

**Fundação Getúlio Vargas**  
**Escola de Administração de Empresas de São Paulo**  
**Programa de Pós-Graduação em Administração Pública e Governo**

**MARCELO TYSZLER**

**Econometria Espacial: Discutindo Medidas para a Matriz de  
Ponderação Espacial**

São Paulo

2006



**Fundação Getúlio Vargas**  
**Escola de Administração de Empresas de São Paulo**  
**Programa de Pós-Graduação em Administração Pública e Governo**

Marcelo Tyszler

Econometria Espacial: Discutindo Medidas para a Matriz de Ponderação Espacial

Dissertação Apresentada à Escola de  
Administração de Empresas de São Paulo  
da Fundação Getúlio Vargas, como requisito  
para a obtenção do título de Mestre em  
Administração Pública e Governo

Linha de Pesquisa: Finanças Públicas  
Orientador: Ciro Biderman

São Paulo

2006



Tyszler, Marcelo.

Econometria Espacial: Discutindo Medidas para a Matriz de Ponderação Espacial / Marcelo Tyszler. - 2006.

155 f.

Orientador: Ciro Biderman.

Dissertação (mestrado) - Escola de Administração de Empresas de São Paulo.

1. Econometria. 2. Método de Monte Carlo. I. Biderman, Ciro. II. Dissertação (mestrado) - Escola de Administração de Empresas de São Paulo. III. Título.

CDU 330.115



Marcelo Tyszler

Econometria Espacial: Discutindo Medidas para a Matriz de Ponderação Espacial

Dissertação Apresentada à Escola de  
Administração de Empresas de São Paulo  
da Fundação Getúlio Vargas, como requisito  
para a obtenção do título de Mestre em  
Administração Pública e Governo

Data de Aprovação: \_\_\_\_/\_\_\_\_/\_\_\_\_

**Banca Examinadora:**

---

Prof. Dr. Ciro Biderman (Orientador)  
FG/-EAESP/ FGV-EESP

---

Prof. Dr. Paulo Arvate  
FG/-EAESP/ FGV-EESP

---

Prof. Dr. Danilo Iglioni  
Cambridge University / USP





## RESUMO

Com o crescente uso das teorias e métodos vindos das áreas de economia regional e urbana, economia e econometria espacial na fronteira entre as áreas de economia e políticas públicas, estudar esse arcabouço teórico e metodológico torna-se relevante no campo da análise de políticas públicas.

O objetivo geral dessa dissertação é discutir de modo crítico aspectos da literatura de econometria espacial e sua aplicabilidade, com especial foco na comparação entre os diferentes modelos e matrizes de ponderação espacial.

Esse estudo procura analisar diferentes formas de medidas que permitem a construção de uma matriz de ponderação espacial, bem como testar por meio de simulação experimental e em uma aplicação empírica a robustez de algumas dessas matrizes e de diferentes modelos.

Dadas as indicações da literatura e as evidências mostradas nessa dissertação, a estratégia mais confiável para a busca de um modelo parece ser utilizar as estimativas dos modelos, em conjunto com as estatísticas de Moran, para identificar presença de algum tipo de correlação espacial, e do teste LM da variável dependente, para indicar esse tipo de correlação especificamente.

Observa-se que, de modo geral, não há muita diferença em termos de esperança dos parâmetros não espaciais, quando comparadas a estimações tradicionais com as que utilizam as técnicas espaciais. Com isso, a indicação que se pode fazer é que o uso das ferramentas de econometria espacial deve ser aplicado quando o interesse da análise é justamente na estimação dos parâmetros de correlação espacial.

Comparando o uso da matriz de contigüidade no lugar da matriz de distâncias, é possível notar que a matriz de contigüidade não é, de forma geral, uma boa substituição para a matriz de distâncias. Ela capta parcialmente o efeito de *rho* e *lambda*, além de captar erroneamente *lambda*, quando este inexistente.

**PALAVRAS-CHAVE:** Econometria Espacial; Matriz de Ponderação Espacial; Simulação de Monte Carlo

## ABSTRACT

Considering the increasing use of methods, theories and techniques from urban and regional sciences, spatial economics and econometrics at overlapping areas of economics and public policy, it becomes relevant to the field of public policy analysis to understand and comprehend such set of knowledge.

The main objective of this dissertation is to critically discuss aspects of the spatial econometrics literature and its applicability, with a special focus on the comparison of different methods and spatial weight matrices.

This study tries to analyze different alternative measures to construct spatial weight matrices, as well as test by experimental simulation and empirical application the robustness of some of these alternative models.

Given the indications at the literature and evidences show at this study, the more reliable strategy for a spatial specification search seems to be the use of model's estimates in conjunction with a few specification tests. Moran's  $I$  is a good test to identify the presence of spatial correlation, while the LM test on the dependent variable is a good test to identify this specific kind of spatial correlation.

It is possible to notice that, in general terms, there is not so much difference in terms of expectation on non-spatial coefficients, exception for the constant term, when spatial estimates are compared with traditional ones. From this, the indication that can be made is that the use of spatial econometrics techniques should take place when the estimation of the spatial correlation parameters are the analysis target.

Comparing the use of the contiguity matrix in place of a distance based matrix, it is possible to verify that the contiguity matrix is not, in general terms, a good substitute. When used on spatial regressions, the spatial coefficients are partially estimated and the error autocorrelation, when non-existent, is wrongly identified.

**KEYWORDS:** Spatial Econometrics; Spatial Weight Matrix; Monte Carlo Simulation

## Sumário

<b>RESUMO</b> .....	i
<b>ABSTRACT</b> .....	ii
<b>Índice de Figuras</b> .....	iv
<b>Índice de Tabelas</b> .....	iv
<b>Notação utilizada</b> .....	vi
<b>PARTE I – INTRODUÇÃO, MOTIVAÇÃO E CONTEXTUALIZAÇÃO</b> .....	1
1. <b>Introdução</b> .....	1
2. <b>Motivação e Justificativa</b> .....	3
3. <b>Introdução à Análise Exploratória de Dados Espaciais</b> .....	7
4. <b>Questões de Investigação</b> .....	15
<b>PARTE II – ECONOMETRIA ESPACIAL</b> .....	16
5. <b>Modelos Espaciais, Heterogeneidade e Dependência Espacial</b> .....	16
6. <b>O Problema da Estimação</b> .....	21
7. <b>Estimando coeficientes em modelos espaciais</b> .....	25
8. <b>Testando a Presença do Componente Espacial e Especificando o Modelo</b> .	31
9. <b>Tópicos Avançados: Painel e Modelos semi-paramétricos</b> .....	46
<b>PARTE III – PESOS ESPACIAIS E SUAS APLICAÇÕES</b> .....	48
10. <b>Distâncias Geográficas</b> .....	48
11. <b>Distâncias Sócio-Econômicas</b> .....	53
12. <b>Medidas de Dissimilaridades</b> .....	55
<b>PARTE IV: SIMULAÇÕES DE MONTE CARLO</b> .....	57
13. <b>O Método de Monte Carlo</b> .....	57
14. <b>Definição Teórico-Operacional da Simulação</b> .....	61
15. <b>Resultados e Conclusões</b> .....	65
<b>PARTE V: APLICAÇÃO EMPÍRICA</b> .....	91
16. <b>Contextualização da análise empírica</b> .....	91
17. <b>Fundamentos e Especificação do Modelo da Análise Empírica</b> .....	92
18. <b>Variáveis Utilizadas</b> .....	101
19. <b>Resultados das Regressões</b> .....	104
20. <b>Análise Crítica da Mudança de Matrizes e Modelos</b> .....	110
<b>PARTE VI – TESTANDO DISTÂNCIAS ECONÔMICAS E UMA ESTRUTURA ALTERNATIVA DA DINÂMICA ESPACIAL</b> .....	116
21. <b>Influências Opostas</b> .....	116
22. <b>Influências Opostas na Simulação</b> .....	119
23. <b>Análise Empírica com Diferentes Medidas Geográficas e Sócio-Econômicas</b> .....	122
<b>PARTE VII: LIMITAÇÕES E CONCLUSÕES</b> .....	126
24. <b>Limitações e Sugestões para Próximos Estudos</b> .....	126
25. <b>Conclusões</b> .....	128
<b>REFERÊNCIAS</b> .....	131
<b>APÊNDICES</b> .....	134
<b>Rotina base de simulação – Programação MatLab</b> .....	134
<b>Testes de Dependência Espacial – Programação MatLab</b> .....	139

## Índice de Figuras

Figura 3. 1 - Mapa, Histograma e Estatísticas Descritivas para $\ln(\text{Renda Per Capita})$ 1991 .....	8
Figura 3. 2 - Mapa, Histograma e Estatísticas Descritivas para $\ln(\text{Renda Per Capita})$ 2000 .....	8
Figura 3. 3 - Mapa, Histograma e Estatísticas Descritivas para $\ln(\text{Emprego Formal})$ 1991 .....	9
Figura 3. 4 - Mapa, Histograma e Estatísticas Descritivas para $\ln(\text{Emprego Formal})$ 2000 .....	9
Figura 3. 5 - Mapa, Diagrama de Dispersão de Moran, Estatísticas descritivas para taxa de crescimento do emprego formal (em $\ln$ ) 1991-2000 .....	11
Figura 3. 6 - Diagrama de Moran com $W_{dist}$ .....	14
Figura 3. 7 - Diagrama de Moran com $W_{contig}$ .....	14
 Figura 8. 1- Diagrama de Dispersão de Moran .....	 33
 Figura 19. 1 - Mapa, Diagrama de Dispersão e Estatísticas Descritivas para Taxa de Variação no Emprego Formal ( $\ln$ ), 1991-2000 .....	 105

## Índice de Tabelas

Tabela 1 – Notação Utilizada .....	vi
 Tabela 8. 1 - Valores Médios e Desvios dos Resultados das Simulações dos Testes de Especificação. SAR, 100 iterações, $n = 27$ .....	 40
Tabela 8. 2 - Valores Médios e Desvios dos Resultados das Simulações dos Testes de Especificação. SAR, 100 iterações, $n = 1000$ .....	41
Tabela 8. 3 - Valores Médios e Desvios dos Resultados das Simulações dos Testes de Especificação. Nenhuma Correlação Espacial, 100 iterações, $n = 27$ .....	42
Tabela 8. 4 - Valores Médios e Desvios dos Resultados das Simulações dos Testes de Especificação. Nenhuma Correlação Espacial, 100 iterações, $n = 1000$ .....	43
Tabela 8. 5 - Valores Médios e Desvios dos Resultados das Simulações dos Testes de Especificação. SEM, 100 iterações, $n = 27$ .....	43
Tabela 8. 6- Valores Médios e Desvios dos Resultados das .....	43
 Tabela 15. 1 - Resultados das Simulações, Sem Correlação Espacial, $n = 27$ .....	 66
Tabela 15. 2 - Resultados das Simulações, SAR, $n = 27$ .....	67
Tabela 15. 3 - Resultados das Simulações, SEM, $n = 27$ .....	68
Tabela 15. 4 - Resultados das Simulações, SAC1, $n = 27$ .....	69
Tabela 15. 5 - Resultados das Simulações, SAC2, $n = 27$ .....	70
Tabela 15. 6 - Resultados das Simulações, Sem Correlação Espacial, $n = 1000$ .....	71
Tabela 15. 7 - Resultados das Simulações, SAR, $n = 1000$ .....	72
Tabela 15. 8 - Resultados das Simulações, SEM, $n = 1000$ .....	73
Tabela 15. 9 - Resultados das Simulações, SAC1, $n = 1000$ .....	74
Tabela 15. 10 - Resultados das Simulações, SAC2, $n = 1000$ .....	75
Tabela 15. 11 - Resultados das Simulações, Sem Correlação Espacial, $n = 4267$ .....	76
Tabela 15. 12 - Resultados das Simulações, SAR, $n = 4267$ .....	77
Tabela 15. 13 - Resultados das Simulações, SEM, $n = 4267$ .....	78

Tabela 15. 14 - Resultados das Simulações, SAC1, $n = 4267$ .....	79
Tabela 15. 15 - Resultados das Simulações, SAC1, $n = 4267$ .....	80
Tabela 15. 16 - Poder de explicação de X sobre Wy, valores médios e desvios em 500 simulações .....	83
Tabela 15. 17 - Correlação e ajuste entre Wdist*y e Wcontig*y, valores médios e desvios em 500 simulações .....	86
Tabela 15. 18 - Simulação de Dados Gerados com a Matriz de Contigüidade. SAR, 500 iterações, $n = 1000$ .....	87
Tabela 15. 19 - Simulação de Dados Estimados com a Matriz de Contigüidade. 500 iterações, $n = 4267$ , mapa aleatório .....	89
Tabela 19. 1 - Resultados das Regressões - Parte I .....	106
Tabela 19. 2 - Resultados das Regressões - Parte II.....	107
Tabela 20. 1 - Comparação dos Resultados das Regressões com Diferentes Matrizes - Parte I.....	111
Tabela 20. 2 - Comparação dos Resultados das Regressões com Diferentes Matrizes - Parte II .....	112
Tabela 20. 3 – Testes de Especificação Espacial, Wdist .....	113
Tabela 20. 4 – Testes de Especificação Espacial, Wcontig .....	113
Tabela 20. 5 – Ajuste de X em W*y .....	114
Tabela 20. 6 - Ajuste entre Wdist*y e Wcontig*y .....	115
Tabela 22. 1 - Resultados das Simulações com Influências Opostas, SAR, $n = 1000$ .	120
Tabela 23. 1 - Resultado das Regressões - Medidas Geográficas e Sócio-Econômicas - Parte I.....	123
Tabela 23. 2 - Resultado das Regressões - Medidas Geográficas e Sócio-Econômicas - Parte II .....	124

## Notação utilizada

Nas equações apresentadas ao longo dessa dissertação, a seguinte notação será utilizada:

Formato	Descrição	Exemplo
Minúsculo itálico	Escalar	$y$
Minúsculo negrito	Vetor Coluna Vetor Linha	$\mathbf{y}$ $\mathbf{y'}$
Maiúsculo itálico	Matriz	$X$
Minúsculo itálico cm Subscrito	Observação específica de uma unidade	$y_i$

**Tabela 1** – Notação Utilizada

## PARTE I – INTRODUÇÃO, MOTIVAÇÃO E CONTEXTUALIZAÇÃO

### 1. Introdução

A presente dissertação é parte dos requisitos necessários para a obtenção do título de Mestre em Administração Pública e Governo pela Escola de Administração de Empresas da Fundação Getúlio Vargas de São Paulo.

O objetivo geral dessa dissertação é discutir de modo crítico aspectos da literatura de econometria espacial e sua aplicabilidade, com especial foco na comparação entre os diferentes modelos e matrizes de ponderação espacial.

Na literatura de econometria espacial admite-se que a escolha de uma determinada matriz de ponderação espacial permite escolhas *ad-hoc* por parte dos pesquisadores. No entanto, algumas medidas para a construção de matrizes são mais usuais e preponderantes. Ainda, podemos considerar determinadas medidas como simplificações de outras.

Essa dissertação procura, então, discutir essas diferentes formas de medidas que permitem a construção de uma matriz de ponderação espacial, bem como testar por meio de simulação experimental e em uma aplicação empírica a robustez de algumas dessas matrizes e de diferentes modelos.

Adicionalmente, ao considerar o uso da econometria espacial no estudo de políticas públicas, são tratados alguns aspectos peculiares a espaços sócio-econômicos que usualmente não emergem da discussão de espaços puramente geográficos. Em particular, procura-se estudar de forma exploratória, por simulação e com dados empíricos, uma estrutura específica de matriz espacial que decompõe a influência espacial.

Essa dissertação está dividida nas seguintes partes: Parte I: Parte I – Introdução, Motivação e Contextualização, Parte II – Econometria Espacial, Parte III – Pesos Espaciais e Suas Aplicações, Parte IV: Simulações de Monte Carlo, Parte V: Aplicação Empírica, Parte VI – Testando Distâncias Econômicas e uma Estrutura Alternativa da Dinâmica Espacial, Parte VII: Limitações e Conclusões.

A primeira parte, além dessa introdução apresenta a motivação e justificativa desse estudo. Inclui ainda uma introdução à análise exploratória de dados espaciais, bem como explicita as questões de interesse.

A segunda parte é dedicada à revisão crítica dos aspectos da literatura de econometria espacial relevantes para essa dissertação. São apresentados os modelos, testes e técnicas de estimação.

