}(t-1) = \sum_j n_{kij}(t-1)(t-1).$$

Para testar  $H_o : \forall k : p_{kij} = p_{ij} (k=1, \dots, K)$  contra a hipótese alternativa  $H_a : \exists k : p_{kij} \neq p_{ij}$ , as probabilidades  $p_{kij}$  são estimadas como  $\hat{p}_{kij} = n_{kij} / n_{ki}$ , onde  $n_{kij} = \sum_{t=2}^T n_{kij}(t)$  e  $n_{ki} = \sum_{t=2}^T n_{ki}(t-1)$ . O teste de razão de verossimilhança (LR) é estimado da seguinte forma:

$$LR^{(O(1))} = 2 \sum_{k=1}^K \sum_{i=1}^K \sum_{j \in C_{ki}} n_{kij} \sum_{j \in A_i} n_{ij} \ln(\hat{p}_{kij} / \hat{p}_{ij}) : asy \chi^2 \left[ \sum_{i=1}^K (c_i - 1)(d_i - 1) \right] \quad (12)$$

De forma similar para ordens superiores  $C_i = \{j : \hat{p}_{ij} > 0\}$ ,  $c_i = \#C_i$ ,  $C_{ki} = \{j : \hat{p}_{kij} > 0\}$  e  $d_i = D_i = \#\{k : n_{ki} > 0\}$ .

Se rejeitar que a cadeia markoviana é de ordem 0 e 1 prossegue-se testando para ordens superiores. Contudo, uma vez que o número de parâmetros a ser estimados cresce exponencialmente com o número de defasagens e o número de observações diminui linearmente para um dado conjunto de dados, o poder do teste diminui rapidamente. Posto isto, deve-se definir a priori uma ordem a ser testada.

### 3. Resultados e discussões

Nesta seção serão apresentados e discutidos os resultados explorando cada uma das metodologias empregadas. Inicialmente faz-se uma breve apresentação das estatísticas descritivas.

A tabela 1 apresenta as estatísticas descritivas da distribuição das cidades brasileiras no período 1910-2010. O crescimento acelerado das médias e medianas ao longo dos cem anos analisados indica a rápida expansão da população brasileira neste período. Entre 1910 e 2000 há um aumento monotonicamente crescente da dispersão

medido pelo coeficiente de variação, mas apresenta uma reversão na tendência na última década.

Tabela 1 Análise Descritiva

Ano	Média	Desvio-padrão	Mediana	Coeficiente de Variação
1910	54009.71	94149.76	26977	1.74
1920	70688.29	130965.9	33369	1.85
1940	95335.39	219925	41091	2.31
1950	120245.3	298493.8	51745	2.48
1960	163481	441705.1	61635	2.71
1970	215578.1	630807.8	73813	2.93
1980	275425.8	834924.9	87347	3.03
1991	339689.3	1014843	104553	2.97
2000	392667.3	1167194	120611	2.97
2010	439823.	1296059	131271	2.95

Fonte: Elaborado com base nos dados do IPEADATA. Elaboração do autor.

### 3.1 A evolução da forma da distribuição da população nas cidades Brasileiras

Usando as observações disponíveis para o período 1910-2010 da população das cidades brasileiras, analisou-se a forma da distribuição do tamanho das cidades. A figura 1 apresenta as estimações das funções de densidade não-paramétricas de Kernel da população das cidades brasileiras para 1910, 1940, 1980 e 2010. Observa-se uma característica de unimodalidade no início do século, até a década de 80. Este resultado sugere divergência do tamanho das cidades. Em 2010, há, contudo, uma ligeira indicação da formação de outra moda entre as cidades localizadas na cauda superior da distribuição o que compreende as maiores cidades. Desta forma, a densidade começa a captar um possível efeito de redistribuição populacional das cidades brasileiras entre o grupo das maiores cidades, que no caso brasileiro é formado na sua maioria pelas capitais dos estados. Justo *et al* (2010) mostra, por exemplo, que algumas cidades nordestinas, não somente as capitais, mas, sobretudo as cidades de porte médio já apresentam um fluxo de entrada de migrantes maior que a saída. Quando se analisa a migração de retorno, este comportamento é ainda mais evidente. Sabe-se, contudo, que no caso brasileiro a migração é um fator importante no crescimento das cidades.

