



O FATOR “AGORA É LULA” NA ELEIÇÃO PRESIDENCIAL DE 2002

Eduardo Simões de Almeida
Fernando Salgueiro Perobelli
Pedro Guilherme Costa Ferreira
Weslem Rodrigues Faria

TD. Mestrado em Economia Aplicada
FEA/UFJF 001/2007

Juiz de Fora
2007

O FATOR “AGORA É LULA” NA ELEIÇÃO PRESIDENCIAL DE 2002

Eduardo Simões de Almeida

Professor Adjunto da UFJF e Pesquisador do CNPq

eduardo.almeida@ufjf.edu.br

Fernando Salgueiro Perobelli

Professor Adjunto da UFJF e Pesquisador do CNPq

fernando.perobelli@ufjf.edu.br

Pedro Guilherme Costa Ferreira

Mestrando em Economia Aplicada da UFJF

pgcferreira@gmail.com

Weslem Rodrigues Faria

Mestrando em Economia do Cedeplar/UFGM

weslemrodrigues@click21.com.br

O FATOR “AGORA É LULA” NA ELEIÇÃO PRESIDENCIAL DE 2002

RESUMO

Após três derrotas em eleições presidenciais, o candidato do partido dos trabalhadores venceu finalmente em 2002. O principal interesse deste trabalho é quantificar a importância do fator “Agora é Lula” nas eleições presidenciais de 2002. Esse fator transcende aspectos sócio-econômicos, político e ideológicos. Para identificar tal fator, procedeu-se a quantificação dos outros fatores determinantes para a vitória do candidato Lula em nível municipal. Para fazer isso, metodologicamente, extraíram-se quatro determinantes por meio da análise fatorial, a saber, o fator sócio-econômico, o fator demográfico, o fator de vulnerabilidade à violência e o fator político. A votação do candidato Lula por município foi regredida contra esses fatores como variáveis explicativas, controlando-se para os efeitos espaciais. Os resultados obtidos mostram que a maior parte da explicação para a votação do candidato do PT deveu-se ao fator “Agora é Lula”, responsável por 53% desta explicação. Ademais, Lula teve mais votos nos municípios mais desenvolvidos em termos sócio-econômicos. Por fim, os efeitos espaciais mostraram-se muito importantes para se analisar a votação do candidato Lula em 2002.

Palavras-chave: eleições presidenciais, análise fatorial, econometria espacial, autocorrelação espacial.

ABSTRACT

After three defeats in presidential elections, the candidate of party of workers won in 2001 at last. The main interest of this paper is to quantify the relevance of the “It’s Lula’s Time” factor in Brazilian presidential pools. This factor transcends socio-economic and political-ideological features. To identify such a factor, the other determinant factors were quantified to explain the victory of the candidate Lula at the municipality level. To do this, four determinants were extracted using factor analysis, namely, socio-economic factor, demographic factor, vulnerability factor and political factor. The votes for the candidate Lula at the municipality were regressed on those factors as explanatory variables, controlling for spatial effects. The findings show the majority of the explanation of votes for Lula is due to “It’s Lula’s time” factor, which stands for 53% of that explanation. Furthermore, Lula had more votes at the more developed municipalities in socio-economic terms. At last, the spatial effects matter when analysing the votes for the candidate Lula in 2002 presidential elections.

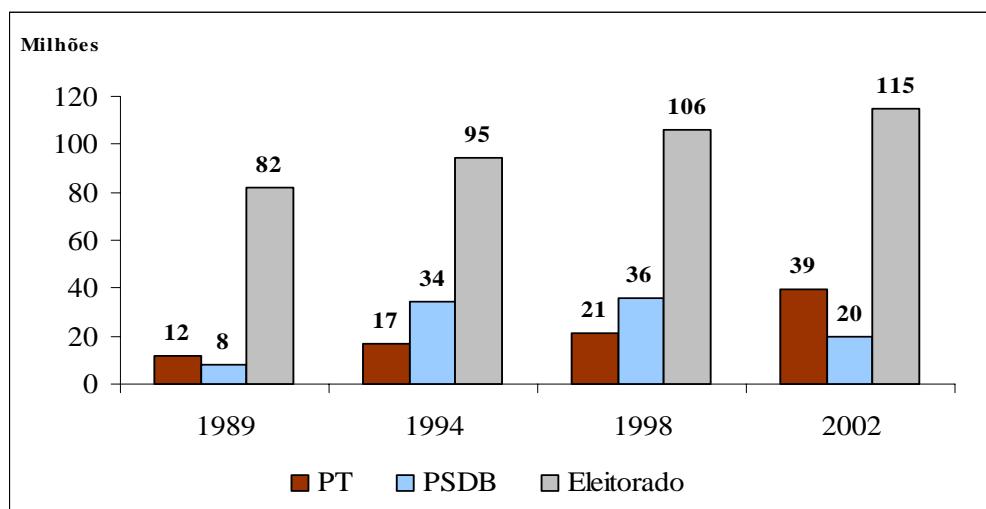
Key words: presidential elections, factor analysis, spatial econometrics, spatial autocorrelation, spatial heterogeneity.

Classificação JEL: R10; C21.

1. INTRODUÇÃO

Antes de ser eleito, Luis Ignácio Lula da Silva foi derrotado à presidência da república em três ocasiões: as eleições de 1989, 1994 e 1998. Entretanto, as eleições presidenciais de 2002 brindariam, finalmente, Lula com a vitória. O gráfico 1 retrata a evolução dos votos para os candidatos a presidência do Partido dos Trabalhadores (PT) e do seu grande rival político, o Partido da Social-Democracia Brasileira (PSDB), no primeiro turno das eleições referentes aos anos de 1989, 1994, 1998 e 2002.

Figura 1. Evolução dos votos para candidatos à presidência do PT e PSDB: 1989-2002 (em milhões de votos)



Fonte: IPEADATA

O candidato à presidência do PT foi Luís Inácio Lula da Silva (Lula) em todos os anos, dos quais ficou à frente do candidato do PSDB no primeiro turno em duas oportunidades (1989 e 2002). O PSDB, por sua vez, apresentou crescimento significativo de votos de 1989 para 1994 (341%). As eleições de 2002 representaram uma reversão da preferência do eleitorado em relação aos anos de 1994 e 1998. Lula teve crescimento de mais de 83% na sua votação, enquanto que o PSDB apresentou redução de mais de 45%.

A eleição de 1998 para presidente foi marcada por uma redução do debate sobre as questões públicas, tendo em vista que as campanhas se mostravam tímidas com a eliminação de notícias eleitorais por parte importante da imprensa. Além disso, o relativo sucesso do Plano Real desde sua aplicação em 1994, dentre outros fatores, ajudou o candidato Fernando Henrique Cardoso a se reeleger em 1998 (Miguel, 2003).

O cenário para as eleições de 2002, no entanto, configurou-se de forma bastante diferente. Grande parte da população se mostrava impaciente com os resultados da economia, como o baixo crescimento econômico e o alto índice de desemprego, bem como pelas medidas de baixa popularidade do então presidente em exercício, Fernando Henrique Cardoso, como a realização de privatizações, a firmação de acordos com organismos financeiros internacionais e a tentativa de aumentar a flexibilização das leis trabalhistas.