A terceira parte continua a revisão da literatura, porém focando especificamente a matriz de ponderação espacial e suas diferentes formas de construção.

Na quarta parte é dedicada à simulação experimental, apresentando o método, sua operacionalização e resultados.

A quinta parte realiza análises semelhantes, porém com dados empíricos. São apresentados os fundamentos da especificação utilizada, os resultados encontrados e a análise da mudança da escolha de matrizes de ponderação espacial.

Uma análise exploratória é realizada na sexta parte, testando algumas medidas sócias econômicas e alternativas de dinâmica espacial, também de forma simulada e empírica.

Por fim, na última parte são apresentadas as limitações e sugestões de futuros estudos, bem como resumidos os principais resultados dessa dissertação.

Incluem-se posteriormente às referências as rotinas de Matlab criadas pelo autor para a execução operacional das simulações e cálculos dessa dissertação



## 2. Motivação e Justificativa

Em análises econométricas, simplificações são feitas na criação de modelos que permitem estudar o mundo. Estas simplificações são mais ou menos fortes de acordo com diversos fatores como disponibilidade de dados e/ou instrumentos, soluções teóricas para determinados problemas ou mesmo condições computacionais de cálculo.

Um modelo nunca deve reproduzir a realidade com perfeição, pois então seria tão extenso que perderia sua capacidade de sistematização. Um modelo deve, no entanto, captar a essência das relações entre as variáveis e dar condições para a realização de inferências e previsões sobre determinadas questões de interesse.

O ponto de partida de quase todos os modelos de análise econométrica é o modelo de regressão linear clássico, estimado pelo método dos mínimos quadrados ordinários (OLS).

O modelo de regressão linear clássico se apóia em alguns pressupostos. Um dos princípios mais importantes desse modelo é o da não correlação das variáveis independentes com o termo de erro (exogeneidade), o que gera, por consequência, estimadores não viesados e consistentes. Além disso, assume-se que as observações são independentes e igualmente distribuídas (i.i.d.). Por fim, para obter maior eficiência, assume-se uma estrutura específica do erro, diagonal, sem alterar a esperança condicional da variância. Assim, tem-se o erro homoscedástico.

Para a grande maioria dos estudos, simplificações são aceitáveis e não ferem esses princípios. No entanto, existe uma classe importante de estudos em que as simplificações necessárias para a criação de modelos verossímeis ferem esses princípios. Quando a chamada componente espacial dos dados se torna relevante, sua omissão gera um importante viés de omissão de variável. Em particular, as observações deixam de ser independentes, justamente por apresentarem um padrão de dependência espacial. Sendo assim, soluções precisam ser tomadas para resolver esses problemas.

Modelos espaciais originaram-se e ainda são usados em muitos estudos da área de economia urbana e regional, e recentemente vêm sendo aplicados a estudos de políticas públicas em que a racionalidade das aglomerações é relevante. Exemplos incluem estudos sobre crime e interação fiscal entre Estados. O crescimento do uso de econometria espacial em problemas de políticas públicas pode ser observado nos trabalhos de Boarnet and Glazer (2002), Case and Rosen (1993), Revelli (2002), Brueckner (2003), entre outros.

O aspecto central para o estudo dessa dissertação é que em estudos que consideram relevante a componente espacial, e têm como intuito medir o grau dessa dependência espacial, algum instrumento específico deve ser utilizado para captar a estrutura de correlação espacial. O instrumento criado para lidar com isso de forma econométrica é a matriz de pesos espaciais, usualmente chamada de  $W$ , que resume a estrutura de correlação espacial.

A matriz de pesos espaciais é uma matriz quadrada ( $n \times n$ ) que contém os pesos espaciais de cada unidade sobre outra. Assim, o elemento  $w_{i,j}$  indica o peso espacial que a unidade  $j$  exerce sobre a unidade  $i$ . Quando a matriz  $W$  é construída, independente do critério utilizado, ela é tratada como um fator exógeno. A maioria dos modelos espaciais consideram a estrutura exógena (conhecida e fixa). Apesar de existirem estudos de matriz endógena, para o escopo dessa dissertação, o foco será dado aos modelos de matriz exógena.

É importante destacar, porém, que sem a imposição exógena, ao menos inicial, de elementos que compõe a matriz, não há como serem estimados todos os parâmetros. Alguns comentários serão feitos sobre modelos semi-paramétricos e de matrizes endógenas, porém o foco se restringirá às matrizes exógenas, visto que é o que prepondera na literatura atual.

Como essa matriz é exógena, ela deve ser determinada a priori pelo pesquisador. Para se construir essa matriz deve existir alguma razão teórica substanciada para se argumentar que ela deva ser construída de tal forma. Além disso, por questões operacionais, muitas vezes uma estrutura ideal dessa matriz deve ser posta de lado dando lugar a alguma simplificação. Fatores que podem exigir tal simplificação podem ser computacionais, em termos de capacidade de cálculo, ou da falta de dados para o cálculo do peso espacial inicialmente desejado. Pode-se notar que essa discricionariedade do pesquisador dá margem a seleções *ad-hoc*, o que certamente não seria desejado. De acordo com Ertur e Gallo (2003), Anselin (1999, 1980), Florax e Rey *apud* Ertur e Gallo (2003), a escolha apropriada de uma matriz de pesos espaciais é um dos aspectos mais difíceis e controversos aspectos metodológicos tanto na análise exploratória de dados espaciais como no próprio campo da econometria espacial.

Existem modelos mais tradicionais que medem as distâncias e pesos espaciais a partir de atributos físicos e geográficos, como vizinhança, distância geográfica ou tempo de deslocamento e modelos, especialmente aplicado a estudos de políticas públicas, em que as distâncias e pesos não estão em um espaço geográfico, mas sim em um espaço sócio-econômico. No entanto, mesmo em medidas sócio-econômicas, o conceito central é o de distância, sendo o peso de uma observação em outra medido de acordo a sua distância, seja ela geográfica ou econômica.

Uma questão essencial para essa dissertação é como escolher a estrutura de ponderação espacial, e qual a consequência de cada escolha. Quando observamos as medidas geográficas, uma questão surge de imediato: quanto uma medida simplificada é capaz de captar relações que são geradas por uma medida mais complexa? Sabemos que por aspectos já indicados, como operacionais, computacionais ou mesmo de informação precisa, medidas simplificadas têm que ser utilizadas. A questão é se isso pode ser prejudicial para a estimação de parâmetros e para a compreensão das relações entre as variáveis de interesse.

Por outro lado, a crescente inclusão da componente espacial em estudos de políticas públicas começa, justamente, a levantar uma outra questão particular sobre os modelos: como medir ou definir essa estrutura de correlação espacial quando espaços geográficos e sócio-econômicos se misturam.

Nos textos de econometria espacial muitas vezes é feita uma analogia com as técnicas de econometria de séries de tempo, sempre se argumentando que a multidirecionalidade da dependência espacial impede que as soluções utilizadas em séries temporais (em geral assumindo defasagem unidirecional no tempo) sejam aplicáveis. Quando esses modelos são levados para campos em que as distâncias sócio-econômicas são relevantes, surgem outras questões que não são levantadas quando a mensuração é feita por atributos geográficos.

Em geral, nos estudos geográficos, as distâncias são sempre medidas em módulo, tendo-se um pressuposto implícito que o que importa é a distância relativa e não a posição que cada unidade ocupa. O uso do módulo também se faz necessário pela bidimensionalidade da posição no espaço geográfico, que impede uma ordenação única, ou um cálculo sem o módulo. Já em estudos sócio-econômicos, como a interação fiscal, por exemplo, a posição relativa em determinado espaço, como o de alíquotas de tributação, ou nível de população não só permite uma ordenação como também pode ser relevante.

Sendo assim, essa dissertação procura concentrar sua atenção no aspecto metodológico de econometria espacial, em particular nas matrizes de ponderação espacial. Esse aspecto mostra-se como relevante pois questiona a robustez e a conseqüente adequação de determinadas escolhas para a estimação de parâmetros em modelos cada vez mais utilizados na fronteira da economia regional e urbana com o estudo de políticas públicas.

### 3. Introdução à Análise Exploratória de Dados Espaciais

De acordo com Ertur e Gallo (2003), a análise exploratória de dados espaciais (ESDA) é um conjunto de técnicas voltadas a descrever e visualizar distribuições espaciais, identificar localizações atípicas ou *outliers* espaciais, detectar padrões de associação espacial e *clusters* além de sugerir regimes espaciais ou outras formas de heterogeneidade espacial.

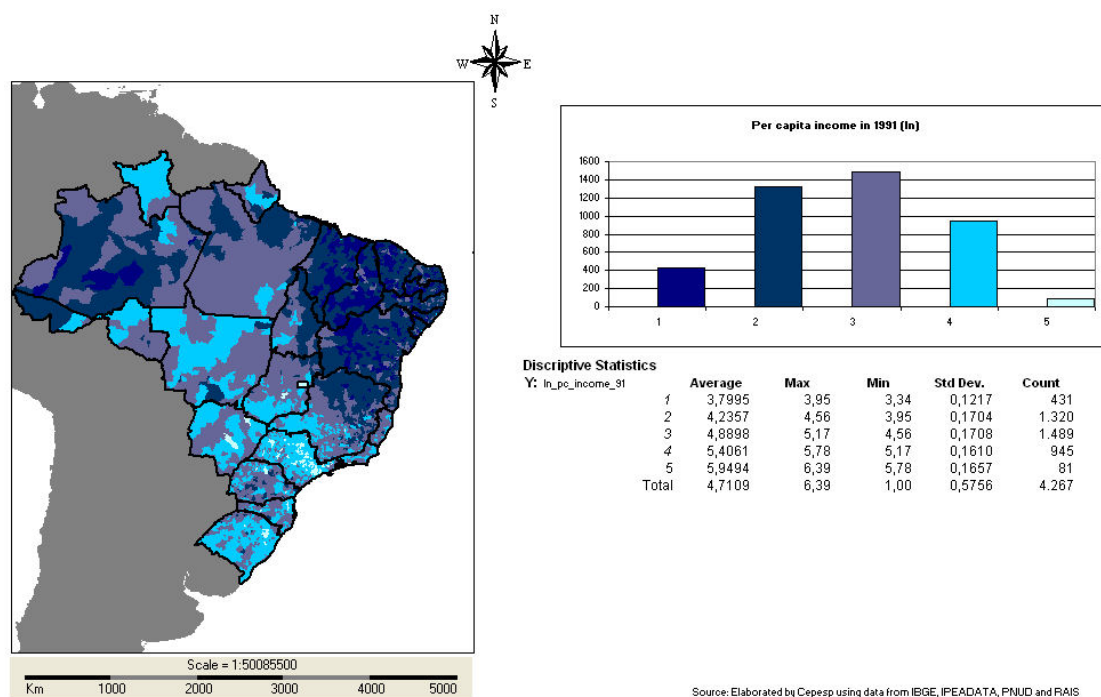
De forma bem geral, a autocorrelação espacial pode ser definida como a coincidência de similaridade de valores com a similaridade de localização (Anselin, 1999; Ertur e Gallo, 2003).

Uma primeira análise descritiva possível é traçar mapas com as unidades de observação divididas em faixas de valores para uma determinada variável. A figura (3.1) a seguir exibe um Mapa, Histograma e Estatísticas Descritivas para faixas de  $\ln(\text{renda per capita})$  no Brasil em 1991 e a figura (3.2) o mesmo para os dados de 2000. As unidades de observação são Áreas Mínimas Comparáveis (AMC), uma agregação feita pelo Ipea que permite comparar as áreas municipais de 1991 a 2000<sup>1</sup>. Cada faixa do histograma corresponde a um quinto da amplitude dos dados.

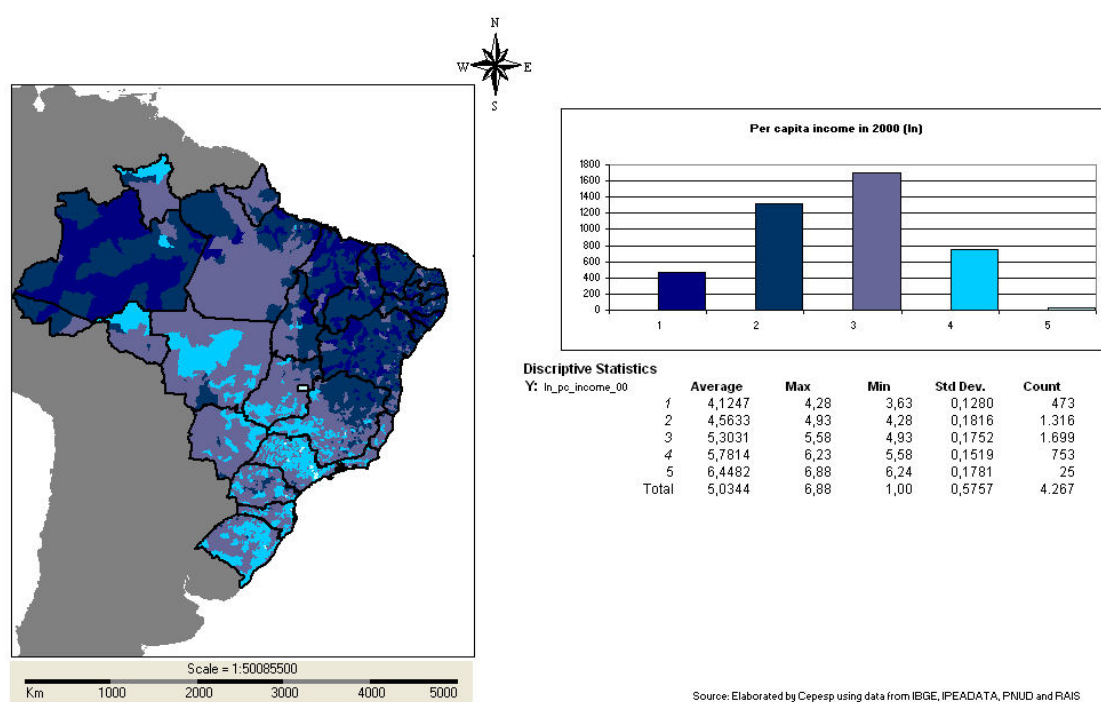
Nos mapas de renda é possível notar a concentração, já conhecida, de AMCs das faixas mais altas nas regiões sul e sudeste. É possível notar também que não há grande alteração nessa distribuição de um ano ao outro, sendo as regiões mais pobres nordeste e Amazonas.