O deslocamento da função de densidade mostra as altas taxas de crescimento populacional das cidades brasileiras ao entre 1910 e 2010. Le Gallo and Chasco (2010)

em estudo similar para o sistema urbano espanhol encontraram distribuição unimodal até a década de 80 e multimodalidade nas últimas décadas.

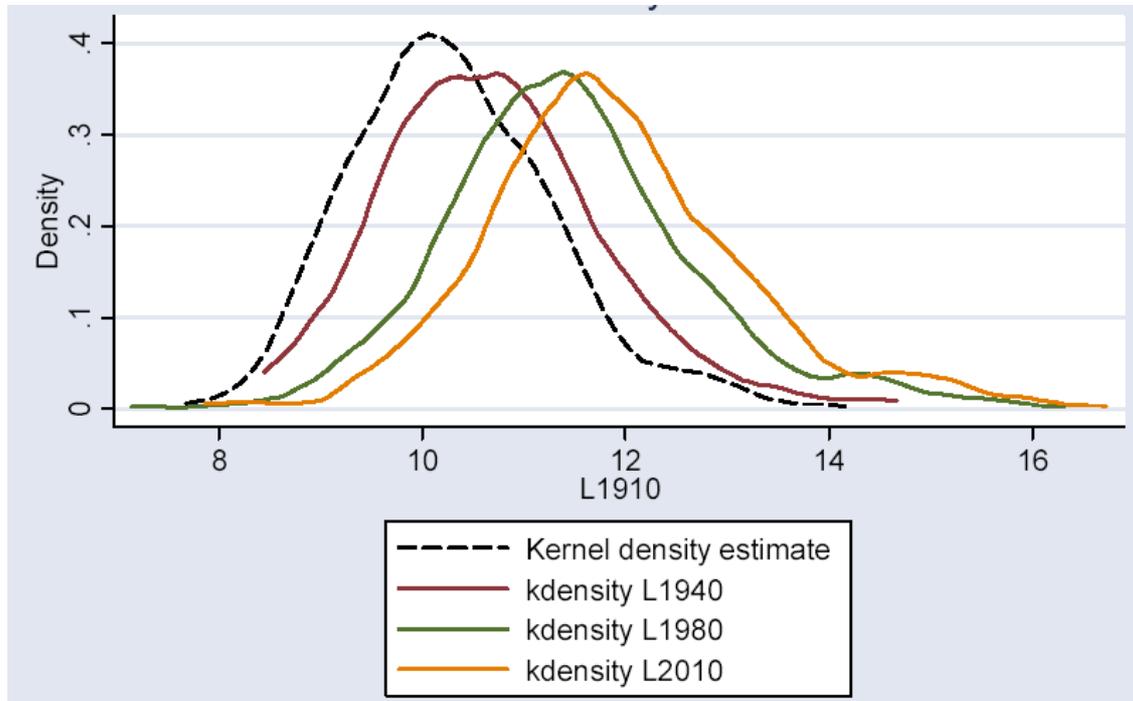


Figura 1 Função de densidade de Kernel para a população das cidades brasileiras: 1910-2010.

Fonte: Elaborado com dados do IPEADATA. Elaboração do autor.

### 3.2 A regra da ordem de tamanho: A lei de Zipf

Continuando a explorar a evolução urbana brasileira no século 1910-2010, explora-se aqui a lei de Zipf. Como dito anteriormente neste trabalho trabalha-se com as AMC que compreendem para o período analisado 431 observações. Isto foge da escolha aleatória dos pesquisadores em definir a priori um tamanho de cidades. A literatura tem mostrado que os resultados são condicionados a esta escolha. Neste sentido, os trabalhos de Justo (2007), Oliveira (2005) evidenciam este problema.