Após várias tentativas, Lula venceu o pleito presidencial de 2002. Afinal, quais foram os fatores determinantes para a vitória de Lula em 2002? A sabedoria convencional afirma que os municípios menos desenvolvidos votaram em Lula, em 2002, na esperança de que a chegada à presidência de uma pessoa que enfatizava em seus discursos questões de pobreza, desigualdade e emprego pudesse, de fato, implementar mudanças reais. O presente trabalho mostrará que isso efetivamente não ocorreu. Lula conseguiu os maiores percentuais de votos em municípios mais desenvolvidos. E, mesmo assim, o componente sócio-econômico não foi o principal fator determinante da vitória do candidato do Partido dos Trabalhadores.

O determinante predominante foi o chamado fator “Agora é Lula”, que transcendeu a aspectos sócio-econômicos, demográficos e mesmo político-ideológicos. Esse fator é constituído por uma miscelânea de aspectos intangíveis e não-observáveis, que deram lugar especificamente nas eleições de 2002. Em primeiro lugar, encontra-se o inegável carisma pessoal de Lula, capaz de galvanizar a atenção do público. Em segundo lugar, destaca-se a persistência do candidato que lhe legou uma espécie de “*top of mind*” eleitoral: quando se questionava, em pesquisas de intenção de voto, sobre possíveis candidatos à presidência, vinha à mente de qualquer eleitor o nome de Lula. Em terceiro lugar, o fator “Agora é Lula” foi alimentado pelo tom despolitizante empreendido à sua campanha eleitoral de 2002, consagrada pelo mote “Lulinha paz e amor”. Por último, em função de seguidas derrotas em pleitos eleitorais anteriores, havia a sensação difusa e generalizada no eleitorado de que tinha chegado, finalmente, a vez de Lula chegar à presidência. Assim sendo, o principal interesse deste trabalho é procurar identificar e quantificar a importância desse fator “Agora é Lula” na eleição do candidato à presidente do PT em 2002.

Outro aspecto importante analisado aqui é a influência dos efeitos espaciais para a intenção de voto do candidato Lula em 2002. A votação de Lula exibiu uma forte dependência espacial e uma elevada heterogeneidade espacial, que se desconsideradas na análise, representariam sérias implicações à inferência estatística e à modelagem econométrica, com graves prejuízos sobre os parâmetros estimados em termos de consistência e de eficiência. Mostra-se, assim, neste trabalho que a discussão das intenções de voto de Lula não pode estar divorciada do espaço em virtude da ocorrência de um processo de difusão

de informação eleitoral, que é espacial e regionalmente circunscrita, e do efeito de vizinhança na determinação da intenção de voto.

Além disso, nossa análise quantitativa reserva outros resultados surpreendentes. A alegação de que Lula ganhou a despeito do PT não se confirma. Os resultados confirmam que o fator político-ideológico foi muito relevante para a votação do candidato à presidente do PT. Isso confirma identidade ideológica do PT, como um partido de base de militantes e com uma consistência ideológica muito acima da média dos outros partidos do espectro político brasileiro. Esse fator constituiu-se num elemento muito relevante na explicação das intenções de voto de Lula, contribuindo sobremaneira para a sua eleição.

Para mostrar todos esses aspectos, é preciso adotar um conjunto de três métodos quantitativos interligados entre si. A importância do espaço será investigada pela análise exploratória de dados espaciais (AEDE) e técnicas econométrico-espaciais, um ferramental ainda pouco usado nos trabalhos aplicados no Brasil quando se usa dados regionais, apesar de sua extrema utilidade.

Por sua vez, os determinantes da votação de Lula serão extraídos da análise fatorial dentre um denso conjunto de potenciais variáveis explicativas para a sua intenção de voto em nível municipal. Evidentemente, esses determinantes serão, posteriormente, introduzidos no lado direito das regressões para explicarem a votação de Lula, controlando-se devidamente para os efeitos espaciais.

Foi adotada a abordagem da análise fatorial para contornar o surgimento de dois problemas. Em primeiro lugar, devido à carência de análises mais quantitativas, a teoria política não sugere um conjunto de variáveis que poderia explicar a votação eleitoral de um candidato. Em segundo lugar, a colocação de variáveis altamente correlacionadas entre si provocaria o surgimento do problema da multicolinearidade. Em virtude desses dois problemas, o método da análise fatorial será adotado a fim de extrair os fatores determinantes das votações dentre um conjunto de 58 variáveis potencialmente explicativas. Por intermédio desse método, será possível sumarizar em fatores um conjunto de quase sessenta variáveis, supostamente relevantes para explicar a votação dos candidatos a presidente em nível municipal.

Após a extração dos fatores, adotar-se-ão as técnicas da econometria espacial para regredir as votações de cada um dos candidatos contra os mesmos fatores, controlando-se para os efeitos espaciais, ou seja, a dependência espacial e a heterogeneidade espacial. Foram estimados o modelo de defasagem espacial e o modelo de erro espacial, tradicionais na literatura de econometria espacial, além do modelo de defasagem espacial com erro espacial, que ainda raramente é encontrado na literatura por suas dificuldades de identificação de parâmetros e de estimação. Para estimar esse último modelo foi adotado o método de

mínimos quadrados em dois estágios espacial generalizado de Kelejian e Prucha (1998). Estimou-se, também, um modelo de regimes espaciais com controle para a dependência espacial.

O trabalho está organizado em três seções, além desta de natureza eminentemente introdutória. Os dados usados no estudo são apresentados e discutidos na próxima seção. Ainda nessa parte do trabalho, é mostrado que a hipótese de aleatoriedade espacial nos dados de votação do Lula é claramente rejeitada pelos tradicionais testes de autocorrelação espacial, bem como todos os passos metodológicos necessários para a extração dos fatores determinantes. A terceira seção é reservada para a apresentação da especificação do modelo econométrico espacial adotado, as regressões estimadas, os resultados obtidos e sua subsequente discussão. Na derradeira seção, são tecidas as considerações finais e recuperadas as conclusões de relevo surgidas ao longo do trabalho.

2. DADOS

2.1. Base de Dados

O presente trabalho faz uso de dados, em nível municipal, sobre eleições presidenciais do primeiro turno disponíveis no sítio www.ipeadata.gov.br, cuja fonte primária é o Tribunal Superior Eleitoral para o ano de 2002.¹ Foram usados os dados da votação do primeiro turno, pois os eleitores têm mais opções no espectro político para fazerem a sua escolha, refletindo suas preferências com mais precisão e nitidez político-ideológica. A eleição presidencial de 2002 do segundo turno assumiu um caráter mais plebiscitário, em que o governo de plantão da época, depois de oito anos no poder, foi avaliado. É importante observar que os partidos que lançaram candidatos realizaram coligações de acordo com a regulamentação da Justiça Eleitoral, isto é, em 2002, houve verticalização eleitoral.²

A variável dependente é a votação percentual no primeiro turno do candidato à presidência do PT no município. Além disso, o trabalho utiliza um conjunto de 58 variáveis para a extração dos fatores. Essas variáveis foram coletadas no Atlas de Desenvolvimento Humano no Brasil (PNDU, 2006) para os municípios brasileiros referente ao ano de 2000 e no sítio ipeadata.gov.br. A relação das variáveis potencialmente explicativas está no anexo deste trabalho.