---

<sup>1</sup> O conceito e justificativa das Áreas mínimas comparáveis são apresentados no capítulo 18



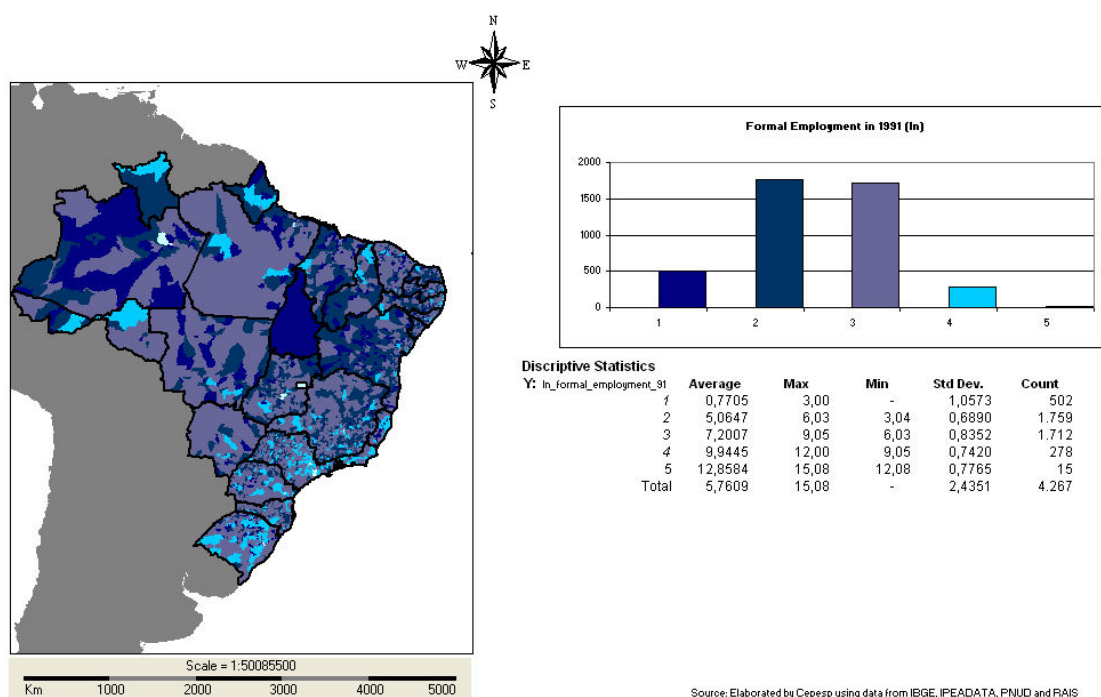
**Figura 3. 1 - Mapa, Histograma e Estatísticas Descritivas para ln(Renda Per Capita) 1991**



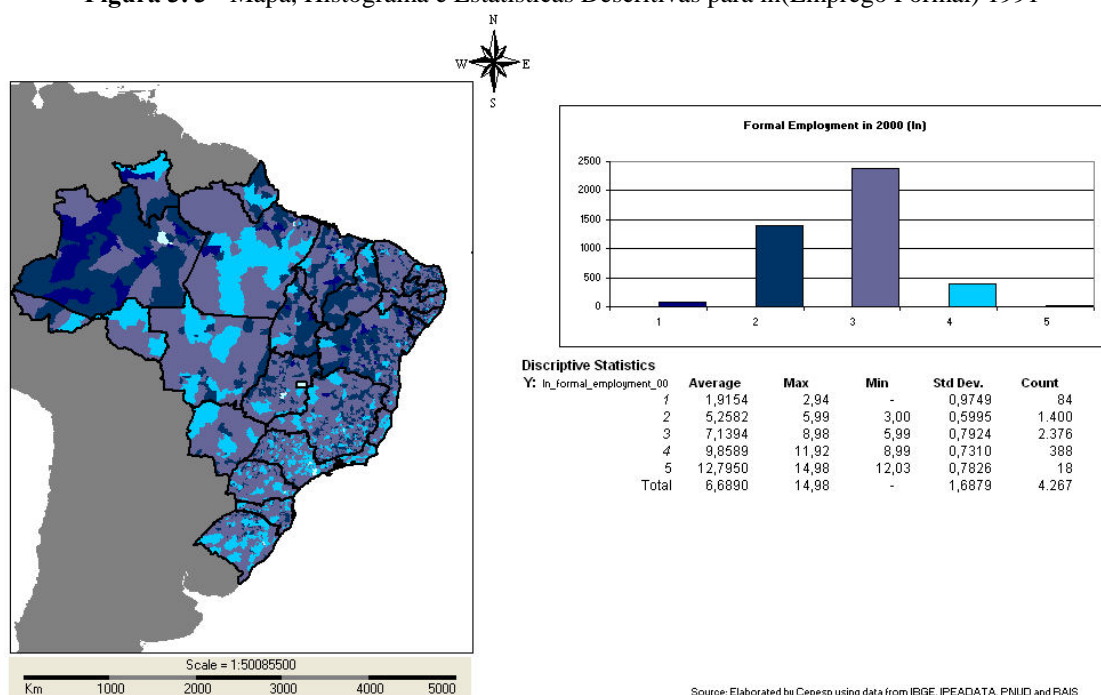
**Figura 3. 2 - Mapa, Histograma e Estatísticas Descritivas para ln(Renda Per Capita) 2000**

Na parte dessa dissertação dedicada à análise empírica, a variável de análise será o emprego formal, mais precisamente a taxa de variação do emprego formal nos período de 1991 a 2000.

As figuras (3.3) e (3.4) a seguir, exibem a mesma estrutura de informação, porém para a variável de análise,  $\ln(\text{emprego formal})$ .



**Figura 3. 3 - Mapa, Histograma e Estatísticas Descritivas para  $\ln(\text{Emprego Formal})$  1991**



**Figura 3. 4 - Mapa, Histograma e Estatísticas Descritivas para  $\ln(\text{Emprego Formal})$  2000**

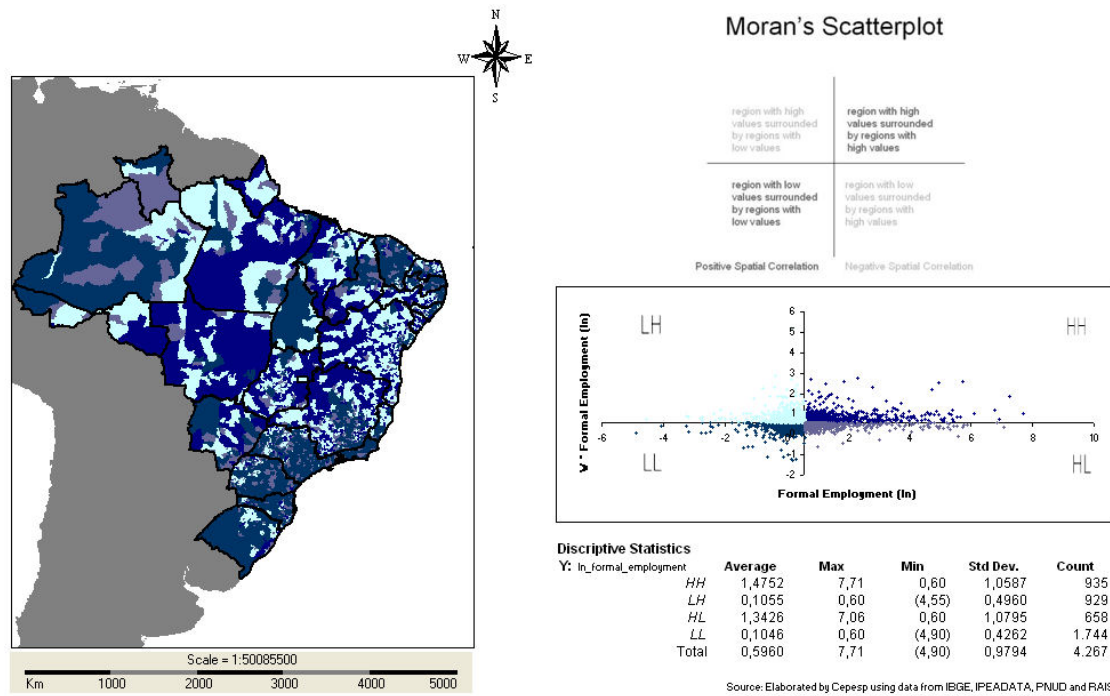
É possível notar, também nos mapas de emprego, uma concentração espacial, porém menos evidente que nos mapas de renda per capita. Novamente, sul e sudeste aglutinam os maiores empregos e nordeste e Amazonas ficam com AMCs com menos empregos.

Já a diferença entre os mapas é mais visível, em especial na região de Tocantins, no centro do país, indicando uma ascensão nas faixas do histograma. Podemos notar também uma maior concentração das AMCs na faixa central e uma diminuição da diferença entre os maiores e menores valores. Isso chama a atenção para uma eventual convergência no emprego. Esta análise será mais explorada na parte destinada à análise empírica.

Observando a variação entre 1991 e 2000 podemos então concentrar a atenção na variável de interesse, a taxa de variação do emprego entre 1991 e 2000. Do ponto de vista da análise espacial, há um interesse em saber se também há padrões de aglomeração espacial na taxa de crescimento do emprego.

A medida mais usual para a mensuração de correlação espacial é o  $I$  de Moran. Os detalhes dessa estatística serão apresentados no capítulo 8, porém ela é facilmente compreendida a partir do gráfico de dispersão de Moran.





**Figura 3. 5** - Mapa, Diagrama de Dispersão de Moran, Estatísticas descritivas para taxa de crescimento do emprego formal (em ln) 1991-2000

O diagrama de dispersão de Moran é um diagrama usual de dispersão no qual a variável na abscissa é a variável em análise e a variável na ordenada é o cálculo da variável em análise defasada espacialmente.

Como já indicado no capítulo 2, para a análise espacial deve-se proceder à construção da matriz de pesos espaciais  $W$ . Essa matriz é quadrada ( $n \times n$ ) contendo os pesos espaciais de cada unidade sobre outra. Assim, o elemento  $w_{i,j}$  indica o peso espacial que a unidade  $j$  exerce sobre a unidade  $i$ . Dessa forma, a variável defasada espacialmente pode ser obtida através do produto matricial de  $W\mathbf{x}$ , onde  $\mathbf{x}$  é o vetor de variáveis originais. Com isso, o elemento  $(W\mathbf{x})_i$  pode ser obtido como:

$$(W\mathbf{x})_i = \sum_{j=1}^n w_{ij}x_j \quad (3.1)$$

Por (3.1) é possível ver que a variável defasada espacialmente é a somatória do produto das observações de todas as demais pelo peso que cada unidade  $j$  exerce em  $i$ . Com isso, a variável defasada espacialmente, torna-se um vetor que contém para cada unidade a ponderação espacial dos valores dos vizinhos. Conforme indicado no quadro explicativo na parte superior da figura (3.5), nos quadrantes do diagrama de dispersão H significa *high* e L *low*. Sendo assim, podemos descrever os quadrantes: HH apresenta observações com valores acima da média e para qual a defasagem espacial também está acima da média; HL apresenta observações com valores acima da média para o qual a defasagem espacial está abaixo da média; LL contém observações com valores abaixo da média para o qual a defasagem espacial está abaixo da média; LH contém observações com valores abaixo da média para o qual a defasagem espacial está acima da média.

A estatística  $I$  de Moran é simplesmente o valor do coeficiente de inclinação de uma reta de regressão para o diagrama apresentado. Note que traçado desta maneira, a indicação é de correlação espacial absoluta. A estatística de Moran, para análises econométricas, no entanto, é aplicada aos resíduos de uma regressão usual, com demais variáveis de controle, indicando então correlação espacial condicional.

Considerando uma reta de regressão no diagrama apresentado, quanto maior a aderência dessa reta, maior a evidência de correlação espacial. Quanto mais próximo de +1, mais forte a correlação espacial positiva e quanto mais próximo de -1, mais forte a correlação espacial negativa. Dessa forma, aglutinações ao longo dos quadrantes LL e HH indicam correlação positiva, ou seja, vizinhos parecidos, valores parecidos. Aglutinações ao longo de LH e HL indicam o inverso. Note que em ambos os casos, mais do que o sinal da correlação a indicação é da não aleatoriedade do comportamento espacial.

O valor da estatística de Moran para esses dados é 0.063 e seu valor padronizado 10.93 com um p-valor próximo de 0. A Hipótese nula é de não correlação espacial. O cálculo do valor padronizado utiliza a estimativa da esperança da estatística e do desvio padrão do  $I$  de Moran. O p-valor é obtido, então, de uma distribuição normal. As formas para cálculo da esperança e variância dessa estatística são apresentadas no capítulo 8.

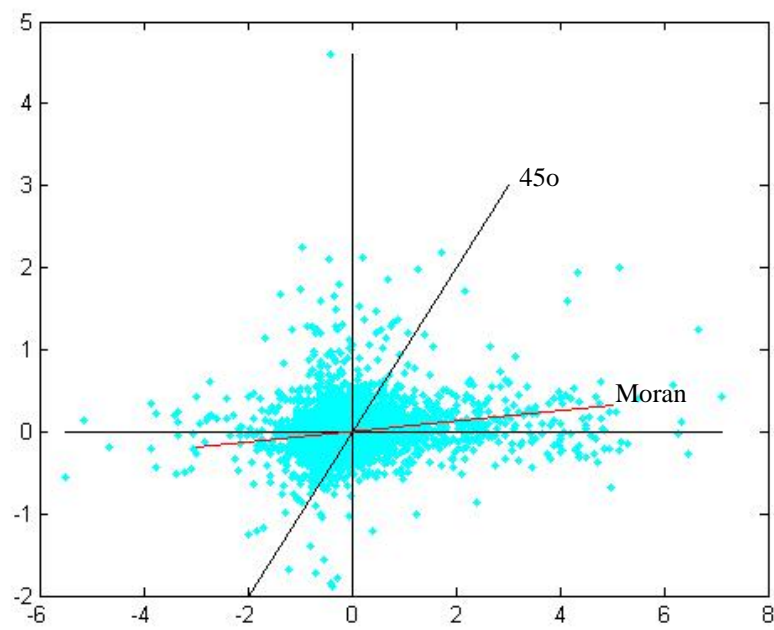
A figura (3.6) detalha o mesmo diagrama incluindo a reta de  $45^\circ$  como referência além da reta de ajuste com a inclinação de  $I$  de Moran.

Pela figura (3.5) é possível notar mais uma vez alguns blocos regionais na taxa de crescimento do emprego. Observando o diagrama de dispersão ali incluído, e a figura (3.6), nota-se uma concentração no quadrante LL, seguido do HH, indicadores de correlação espacial positiva.

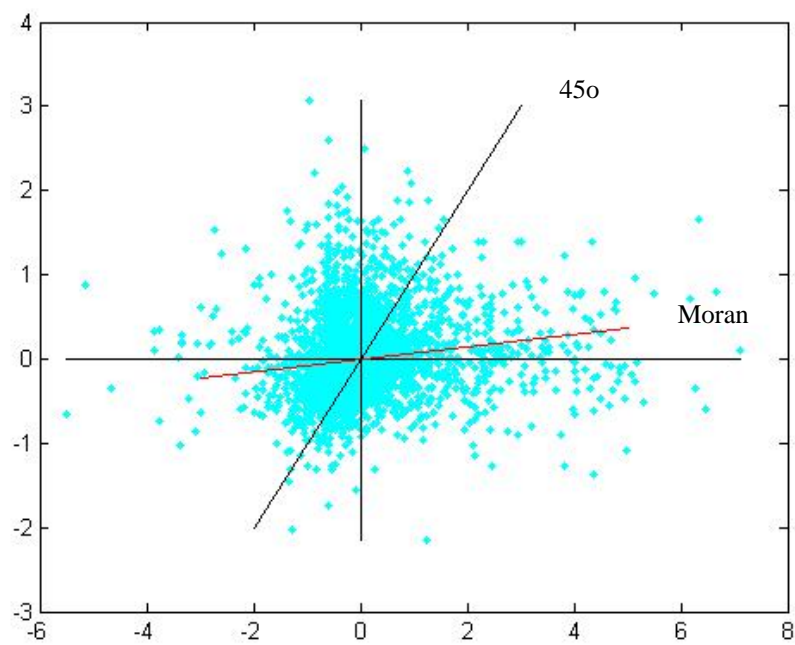
A aderência absoluta não se mostra tão forte nesses diagramas, mas já é possível notar a correlação bem como alguns *outliers* espaciais, que são os pontos longe do padrão do diagrama. Além disso, o p-valor da estatística é alto, indicando força na rejeição da hipótese nula, apesar do valor da inclinação estimada ser da ordem de 0.063

Um aspecto essencial para a ESDA e para o cálculo dessas estatísticas é o cálculo da variável defasada espacialmente, e, por consequência, a definição da matriz de ponderação  $W$ . Alterar essas matrizes pode significar alterar a significância e os valores dessas estatísticas. Ao longo dessa dissertação, a comparação usual será dos valores obtidos a utilizando uma matriz  $W$  construída a partir de distâncias (Dist) com os valores obtidos utilizando uma matriz  $W$  construída pelo critério de fronteira comum, conhecida como matriz de contigüidade (Contig).

A figura (3.7) apresenta os mesmo elementos da figura (3.6), porém calculados com a matriz de contigüidade. Nesse caso, o valor da estatística de Moran para esses dados é 0.07 e seu valor padronizado 7.96 com um p-valor próximo de 0. Note que é possível notar algumas diferenças no diagrama e o p-valor cai com o uso da matriz de contigüidade.



**Figura 3. 6** - Diagrama de Moran com Wdist



**Figura 3. 7** - Diagrama de Moran com Wcontig

#### **4. Questões de Investigação**

Conforme já indicado anteriormente, o foco de investigação dessa dissertação será a matriz de ponderação espacial. Sendo ela imposta exogenamente pelo investigador, algumas questões surgem de imediato sobre a adequação da escolha dessa matriz.

Existem diferentes aspectos passíveis de serem medidos, como medidas geográficas ou medidas sócio-econômicas, e dentro de cada aspecto, diferentes formas de se medir. Por exemplo, ao se medir distância geográfica, pode-se medir a partir do inverso da distância euclidiana, do inverso do quadrado da distância euclidiana, ou a simplificação mais comumente utilizada, a matriz de fronteiras/contigüidade. Ao se medir distâncias sócio-econômicas, pode-se utilizar a diferença na composição demográfica, diferenças de indicadores econômicos ou sociais, o quadrado dessa diferença, entre outros.