A figura 2 apresenta a distribuição populacional das cidades brasileiras em 1910 (à esquerda) e 2010 (à direita). Percebe-se que após um século há uma concentração espacial da população das cidades brasileiras e uma tendência de aglomeração em cidades da Região Sudeste e no litoral e interior da Região Nordeste. Este resultado está em linha com os apontados pela estimação das funções de densidade de Kernel apresentados anteriormente.

Por outro lado, esta dinâmica da distribuição espacial da população das cidades brasileiras remete a necessidade de testar e incorporar nos modelos econométricos, caso os testes assim os indiquem, o efeito de vizinhança nos modelos que busca apreender a trajetória de crescimento das cidades brasileiras entre 1910 e 2010.

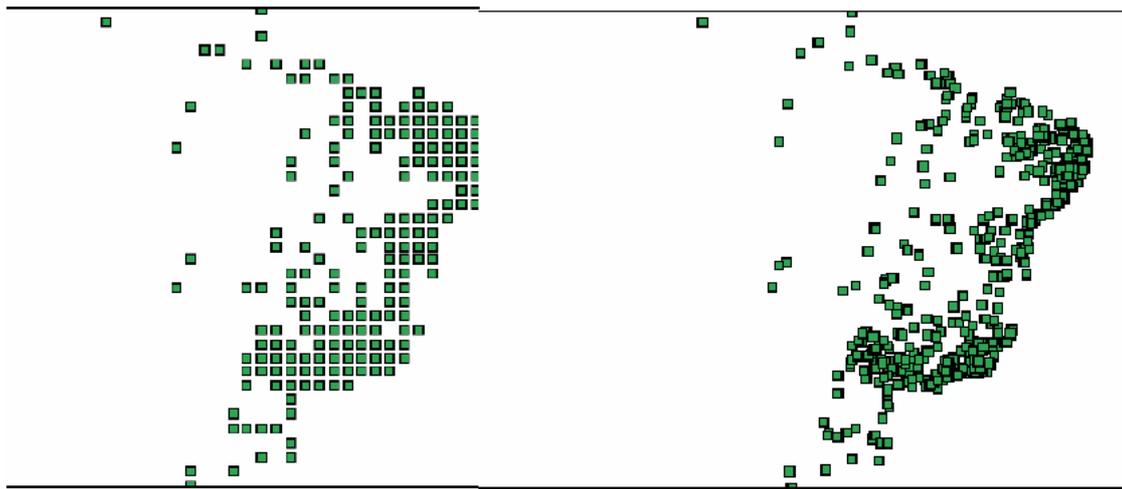


Figura 2 Distribuição espacial da população das cidades brasileiras: 1910-2010.

Fonte: Elaborado com base nos dados do IPEADATA. Elaboração do autor.

Posto isto, parte-se para testar o efeito espacial através da estatística I de Moran. A figura 3 apresenta o resultado do teste da estatística I de Moran da equação 3 para 1910 e 2010. O valor da estatística confirma a necessidade de incorporar o efeito de vizinhança na estimação dos modelos que buscam explicar a dinâmica de crescimento das cidades brasileiras<sup>5</sup>.

<sup>5</sup> Foram realizados os testes para todas as décadas. Apresenta-se apenas para a primeira e a última década por falta de espaço. Também foram estimados outros testes para indicar o efeito espacial, tais como os testes de Moran sob a hipótese da aleatoriedade entre outros.