¹ A lista de partidos que disputaram as eleições presidenciais de 2002 é apresentada a seguir, bem como os nomes dos candidatos que os representaram: a) PCO – Partido da Causa Operária: Rui Costa Pimenta; b) PPS – Partido Popular Socialista: Ciro Ferreira Gomes; c) PSB – Partido Socialista Brasileiro: Anthony William Garotinho Matheus de Oliveira; d) PSDB – Partido da Social Democracia Brasileira: José Serra; e) PSTU – Partido Socialista dos Trabalhadores Unificado: José Maria de Almeida e f) PT – Partido dos Trabalhadores: Luís Inácio Lula da Silva.

² Nas eleições de 2002, tanto o PCO quanto o PSTU lançaram candidatura sem realizar coligações, ao contrário dos outros partidos. O PPS coligou ao PTB e ao PDT, o PSB coligou com o PGT e com o PTC, o PSDB coligou apenas com o PMDB. O PT foi o partido com maior coligação: PL, PC do B, PMN e PCB.

2.2. Análise Exploratória Espacial da Variável Dependente

O primeiro passo a ser dado para averiguar a presença de autocorrelação espacial nos dados da votação de Lula está em analisar o índice I de Moran. Esta estatística fornece a indicação formal do grau de associação linear entre os vetores de valores observados no tempo t (z_t) e a média ponderada dos valores da vizinhança, ou as defasagens espaciais (Wz_t). Segundo Cliff e Ord (1981), em termos formais, a estatística I de Moran pode ser expressa como:

$$I_t = \left(\frac{n}{S_o} \right) \left(\frac{\bar{z}_t' W z_t}{\bar{z}_t' \bar{z}_t} \right) \quad t = 1, \dots, n \quad (1)$$

em que z_t é o vetor de n observações para o ano t na forma de desvio em relação à média. W é a matriz de pesos espaciais: os elementos w_{ii} na diagonal são iguais a zero, enquanto que os elementos w_{ij} indicam a forma como a região i está espacialmente conectada com a região j . O termo S_o é um escalar igual à soma de todos os elementos de W .

Quando a matriz de pesos espaciais é normalizada na linha, isto é, quando os elementos de cada linha somam um, a expressão (1) será da seguinte forma:

$$I_t = \left(\frac{\bar{z}_t' W z_t}{\bar{z}_t' \bar{z}_t} \right) \quad t = 1, \dots, n \quad (2)$$

Valores de I maiores (ou menores) do que o valor esperado $E(I) = -1/(n-1)$ significam que há autocorrelação positiva (ou negativa). Por exemplo, uma indicação de autocorrelação espacial positiva significa dizer que municípios que apresentam elevado percentual de votos para o candidato do PT são vizinhos de outros municípios que também apresentam elevado percentual de votos para este mesmo partido ou, alternativamente, que municípios com pequeno número de votos para determinado partido são circundados por outros municípios que também apresentam pequeno número de votos para o mesmo partido.

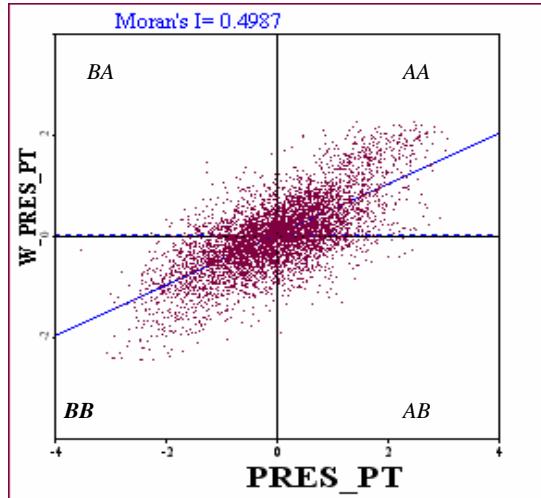
Segundo Anselin (1996), o diagrama de dispersão de Moran é uma das formas de interpretar a estatística I de Moran. Desse modo, trata-se de uma representação do coeficiente de regressão e permite visualizar a correlação linear entre z e Wz por meio do gráfico de duas variáveis. No caso específico da estatística I de

Moran tem-se o gráfico de Wz e z . O coeficiente I de Moran será a inclinação da curva de regressão de Wz contra z e indicará o seu grau de ajustamento.

O diagrama de dispersão de Moran (Figura 2) apresenta no eixo horizontal o percentual de votos que o partido dos trabalhadores obteve para presidente em determinado município para 2002 (a variável de interesse) respectivamente e, na vertical, a média do percentual de votos do PT nos municípios considerados vizinhos.

Outra vantagem do diagrama de dispersão de Moran é poder classificar os municípios de acordo com quatro diferentes agrupamentos de dados conforme o quadrante do diagrama. Esses agrupamentos podem ser: baixo-alto (BA), que indica que municípios que apresentam pequeno percentual de votos para o candidato Lula são vizinhos de municípios que apresentam um grande percentual de votos para esse candidato; o agrupamento alto-alto (AA) denota que municípios que apresentam grande percentual de votos para o candidato do PT são vizinhos de municípios que também apresentam um grande percentual de votos; o agrupamento alto-baixo (AB) informa que municípios que apresentam grande percentual de votos para o PT são vizinhos de municípios que apresentam pequeno percentual de votos; e, finalmente, o agrupamento baixo-baixo (BB) implica que municípios que apresentam pequeno percentual de votos para determinado partido são vizinhos de outros municípios que apresentam pequeno percentual de votos.

Figura 2
Diagrama de Dispersão de Moran dos Votos PT para Presidente em 2002



Fonte: elaboração própria dos autores

O diagrama acima indica que a grande parte dos municípios brasileiros está localizada nos quadrantes AA e BB na votação do candidato do PT. O *I* de Moran, por sua vez, apresenta valores positivos e significativos estatisticamente, indicando a presença de autocorrelação espacial positiva de 0,50.

A estatística global do *I* de Moran, por sua vez, pode esconder padrões locais de autocorrelação espacial. A fim de observar a existência de *clusters* espaciais locais de valores altos ou baixos e quais as regiões que mais contribuem para a existência de autocorrelação espacial, deve-se implementar as medidas de autocorrelação espacial local.

Segundo Anselin (1995), um “*Local Indicator of Spatial Association (LISA)*” será qualquer estatística que satisfaça a dois critérios: a) um indicador LISA deve possuir, para cada observação, uma indicação de *clusters* espaciais significantes de valores similares em torno da observação (*e.g.* região) e b) o somatório dos LISAs, para todas as regiões, é proporcional ao indicador de autocorrelação espacial global.

Segundo Le Gallo e Ertur (2003), a estatística LISA, baseada no *I* de Moran local, pode ser especificada da seguinte forma³:

$$I_{i,t} = \frac{(x_{i,t} - \mu_t)}{m_o} \sum_j w_{ij} (x_{j,t} - \mu_t) \text{ com } m_o = \frac{(x_{i,t} - \mu_t)^2}{n} \quad (3)$$

onde $x_{i,t}$ é a observação de uma variável de interesse na região i para o ano t , μ_t é a média das observações entre as regiões no ano t para a qual o somatório em relação a j é tal que somente os valores vizinhos de j são incluídos.