Uma questão específica de interesse é o quanto uma medida simplificada é capaz de captar os efeitos que seriam produzidos por uma medida mais complexa e vice-versa. Outra questão é qual a diferença obtida ao se optar por diferentes medidas para a composição da ponderação espacial. Além disso, procura-se entender se os diferentes modelos são capazes de diferenciar a forma de geração de dados, ou seja, se há uma forma para se escolher o modelo mais adequado.

## PARTE II – ECONOMETRIA ESPACIAL

**5. Modelos Espaciais, Heterogeneidade e Dependência Espacial**

De acordo com Anselin (1988), Lesage (1998a), Lesage (1999), Elhorst (2000) e Elhorst (2001), o uso de modelos tradicionais da economia e da econometria para análise de problemas que contenham algum componente espacial sofre de algumas limitações. Os dados coletados em situações onde esse componente espacial é relevante possuem características que ferem pressupostos básicos desses modelos. Dividem-se, de modo geral, as peculiaridades desses dados em dois grandes grupos: heterogeneidade espacial e a dependência ou autocorrelação espacial.

O primeiro tipo de problema citado, a heterogeneidade espacial, diz respeito à falta de estabilidade de comportamento ao longo do espaço, como por exemplo, estados ricos e pobres aglutinados em diferentes regiões de um país. Em termos de modelagem econométrica, segundo Elhorst (2000), a heterogeneidade espacial significa que os parâmetros não são homogêneos ao longo do *set* de dados, variando com a unidade. Segundo Anselin (1988), esse tipo de problema costuma ser resolvido por técnicas tradicionais de econometria.

Já a dependência ou autocorrelação espacial surge ao se questionar o princípio de independência entre os dados coletados, normalmente assumido em sets de *cross-section* quando analisados por técnicas tradicionais. De acordo com a primeira lei da geografia por Tobler *apud* Anselin (1988), tudo está relacionado a todo o resto, porém elementos mais próximos estão mais relacionados que elementos mais distantes<sup>2</sup> - de onde partem as frases de início de cada capítulo na parte seguinte.

Segundo Anselin(1999) e Ertur e Gallo(2003), a autocorrelação espacial pode ser definida como a coincidência de similaridade de valores com a similaridade de localização

---

<sup>2</sup> Do original: everything is related to everything else, but near things are more related than distant things

Nesse sentido, a dependência espacial é determinada por uma noção de espaço relativo ou distância relativa. De forma mais geral, pode-se considerar espaços diferentes do sentido Euclidiano estrito, incluindo-se distâncias sociais, políticas, etc, de forma a se visualizar o potencial para a aplicação em diversas áreas das ciências sociais.

Anselin (1988) indica que a dependência espacial pode surgir devido a alguns problemas de mensuração, como divisão das unidades espaciais, problemas em sua agregação, porém, principalmente devido a externalidades espaciais e a efeitos de *spill-over*. Elhorst (2000) indica, como exemplo, que um agente pode mudar suas decisões de acordo com condições de mercado na sua região em comparação com outras regiões e com a distância dessas regiões. Como sempre referenciado nos textos de econometria espacial, a dependência espacial é, de certa forma, similar à dependência temporal da análise de séries temporais; porém a grande diferença é que a influência espacial se dá de forma recíproca (a unidade A influencia B e vice-versa), e não como nas séries temporais onde ocorre sem a reciprocidade (t-1 influencia t, mas t não influencia t-1).

Nos modelos que consideram a componente espacial, a dependência espacial pode ser encontrada nas variáveis dependentes, no termo de erro ou mesmo nas variáveis independentes. As influências espaciais são, usualmente, resumidas em uma matriz de pesos  $W$ , onde o elemento  $w_{ij}$  indica a influência da unidade  $j$  na unidade  $i$ . Além disso,  $w_{ii} = 0$ , ou seja, a própria unidade não se influencia, e as linhas são usualmente normalizadas para a soma ser 1. Com isso, assume-se que a soma de todas as influências são iguais para cada unidade, além de se permitir a comparabilidade (Case and Rosen, 1993; Anselin, 1988; Lesage, 1998; Lesage 1999a). Uma outra implicação da matriz normalizada é restringir a amplitude de possibilidade dos parâmetros espaciais para o intervalo de (-1) a (+1).

Como indicado no capítulo 3, introdutório à ESDA, qualquer variável defasada espacialmente pode ser obtida através do produto matricial de  $W\mathbf{x}$ , onde  $\mathbf{x}$  é o vetor de variáveis originais. Conforme (3.1), o elemento  $(W\mathbf{x})_i$  pode ser obtido como:

$$(W\mathbf{x})_i = \sum_{j=1}^n w_{ij}x_j \quad (3.1)$$

De forma equivalente, qualquer matriz genérica  $X$ , de  $n$  observações e qualquer número de variáveis, se pré-multiplicada por  $W$  gera uma nova matriz  $WX = W*X$  contendo cada variável defasada espacialmente.

Anselin (1988) afirma que quando o termo de erro em um modelo de regressão apresenta dependência espacial, o pressuposto padrão de uma matriz de covariância de erro esférica não se mantém. Isso indica uma perda de eficiência, e a invalidade do pressuposto da homoscedasticidade, porém, não gera viés nem inconsistência nos estimadores. Já quando a variável dependente possui correlação espacial, ignorar esse aspecto pode ocasionar um viés de omissão de variável, o que é prejudicial para a estimação e para a inferência..

A presença da algum grau de dependência espacial é verificado através de alguns testes específicos, como o  $I$  de Moran, aplicado a resíduos de uma regressão linear, ou o teste de Wald, *Likelihood* Ratio, e o Multiplicador de Lagrange (LM), baseados na estimação por máxima verossimilhança. Esses testes serão descritos nos capítulos seguintes.

Os principais modelos de econometria espacial partem do chamado modelo completo (SAC – *Spatial mixed regressive Auto-regressive Complete*):

$$\begin{aligned} \mathbf{y} &= \rho W_1 \mathbf{y} + X\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda W_2 \mathbf{u} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \end{aligned} \tag{5.1}$$

Onde:

$\mathbf{y}$  = vetor de variáveis dependentes

$W_1$  e  $W_2$  = matrizes de pesos espacial

$\rho$  e  $\lambda$  = coeficientes de correlação espacial na variável dependente e no termo de erro respectivamente

$\boldsymbol{\beta}$  = vetor de coeficientes

$X$  = matriz de variáveis independentes

Assumindo valor igual a 0 para alguns dos parâmetros, é possível chegar a modelos mais simplificados.



Com  $\rho = 0$  e  $\lambda = 0$ , teríamos um modelo usual, sem a presença do componente espacial.

Com  $\lambda = 0$  e  $X = 0$ , temos apenas  $\mathbf{y}$  perfeitamente autocorrelacionado espacialmente, chamado de modelo de autocorrelação de primeira ordem (FAR – *First order Auto-Regressive*):

$$\mathbf{y} = \rho W_1 \mathbf{y} + \boldsymbol{\varepsilon} \quad (5.2)$$

Com  $\lambda = 0$ , temos o modelo chamado semi autocorrelacionado-regressivo (SAR – *Spatial mixed regressive Auto-Regressive*):

$$\mathbf{y} = \rho W_1 \mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (5.3)$$

Com  $\rho = 0$ , temos o modelo de erro autocorrelacionado espacialmente (SEM – *Spatial Error Model*):

$$\begin{aligned} \mathbf{y} &= X\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda W_2 \mathbf{u} + \boldsymbol{\varepsilon} \end{aligned} \quad (5.4)$$

Além desses modelos, há também um outro chamado Durbin Model, que incorpora a dependência espacial nas variáveis independentes.

$$\mathbf{y} = \rho_1 W_1 \mathbf{y} + \rho_2 W_1 X\boldsymbol{\beta} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (5.5)$$

Ainda, alguns modelos podem utilizar apenas a correlação espacial nas variáveis explicativas:

$$\mathbf{y} = X\boldsymbol{\beta} + \rho W_1 Z\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (5.6)$$

Onde  $Z$  é uma matriz com um *sub-set* de  $X$ .

Este modelo, no entanto, como será visto no capítulo 7, deve ser usado de forma criteriosa.

Os modelos mais utilizados na literatura são o SAC (5.1), SAR (5.3) e SEM (5.4). Estes serão os modelos utilizados nas simulações e na análise empírica dessa dissertação.

Em termos econométricos, os autores indicam que os estimadores de modelos espaciais obtidos pelo método tradicional dos mínimos quadrados são inapropriados. Segundo Elhorst(2000) e Lesage(1999a) e Anselin(1988), caso haja autocorrelação no termo de erro, os estimadores de mínimos quadrados continuam sem viés, porém perdem sua eficiência. No caso de autocorrelação na variável dependente, se estimados por OLS, os estimadores se tornam viesados e perdem a consistência. Por isso, métodos alternativos precisam ser utilizados para a estimação, como máxima verossimilhança, GMM ou variáveis instrumentais. Esses tópicos serão explorados nos capítulos seguintes.

## 6. O Problema da Estimação

Se considerarmos o modelo mais simples de todos, o FAR (5.2), considerando  $\mathbf{y}$  como desvios da média, pode-se ver como a estimação por OLS será viesada e inconsistente:

$$\mathbf{y} = \rho W_1 \mathbf{y} + \boldsymbol{\varepsilon} \quad (5.2)$$

Se considerarmos  $r$  um estimador OLS para  $\rho$ , este deve minimizar a soma dos quadrados dos resíduos, dado por  $Q$ :

$$\begin{aligned} Q &= (\mathbf{y} - r W_1 \mathbf{y})' (\mathbf{y} - r W_1 \mathbf{y}) \\ Q &= \mathbf{y}' \mathbf{y} - \mathbf{y}' r W_1 \mathbf{y} - (r W_1 \mathbf{y})' \mathbf{y} + (r W_1 \mathbf{y})' (r W_1 \mathbf{y}) \end{aligned}$$

como  $r$  e  $(r W_1 \mathbf{y})' \mathbf{y}$  são escalares:

$$Q = \mathbf{y}' \mathbf{y} - 2r (W_1 \mathbf{y})' \mathbf{y} + (W_1 \mathbf{y})' (W_1 \mathbf{y}) r^2 \quad (6.1)$$

Para minimizar essa expressão, basta tomar a derivada parcial em relação a  $r$  e igualar a 0:

$$\begin{aligned} \frac{\partial Q}{\partial r} &= -2(W_1 \mathbf{y})' \mathbf{y} + 2(W_1 \mathbf{y})' (W_1 \mathbf{y}) r = 0 \\ 2(W_1 \mathbf{y})' (W_1 \mathbf{y}) r &= 2(W_1 \mathbf{y})' \mathbf{y} \\ r &= [(W_1 \mathbf{y})' (W_1 \mathbf{y})]^{-1} (W_1 \mathbf{y})' \mathbf{y} \end{aligned} \quad (6.2)$$

Note que este estimador é análogo ao estimar usual  $\mathbf{b}$  obtido por OLS:

$$\mathbf{b} = [(X)'(X)]^{-1} (X)' \mathbf{y}$$

Nesse caso particular,  $X = W_1 \mathbf{y}$

Para a prova usual de não viés, a expressão da população (5.2) é substituída no  $\mathbf{y}$  à direita da expressão para o estimador (6.2). Fazendo o mesmo, temos:

$$\begin{aligned} r &= [(W_1 \mathbf{y})' (W_1 \mathbf{y})]^{-1} (W_1 \mathbf{y})' (\rho W_1 \mathbf{y} + \boldsymbol{\varepsilon}) \\ r &= [(W_1 \mathbf{y})' (W_1 \mathbf{y})]^{-1} [(W_1 \mathbf{y})' (W_1 \mathbf{y})] \rho + [(W_1 \mathbf{y})' (W_1 \mathbf{y})]^{-1} (W_1 \mathbf{y})' \boldsymbol{\varepsilon} \\ r &= \rho + [(W_1 \mathbf{y})' (W_1 \mathbf{y})]^{-1} (W_1 \mathbf{y})' \boldsymbol{\varepsilon} \end{aligned} \quad (6.3)$$

Tomando-se esperanças temos que:

$$E(r) = \rho + E\{[(W_1\mathbf{y})'(W_1\mathbf{y})]^{-1}(W_1\mathbf{y})'\boldsymbol{\varepsilon}\}$$

Pela lei das expectativas iteradas:

$$\begin{aligned} E(r) &= \rho + E\{E\{[(W_1\mathbf{y})'(W_1\mathbf{y})]^{-1}(W_1\mathbf{y})'\boldsymbol{\varepsilon} \mid W_1\mathbf{y}\}\} \\ E(r) &= \rho + E\{[(W_1\mathbf{y})'(W_1\mathbf{y})]^{-1}(W_1\mathbf{y})'E(\boldsymbol{\varepsilon} \mid W_1\mathbf{y})\} \end{aligned} \quad (6.4)$$

Assim, para que o segundo termo seja = 0, como no caso tradicional OLS, é necessário que:

$$E(\boldsymbol{\varepsilon} \mid W_1\mathbf{y}) = 0$$

Porém:

$$E(\boldsymbol{\varepsilon} \mid W_1\mathbf{y}) = E(\mathbf{y} - \rho W_1\mathbf{y} \mid W_1\mathbf{y}) = E(\mathbf{y} \mid W_1\mathbf{y}) - \rho W_1\mathbf{y} = \rho W_1\mathbf{y} + \boldsymbol{\varepsilon} - \rho W_1\mathbf{y} = \boldsymbol{\varepsilon} \quad (6.5)$$

Dessa forma, pode-se notar que, exceto no caso em que  $\rho = 0$ <sup>3</sup>, a esperança condicional do erro não é 0, e o erro é correlacionado, provocando um viés no estimador.

Para checar a consistência, temos que checar se o limite em probabilidade do estimador  $r$  é o parâmetro da população.

Pode-se observar que:

$$[(W_1\mathbf{y})'(W_1\mathbf{y})] = \sum_{i=1}^N (W_1\mathbf{y})_i^2$$

e

$$(W_1\mathbf{y})'\boldsymbol{\varepsilon} = \sum_{i=1}^N (W_1\mathbf{y})_i \varepsilon_i$$

Com isso é possível reescrever a equação de  $r$  (6.3) como:

---

<sup>3</sup> Note que se  $\rho = 0$ , no modelo original teríamos apenas  $\mathbf{y} = \boldsymbol{\varepsilon}$  e

$E(\boldsymbol{\varepsilon} \mid W_1\mathbf{y}) = E(\mathbf{y} - \rho W_1\mathbf{y} \mid W_1\mathbf{y}) = E(\mathbf{y} \mid W_1\mathbf{y}) - 0 = \mathbf{y}$  e nesse modelo,  $E(\mathbf{y}) = 0$

$$r = \rho + \left[ \frac{\sum_{i=1}^N (W_1 \mathbf{y})_i^2}{N} \right]^{-1} \left[ \frac{\sum_{i=1}^N (W_1 \mathbf{y})_i \varepsilon_i}{N} \right] \quad (6.6)$$

Tomando o *plim* de ambos os lados e utilizando a Lei dos Grandes números:

$$p \lim(r) = \rho + E[(W_1 \mathbf{y})_i^2]^{-1} E[(W_1 \mathbf{y})_i \varepsilon_i] \quad (6.7)$$

Nas demonstrações de estimadores OLS, a segunda esperança do último termo da equação (6.7) seria igual a 0, o que garantiria a consistência. No caso espacial, como já visto na demonstração de viés (6.4 e 6.5), esta esperança não é zero, exceto no caso trivial em que  $\rho = 0$ .

Assim, na presença de correlação espacial, o estimador OLS é viesado e inconsistente.

Quando a correlação espacial está presente no termo de erro, não há viés e inconsistência, porém o estimador deixa de ser o mais eficiente, pois a matriz de variância-covariância perde sua estrutura diagonal.

Tomando por base o modelo SEM (5.4):

$$\begin{aligned} \mathbf{y} &= X\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda W_2 \mathbf{u} + \boldsymbol{\varepsilon} \end{aligned} \quad (5.4)$$

É possível reescrever  $\mathbf{u}$  como:

$$\mathbf{u} = (I_n - \lambda W_2)^{-1} \boldsymbol{\varepsilon} \quad (6.8)$$

Sendo  $I_n$  a matriz identidade ( $n \times n$ ).