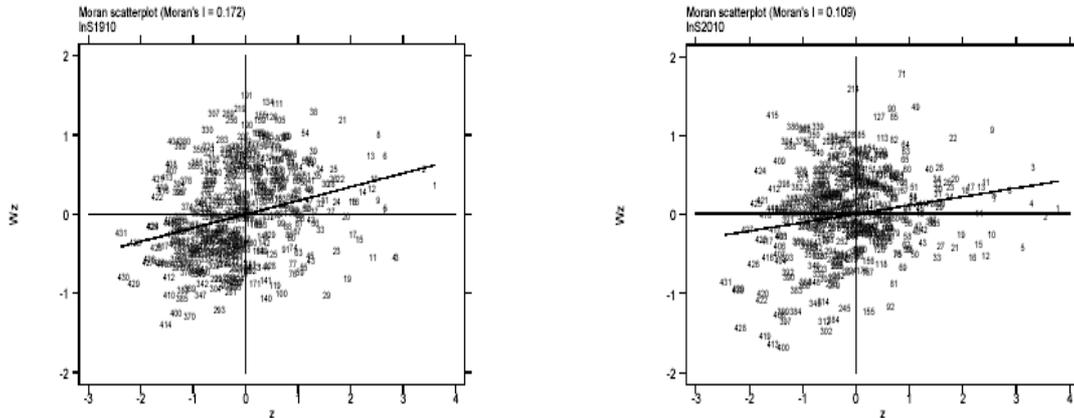


Figura 3 Gráfico de Dispersão de Moran: 1910 e 2010.

Fonte: Elaborado com base nos dados do IPEADATA. Elaboração do autor.

A tabela 2 apresenta o resultado das estimações do modelo que testa a ordem do tamanho das cidades brasileiras em cada década no período 1910-2010. Os resultados das estimações da equação por (MQO) da equação (2) indicam a não normalidade dos resíduos e exibem heteroscedasticidade como mostram os testes de Jarque-Bera, Koenker-Basset, respectivamente. Desta forma, infere-se que ambos os efeitos estão presentes nas 10 estimações.

Assim, parte-se para estimação de modelos espaciais. O resultado na coluna (SEM) e (SLM) são as estimações da equação (2) modificadas pelas equações (4) e (5). Isto é, do Modelo de Erro Espacial e o Modelo de Defasagem Espacial.

Tabela 2 Regressões do rank do tamanho das cidades brasileiras: 1910-2010

Ano	MQO					SEM			SLM		
	$\hat{\beta}_1$	$\hat{\beta}_2$	KB	JB	$\bar{R}^2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\lambda}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$
1910	14.21	0.93	25	904	0.91	13.76	0.89	0.25	16.2	0.98	0.32
1920	14.04	0.90	24	828	0.90	13.61	0.86	0.26	16.08	0.95	0.31
1940	13.70	0.85	20	572	0.90	13.34	0.82	0.27	15.59	0.89	0.31
1950	13.54	0.82	19	623	0.91	13.11	0.79	0.27	15.57	0.87	0.34
1960	13.42	0.80	18	726	0.91	12.98	0.76	0.26	15.31	0.84	0.31
1970	13.24	0.77	17	704	0.92	12.79	0.73	0.26	15.10	0.81	0.30
1980	13.16	0.75	16	1146	0.91	12.69	0.71	0.24	15.02	0.78	0.32
1991	13.12	0.73	15	1636	0.92	12.61	0.69	0.24	15.00	0.77	0.31
2000	13.15	0.73	14	1843	0.92	12.63	0.68	0.25	15.03	0.76	0.30

2010	13.16	0.72	12	2024	0.91	12.62	0.67	0.24	15.09	0.75	0.31
LR						23.691			10.118		
AIC						132			147		
BIC						140			159		

Fonte: Elaborado com base nos dados do IPEADATA. Elaboração do autor.

\* todos os coeficientes são significativos a 1%. MQO (estimação por Mínimos Quadrados Ordinários). JB – Testes Jarque-Bera para normalidade dos resíduos. KB – Teste Koenjer-Basset para heteroscedasticidade. AIC e BIC – (Critério de Informação de Akaike e Critério de informação Bayesiana de Schwarz).