A estatística pode ser interpretada da seguinte forma: valores positivos de $I_{i,t}$ indicam que existem *clusters* espaciais com valores similares (alto ou baixo), ao passo que valores negativos significam que existem *clusters* espaciais com valores diferentes (dissimilares) entre a região e seus vizinhos.

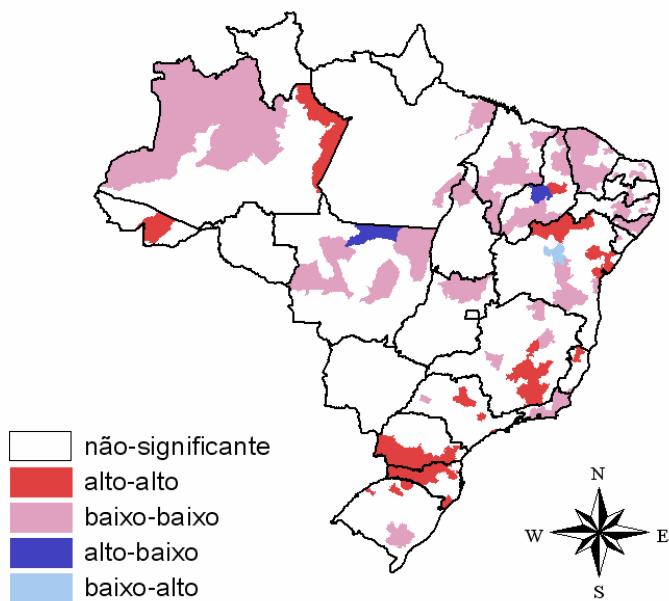
De acordo com Anselin (1995), a estatística LISA é utilizada para medir a hipótese nula de ausência de associação espacial local. É importante salientar que, assim como a distribuição para as estatísticas globais, a distribuição genérica para a estatística LISA também é de difícil apuração. Portanto, para

³ Existem outras estatísticas do tipo LISA na literatura, além do *I* de Moran local, tais como o *c* de Geary local e o Gama local. Para maiores detalhes a respeito disso, consulte Anselin (1995).

solucionar tal problema, deve-se trabalhar com resultados assintóticos. Logo, a alternativa é a utilização de uma aleatorização condicional ou uma permutação que permita auferir pseudoníveis de significância.⁴

Os indicadores locais de associação espacial (LISA) da votação do candidato do PT à presidência da república no ano de 2002 para os municípios do Brasil, estão representados no mapa de *clusters* espaciais (Figura 3), para um nível de significância de 5 %.

Figura 3
Mapa de *Clusters* para a Votação do PT para Presidente em 2002



Fonte: elaboração própria dos autores

Da observação do mapa de clusters espaciais (figura 3), destacam-se os *clusters* de cor rosa que demonstram os municípios que apresentaram baixa votação para este candidato e que são rodeados por municípios que tiveram baixa votação. Tal candidato apresentou votação mais significativa nas Regiões Sul e Sudeste, sendo que *clusters* espaciais alto-alto de coloração vermelha indicam este resultado.

⁴ Os níveis de significância das distribuições marginais serão aproximados por intermédio das desigualdades de Bonferroni ou por meio da estrutura proposta por Sidák (1967) *apud in* Anselin (1995). “*This means that when the overall significance associated with the multiple comparisons (correlated tests) is set to α , and there are m comparisons, then the individual significance α_i should be set to either α/m (Bonferroni) or $1 - (1 - \alpha)^{1/m}$ (Sidák)*” (Anselin, 1995 pp. 96). Neste trabalho, a análise de Bonferroni será utilizada.

Em suma, esta seção mostrou como os dados de votação do candidato Lula estão fortemente autocorrelacionados no espaço, tanto global quanto localmente.

3. OBTENÇÃO DOS FATORES DETERMINANTES

Ao se dispor de observações de diversas variáveis que podem estar vinculadas à intenção de voto de um candidato, obtidas em uma mesma unidade amostral (neste estudo em particular, municípios), é possível que a interpretação e análise dos dados tornem-se uma tarefa complexa, sendo desejável, pois, uma simplificação dos mesmos. Não obstante, deve-se conservar o máximo de informações fornecidas pelas variáveis originais. Visando contornar tal questão e objetivando reduzir a dimensão do vetor de dados, utilizam-se técnicas de sumarização de informações.

O método de análise fatorial é uma técnica estatística multivariada de sumarização usada para identificar um número reduzido de padrões de características (chamados fatores), relativos a um conjunto de variáveis correlacionadas entre si. Segundo Gontijo (1998) e Schilderink (1978), esse método tem por objetivo reunir os dados empíricos não-ordenados das variáveis por intermédio de uma combinação linear, a fim de que:

- a) Um número menor de variáveis (ou fatores) seja obtido a partir das variáveis escolhidas (matriz dos dados originais), sem perda de informações, que serão reproduzidas de maneira resumida no modelo final;
- b) A obtenção dos fatores permita a reprodução de padrões de relações separadas entre grupos de variáveis;
- c) Cada padrão de relações tenha condições de ser interpretado de maneira lógica.

Para uma dada variável “ X_i ” padronizada, o modelo é, de forma geral, escrito como (Norusis, 1994):

$$X_i = A_{i1}F_1 + A_{i2}F_2 + \dots + A_{ik}F_k + U_i + E_i \quad (4)$$

no qual:

A_{ik} – cargas fatoriais, usadas para combinar linearmente os fatores comuns.

F_k – fatores comuns ($1 \leq k \leq N$).

U_i – fator único ($i = 1, 2, \dots, N$).

E_i – fator de erro

N – número de variáveis.

O modelo matemático de análise factorial é, em alto grau, similar ao modelo de regressão múltipla. A principal diferença está no fato de que as variáveis independentes da equação de regressão múltipla têm suas observações conhecidas no momento da estimativa, enquanto que no modelo de análise factorial isto não se verifica para os fatores. Na análise factorial, os “ F_k ’s” fatores, utilizados na caracterização de um conjunto de variáveis, são inferidos das variáveis observadas e podem ser estimados como combinações lineares das mesmas.

Os passos a serem seguidos para a efetivação do método de análise factorial são, a saber:

- i) A montagem da matriz de correlação para todas as variáveis. A adequação do modelo é, portanto, avaliada neste passo;
- ii) A extração dos fatores iniciais. Definem-se as cargas fatoriais de cada variável, raiz característica associada aos respectivos fatores e as proporções da variância total do conjunto de variáveis;
- iii) A rotação ou transformação dos fatores, que é realizada com o objetivo de definir de maneira mais apropriada as relações entre as variáveis e os fatores;
- iv) O cálculo dos escores fatoriais para cada observação, que podem ser utilizados em uma ampla variedade de análises.

De acordo com Queiroz (1984), a análise de fatores consiste na substituição de “ N ” variáveis aleatórias por um número menor de variáveis hipotéticas denominadas fatores, considerando para isso as relações internas apresentadas pelo conjunto de variáveis. Esses fatores podem ser correlacionados (oblíquos) ou não correlacionados (ortogonais) e devem explicar parte significativa da variância apresentada pelo conjunto das variáveis originais. As variáveis dos grupos escolhidos devem estar bem correlacionadas entre si para que se extraiam fatores comuns significativos.