Continuando com o pressuposto de que  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ , a matriz de variância-covariância:

$$\begin{aligned}\Omega &= E[\mathbf{u}\mathbf{u}'] = [(I_n - \lambda W_2)^{-1} \boldsymbol{\varepsilon}][(I_n - \lambda W_2)^{-1} \boldsymbol{\varepsilon}]' \\ &= (I_n - \lambda W_2)^{-1} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' (I_n - \lambda W_2)^{-1} \\ &= \sigma^2 (I_n - \lambda W_2)^{-1} [(I_n - \lambda W_2)^{-1}]'\end{aligned}\tag{6.9}$$

Assim, é possível ver que o termo  $(I_n - \lambda W_2)^{-1} [(I_n - \lambda W_2)^{-1}]'$  dará uma forma não diagonal à matriz, o que reduz sua eficiência, se  $\mathbf{u}$  for estimado assumindo-se  $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ . No entanto, como já colocado, esse pressuposto está relacionado apenas com a eficiência e não gera viés nem inconsistência na estimação.

O capítulo seguinte apresenta, então, os métodos para solucionar os problemas aqui apresentados.

## 7. Estimando coeficientes em modelos espaciais

Como a estimação usual por OLS não atende a requisitos mínimos quando existe a presença de dependência espacial na variável dependente e como os parâmetros perdem eficiência se a dependência ocorrer apenas no erro, é necessária a busca de formas adequadas para estimação dos parâmetros.

Os métodos aplicados são a estimação por máxima verossimilhança (ML), variáveis instrumentais (IV) e método generalizado dos momentos (GMM).

Para estimar um modelo por ML é necessário, em primeiro lugar, encontrar sua função conjunta de probabilidade, para então maximizá-la em um vetor de parâmetros.

O modelo mais amplo, o SAC (5.1) possui a seguinte forma:

$$\begin{aligned} \mathbf{y} &= \rho W_1 \mathbf{y} + X\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda W_2 \mathbf{u} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \end{aligned} \tag{5.1}$$

Onde podemos chamar a matriz de variância-covariância do resíduo por uma matriz  $\Omega$ .

O vetor de parâmetros a ser estimado é  $\theta = [\rho, \boldsymbol{\beta}', \lambda, \sigma^2]$ , já focando o caso homoscedástico por simplificação.

A equação (5.1) pode ser reescrita como

$$\begin{aligned} (I - \rho W_1) \mathbf{y} &= X\boldsymbol{\beta} + \mathbf{u} \\ (I - \lambda W_2) \mathbf{u} &= \boldsymbol{\varepsilon} \end{aligned} \tag{7.1}$$

Fazendo as seguintes simplificações de notação, conforme Anselin (1988):

$$(I - \rho W_1) = A \tag{7.2}$$

$$(I - \lambda W_2) = B \tag{7.3}$$

Podemos agora reescrever (7.1) como:

$$A\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u} \quad (7.4)$$

$$B\mathbf{u} = \boldsymbol{\varepsilon} \quad (7.5)$$

Além disso, considerando o erro homoscedástico e a matriz de variância-covariância do resíduo  $\Omega$ , podemos encontrar um vetor  $\mathbf{v}$  de erros homoscedásticos padronizados, dado por:

$$\begin{aligned} \mathbf{v} &= \Omega^{-1/2} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \Omega^{1/2} \mathbf{v} \end{aligned} \quad (7.6)$$

Utilizando (7.4), (7.5) e (7.6) e escrevendo a equação (7.1) em função de  $\mathbf{v}$ , obtemos:

$$\begin{aligned} A\mathbf{y} &= X\boldsymbol{\beta} + B^{-1}\Omega^{1/2}\mathbf{v} \\ \mathbf{v} &= \Omega^{-1/2}B(A\mathbf{y} - X\boldsymbol{\beta}) \\ \mathbf{v} &\sim N(0, I) \end{aligned} \quad (7.7)$$

Como  $\mathbf{v}$  possui distribuição normal padrão, poderíamos encontrar a função de log-verossimilhança usual derivada a partir da distribuição normal. Utilizando o resultado de Greene (2003: 493) teríamos:

$$l = \ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma_v^2) - \frac{1}{2\sigma_v^2} (\mathbf{v}'\mathbf{v}) \quad (7.8)$$

Como  $\mathbf{v}$  possui uma distribuição normal padronizada, sabemos que  $\sigma_v^2=1$ , e assim, a expressão (7.8) simplifica para:

$$l = \ln L = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} (\mathbf{v}'\mathbf{v}) \quad (7.9)$$

No entanto, para encontrarmos uma função de verossimilhança adequada para a estimação do vetor de parâmetros  $\boldsymbol{\theta}$ , precisaríamos partir da distribuição de  $\mathbf{y}$  e não de  $\mathbf{v}$ . A transformação pode ser feita, a partir da técnica de transformações de variáveis indicada em Greene (2003: 858), multiplicando  $L$ , pelo termo Jacobiano, ou de forma equivalente somando  $l$  ao  $\ln$  do termo Jacobiano definido como:



$$J = \left| \frac{\partial \mathbf{v}}{\partial \mathbf{y}} \right| = \left| \Omega^{-1/2} B A \right| = \left| \Omega^{-1/2} \right| \times |B| \times |A| \quad (7.10)$$

Somando o logaritmo natural de (7.10) à expressão (7.9) e rearranjando os termos, encontramos a função de log-verossimilhança para a estimação de modelos espaciais:

$$l = \ln L = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln|\Omega| + \ln|B| + \ln|A| - \frac{1}{2} (\mathbf{v}' \mathbf{v}) \quad (7.11)$$

$$\mathbf{v}' \mathbf{v} = (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' B' \Omega^{-1} B (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Considerando que  $\Omega = \sigma^2 \mathbf{I}$  e que  $|\sigma^2 \mathbf{I}| = \sigma^{2n} |\mathbf{I}| = \sigma^{2n}$  podemos ainda escrever (7.11) como:

$$l = \ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln|B| + \ln|A| - \frac{(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' B' B (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2} \quad (7.12)$$

A partir da expressão (7.11) ou (7.12) são derivadas as funções de modelos mais simplificados, assumindo valor 0 para alguns parâmetros, e são derivadas as estatísticas de testes que utilizam a estimação de verossimilhança.

Para uma referência de possíveis variações, Elhorst (2000) apresenta diversas alterações da função de máxima-verossimilhança, incluindo inclusive a componente temporal.

Uma alternativa para estimar as regressões espaciais é utilizar instrumentos de forma a eliminar a simultaneidade produzida pela presença da variável dependente defasada espacialmente. Este método é indicado por Brueckner (2003) e Anselin (1988, 1999). O instrumento indicado para a estimação é  $WX$ .

Um bom instrumento, em termos econométricos, deve ser correlacionado com a variável endógena e não deve estar correlacionada com o erro.

Se considerarmos que:

$$\text{Corr}(X, \mathbf{y}) \neq 0 \quad (7.13)$$

$$\text{Cov}(X, \boldsymbol{\varepsilon}) = E(X' \boldsymbol{\varepsilon}) = \mathbf{0} \quad (7.14)$$

Podemos deduzir que:

$$\text{Corr}(WX, W\mathbf{y}) \neq 0 \quad (7.15)$$

$$\text{Cov}(WX, \boldsymbol{\varepsilon}) = W' E(X' \boldsymbol{\varepsilon}) = \mathbf{0} \quad (7.16)$$

Assim,  $WX$  atende aos requisitos necessários para um bom instrumento. Kelejian and Prucha (1998) e Kelejian and Robinson (1993) trazem as discussões teóricas e técnicas para a estimação correta das diferentes formas de dependência espacial através de regressão em dois estágios.

No caso do modelo SAR (5.3), bastaria realizar uma regressão tradicional de variáveis instrumentais (IV) com  $WX$  e  $X$  como instrumentos para  $W\mathbf{y}$ . Nesse caso, os parâmetros seriam estimados por OLS a partir da seguinte especificação:

$$\mathbf{y} = \rho \hat{\mathbf{z}} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (7.17)$$

Em que  $\hat{\mathbf{z}}$  é o vetor das estimações de  $W\mathbf{y}$  a partir da regressão, também estimada por OLS:

$$W\mathbf{y} = X\boldsymbol{\gamma} + WX\boldsymbol{\pi} + \boldsymbol{\xi} \quad (7.18)$$

Desse modo:

$$\hat{\mathbf{z}} = X\hat{\boldsymbol{\gamma}} + WX\hat{\boldsymbol{\pi}} \quad (7.19)$$

Por essa técnica, a endogeneidade de  $W\mathbf{y}$  seria eliminada no primeiro estágio. Essa é exatamente a especificação sugerida por Brueckner (2003).

Note que essa técnica não estima o coeficiente de correlação espacial nos resíduos, se houver. No entanto isso não gera problemas de viés nem de consistência, devido à natureza do impacto da dependência espacial do erro. De acordo com Kelejian and Prucha (1998) e Brueckner (2003), as estimativas dos parâmetros de (7.17) serão consistentes mesmo na presença de uma estrutura de erro autocorrelacionado espacialmente.

Em alguns casos é usada uma versão simplificada do modelo, incluindo-se  $WX$  e deixando  $WY$  incluído no termo de erro. Essa abordagem é similar à equação (5.6), porém com  $Z$  praticamente igual a  $X$ . Isso, no entanto, deve ser feito de forma criteriosa, pois pode gerar um viés de omissão de variável mais forte do que ignorar todas as variáveis espaciais. Caso seja válido o argumento em (7.15) e (7.16), em favor do uso de  $WX$  como um bom instrumento, incluí-lo, deixando  $WY$  no erro estaria teoricamente e tecnicamente incorreto, visto que teríamos um viés de omissão de variável pela condição (7.15) caso  $\rho \neq 0$ . No entanto, caso a defasagem espacial de apenas algumas variáveis de  $X$  seja incluído, não havendo correlação desta com  $WY$ , a simplificação seria válida.

Por fim é utilizada a estimação pelo método generalizado dos momentos (GMM), em especial a partir de Kelejian and Prucha (1995 e 1998). O método dos momentos utiliza também uma especificação de dois estágios, porém colocando mais defasagens espaciais de  $X$ . Assim, o primeiro estágio da regressão base do método de GMM segue:

$$W_1y = X\beta + \pi WX\beta + \eta W^2X\beta + \xi \quad (7.20)$$

De acordo com Anselin(1999), defasagens superiores podem ser adicionadas, mas isso deve ser feito criteriosamente. O detalhamento teórico e técnico dos modelos de GMM, em especial para a estimação do coeficiente de correlação espacial nos resíduos foge ao escopo dessa dissertação.

Vale ressaltar algumas implicações teóricas e, ainda mais, práticas do contraste entre as estimações de ML versus as estimações por IV e GMM. IV e GMM são modelos em princípio mais robustos que os modelos de máxima verossimilhança, pois não assumem, a priori, a normalidade dos resíduos, além de permitirem versões robustas das estimações.

Um aspecto prático muito importante é que a maioria dos pacotes estatísticos ainda não possui rotinas e algoritmos para a estimação de modelos espaciais. Essas estimativas devem, então, ser feitas em softwares específicos, como GeoDa e SpaceStat, ou através do uso da biblioteca de funções de Matlab disponibilizada por LeSage(1999), o que é feito nos aspectos operacionais dessa dissertação.

A estimação por IV, ou do modelo SAR-GMM, no entanto pode ser facilmente realizada em qualquer software estatístico desde que as variáveis defasadas espacialmente tenham sido computadas. Por fim, a estimação por ML, como poderá ser comprovado nas partes dessa dissertação dedicada às simulações e análises empíricas, é visivelmente mais custosa em termos de tempo do que as demais. A única exceção é a estimação do modelo SEM por GMM que é mais custosa em termos de capacidade de memória e exige muito tempo de acordo com o tamanho da amostra.

## 8. Testando a Presença do Componente Espacial e Especificando o Modelo

Como usual nas especificações econométricas, existem testes para verificar a adequação de determinada especificação e/ou seus pressupostos para que a inferência seja válida. No caso dos modelos espaciais existem testes baseados na estimação por OLS outros na estimação por máxima-verossimilhança.

Certamente o teste mais usado e citado para verificar a presença do componente espacial no erro é o  $I$  de Moran (Moran, 1950 e Anselin, 1988 e 1999). A estatística  $I$  de Moran tem a seguinte forma:

$$I = \frac{n}{s} \frac{\mathbf{u}'W\mathbf{u}}{\mathbf{u}'\mathbf{u}} \sim N(E(I), Var(I)) \quad (8.1)$$

Onde  $\mathbf{u}$  é o vetor de erros de uma regressão OLS,  $n$  é o número de observações e  $s$  é a soma de todos os elementos da matriz  $W$ . Como  $s = \sum_i \sum_j w_{ij}$ , no caso de uma matriz padronizada para a soma de linha igualar 1,  $\sum_j w_{ij} = 1$  e, portanto,  $s = \sum_i 1 = n$ .

Assim, o termo  $\frac{n}{s}$  simplifica para 1 e a estatística  $I$  pode ser escrita como:

$$I = \frac{\mathbf{u}'W\mathbf{u}}{\mathbf{u}'\mathbf{u}} \quad (8.2)$$

A inferência a partir da estatística  $I$  de Moran é feita normalizando-se a variável e utilizando as inferências usuais. A derivação das expressões para a esperança e variância tornam-se mais complexas pela presença do termo de erro. A partir de Anselin(1988:102), assumem a seguinte forma:

$$E(I) = \frac{n}{s} \frac{tr(MW)}{(n-k)} \quad (8.3)$$

$$Var(I) = \left(\frac{n}{s}\right)^2 \frac{tr(MWMW') + tr[(MW)^2] + [tr(MW)]^2}{(n-k)(n-k+2)} - [E(I)]^2 \quad (8.4)$$

Onde:

$M = I_n - X(X'X)^{-1}X'$ , matriz de projeção, com  $I_n$  sendo a matriz identidade.

$k$  é o número de variáveis explicativas

$tr$  indica o operador traço

A estatística  $I$  de Moran tem aspectos positivos e negativos evidenciados em alguns estudos (Acosta el, 2003, Keleijian e Robinson, 1998, Fingleton, 1999, Anselin e Florax, 1995). Tais estudos mostram que o teste de Moran apresenta performance melhores que outros indicadores em termos de poder do teste para identificar correlações espaciais, porém não é capaz de diferenciar a correlação espacial nos erros da correlação espacial na variável dependente.

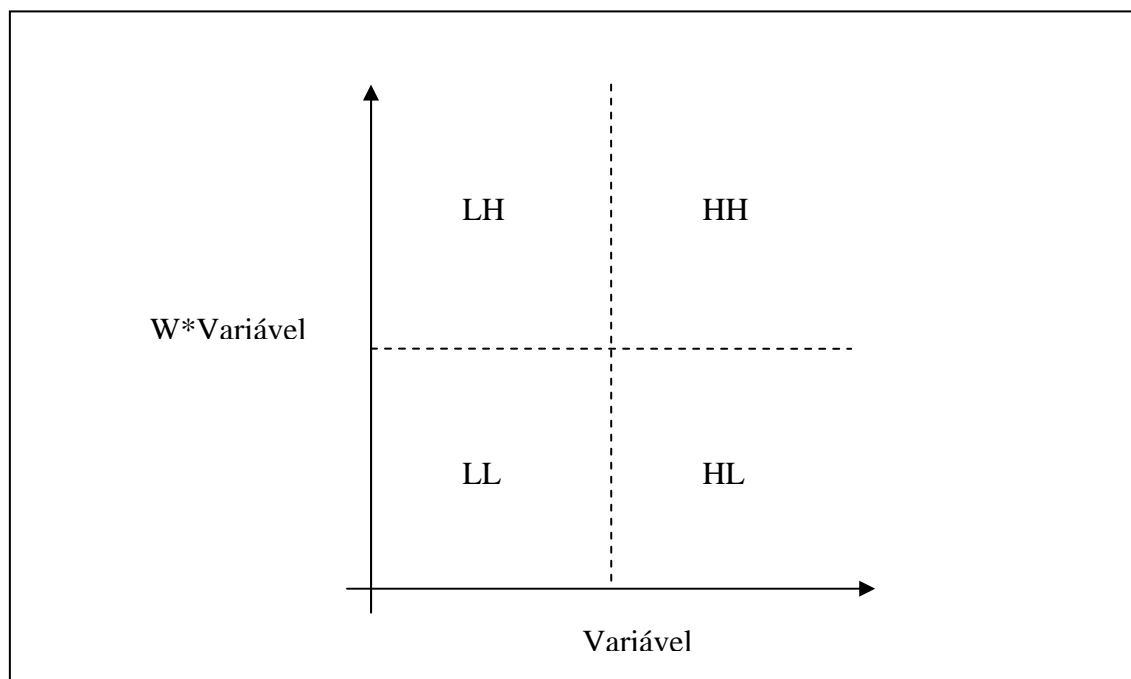
Esse fato, apesar de ser ruim para a especificação adequada do modelo, é simples de ser compreendido: caso a correlação espacial exista na variável dependente, como no modelo SAR, e não no resíduo, como no modelo SEM, após a regressão de um modelo que não considere defasagens espaciais, a variável dependente defasada estará presente no termo de erro dessa regressão. Assim, haverá algum sinal de correlação espacial no resíduo, mesmo que isso tenha sido gerado pela dependência em  $Y$  e não nas demais variáveis não-observáveis incluídas no erro.