A estatística LM-Error e LM-Lag são utilizadas para a escolha do modelo espacial mais adequando ao conjunto de dados. Neste caso os resultados apontam que o modelo de erro espacial é o mais adequado. O valor mais alto da estatística do teste LM confirma a superioridade do Modelo de Erro Espacial. O valor das estatísticas AIC e BIC também confirma esta escolha. Os expoentes de Pareto corrigidos no modelo espacial são menores que os estimados por MQO. Isto é, a divergência na distribuição das cidades brasileiras no último século é ainda mais intensa após a correção espacial. Le Gallo and Chasco (2009) encontraram resultados similares, mas com magnitudes diferentes para o sistema urbano espanhol. Lá os expoentes de Pareto são menores e também apresentaram efeito espacial significativo para as cidades espanholas.

O gráfico 1 apresenta a evolução ao longo do tempo das três estimações do expoente de Pareto nos últimos 100 anos. Observa-se que as estimações maiores são as encontrada pelo Modelo de Defasagem espacial seguido na ordem pelas estimações de Mínimos Quadrados Ordinários e o Modelo de Erro espacial. Percebe-se também uma tendência monotonicamente decrescente das estimações dos três modelos. Contudo, há uma desaceleração nas últimas décadas. Este último resultado indica uma possível desconcentração populacional nas grandes metrópoles brasileiras seja pela redução na taxa de crescimento das maiores cidades e/ou pelo aumento nas taxas de crescimento das cidades de porte médio que têm atraído um maior número de migrantes como mostrado por Justo (2010). Estes resultados são corroborados pelas estimações das funções de densidade de Kernel apresentados anteriormente.

Le Gallo e Chasco (2009) também analisando o comportamento da distribuição populacional nos últimos 100 anos entre os municípios da Espanha encontraram resultado nesta direção, sendo que lá há uma inflexão no final do período.

A análise da Lei de Zipf nos remete a observar que até a década de 80 o aumento da concentração urbana nas maiores cidades se dá de forma mais acelerada e perde força

a partir de então. Os valores dos coeficientes de variação já apontavam este comportamento.

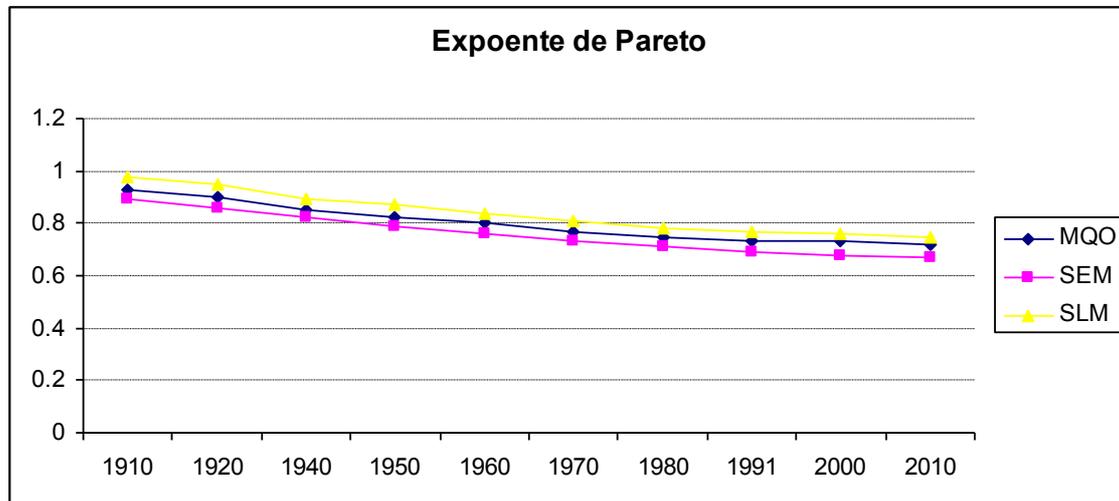


Gráfico 1 Evolução das estimações do expoente de Pareto (N= 431). Estimação por MQO (Mínimos Quadrados Ordinários), Modelo de Erro Espacial (SEM) e Modelo de Defasagem Espacial (SLM).

Fonte: Elaborado com base nos dados da tabela 1. Elaboração do autor.