O índice Kaiser-Meyer-Olkin (KMO) permite verificar a consistência dos dados originais, comparando as magnitudes dos coeficientes de correlação “ r_{ij} ” observados aos coeficientes de correlação parcial “ a_{ij} ” (Norusis, 1994). O índice é computado por:

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum a_{ij}^2} \quad (5)$$

Já o índice Bartlett’s Test of Sphericity (BTS), por sua vez, é usado para testar a hipótese de a matriz de correlação ser uma matriz de identidade. Se a matriz de correlação não for uma identidade, o valor da

estatística será alto e o nível de significância associado baixo, de forma que o uso do modelo de análise fatorial será apropriado.

Aplicando a análise fatorial para estudar os determinantes de votação do candidato do PT foram escolhidas 26 variáveis dentro de um conjunto de 58 variáveis potencialmente explicativas (veja a relação delas no anexo). Essas 26 variáveis podem ser classificadas em grupos com estreita relação com fatores sócio-econômicos, com fatores de vulnerabilidade, com questões políticas e com fatores demográficos. Após a utilização da análise fatorial tais variáveis foram agrupadas em fatores de acordo com o tipo de grupo com os quais estavam mais correlacionadas, obtendo-se um total de 4 fatores (ver tabela 1).

Tabela 1. Descrição das Variáveis Utilizadas no Modelo

Indicador	Descrição	Unidade/Año	Fator
X001	Índice de Desenvolvimento Humano (IDH)	2000	F1
X002	Taxa de Alfabetização	2000	F1
X003	Renda per capita	2000	F1
X004	Média de anos de estudo das pessoas de 25 anos ou mais de estudo	2000	F1
X005	Percentual de pessoas que vivem em domicílios com água encanada	2000	F1
X006	Percentual de pessoas que vivem em domicílios com carro	2000	F1
X007	Percentual de pessoas que vivem em domicílios com energia elétrica e TV	2000	F1
X008	Percentual de pessoas que vivem em domicílios urbanos com serviço de coleta de lixo	2000	F1
X009	Percentual de pessoas que vivem em domicílios com telefone	2000	F1
X010	Número de médicos residentes por mil habitantes	2000	F1
X011	Índice de Desenvolvimento Humano Municipal - Renda	2000	F1
X012	Índice de Desenvolvimento Humano - Longevidade	2000	F1
X013	Índice de Desenvolvimento Humano – Educação	2000	F1
X014	População total	2000	F2
X015	População Urbana	2000	F2
X016	Mulheres de 25 anos ou mais de idade	2000	F2
X017	População de 25 anos ou mais de idade	2000	F2
X018	Taxa de Homicídios	1998	F3
X019	Taxa de Homicídios	1999	F3
X020	Taxa de Homicídios	2000	F3
X021	Taxa de Homicídios	2001	F3
X022	Taxa de Homicídios	2002	F3
X023	Votação percentual do candidato à presidência do PT por município	1998	F4
X024	Votação percentual do candidato à presidência do PSDB por município	1998	F4
X025	Votação percentual para deputado federal da coligação do PT nas eleições	2002	F4
X026	Votação percentual para deputado federal da coligação do PSDB nas eleições	2002	F4

Fonte: resultados da pesquisa.

Os quatro fatores que possuem raiz unitária acima da unidade e explicam conjuntamente 74,47% da variância total do modelo. O fator 1, referente às condições sócio-econômicas, explica cerca de 37% da variância total como pode ser verificado na Tabela 2.

Tabela 2. Valores da Raiz Característica e Percentagem da Variância Total Explicada pelos Quatro Fatores

Fator	Raiz Característica	% da variância	% acumulado
1	9,723	37,396	37,396
2	4,108	15,800	53,196
3	3,187	12,258	65,454
4	2,343	9,013	74,467

Fonte: resultados da pesquisa.

Tabela 3. Resultados dos Testes de KMO e BTS

Kaiser-Meyer-Olklin	0,823
Bartlett's Test of Sphericity	305666,7
Nível de significância	0,000

Fonte: resultados da pesquisa.

A Tabela 3 apresenta os valores dos testes de Kaiser-Meyer-Olkin, do Bartlett's Test of Sphericity e o nível de significância. Pelo KMO obteve-se o valor 0,823, índice considerado muito bom, o que indica que os dados originais são consistentes. Quanto ao BTS, verifica-se que, como este teste apresentou um valor elevado e o nível de significância foi desprezível, é improvável que a matriz de correlação seja uma identidade.

Tabela 4. Cargas Fatoriais

Variáveis	Fatores			
	1	2	3	4
X001	0,971			
X002	0,915			
X003	0,909			
X004	0,909			
X005	0,906			
X006	0,922			
X007	0,831			
X008	0,692			
X009	0,830			
X010	0,434			
X011	0,959			
X012	0,844			
X013	0,907			
X014		0,987		
X015		0,986		
X016		0,988		
X017		0,988		
X018			0,758	
X019			0,762	
X020			0,777	
X021			0,800	
X022			0,785	
X023				0,847
X024				-0,788
X025				0,677
X026				-0,633

Fonte: resultados da pesquisa.

A Tabela 4 mostra as cargas fatoriais, as quais determinam a associação entre cada variável e os fatores.

Ao se fazer uma análise mais detalhada dos fatores, verifica-se que:

- a) **Fator 1 (F1)** – é identificado como inerente aos aspectos sócio-econômicos e representa 37,4% da variância total do modelo. Todas as variáveis estão positivamente relacionadas com esse fator sócio-econômico e apresenta, em sua grande maioria, alta correlação com o mesmo.
- b) **Fator 2 (F2)** – é o fator demográfico, sendo responsável por 15,8% da variância total. As quatro variáveis que o formam estão altamente correlacionadas, ou seja, acima de 0,98.
- c) **Fator 3 (F3)** – é aquele que caracteriza o grau de vulnerabilidade do município à violência. As variáveis estão positivamente relacionadas com o fator de vulnerabilidade.
- d) **Fator 4 (F4)** – está associado às características políticas e representa 9,0% da variância total do modelo.

Assim sendo, foi possível classificar os municípios de acordo com os quatro fatores anteriormente mencionados. A análise fatorial permitiu construir fatores consistentes e que congregam os quatro principais fatores encontrados na literatura que explicam o resultado das intenções de votos em eleições, ou seja, questões econômicas, sociais, políticas, demográficas e de vulnerabilidade à violência.

4. RESULTADOS E DISCUSSÃO

Adotando a metodologia da econometria espacial (Anselin, 1988), os fatores extraídos da análise fatorial da seção anterior serão incluídos como variáveis independentes nas regressões para explicar a votação do candidato do PT nas últimas eleições presidenciais no país (a variável dependente). Usando o procedimento proposto por Florax *et al.* (2002), o primeiro passo foi regredir pelo método dos Mínimos Quadrados Ordinários (MQO) o percentual da votação de Lula no município contra o fator sócio-econômico (F1), o fator demográfico (F2), o fator de vulnerabilidade (F3) e o fator político (F4).

Os resultados estão reportados na tabela 5. O diagnóstico revela que, pela averiguação do teste Jarque-Bera, os erros não são distribuídos normalmente. Ademais, há evidências, por meio do teste de Koenker-Bassett (KB) de heterocedasticidade. Por outro lado, por se terem extraídas as variáveis independentes da análise fatorial, não há indicação de problema de multicolinearidade.