Com isso, pode-se concluir que o  $I$  de Moran é um bom identificador de correlação espacial, porém não um bom identificador de qual forma de correlação espacial está presente.

Uma aplicação derivada do  $I$  de Moran é o gráfico de dispersão de Moran, como já indicado no capítulo 3, de ESDA. Note que (8.2) é o equivalente ao estimador da regressão:  $Wu = Iu + \xi$ . Poderíamos então, traçar um diagrama de dispersão de  $u$  contra  $Wu$ . Em geral, no entanto, são feitos diagramas exploratórios de dispersão de  $y$  contra  $Wy$ . De modo geral, pode-se fazer um diagrama de qualquer variável contra sua defasagem espacial.

Se calculada a estatística ou feito o diagrama com a variável bruta, a evidência buscada é de correlação espacial absoluta, já se feito com os resíduos de uma regressão, a busca é de correlação espacial condicional.

Considerando a variável de estudo como desvios em relação à sua média, é possível traçar quatro quadrantes, conforme o quadro esquemático abaixo:



**Figura 8. 1-** Diagrama de Dispersão de Moran

Nos quadrantes acima, H significa *high* e L *low*. HH apresenta observações com valores acima da média e para qual a defasagem espacial também está acima da média; HL apresenta observações com valores acima da média para o qual a defasagem espacial está abaixo da média; LL contém as observações com valores abaixo da média para o qual a defasagem espacial está abaixo da média; por fim, LH engloba as observações com valores abaixo da média para o qual a defasagem espacial está acima da média.

Com esses quadrantes podemos formar dois grupos: LL e HH; LH e HL. No primeiro grupo temos evidências de correlação espacial positiva, uma vez que a variação da variável em estudo caminha no mesmo sentido da variação da variável defasada espacialmente. No outro grupo, a evidência é inversa. A inclinação de uma reta de regressão traçada seria o valor de estatística de  $I$  de Moran para a variável em questão. Se for maior que zero, há evidência de correlação positiva (a reta estará nos quadrantes do primeiro grupo), e se for menor que zero, teremos o inverso. O quão melhor for o *fit* dos pontos à reta, mais forte a evidência da correlação espacial. No entanto, para se obter uma estimativa correta do coeficiente de correlação espacial, é necessário recorrer a alguns dos métodos apresentando anteriormente devido aos problemas também já mencionados nessa dissertação.

Existe um outro grupo de testes que é baseado na estimação por máxima verossimilhança. Estes são o testes de Wald, a razão de máxima verossimilhança (*likelihood ratio* - LR) e os testes do multiplicador de Lagrange(LM).

Anselin (1988) indica que os dois primeiros testes, Wald e LR receberam mais atenção da literatura de econometria espacial até então. No entanto, o próprio Anselin em texto posteriores (Anselin 1996 e 1999), e Acosta et al (2003) indicam que a literatura mais recente focou os testes LM em diferentes formas e este é o teste indicado por Anselin (2003) para a escolha da especificação mais adequada.

A grande vantagem do teste LM é que, apesar de ser baseado na lógica da estimação por máxima verossimilhança, ele exige apenas os resíduos e parâmetros de uma regressão OLS usual, o que reduz em muito o custo de cálculo das estatísticas.

O teste de Moran, como visto, possui como hipótese nula a inexistência de dependência espacial e como hipótese alternativa alguma estrutura de dependência espacial não claramente definida. Já os testes baseados na estimação de máxima verossimilhança, ao contrário, são estruturados de forma específica em cima de hipóteses nula e alternativa sobre os parâmetros de correlação espacial  $\rho$  e  $\lambda$ , sendo nula a hipótese de que um deles é igual a zero.



O teste de Wald tem com  $H_0: \lambda=0$ , sendo a hipótese alternativa o modelo SEM(5.4). A sua expressão para esse modelo espacial, é:

$$Wald = \hat{\lambda}^2 \left[ t_2 + t_3 - \frac{(t_1)^2}{n} \right] \sim \chi_1^2 \quad (8.5)$$

Onde, a partir de (7.3):

$$(I_n - \hat{\lambda}W) = \hat{B}$$

Temos:

$$t_1 = tr(W)\hat{B}^{-1}$$

$$t_2 = tr(W\hat{B}^{-1})^2$$

$$t_3 = tr[(W\hat{B}^{-1})'(W\hat{B}^{-1})]$$

Onde  $I_n$  é a matriz identidade e  $tr$  indica o operador traço.

A estatística de Wald possui distribuição qui-quadrado com 1 grau de liberdade, sendo feita a inferência a partir de então.

O teste LR para a presença da dependência espacial no erro parte da função de log-verossimilhança dos modelos espaciais (7.12) já para o caso em que  $\rho=0$ :

$$l = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln|B| + -\frac{(\mathbf{y} - X\boldsymbol{\beta})' B' B (\mathbf{y} - X\boldsymbol{\beta})}{2\sigma^2} \quad (8.6)$$

Sob  $H_0: \lambda=0$ , sendo (8.6) reescrita como:

$$l = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + -\frac{(\mathbf{y} - X\boldsymbol{\beta})' (\mathbf{y} - X\boldsymbol{\beta})}{2\sigma^2} \quad (8.7)$$

Dessa forma, o teste LR se baseia no cálculo da diferença entre (8.6) e (8.7) utilizando os parâmetros obtidos a partir de uma regressão SEM por ML em (8.6) e OLS em (8.7).

Como, no modelo SEM (5.4) temos que:

$$\mathbf{u} = \lambda W\mathbf{u} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} = (I_n - \lambda W)\mathbf{u} \quad (8.8)$$

A variância estimada,  $s_1^2$  seria:

$$\frac{\mathbf{\varepsilon}'\mathbf{\varepsilon}}{n} = \frac{\mathbf{u}'(I_n - \lambda W)'(I_n - \lambda W)\mathbf{u}}{n} = \frac{(\mathbf{y} - X\mathbf{\beta})' B' B(\mathbf{y} - X\mathbf{\beta})}{n} \quad (8.9)$$

Dessa forma, a expressão (8.6) avaliada com os parâmetros estimados por ML, com  $s_1^2$  no lugar de  $\sigma^2$ , se torna:

$$l = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln s_1^2 + \ln|B| + -\frac{n}{2} \quad (8.10)$$

Por transformação semelhante da variância estimada por OLS, chamando-a de  $s_0^2$ , é possível reescrever (8.7) utilizando os parâmetros OLS como:

$$l = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln s_0^2 + -\frac{n}{2} \quad (8.11)$$

Dessa forma, fazendo a duas vezes a diferença de (8.10) e (8.11), e ajustando os termos, temos a estatística LR:

$$LR = -n(\ln s_0^2 - \ln s_1^2) + 2\ln|B| \sim \chi_1^2 \quad (8.12)$$

Assim como a estatística de Wald, LR possui distribuição qui-quadrado com 1 grau de liberdade, sendo feita a inferência a partir dessa distribuição.

Como já foi mencionado e é possível ver pelas expressões desses últimos dois testes, eles utilizam valores de parâmetros e erros estimados por ML. A família de testes do multiplicador de Lagrange (LM) utiliza apenas os erros e parâmetros da regressão OLS.

A estatística LM é baseada nas condições de primeira ordem da otimização de um lagrangiano da função de log-verossimilhança. Sua derivação é mais complexa do que o escopo dessa dissertação permite. Porém, sua expressão simplifica e pode ser facilmente calculada a partir de parâmetros e resíduos de uma regressão OLS.

O teste LM para a presença do parâmetro de dependência espacial no erro pode ser dado por:

$$LM_{\lambda} = \left( \frac{\frac{\boldsymbol{\varepsilon}' W \boldsymbol{\varepsilon}}{\frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n}}}{\frac{1}{tr(W^2 + W'W)}} \right)^2 \quad (8.13)$$

O teste LM para a presença do parâmetro de dependência espacial na variável dependente pode ser dado por:

$$LM_{\rho} = \left( \frac{\frac{\boldsymbol{\varepsilon}' W \mathbf{y}}{\frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n}}}{\frac{1}{tr(W^2 + W'W) + \frac{[(WX\boldsymbol{\beta})'(I_n - X(X'X)^{-1}X')(WX\boldsymbol{\beta})]}{\frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n}}}} \right)^2 \quad (8.14)$$

Ambos os testes também possuem distribuição qui-quadrado com 1 grau de liberdade.

Uma importante alteração dos testes (8.13) e (8.14) são suas versões robustas. As versões robustas desses testes foram desenvolvidas por Anselin et al. (1996) e Bera and Yoon (1993). Tais versões procuram considerar o possível efeito da dependência espacial que o teste não considera, ou seja, influência espacial na variável dependente para o teste de LM no erro e influência da dependência espacial no erro no teste de LM para dependência espacial da variável dependente.

Se considerarmos, de forma a simplificar a notação:

$$d_{\lambda} = \left( \frac{\frac{\boldsymbol{\varepsilon}' W \boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}}{n} \right)$$

$$d_{\rho} = \left( \frac{\frac{\boldsymbol{\varepsilon}' W \mathbf{y}}{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}}{n} \right)$$

$$T = \text{tr}(W^2 + W'W)$$

$$D = (WX\boldsymbol{\beta})'(I_n - X(X'X)^{-1}X')(WX\boldsymbol{\beta}) + T \frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n}$$

A equação (8.13), poderia ser reescrita:

$$LM_{\lambda} = \frac{d_{\lambda}^2}{T} \quad (8.15)$$

A versão robusta de (8.15) fica:

$$LM_{\lambda}^{rob} = \frac{\left( d_{\lambda} - T \frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n} D^{-1} d_{\rho} \right)^2}{T \left( 1 - T \frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n} D^{-1} \right)} \quad (8.16)$$

A equação (8.14), poderia ser reescrita, a partir de algumas manipulações algébricas como:

$$LM_{\rho} = \frac{\frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n} d_{\rho}^2}{D} \quad (8.17)$$

A versão robusta de (8.17) fica:

$$LM_{\rho} = \frac{\frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n} (d_{\rho} - d_{\lambda})^2}{D - T \frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n}} \quad (8.18)$$

Note que as versões (8.16) e (8.18) procuram retirar da estatística a possível influência da outro termo de correlação espacial, que não o efetivamente testado.

A programação em Matlab de LeSage(1999) inclui apenas os testes Moran e LM do erro. A programação para os testes LM da variável dependente bem como para as versões robustas de LM do erro e da variável dependente foram construídas em Matlab e se encontram no apêndice.

Acosta et al. (2003) indica que o teste LM, na presença de ambos os parâmetros, apresenta uma espécie de “erro Tipo-III”, em que hipótese nula seria corretamente rejeitada porém em favor de uma hipótese alternativa incorretamente especificada. Por exemplo, uma rejeição por um teste LM de  $\rho$ , rejeita  $\rho=0$  (um modelo convencional) em favor de  $\rho \neq 0$  (um modelo SAR-5.3), porém assumindo  $\lambda=0$ , quando isso pode não ser verdade, o que implicaria um modelo SAC(5.1).

Acosta et al. (2003) realiza simulações de Monte Carlo para comparar os testes (8.13) e (8.14) e suas versões robustas (8.16) e (8.18) entre si, em termos de poder e tamanho dos testes. As versões robustas de fato se mostram bem superiores às versões originais. Em particular, para valores medianos de um parâmetro, o teste original incorreto (LM de erro para testar  $\rho$  e LM da dependente para testar  $\lambda$ ) rejeita a hipótese nula com uma porcentagem muita acima dos 5% usualmente aceitos, sendo o teste de correlação do erro pior. Já o teste robusto se mostra mais estável e com níveis de rejeição da hipótese nula aceitáveis.

Uma análise mais crítica de suas tabelas de resultados indica, para o testes nas versão não robustas, um problema semelhante ao da estatística de Moran, porém em menor escala. Dessa forma, muitas vezes os testes não são capazes de diferenciar os modelos entre si. Esse aspecto é substancialmente melhorado nas versões robustas.

Porém, quando valores dos coeficientes espaciais são pequenos, os testes robustos se mostram muito exigentes, em especial quando o parâmetro do outro coeficiente se torna alto, o que certa forma acabada encobrindo o valor do parâmetro testado. Esse efeito é justamente o oposto do observado no teste não robusto, em que um alto valor de um parâmetro parece aumentar a taxa de rejeição no teste incorreto.

No entanto, aparentemente os testes, em especial o teste do erro, parecem ter uma tendência de rejeitar a hipótese nula com mais frequência quando o tamanho da amostra aumenta, o que significaria uma dificuldade de diferenciar os possíveis modelos.

Foram geradas de forma experimental, de acordo com os detalhes a serem explicitados no capítulo 14, dados aleatórios e calculadas as estatísticas  $I$  de Moran, LM do Erro, LM do Erro Robusto, LM da Variável Dependente e LM da Variável Dependente Robusto. Os dados foram gerados de acordo como um modelo SAR, ou seja, com dependência espacial na variável dependente e sem dependência espacial no resíduo.

As tabelas a seguir apresentam resultados médios e os desvios para 100 iterações com dois tamanhos de amostra, uma pequena onde  $n=27$  e outra onde  $n=1000$ . Como a matriz utilizada no cálculo do teste é a mesma utilizada na geração de dados, esperava-se uma precisão boa dos testes, em particular esperava-se p-valor baixo para Moran e LM da variável dependente, e p-valor mais alto para LM do Erro. Apenas para referência, o valor crítico ao nível de 1% da distribuição qui-quadrado com 1 grau de liberdade é 6.635, e o da distribuição normal, ao nível de 5%, é 1.96.

	Estatística	Estatística Padronizada	P-valor
Moran	0.28	3.17	0.04
	0.13	1.36	0.08
LM Erro		7.31	0.10
		6.47	0.16
LM Erro Robusto		1.34	0.49
		1.81	0.29
LM Var. Dep.		28.14	0.00
		5.70	0.00
LM Var. Dep. Robusto		22.16	0.00
		2.45	0.00

**Tabela 8. 1** - Valores Médios e Desvios dos Resultados das Simulações dos Testes de Especificação.  
SAR, 100 iterações,  $n = 27$

	Estatística	Estatística Padronizada	P-valor
Moran	0.38 0.04	24.57 2.3406	0 0
LM Erro		600.56 115.79	0 0
LM Erro Robusto		220.56 60.757	0 0
LM Var. Dep.		1174.2 61.873	0 0
LM Var. Dep. Robusto		794.21 13.111	0 0

**Tabela 8. 2** - Valores Médios e Desvios dos Resultados das Simulações dos Testes de Especificação. SAR, 100 iterações, n = 1000

Note que o simples aumento do tamanho da amostra faz com que a ordem de grandeza das estatísticas LM aumente. Veja que a estatística de Moran não padronizada não muda sua ordem de grandeza, mas a padronizada sim. Na amostra com 27 observações, o teste robusto claramente não consegue rejeitar a hipótese nula da presença do componente espacial no erro ser igual a 0, porém ao aumentar o tamanho da amostra, como a estatística cresce, isso não ocorre mais.

De forma também a verificar se esse comportamento é generalizado ou não, fez mais dois conjuntos de testes, resumidos nas tabelas a seguir. As tabelas (8.3) e (8.4) refazem os testes para um conjunto de dados gerados sem nenhuma correlação espacial. Já as tabelas (8.5) e (8.6) exibem os resultados de dados gerados por um modelo SEM, ou seja, sem dependência espacial na variável dependente e com dependência espacial no resíduo. Nesses casos, esperava-se p-valor alto para todos os testes em (8.3) e (8.4) e para os testes de LM Variável Dependente nas tabelas (8.5) e (8.6). Ainda para as tabelas (8.5) e (8.6) esperava-se um p-valor baixo para Moran e LM do Erro.