### 3.3 Mobilidade Dentro do Sistema Urbano Brasileiro: 1910-2010.

Os resultados obtidos até aqui com a estimação das funções de densidade de Kernel e a análise da Lei de Zipf ainda que considerando o efeito espacial, não permite apreender informações a respeito da possibilidade de movimentos dentro da distribuição ao longo do tempo. Desta forma, parte-se para a análise do processo estocástico markoviano.

A utilização da metodologia explorada na seção anterior, contudo, pode trazer algumas incorreções a depender da divisão da distribuição em um dado conjunto de classes utilizada para encontrar a matriz de transição de probabilidades. Desta forma, testaram-se várias possibilidades e optou-se pela distribuição discreta que mais se aproximou da distribuição contínua.

Os resultados da matriz de transição de probabilidade são apresentados na tabela 3<sup>6</sup>. As probabilidades na diagonal mostram a baixa mobilidade interclasse, isto é, alta

<sup>6</sup> O teste de comparação das matrizes usando a equação 11 indica que o processo é de ordem 1. O valor da estatística LR = 15.000 e o Prob = 0.0000.

persistência das cidades permanecerem na sua própria classe entre uma década e outra ao longo dos últimos cem anos. Se os elementos da diagonal principal fossem todos iguais a 1 podiam ser interpretados como crescimento paralelo como sugerem Eaton and Eckstein (1997)<sup>7</sup>. No entanto, como não são exatamente iguais a 1, interpreta-se como a propensão de cada cidade mover para outra célula. Em particular, os resultados apontam que as maiores e as menores cidades (classe 6 e 1) apresentam maior persistência enquanto as cidades de porte médio são mais prováveis de se moverem dentro da distribuição (classes 3, 4 e 5). Contudo, além da baixa mobilidade constatada nos resultados quando há movimentos entre as classes é apenas para a classe seguinte<sup>8</sup>.

Tabela 3 Matriz de probabilidades de transição, 1910 -2010: População das Áreas Mínimas Comparáveis do Brasil.

	1 <20%*	2 <50%	3 <80%	4 <135%	5 <185%	6 >185%
1	0.9974	0.0026	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.9969	0.0031	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.9815	0.0185	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.9773	0.0227	0.0000
5	0.0000	0.0000	0.0000	0.0000	0.9524	0.0476
6	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Fonte: Elaborado com dados do IPEADATA. Elaboração do autor.

\* Nota: Os valores em percentual referem-se à média populacional da distribuição.

A chance de cidades pertencentes à classe 4 passarem para a classe 5 é de 2,27% enquanto as cidades pertencentes à classe 5 passarem para a classe 6 é de 4,76%. No limite inferior, contudo, a chance de uma cidade pertencente à classe 1 passar para a classe dois na década seguinte é de apenas 0,26%.

Considerou-se também a distribuição ergótica que pode ser interpretada como equilíbrio de longo prazo da distribuição do tamanho das cidades. Explicitamente, dada uma matriz de transição regular, com o decorrer de muitas décadas, haverá um tempo em que a distribuição das cidades não mais mudará. Isto é a distribuição ergótica ou

<sup>7</sup> A idéia dos autores é que as cidades iniciariam com tamanhos diferentes, mas cresceriam à taxa semelhante.

<sup>8</sup> Le Gallo and Chasco (2009) encontraram também baixa mobilidade e alta persistência para as cidades espanholas, mas em magnitudes diferentes. Lá há mobilidade nas classes intermediárias para além da classe imediatamente superior e também para classes inferiores. Contudo na dinâmica de crescimento urbano espanhol no final do século passado o comportamento do expoente de Pareto inverte a tendência.

limite da distribuição. Ela é usada para identificar a forma de convergência na distribuição. Concentração de frequência em certas classes implica em convergência. Por exemplo, se a concentração é na classe intermediária, haveria convergência para o meio da distribuição. Se, no entanto, houver concentração em várias classes, a distribuição limite seria multimodal e poderia ser interpretada como uma tendência de diferentes clubes de convergência. Finalmente se houver dispersão na distribuição ergótica entre as diversas classes seria interpretada como divergência.