Tabela 5. Resultados da Regressão por MQO*

Coeficientes	
Constante	0,191 (0,000)
F1	0,228 (0,000)
F2	0,034 (0,634)
F3	-0,069 (0,000)
F4	0,408 (0,000)
<i>I</i> de Moran	0,237 (0,000)
ML - Erro	835,390 (0,000)
ML - Defasagem	301,847 (0,000)
MLR - Erro	624,552 (0,000)
MLR - Defasagem	91,004 (0,000)
Teste SARMA	926,399 (0,000)
Teste Koenker - Bassett	122,427 (0,000)
Teste de Jarque - Bera	17,481 (0,000)
AIC	-10369,700
SC	-10336,700
LIK**	5189,870
n	5507
R ² Ajustado	0,395

Fonte: resultados da pesquisa.

* Os resultados entre parênteses indicam o valor da probabilidade.

** LIK é o valor da função de máxima verossimilhança.

Todavia, há fortes sinais de que os erros estejam autocorrelacionados espacialmente tanto pelo I de Moran quanto pelos multiplicadores de Lagrange, estatisticamente significativos. Todos os testes de multiplicador de Lagrange mostram-se significativos. Em que pese isso, o valor mais alto do teste de multiplicador de Lagrange na versão robusta do erro indica que um bom modelo é o modelo de erro espacial. Entretanto, o teste SARMA, que indica um modelo de ordem superior, é altamente significativo.⁵ Ademais, o valor do teste SARMA é maior ainda do que o valor do teste ML robusto do erro. Assim, será estimado o modelo geral, especificado como segue:

$$PT = \beta_0 + \rho W_1 PT + \beta_1 F1 + \beta_2 F2 + \beta_3 F3 + \beta_4 F4 + u \quad (6)$$

$$u = \lambda W_2 u + \varepsilon$$

$$\varepsilon \sim (0, \Omega)$$

em que PT é o percentual da votação de Lula por município, W_1 é a matriz de pesos espaciais binários com a convenção de contigüidade torre, W_2 é a matriz dos 5 vizinhos mais próximos,⁶ λ e r são os parâmetros espaciais a serem estimados e ε é o termo de erro com média zero. A matriz Ω representa a matriz de variância-covariância cujos termos da diagonal principal não são constantes.

Serão estimados o modelo de defasagem espacial – cuja especificação é dada quando a expressão (6) possui a restrição $\lambda = 0$ –, o modelo de erro espacial – com a restrição de que $\rho = 0$ – e o modelo de defasagem espacial com erro espacial, que coincide com o modelo geral irrestrito. A razão para isso reside no fato de que é preciso se certificar que não há realmente o componente da defasagem espacial no modelo, pois a sua ausência, quando for relevante, conduz a estimativas inconsistentes.

Como os resíduos não são normais, o modelo de defasagem espacial será estimado pelo método das variáveis instrumentais,⁷ usando a matriz de variância-covariância consistente de White para corrigir a heterocedasticidade. Por sua vez, o modelo de erro espacial foi estimado pelo método generalizado de momentos (MGM), proposto por Kelejian e Prucha (1999), que prescinde do pressuposto da normalidade

⁵ Para detalhes técnicos sobre o teste SARMA, consulte Anselin e Bera (1998).

⁶ A seleção de cinco vizinhos mais próximos foi obtida usando o seguinte procedimento de busca. Foram tentadas várias matrizes W com $k=5, 10, 15, 20, 25, 30$ e 35 vizinhos mais próximos. Para cada uma das matrizes foi computado o I de Moran. Foi escolhida a matriz W que deu origem ao valor mais alto do I de Moran (no caso, cinco vizinhos mais próximos).

⁷ Os instrumentos utilizados foram os fatores $F1, F2, F3$ e $F4$ defasados espacialmente, usando a matriz torre para efetuar a operação de defasagem.

na sua estimação, com uma especificação para o erro heterocedástico na forma de grupos (o indicador dos grupos foram as cinco macro-regiões).⁸ O modelo de defasagem espacial com erro espacial foi estimado pelo método dos mínimos quadrados em dois estágios espacial generalizado (MQ2EEG), proposto por Kelejian e Prucha (1998). Neste artigo, será estendido o procedimento de Kelejian e Prucha, usando o MQ2EEG com erro heterocedástico em grupos. Os resultados da estimação dos modelos espaciais estão reportados na tabela 6.

Tabela 6. Estimação dos Modelos Espaciais

Coeficientes	Erro Espacial	Defasagem Espacial	Defasagem com Erro Espacial
Constante	0,181 (0,000)	0,286 (0,000)	0,134 (0,000)
F1	0,214 (0,000)	0,274 (0,000)	0,399 (0,000)
F2	0,047 (0,453)	-0,040 (0,691)	0,321 (0,000)
F3	-0,029 (0,017)	-0,101 (0,000)	0,047 (0,001)
F4	0,441 (0,000)	0,416 (0,000)	0,522 (0,000)
λ	0,4369		0,177
ρ		-0,259 (0,000)	-0,304 (0,000)
n	5507	5507	5507
Pseudo R ²	0,433	0,414	0,596

Fonte: resultados da pesquisa.

⁸ A matriz de pesos espaciais usada na estimação foi baseada na convenção de contigüidade da torre. Consulte Anselin (1988) para maiores detalhes técnicos a respeito de matrizes de pesos espaciais.

Os resíduos do modelo de defasagem espacial e do modelo de defasagem com erro espacial heterocedástico contiveram ainda evidências de autocorrelação espacial. Por isso, o modelo mais apropriado foi considerado o modelo de erro espacial. Segue, então, a análise dos resultados para esse modelo. O coeficiente estimado que acompanha o fator sócio-econômico (F1) assume sinal positivo e é altamente significativo do ponto de vista estatístico. Esse resultado é surpreendente, pois revela que a votação de Lula foi maior em municípios com melhores condições sócio-econômicas. Assim sendo, os grotões pobres do país não deram a vitória a Lula.

A estimativa do fator demográfico (F2) revela-se diretamente relacionada com a votação de Lula, com elevado nível de significância. Por sua vez, o coeficiente estimado para o fator de vulnerabilidade à violência (F3), apesar de ter sinal negativo, não se mostrou significativo estatisticamente. A interpretação disso é que o candidato do PT não conseguiu capitalizar a situação de insegurança a que estavam submetidos os municípios. A imagem de Lula (“Lulinha paz e amor”) não está associada à eficácia na resolução das questões ligadas à violência.

O fator político-ideológico desempenha um papel importante na explicação da votação do candidato do PT. O valor do coeficiente estimado (0,44), altamente significativo, é superior ao valor dos coeficientes que acompanham o fator sócio-econômico e o fator demográfico, mostrando-se mais relevante que esses dois últimos.

No método generalizado dos momentos proposto por Kelejian e Prucha (1999), o parâmetro espacial estimado é considerado como um termo de distúrbio (*nuisance*) e não apresenta nível de significância. Contudo, a correção da autocorrelação espacial nos resíduos aumenta a eficiência das estimativas.