	Estatística	Estatística Padronizada	P-valor
Moran	-0.05 0.10	-0.11 0.96	0.53 0.27
LM Erro		0.93 1.52	0.51 0.27
LM Erro Robusto		0.94 1.54	0.38 0.27
LM Var. Dep.		1.08 1.52	0.50 0.30
LM Var. Dep. Robusto		1.10 1.50	0.49 0.30

**Tabela 8. 3** - Valores Médios e Desvios dos Resultados das  
Simulações dos Testes de Especificação.  
Nenhuma Correlação Espacial, 100 iterações,  $n = 27$



	Estatística	Estatística Padronizada	P-valor
Moran	0.00	-0.02	0.50
	0.01	0.97	0.29
LM Erro		0.93	0.50
		1.14	0.29
LM Erro Robusto		0.91	0.44
		1.10	0.29
LM Var. Dep.		0.87	0.55
		1.45	0.29
LM Var. Dep. Robusto		0.85	0.55
		1.40	0.29

**Tabela 8. 4** - Valores Médios e Desvios dos Resultados das Simulações dos Testes de Especificação.  
Nenhuma Correlação Espacial, 100 iterações, n = 1000

	Estatística	Estatística Padronizada	P-valor
Moran	0.09	1.25	0.34
	0.13	1.32	0.32
LM Erro		1.85	0.43
		3.14	0.30
LM Erro Robusto		1.78	0.10
		2.93	0.30
LM Var. Dep.		1.58	0.43
		2.10	0.33
LM Var. Dep. Robusto		1.51	0.42
		1.80	0.32

**Tabela 8. 5** - Valores Médios e Desvios dos Resultados das Simulações dos Testes de Especificação.  
SEM, 100 iterações, n = 27

	Estatística	Estatística Padronizada	P-valor
Moran	0.15	9.57	0.00
	0.02	1.43	0.00
LM Erro		91.48	0.00
		27.10	0.00
LM Erro Robusto		90.27	0.00
		26.74	0.00
LM Var. Dep.		2.34	0.36
		2.99	0.30
LM Var. Dep. Robusto		1.13	0.45
		1.51	0.27

**Tabela 8. 6**- Valores Médios e Desvios dos Resultados das Simulações dos Testes de Especificação.  
SEM, 100 iterações, n = 1000

Com mais esse conjunto de tabelas é possível verificar que, conforme indicações da literatura, o teste LM da variável dependente parece ser de fato mais robusto. Já o teste LM do erro, parece sofrer com o aumento do tamanho da amostra. Assim, o teste LM da variável dependente parece ser capaz de diferenciar quando há ou não esta correlação, já o teste LM do erro, não demonstra tanto sucesso, mesmo na versão robusta.

Apesar de ser apenas um exercício exploratório de onde não se pode generalizar maiores críticas a testes utilizados constantemente na literatura, podemos inferir problemas potenciais com esses testes. Certamente, esta área ainda é uma que ainda necessita de maior investigação, bem como testes que permitam melhor indicar qual o melhor modelo a ser escolhido.

Existem ainda outras variações de testes, como, por exemplo, o teste K-R de Kelejian e Robinson e o teste de LM no erro, utilizando parâmetros e resíduos estimados por ML com o pressuposto de um modelo SAR (6.3). Outras variações não cabem ser mencionadas aqui.

Anselin(2003) sugere duas possíveis abordagens de como encontrar a especificação adequada de um modelo. Uma seria a abordagem *forward* e a outra, *backward*.

A abordagem *forward* parte de um modelo simples para um mais completo, de acordo com o seguinte procedimento: (1) obter os parâmetros e resíduos por uma regressão OLS; (2) calcular os testes (8.13) e (8.14) e suas significâncias. Caso nenhum ou apenas um deles seja significativo, já se sabe qual modelo utilizar. Caso ambos sejam significantes, ele sugere verificar os valores e significância para as versões robustas desses testes, optando pelo modelo de maior significância. Curiosamente nessa sugestão não se fala do modelo completo, SAC. Podemos assumir que se as versões robustas não forem claramente distintas em termos de significância o modelo SAC deveria então ser aplicado.

A abordagem *backward* faz o inverso. Estima um modelo completo, e a partir da significância e valor dos parâmetros elimina-os buscando modelos mais simples.

Dadas as indicações da literatura e as evidências mostradas nessa dissertação, a estratégia mais confiável para a busca de um modelo parece ser utilizar as estimativas dos modelos, em conjunto com as estatísticas de Moran, para identificar presença de algum tipo de correlação espacial, e do LM da variável dependente, para indicar esse tipo de correlação especificamente. Como visto, o teste de LM do erro não parece ser tão confiável na diferenciação dos modelos.

Em particular, uma abordagem *forward* poderia iniciar com um modelo SAR e então avançar para um modelo SAC, no caso de existência da correlação espacial na variável dependente ou para um modelo SEM, caso isso não seja observado. Uma abordagem *backward* poderia partir do modelo SAC e, de acordo com o observado, reduzir para um modelo SAR ou SEM.

Os testes parecem ser bons para identificar a presença de componentes espaciais, mas não para diferenciá-los de forma confiável. Já os modelos SAR e SAC parecem gerar bons indicadores de qual o melhor modelo a ser utilizado. Serão feitos mais comentários sobre isso na parte dedicada à simulação.

Vale ressaltar também que todos estes testes dependem muito da escolha adequada da matriz de pesos espaciais, que pode alterar e muito os resultados obtidos. Anselin (1988) indica, e será observado na simulação que uma escolha inadequada da matriz pode levar a resultados não só viesados como inconsistentes, o que acarreta uma série de inferências incorretas.

## 9. Tópicos Avançados: Painel e Modelos semi-paramétricos

Existe atualmente um claro avanço na literatura sobre a análise espacial em painel e no uso de modelos semi-paramétricos. Apesar de não constituírem o foco dessa dissertação, serão feitos alguns comentários sobre esses avanços.

Elhorst (2000 e 2001) indica que dois problemas específicos e relevantes que aparecem quando se criam modelos no tempo e no espaço são a dependência serial entre as observações de cada unidade ao longo do tempo e a dependência espacial entre as observações em cada ponto no tempo.

Para analisar tais situações ele parte de um modelo geral que inclui correlação serial de primeira ordem e autocorrelação espacial nas variáveis dependentes e independentes:

$$\mathbf{y}_t = \tau \mathbf{y}_{t-1} + \delta W \mathbf{y}_t + \eta W \mathbf{y}_{t-1} + X_t \boldsymbol{\beta}_1 + X_{t-1} \boldsymbol{\beta}_2 + W X_t \boldsymbol{\beta}_3 + W X_{t-1} \boldsymbol{\beta}_4 + u_t \quad (9.1)$$

Aqui  $\tau, \delta, \eta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4$  são os regressores. Variáveis com o subscrito  $t-1$  representam uma variável defasada temporalmente e uma variável multiplicada por  $W$  representa uma variável defasada espacialmente.

Assim como na explicação dos modelos de *cross-section* para que incluem dependência espacial, nesses modelos para painel, considerando alguns dos regressores como sendo 0, teremos modelos simplificados. Assim, em seu artigo, Elhorst (2000) apresenta nove modelos para painel com dependência temporal e/ou espacial.

Para a estimação desses modelos, são derivadas as funções de máxima verossimilhança, e para alguns deles, Elhorst *apud* LeSage(1999) fornece a programação em Matlab.

É importante ressaltar, ainda, que essa é uma área em expansão na econometria espacial e ainda não há modelos de painel com componente espacial consolidados na literatura.

Da mesma forma que a área de econometria espacial em painel é uma área de fronteira, modelos semi-paramétricos e com endogeneização da matriz de ponderação espacial também são áreas de fronteira na atualidade. Essa certamente deverá ser uma área em expansão visto as peculiaridades identificadas nessa dissertação com a escolha da matriz de pesos espaciais.

Não cabe no escopo desse estudo aprofundar nesses modelos. É válido no entanto mencionar que Conley(1999) vem desenvolvendo uma série de estudos para o uso de modelos semi-paramétricos, aplicados prioritariamente a distâncias econômicas. Algumas de suas rotinas já estão programadas em Stata, o que é um indicar de uma possível ida dessas metodologias para setores *mainstream* da econometria atual.

### PARTE III – PESOS ESPACIAIS E SUAS APLICAÇÕES

#### 10. Distâncias Geográficas

*Todo mundo influencia, quem está mais perto influencia mais.*

Os pesos espaciais construídos a partir de características geográficas são, certamente, o formato mais utilizado na literatura de econometria espacial. A lógica teórica de construção desses pesos está em atribuir maior peso para unidades geograficamente mais próximas e, conseqüentemente, reduzir o peso de unidades mais distantes.

Uma técnica usual, para qualquer critério de construção de pesos, é normalizar as linhas para sua soma ser igual a 1. Isso permite uma comparabilidade direta entre diferentes pesos da mesma matriz, entre diferentes matrizes e facilita a leitura de  $w_{ij}$  como a porcentagem da influência total em  $i$  que é de responsabilidade de  $j$ . Anselin(1988) indica que apesar de não haver um requisito matemático para a normalização, esse procedimento facilita a leitura dos coeficientes. Além disso, do ponto de vista matemático, isso restringe os parâmetros espaciais para o intervalo entre (-1) a (+1).

No entanto, Anselin (1988) alerta que a normalização não deve ser feita automaticamente. Dependendo do modelo, se os valores não normalizados forem importantes em termos comparativos, a normalização não deveria ser feita. Porém, é valido notar que na literatura encontra-se, de forma preponderante, estudos com a matriz normalizada.

As duas principais formas de se construir essa matriz de pesos espaciais são por fronteiras ou por distâncias geográficas dos centróides. Para a mensuração por fronteiras, cria-se uma Matriz de fronteiras ou contigüidade, onde:

$$w_{ij}^* = \begin{cases} 0, & \text{se não há fronteira entre } i \text{ e } j \\ 1, & \text{se há fronteira entre } i \text{ e } j \end{cases} \quad (10.1)$$

Após a normalização, obtém-se a matriz  $W$ :

$$w_{ij} = \frac{w_{ij}^*}{\sum_j w_{ij}^*} \quad (10.2)$$

Para uma distância por centróides, é calculado o inverso da distância Euclidiana:

$$w_{ij}^* = \begin{cases} \frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases} \quad (10.3)$$

Em que  $x_i, x_j$  e  $y_i, y_j$  são coordenadas dos centróides das unidades  $i$  e  $j$ .

Assim, o elemento  $w_{ij}$  fica, após a normalização, ponderado da mesma forma que em (10.2)

Uma medida usual, derivada da distância Euclidiana, é o inverso da distância Euclidiana ao quadrado. Dessa forma teríamos:

$$w_{ij}^* = \begin{cases} \frac{1}{(x_i - x_j)^2 + (y_i - y_j)^2}, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases} \quad (10.4)$$

A normalização segue a mesma regra de (10.2)

A razão teórica para o uso de uma matriz com o inverso de uma distância ao quadrado (10.4) é penalizar mais rapidamente unidades mais distantes, e não de forma linear como na medida exibida em (10.3).

Entre os dois extremos de uma matriz de contigüidade (10.2) ou uma matriz de distância como (10.3) ou (10.4), existem diferentes combinações.

Algumas simplificações são comuns na literatura. Uma delas é a matriz de vizinhança incluindo-se um certo número  $k$  de vizinhos mais próximos, conhecido na literatura como *k-nearest neighbors weight matrix*. Assim, a matriz  $W(k)$  é definida como:

$$w_{ij}^*(k) = \begin{cases} 0, & \text{se } i = j \\ 1, & \text{se } d_{ij} \leq d_i(k) \\ 0, & \text{se } d_{ij} > d_i(k) \end{cases} \quad (10.5)$$

Aqui,  $d_i(k)$  é a distância de corte, ou seja, a distância do vizinho de ordem  $k$ .

Após a normalização, obtém-se a matriz  $W$ , equivalente à definição (10.2):

$$w_{ij}(k) = \frac{w_{ij}^*(k)}{\sum_j w_{ij}^*(k)} \quad (10.6)$$

Note que mesmo no caso  $k=1$  a matriz (10.6) não será igual à matriz (10.1). Isso se deve ao fato do critério (10.5) forçar cada unidade a ter ao menos 1 vizinho, o que não é garantido no caso (10.1). Esse aspecto é relevante quando existem ilhas importantes. No caso brasileiro, por exemplo, seria de extrema simplificação, por exemplo, dizer que uma ilha turística, como Ilha Bela, no litoral norte paulista, não recebe e nem provoca influência nos demais municípios pelo simples fato de que não possui fronteiras comuns. No caso europeu, por exemplo, a influência dos países da Grã-bretanha seriam ignorados na Europa continental e vice-versa caso a matriz fosse construída a partir de (10.1) e não (10.5).

A versão (10.5) é apresentada por Ertur e Gallo(2003). No entanto, a *k-nearest neighbors weight matrix* é por vezes apresentada com a ponderação construída a partir da distância de não de forma binária. Dessa forma, (10.5) seria reescrito como:



$$w_{ij}^*(k) = \begin{cases} 0, & \text{se } i = j \\ \frac{1}{d_{ij}}, & \text{se } d_{ij} \leq d_i(k) \\ 0, & \text{se } d_{ij} > d_i(k) \end{cases} \quad (10.7)$$

Aqui,  $d_i(k)$  é a distância (ao a distância ao quadrado) de corte, ou seja, a distância do vizinho de ordem  $k$ .

Após a normalização (10.6), obtém-se a matriz  $W$ .

Um outro critério de corte é a distância limite, e não o número de vizinhos. A desvantagem desse método é que unidades com maiores áreas podem ser prejudicadas e acabarem ficando com poucos vizinhos, enquanto que unidades com áreas menores terão o inverso. Por outro lado, restringe e iguala a área de influência para todas as unidades.

Desse modo, o critério para a construção de uma matriz binária seria:

$$w_{ij}^*(d) = \begin{cases} 0, & \text{se } i = j \\ 1, & \text{se } d_{ij} \leq d \\ 0, & \text{se } d_{ij} > d \end{cases} \quad (10.8)$$

Para uma matriz com o inverso da distância (ou a distância ao quadrado), o critério seria:

$$w_{ij}^*(d) = \begin{cases} 0, & \text{se } i = j \\ \frac{1}{d_{ij}}, & \text{se } d_{ij} \leq d \\ 0, & \text{se } d_{ij} > d \end{cases} \quad (10.9)$$

Onde  $d$  é a distância (ao a distância ao quadrado) de corte.

Note que nesses casos (10.8 e 10.9),  $d$  não varia de acordo com  $i$ , sendo o mesmo para toda a amostra.

A normalização, com uma variação mínima de (10.6) seria:

$$w_{ij}(d) = \frac{w_{ij}^*(d)}{\sum_j w_{ij}^*(d)} \quad (10.10)$$

Para muitas análises, no entanto, argumenta-se que seria mais relevante o tempo efetivo de deslocamento do que a medida entre os centróides. Esse argumento certamente é válido, pois o tempo de deslocamento é uma função da distância e de outros fatores como qualidade de sistema de transportes, condições de trânsito e acidentes geográficos. Se considerarmos  $d_{ij}$  como o tempo de deslocamento, ou ainda como o quadrado do tempo de deslocamento, entre as unidades  $i$  e  $j$ , teríamos:

$$w_{ij}^* = \begin{cases} \frac{1}{d_{ij}}, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases} \quad (10.11)$$

A normalização seria a usual (10.2). As variações (10.5), (10.7), (10.8) e (10.9) poderiam ser aplicadas de forma similar.

As matrizes construídas de acordo com (10.1), (10.4) serão utilizadas tanto na simulação quanto na análise empírica dessa dissertação, por serem as mais usuais e as que melhor representam dois extremos na forma de medir um mesmo aspecto.

## 11. Distâncias Sócio-Econômicas

*Todo mundo influencia; quem é mais parecido influencia mais.*

A partir dos conceitos de dependência espacial, alguns autores, destacando-se Case and Rosen (1993), começaram a criar medidas de distâncias não geográficas. Para mensurar a proximidade entre diferentes unidades de observação, foram testadas medidas de distâncias a partir da composição econômica e social, como, por exemplo, porcentagem de negros na população e renda per capita. Chamaremos, então, essas medidas de distâncias sócio-econômicas.