A distribuição ergótica é, então, calculada para o tamanho da população das cidades brasileiras e apresentada na tabela 4. Os resultados sugerem que a distribuição ergótica é mais concentrada nas menores cidades, particularmente na primeira e segunda classe. Percebe-se, não obstante, que há estabilidade na classe mais elevada. Estes resultados sugerem convergência para as classes inferiores, mas, mantendo uma pequena concentração na classe superior, ou seja, as maiores cidades.

Tabela 4 Distribuição inicial versus distribuição ergótica da população das cidades brasileiras:

	1910-2010					
	1	2	3	4	5	6
	<20%	<50%	<80%	<135%	<185%	>185%
Distribuição Inicial	0.4748	0.3846	0.0626	0.0510	0.0244	0.0476
Distribuição Ergótica	0.4567	0.3834	0.0615	0.0499	0.0232	0.0476

Fonte: Elaborado a partir dos dados do IPEADATA. Elaboração do autor.

Como dito anteriormente, a distribuição ergótica pode ser afetada pela forma que se divide a distribuição em classes. Testaram-se outros tipos de estratificação e diferentes números de classes da distribuição das cidades. Os resultados de ligeira convergência, no entanto, foram robustos a estas escolhas.

#### 4. Conclusões

O processo de crescimento das cidades brasileiras nos últimos cem anos foi movido pela industrialização e pela expansão da fronteira agrícola. A primeira fortemente concentrada inicialmente na Região Sudeste e nas últimas décadas com indícios de desconcentração. Já a fronteira agrícola vai deste a Região Norte e Centro-Oeste com a ocupação do cerrado incluindo recentemente regiões do Nordeste. Neste estudo, utilizaram-se as AMC do IPEA para analisar a dinâmica de crescimento das

idades brasileiras no período 1910-2010. A vantagem deste banco de dados é que abrange o maior período disponível para o Brasil neste nível de agregação. Outra contribuição é que os estudos que analisam a ordem de tamanho das cidades brasileiras apontam que o expoente de Pareto é sensível ao tamanho das cidades escolhidas *ad hoc* pelo pesquisador. Desta forma, evita-se este problema. Trabalhou-se, portanto com 431 observações deixando de fora apenas o território de Fernando de Noronha.

A função de densidade da distribuição das cidades brasileiras nos últimos cem anos mostra divergência, mas com ligeira indicação de formação de uma distribuição bimodal na última década. A Lei de Zipf corrobora com o processo de divergência, mas apresentando uma perda de força na última década analisada. Ou seja, as maiores cidades crescem mais rapidamente que as menores cidades. Porém na última década há indicação que as cidades de porte médio dão sinais de crescimento mais acelerado.

Os efeitos de vizinhança mostraram ser importantes na análise da distribuição das cidades brasileiras tendo em vista que os testes apontaram dependência espacial na distribuição. Desta forma, a lei de Zipf foi analisada apreendendo o efeito espacial relevando a contribuição deste artigo nesta literatura.

Finalmente a análise de Markov Chain mostrou pouca mobilidade e alta persistência. Ou seja, a probabilidade de uma cidade mudar de uma faixa para outra entre décadas sucessivas é muito baixa. Por outro lado, a probabilidade das cidades permanecerem na mesma faixa da distribuição do tamanho das cidades é muito alta ao longo do período 1910-2010. Apenas as cidades situadas no meio da distribuição conseguem passar para classes mais elevadas. Estes resultados foram robustos à forma utilizada para divisão da distribuição em classes.

### Referências

AMEMIYA, T. (1985) “**Advanced econometrics**”. Harvard University Press, Cambridge.

ANSELIN, L. (2002) “Under the hood: Issues in the specification and interpretation of spatial regression models”. **Agricultural Economics**, 27, 247–267.