Em modelos nos quais os erros não são esféricos, a tradicional medida de ajuste da regressão, o coeficiente de determinação (R^2), perde sentido, não podendo ser usada para comparar modelos espaciais concorrentes. Com o intuito de contornar isso, é mostrada a medida de um pseudo R^2 expresso como a razão entre a variância dos valores previstos pelo modelo e a variância dos valores observados para a variável dependente (Anselin, 1988).

O primeiro efeito espacial – a autocorrelação espacial – foi acomodado no modelo. A intenção agora é incorporar o segundo efeito espacial – a heterogeneidade espacial – manifestada na diferença dos parâmetros (β) em função das diferenças regionais. Entende-se que a questão regional desempenha papel fundamental nas eleições presidenciais no Brasil. Para isso, pretende-se introduzir uma variável categórica para identificar regimes espaciais, representados pelas macro-regiões brasileiras. Essa variável

categórica atribui o indicador 1 para a região Norte, 2 para o Nordeste, 3 para a região Sul, o indicador 4 para o Sudeste e, finalmente, 5 para o Centro-Oeste.

O modelo de erro espacial com mudança estrutural nos parâmetros, definida pelos regimes espaciais, foi estimado pelo método generalizado dos momentos de Kelejian e Prucha (1999). Os resultados estão apresentados na tabela 7.

Tabela 7. Estimação do Modelo de Erro Espacial com Regimes Espaciais por MGM *

Coeficientes	Norte	Nordeste	Centro-Oeste	Sudeste	Sul
Constante	0,142 (0,000)	0,119 (0,000)	0,216 (0,000)	0,235 (0,000)	0,228 (0,000)
F1	0,349 (0,000)	0,264 (0,000)	0,212 (0,000)	0,170 (0,000)	0,103 (0,001)
F2	-0,834 (0,075)	0,127 (0,635)	-0,439 (0,112)	0,141 (0,035)	-0,368 (0,249)
F3	0,071 (0,043)	-0,001 (0,983)	0,008 (0,750)	-0,194 (0,000)	0,036 (0,288)
F4	0,442 (0,000)	0,508 (0,000)	0,352 (0,000)	0,458 (0,000)	0,448 (0,000)
λ	0,412 (0,000)	0,412 (0,000)	0,412 (0,000)	0,412 (0,000)	0,412 (0,000)
Pseudo R ²	0,470	0,470	0,470	0,470	0,470

Fonte: resultados da pesquisa.

* os resultados entre parênteses indicam probabilidade.

O teste Chow-Wald averigua a hipótese nula da estabilidade estrutural conjunta dos coeficientes da regressão. Essa hipótese é amplamente rejeitada, pois o valor estimado do teste (75,06) é um valor muito extremo na distribuição qui-quadrado. Os testes sobre os coeficientes individuais indicam que há uma significativa diferença entre os coeficientes do intercepto, do fator sócio-econômico (F1) e do fator político-ideológico (F4) entre as regiões brasileiras. Todavia, os testes indicam, também, que não existe diferença significativa entre os coeficientes dos fatores demográficos (F2) e de vulnerabilidade (F3). Assim, para os fatores F1 e F4, existem fortes evidências de que os seus coeficientes variam através das macro-regiões.

Em que pese isso, o coeficiente associado ao fator sócio-econômico sempre se mostra com sinal positivo em todas as regiões, indicando que os municípios que apresentam as melhores condições sócio-econômicas proporcionaram os maiores percentuais de votos para Lula. Surpreendentemente, o coeficiente do fator F1 assume o maior valor na região Nordeste, enquanto ele assume o menor valor na região Centro-Oeste.

Os resultados da regressão mostram que o fator político-ideológico (F4) tem um maior impacto do que o fator sócio-econômico em todas as regiões. O coeficiente do F4 é sempre positivo e assume um valor mais baixo na região Sudeste e um valor mais alto na região Norte.

O pseudo R² tem um valor de 0,47, indicando que 47% da variação nos dados da variável dependente são explicadas pelos quatro fatores. Logo, 53% da variação da variável dependente podem ser interpretados como a influência do fator “Agora é Lula” na votação do candidato do PT. Assim, a maior parte dos votos de Lula não é explicada pelos fatores sócio-econômico, demográfico, de vulnerabilidade ou político-ideológico, mas sim pelo sentimento difuso, amplo e generalizado de que havia chegado a vez do candidato vencer as eleições.

5. CONSIDERAÇÕES FINAIS

Adotando um conjunto de métodos de análise quantitativa, incluindo a análise exploratória de dados espaciais, a econometria espacial e a análise fatorial, este trabalho chegou a resultados que desafiam a sabedoria convencional. Em primeiro lugar, Lula teve suas maiores votações em 2002 em municípios com mais elevado fator sócio-econômico, mais desenvolvidos, portanto, ao contrário do que se supõe.

O fator político-ideológico foi importante para a votação do candidato do PT, mostrando que este partido teve uma grande capacidade de fidelizar os seus votos nas eleições presenciais de 2002.

O componente espacial importou na votação do candidato do PT. Conforme salientado por Darmofal (2006), a ciência política reconhece que certos eventos políticos tendem a se concentrar espacialmente. Os votos de Lula mostraram tal padrão de concentração por causa da difusão da informação entre os eleitores e efeitos de vizinhança que influenciam a intenção de votos nos municípios brasileiros. Foi importante modelar tanto a autocorrelação espacial quanto a heterogeneidade espacial manifestada na diferença dos coeficientes estimados por região. Foram estimados o modelo de defasagem espacial, o modelo de erro espacial e o modelo de defasagem espacial com erro espacial. O modelo mais apropriado

para representar o processo das eleições presidenciais brasileiras de 2002 foi o modelo de erro espacial. Para acomodar a heterogeneidade regional, foi estimado, também, um modelo de erro espacial com regimes espaciais.

Contudo, o principal interesse deste trabalho repousou na intenção de quantificar o chamado efeito “Agora é Lula”. Tal efeito foi identificado como a parte da regressão não explicada pelos fatores determinantes extraídos da análise factorial, a saber, o fator sócio-econômico, o fator demográfico, o fator de vulnerabilidade e o fator político-ideológico, que são compostos por 26 variáveis relacionadas com as intenções de voto. O que não foi explicado por esses fatores foi considerado a influência do fator “Agora é Lula”, que pode ser expresso por um sentimento difuso e generalizado de que havia chegado a vez do candidato do PT assumir a presidência da República. O fator “Agora é Lula” foi alimentado pelo carisma do candidato, pelo tom despoliticizante empreendido na campanha, cujo mote é o personagem “Lulinha Paz e Amor”, a persistência da candidatura que havia perdido em três pleitos anteriores, além do processo de “beatificação” a que havia sido submetido fazia alguns anos.

Uma óbvia extensão deste trabalho é averiguar se os padrões verificados na votação de Lula na eleição de 2002 permaneceram ou se alteraram para as eleições de 2006. Essa extensão, evidentemente, depende, não só da disponibilidade de dados eleitorais, bem como de informações para as variáveis independentes para ser realizado.

REFERÊNCIAS

ANSELIN, L. The Moran scatterplot as an ESDAtool to assess local instability in spatial association. Fisher, M., Scholten, H.J. and Unwin, D. W. (eds). **Spatial analytical perspectives in GIS**. Taylor&Francis. London. p 111-125. 1996.

ANSELIN, L. Local indicators of spatial association – LISA. **Geographical Analysis**. V 27 (2), April. p 93-115. 1995.