BICKENBACH, F., BODE, E. (2003) “Evolution the Markov property in studies of economic convergence”. **Int Reg Sci Rev**, p 363-392.

BLACK, D., HENDERSON, V. (2003) “Urban evolution in the USA”. **J. Econ Geogr**, 3, p 343-372.

- BULLI, S. D. (2001) "Distribution dynamics and cross-country convergence: a new approach". *Scott J Polit Econ*, 48, p 226-246.
- CHI, G. and ZHU, J. (2008) "Spatial regression models for demographic analysis". *Popul. Res. Policy Rev*, 027, p. 17 -42.
- EATON, J., ECKSTEIN, Z. (1997) "City and growth: theory and evidence from France and Japan". *Reg. Sci Urban Econ*, 27, p 443-474.
- DURATON, G. (2006) "Some foundation for Zipf's law: product proliferation and local spillovers". *Reg Sci Urban Econ*, 36, p 542-563.
- EDWARDS, M.E.(2007) "**Regional and Urban economics development: theory and methods**". Auerbach Publications, New York.
- FUJITA, M., KRUGMAN, P., VENABLES, A.J. (2001) "**The Spatial Economy: Cities, Regions, and International Trade**". The MIT Press.
- GROSSMAN, G., HELPAN, E.(1991) "**Innovation and growth in the world economy**". MIT Press, Cambridge.
- HAMILTON, J.D. (1994) "**Time series analysis**". Princeton University Press. Princeton.
- JUSTO, E.R, FERREIRA, R de A., LIMA, C.F., MARTINS, G.N. (2010) "Os determinantes da migração e da migração de retorno intermunicipal no Brasil". **In: Anais do XXVIII Encontro nacional de Economia: ANPEC**, Salvador.
- JUSTO, W.R. (2007) "A Lei de ZIPF e a Lei de GIBRAT: O que os dados têm a dizer sobre as cidades brasileiras". *Scientia Una* (Olinda), v. 8, p. 78-97.
- KEMENY, J., SNELL, L. (1976) "**Finite Markov Chain**". Springer, New York.
- KRUGMAN, P. (1996) "**The self-organizing economy**". Blackwell, Cambridge.
- LANAPA, L., PERDIGUERO, A.M., SNAZ, F. (2004) "La distribución del tamaño de las ciudades en España, 1900-1999". *Revista de Economía Aplicada*, 34,p 5-16.
- LE GALLO, J., CHASCO, C. (2009) "Spatial analysis of urban growth in Spain, 1900-2001". **In: GIUSEPPE, A., BALTAGI, B.H.(Eds). Spatial Econometrics: Methods and Applications**. Physical-Verlag Heidelberg. Heidelberg, Germany, p 59-80.
- LeSAGE, J. e PACE, R.K.(2009) "**Introduction to spatial econometrics**". New York: Champain & Hall/CRC.
- MELLA, J.M., CHASCO, C. (2006) "A spatial econometric analysis of urban growth and territorial dynamics: a case study on Spain". **In: NINKAMP P, REGGIANI A. (eds) Spatial evolution and modeling**. Edward Elgar, pp. 319 -360.
- OLIVEIRA, C. De A. (2005) "A evolução da distribuição do tamanho das cidades

brasileiras: 1936-2000”. Passo Fundo, Texto para Discussão..

OVERMAN, H.G., IOANNIDES, Y. M.(2001) “Cross-sectional evolution of the US city size distribution”. **J Urban Econ**, 49,p 543-566.

PARETO, V.(1897) “**Cours d’Economie politique**”. Rouge et Cie. Paris.

PACHECO, A. I., and TYRRELL, T. J.(2002) “Testing spatial patterns and growth spillover effects in clusters of cities”. **Journal of Geographical Systems**, 4, p.275–285.

VOSS, P. R., and CHI, G. (2006) “Highways and population change”. **Rural Sociology**, 71(1), p.33–58.

ZIPF, G.K. (1949) “**Human behavior and the principle of least effort**”. Addison-Wesley. Cambridge.