ANSELIN, L. **Spatial econometrics: methods and models**. Kluwer Academic, Boston, 1988.

ANSELIN, L. e BERA, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A. e Giles D. E. (eds.) **Handbook of applied economic statistics**. Marcel Dekker, New York, p. 237-289.

CLIFF, A. D. e ORD, J.K. **Spatial processes: models and applications**. Pion, London. 1981

DARMOFAL, D. **Spatial econometrics and political science**. Mimeo., Dept. of Political Science, University of South Carolina, 2006.

FLORAX, R., FOLMER, H. e REY, S. **Specification searches in spatial econometrics: the relevance of Hendry's methodology**. Working paper, Dept. of Spatial Economics, Free University of Amsterdam, 2002.

GONÇALVES, E. **A Distribuição Espacial da Atividade Inovadora Brasileira: Uma Análise Exploratória**. Texto para discussão CEDEPLAR N° 246. Belo Horizonte: UFMG, 2005. 33p

GONTIJO, A. Elementos para uma tipologia do uso do solo agrícola no Brasil: uma aplicação de análise fatorial. **Revista Brasileira de Economia**, Rio de Janeiro, v. 2, n.1, jan/mar. 1998.

KELEJIAN, H. H. e PRUCHA, I. R. A generalized moments estimator for the autoregressive parameter in a spatial model. **International Economic Review**, v. 40, n. 2, 1999.

KELEJIAN, H. H. e PRUCHA, I. R. **A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances**. Mimeo., Department of Economics, University of Maryland, 1998.

LE GALLO, J and ERTHUR, C. **Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995**. Papers in Regional Science v 82(2) p 175-201. 2003.

MESSNER, S. F., ANSELIN, L, BALLER, R. D., HAWKINS, D. F., DEANE, G., and TOLNAY, S. E. The spatial patterning of country homicide rates: an application of exploratory spatial data analysis. **Journal of Quantitative Criminology**, 15(4): 423-450, 1999.

MIGUEL, Luis Felipe. **The Visible Election: the Globo television network discovers politics in 2002**. *Dados*, vol.46, no.2, p. 289-310, 2003.

NORUSIS, M.J. **SPSS Guide**. 1994.

Programa das Nações Unidas para o Desenvolvimento (PNUD). Brasília, 2005. Disponível em: < <http://www.pnud.org.br/> > Acesso em: 10 mar 2006.

QUEIROZ, W.T **Análise de fatores pelo método de máxima verossimilhança: aplicação no estudo de florestas tropicais**. Piracicaba, 1984 ESLQ/USP. 112p (Tese de doutorado).

SCHILDERINCK, J.H.F **Regression and factor analysis applied in econometrics**. Leiden: Martinus Nijhoff Social Sciences Division, p. 59-113, 1978 (Gilburg Studies in Econometrics v. 1).

Tribunal Superior Eleitoral (TSE). Brasília, 2006. Disponível em: < <http://www.tse.gov.br/> > Acesso em: 10 mar 2006.

VARGA, A. **University research and regional innovation: a spatial econometric analysis of academic technology transfers**. Boston/Dordrecht/London: Kluwer Academic Publishers, 1998.

Anexo

Descrição das Variáveis Usadas na Análise Fatorial

DEPFED_PT: votação percentual para deputado federal da coligação do PT nas eleições de 2002.

DEPFED_PSDB: votação percentual para deputado federal da coligação do PSDB nas eleições de 2002.

DEPFED_PSB: votação percentual para deputado federal da coligação do PSB nas eleições de 2002.

DEPFED_PPS: votação percentual para deputado federal da coligação do PPS nas eleições de 2002.

1998/PT: votação percentual do presidente do PT por município em 1998.

1998/PPS: votação percentual do presidente do PPS por município em 1998.

1998/PSDB: votação percentual do presidente do PSDB por município em 1998.

EL 98-02: Efeito Lula, é a diferença percentual entre a votação do PT em 1998 e 2002.

INTPOB1991: intensidade de pobreza em 1991 - linha de R\$ 75,5.

INTPOB2000: intensidade de pobreza em 2000 - linha de R\$ 75,5.

IDH-1991: índice de desenvolvimento humano 1991.

IDH-2000: índice de desenvolvimento humano 2000.

ALF-1991: taxa de alfabetização em 1991.

ALF-2000: taxa de alfabetização em 2000.

GINI2000: índice de GINI 2000.

HOM98: taxa de homicídios em 1998.

HOM99: taxa de homicídios em 1999.

HOM00: taxa de homicídios em 2000.

HOM01: taxa de homicídios em 2001.

HOM02: taxa de homicídios em 2002.

RPC-91: renda per-capita 1991.

RPC-00: renda per capita 2000.

CH – 2000: capital humano 2000.

IPA – 91: intensidade de pobreza 1991 - linha de R\$ 37,75.

IPA – 00: intensidade de pobreza 2000 - linha de R\$ 37,75.

% POP UR: percentual de população urbana.

PR-00: População rural, 2000.

PT-00: População total, 2000.

PU-00: População urbana, 2000.

EVN-00: Esperança de vida ao nascer, 2000.

M5-00: Mortalidade até cinco anos de idade, 2000.

PS40-00: Probabilidade de sobrevivência até 40 anos, 2000.

PS60-00: Probabilidade de sobrevivência até 60 anos, 2000

TFT-00: Taxa de fecundidade total, 2000.

ME25-00: Média de anos de estudo das pessoas de 25 anos ou mais de idade, 2000.

PRTG-00: Percentual de pessoas com mais de 50% da sua renda proveniente de transferências governamentais, 2000.

RTG-00: Percentual da renda proveniente de transferências governamentais, 2000.

RAR-00: Percentual da renda apropriada pelos 10% mais ricos da população, 2000.

II-00: Intensidade da indigência, 2000.

AENC-00: Percentual de pessoas que vivem em domicílios com água encanada, 2000.

BANAG-00: Percentual de pessoas que vivem em domicílios com banheiro e água encanada, 2000.

CAR-00: Percentual de pessoas que vivem em domicílios com carro, 2000.

ELTV-00: Percentual de pessoas que vivem em domicílios com energia elétrica e TV , 2000.

EL-00: Percentual de pessoas que vivem em domicílios com energia elétrica, 2000.

LIXO-00: Percentual de pessoas que vivem em domicílios urbanos com serviço de coleta de lixo, 2000.

TEL-00: Percentual de pessoas que vivem em domicílios com telefone, 2000.

MED-00: Número de médicos residentes por mil habitantes, 2000.

TALF-00: Taxa de alfabetização, 2000.

FREESC-00: Taxa bruta de freqüência à escola, 2000.

IDHRENDA-00: Índice de Desenvolvimento Humano Municipal-Renda, 2000.

IDHLONG-00: Índice de Desenvolvimento Humano Municipal-Longevidade, 2000.

IDHEDU-00: Índice de Desenvolvimento Humano Municipal-Educação, 2000.

%PR: Percentual de população rural, 2000.

%PU: Percentual de população urbana, 2000.

M25-00: Mulheres de 25 anos ou mais de idade, 2000.

P25-00 : População de 25 anos ou mais de idade, 2000.

%M25-00: Percentual Mulheres de 25 anos ou mais de idade, 2000.