O conceito análogo ao de distância passa ser o de similaridade: quem é mais parecido com um determinado objeto de análise influencia mais do que os mais diferentes. Esse tipo de medida de influência passa a ser necessário quando se começam a analisar efeitos em variáveis dependentes em que a simples fronteira ou distância geográfica parece não ser o principal ponto de ponderação de influências. Isso ocorre, por exemplo, quando se analisam decisões de governos subnacionais, em que se pode imaginar que um determinado estado, mais similar, irá influenciar mais do que um vizinho de fronteira, menos similar. Por exemplo, um município líder comercial em sua região pode ter suas decisões altamente influenciadas pelo líder comercial de outra região, mais distante geograficamente, e pouco influenciadas por um município vizinho, mais próximo geograficamente.

Operacionalmente, as distâncias passam a ser medidas como as diferenças de valores para um mesmo indicador de duas localidades. De forma a dar mais importâncias para as menores distâncias, o usual é uma medida como a distância euclidiana invertida. (Case and Rosen, 1993; Boarnet and Glazer, 2002; Hernández-Murillo, 2003).

Inverso da distância euclidiana sócio-econômica:

$$w^*_{ij} = \begin{cases} \frac{1}{|d_i - d_j|}, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases} \quad (11.1)$$

Onde  $d_i$  e  $d_j$  são valores da variável de interesse para  $i$  e  $j$ .

Para obter-se a matriz  $W$ , utiliza-se normalização (10.2), aqui reproduzida:

$$w_{ij} = \frac{w_{ij}^*}{\sum_j w_{ij}^*} \quad (10.2)$$

Assim como o que foi indicado para distâncias geográficas bidimensionais, pode-se calcular a distância euclidiana sócio-econômica invertida n-dimensional, substituindo-se o módulo da diferença da variável de interesse, feito no caso unidimensional, pela raiz quadrada da soma das diferenças quadradas das  $n$  variáveis de interesse.

## 12. Medidas de Dissimilaridades

*Todo mundo influencia, quem é mais diferente influencia mais.*

Utilizando-se ainda de uma analogia com as distâncias geográficas, pode-se fazer uma análise diferente da distância e da similaridade, fazendo com que o peso maior seja dado a quem possui mais diferenças, ou, de forma equivalente, a maior distância. Essa medida de dissimilaridade é válida quando, por exemplo, estamos medindo disputas por alíquotas, e influencia mais um determinado estado aquele que possui a alíquota mais distante, pois estará atraindo ou enviando empresas para a unidade em análise. Outra medida de dissimilaridade é a usado por Hernández-Murillo (2003), em que o peso é dado de acordo com a população dos estados vizinhos, recebendo maior peso os estados mais populosos.

Assim como para similaridades é usada a distância euclidiana invertida, de forma a dar maior peso para as menores distâncias, para uma matriz de dissimilaridades, utiliza-se, a distância euclidiana tradicional, dando assim maior importância para as maiores distâncias.

Distância euclidiana sócio-econômica:

$$w_{ij}^* = \begin{cases} |d_i - d_j|, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases} \quad (12.1)$$

Onde  $d_i$  e  $d_j$  são valores da variável de interesse para  $i$  e  $j$ .

Para obter-se a matriz  $W$ , utiliza-se normalização (10.2), aqui reproduzida:

$$w_{ij} = \frac{w_{ij}^*}{\sum_j w_{ij}^*} \quad (10.2)$$

Do mesmo modo como foi indicado para distância euclidiana sócio-econômica invertida, pode-se calcular a distância euclidiana sócio-econômica  $n$ -dimensional, substituindo-se o módulo da diferença da variável de interesse, feito no caso unidimensional, pela raiz quadrada da soma das diferenças quadradas das  $n$  variáveis de interesse.

## PARTE IV: SIMULAÇÕES DE MONTE CARLO

### 13. O Método de Monte Carlo

De forma a testar as propriedades de diferentes matrizes em diferentes modelos e técnicas de estimação, iremos utilizar o método de Monte Carlo. De acordo com Hendry (1984:939), a experimentação de Monte Carlo consiste em uma abordagem geral em que problemas matemáticos de natureza analítica que se mostrem tecnicamente difíceis de serem tratados, ou cuja análise técnica seria excessivamente trabalhosa, podem ser “resolvidos” (aspas de Hendry (1984)) pela substituição de um problema estocástico equivalente.

A experimentação de Monte Carlo é, como o próprio nome indica, um técnica de características experimentais. De acordo com Shaddish, Cook and Campbell (2002), um experimento pode ser definido como um “teste sob condições controladas que é feito para demonstrar uma verdade conhecida; examinar a validade de uma hipótese ou determinar a eficácia de algo não testado”, ou ainda, de acordo com os mesmo autores, “um estudo no qual uma intervenção é deliberadamente introduzida para se observar seus efeitos”.

No entanto, uma característica essencial em experimentos é o objetivo de traçar relações causais, em geral a respeito de comportamento e decisões de pessoas. No método de Monte Carlo, o objetivo mais geral é testar propriedades de determinadas técnicas, de modo que essa expressão causal, ou ao menos correlacional, torna-se mais sutil.

As duas características de experimentos que são fortemente utilizadas na experimentação de Monte Carlo são o controle e a construção de um contrafactual. O contrafactual está aliado ao princípio de falsiabilidade da ciência, indicando que seria possível provar que determinada afirmativa não é verdadeira, encontrando explicações alternativas. O contrafactual objetiva eliminar essas explicações alternativas. O contrafactual mais perfeito possível seria exatamente a mesma observação experimentada passando por um outro processo de tratamento, sendo testada de outro modo, indicando, por exemplo, que apenas sob um tratamento específico produz-se o efeito observado.

Desse modo, o controle é certamente a característica mais marcante do método de Monte Carlo, mais forte até que em experimentos de um modo geral. Isso se dá, pois, no método de Monte Carlo há um controle absoluto sobre a geração de dados e sobre todo o procedimento que está acontecendo, eliminando-se quase qualquer possibilidade de efeitos ou ações não observáveis e/ou indesejadas.

Sendo assim, se uma técnica experimental busca traçar relações causais, e auxiliar na descoberta da verdade, como no método de Monte Carlo a verdade é conhecida, busca-se identificar as melhores estratégias para se encontrar essa verdade e diferenciá-la corretamente de outras explicações alternativas.

A chamada experimentação de Monte Carlo é um método de simulação, uma vez que é feita simulando um grande número de repetições de um processo que gera números aleatórios de acordo com determinadas distribuições. Bussab (2004) indica que quando “certas questões não podem ser resolvidas analiticamente (...) teremos de recorrer a estudos de simulação para obter aproximações de quantidades de interesse.” Em outras palavras, as simulações são feitas com o intuito de gerar um número grande de amostras da mesma população. Essas simulações são feitas justamente gerando variáveis aleatórias, o que pressupõe que o pesquisador possui um gerador de números aleatórios de acordo com determinada distribuição.

Sendo um método de simulação com características experimentais, ou, de acordo com Hendry (1984) uma “simulação experimental”, o método de Monte Carlo irá compartilhar algumas características e críticas de experimentos e simulações. Em particular, um estudo de Monte Carlo é muito forte na sua validade interna e fraco na sua validade externa. Considerando as definições de Shaddish, Cook and Campbell (2002), de forma resumida, a validade interna indica que as covariações e/ou causalidades observadas entre diferentes tratamentos são de fato causadas pelas mudanças introduzidas. Já a validade externa seria a capacidade de generalizar as inferências experimentais para a mesma população, outras populações e situações.



É importante destacar que a simulação permite um grande benefício ao conjunto de técnicas experimentais. Apesar de ser frágil na validade externa sobre outras populações, a simulação possui grande validade externa sobre a própria população. Uma das maneiras das ciências experimentais fortalecerem sua validade externa é através da replicação. No caso da simulação de Monte Carlo, por construção, há um número grande de replicações que permitem uma inferência mais segura do ponto de vista externo, para a população em análise.

Como há o controle praticamente absoluto do processo de geração de dados e dos testes realizados, um estudo de Monte Carlo tem certamente uma validade interna quase incontestável. No entanto, como dito, é mais difícil generalizar esses resultados para outras populações, sendo que a crítica mais usual é que não há como saber se, nas populações não simuladas, o processo de geração de dados se dá dessa maneira.

É importante ressaltar, porém, conforme Hendry (1984) indica, que a experimentação de Monte Carlo procura apenas resolver por simulação os problemas analíticos partindo dos mesmos pressupostos que esses modelos. Assim, a crítica usual à simulação de Monte Carlo quanto ao modelo de geração de dados seria, ao menos em parte, uma crítica aos pressupostos do modelo, e não à técnica de Monte Carlo.

Em contrapartida, o processo de Monte Carlo possui diversas vantagens. Muitas inferências estatísticas e econométricas só poderiam ser perfeitamente confirmadas no caso de se conhecer a verdade – porém, caso se conhecesse a verdade, não seriam necessárias. A simulação experimental de Monte Carlo permite que a verdade seja criada. Com isso, um importante teste já é feito: se, com a verdade conhecida uma determinada técnica se mostra falha, pode-se com segurança generalizar que ela será falha para inferências em verdades desconhecidas. Se, por outro lado, uma determinada técnica se mostra consistente com uma verdade simulada conhecida, pode-se ter mais segurança quanto à generalização dessa técnica, sendo que só seria falha se os pressupostos e/ou condições de geração de dados na população fossem outros.

Dessa forma, nas palavras de Hendry (1984: 944) a “experimentação de Monte Carlo pode complementar de forma eficiente a análise a estabelecer fórmulas numérico-analíticas que conjuntamente sumarizam os resultados experimentais e resultados analíticos conhecidos de forma a auxiliar na interpretação de evidência empírica e a computar resultados em outros pontos dentro do espaço relevante de parâmetros.” Ele ressalta ainda que estudos de Monte Carlo não devem ser vistos como produto final, mas como insumo para demais análises estatísticas e/ou econométricas.

Com fins econométricos, o método de Monte Carlo é utilizado, usualmente para testar tamanho e poder de testes; viés, eficiência e consistência de estimadores e propriedades de pequenas e grandes amostras. No escopo da econometria espacial, as aplicações costumam focar em propriedades de testes e se estes são capazes de identificar as diferentes possibilidades de estruturas espaciais, além de seus comportamentos sob diferentes tamanhos de amostra e distribuições do erro.

No estudo de Acosta et al. (2003) já citado nessa dissertação, por exemplo, são feitos testes para comparar as propriedades dos diferentes testes LM de correlação espacial. Demais autores como Kelejian and Robinson (1998), Fingleton (1999), Anselin e Florax (1995), entre outros, recorrem ao método de Monte Carlo diversas vezes na literatura para testar propriedades de testes de especificação e poder de algumas estatísticas.

#### 14. Definição Teórico-Operacional da Simulação

Para uma simulação de Monte Carlo o primeiro passo a ser feito é preparar o processo gerador de dados. O modelo mais amplo, o SAC (5.1) possui a seguinte forma:

$$\begin{aligned} \mathbf{y} &= \rho W_1 \mathbf{y} + X\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \lambda W_2 \mathbf{u} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \end{aligned} \quad (5.1)$$

Reescrevendo-o na forma reduzida, obtemos:

$$\begin{aligned} \mathbf{y} &= (I - \rho W_1)^{-1} X\boldsymbol{\beta} + (I - \rho W_1)^{-1} (I - \lambda W_2)^{-1} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \end{aligned} \quad (14.1)$$

Para efeitos práticos das simulações, consideraremos  $X$  uma matriz  $n \times 3$ , sendo a primeira coluna a constante e as duas colunas restantes retiradas de uma distribuição normal,  $\boldsymbol{\beta}$  um vetor constante de parâmetros  $3 \times 1$  e  $W_1 = W_2$ . Assim, aplicando essas definições a (15.1), e escolhendo os valores para os parâmetros, temos a seguinte forma operacional do gerador de dados:

$$\begin{aligned} \mathbf{y} &= (I - \rho W)^{-1} [\mathbf{j} \quad \mathbf{x}_1 \quad \mathbf{x}_2]' [c \quad b_1 \quad b_2]' + (I - \rho W)^{-1} (I - \lambda W)^{-1} \boldsymbol{\varepsilon} \\ \mathbf{j} &= \text{vetor } (n \times 1) \text{ de } 1 \\ \mathbf{x}_1 &\sim N(50 * \mathbf{j}, 5^2 * \mathbf{I}) \\ \mathbf{x}_2 &\sim N(50 * \mathbf{j}, 5^2 * \mathbf{I}) \\ c &= 2 \\ b_1 &= 1.5 \\ b_2 &= 0.8 \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, (0.7)^2 * \mathbf{I}) \end{aligned} \quad (14.2)$$

Como o objetivo dessa dissertação é testar o efeito das diferentes matrizes nas estimações de econometria espacial, as simulações foram preparadas de forma a verificar exatamente isso. Desse modo, são escolhidos os valores dos parâmetros e é feita estimação por cada modelo com diferentes matrizes.

As matrizes escolhidas para a simulação são:

- $W_{contig}$  : Matriz de contigüidade (10.1), normalizada como (10.2)
- $W_{dist}$ : Matriz do Inverso da Distância Euclidiana ao quadrado (10.4), normalizada como (10.2).

As Matrizes de  $W_{contig}$  e  $W_{dist}$  já foram apresentadas e justificadas ao longo dessa dissertação. Como a primeira matriz, é considerada uma simplificação de uma medida mais complexa, esta será usada na estimação, porém não na geração de dados. Vale ressaltar que ambas são as matrizes mais usualmente utilizadas na literatura. Dessa forma, os dados são gerados com a matriz  $W_{dist}$  e a estimação é feita com  $W_{contig}$  e  $W_{dist}$ .

Além disso, serão utilizados 3 tamanhos de amostra: uma pequena, com 27 observações nas quais as matrizes são construídas a partir dos Estados brasileiros; uma grande, com 4267 observações, nas quais as matrizes são construídas a partir das Áreas Mínimas Comparáveis (AMC) dos municípios brasileiros entre 1991 e 2000. Essa escolha foi feita de modo a emular as duas agregações mais relevantes para análises empíricas no Brasil: Estados e Municípios. Além disso, como na seção de aplicação empírica estaremos usando a amostra de 4267 AMCs, a analogia fica mais adequada.

Optou-se também pela inclusão de uma matriz com 1000 observações, com o intuito de obter uma amostra de tamanho intermediário entre os dois extremos anteriores. As matrizes correspondentes foram geradas a partir de coordenadas sorteados de forma aleatória. Após o sorteio inicial, as matrizes foram mantidas as mesmas por toda a simulação.

Por questões computacionais, no caso da amostra de 4267, após a normalização da matriz  $W_{dist}$ , os valores menores que 0.0001 foram igualados a 0.

É necessário também escolher os valores dos parâmetros  $(\rho, \lambda)$  para a simulação. Todas as combinações que envolvem valores entre -1 e 1 seriam válidas, porém é simples verificar que isso tornaria uma análise praticamente inviável. Assim, foram escolhidas 5 combinações: nenhuma correlação espacial: (0,0); correlação espacial na variável dependente: SAR: (0.5, 0); correlação espacial no erro: SEM (0, 0.5); correlação espacial tanto no erro quanto na variável dependente, em sentidos opostos: SAC1 (-0.5, 0.5) e correlação espacial tanto no erro quanto na variável dependente, no mesmo sentido: SAC2 (0.5, 0.5). A razão para o sinal trocado se deve à verificação empírica desse efeito, e a situação com ambos em mesmo sentido serve como verificação experimental da contraposição desse efeito.

A simulação consiste, operacionalmente, nos seguintes passos:

- 1) Escolher  $n$  (27, 1000 ou 4267) e o *set* de matrizes correspondente (Estados, Mapa Aleatório, AMC)
- 2) Atribuir valores para o par  $(\rho, \lambda) = (0,0); (0.5, 0); (0, 0.5), (-0.5, 0.5), (0.5, 0.5)$
- 3) Para cada combinação de valores acima, rodar um determinado número de extrações de  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  e  $\boldsymbol{\varepsilon}$ , estimando os parâmetros por OLS e com cada uma das 2 matrizes por:
  - a. SAR-ML
  - b. SAR-GMM
  - c. SEM-ML
  - d. SEM-GMM
  - e. SAC-ML
  - f. SAC-GMM

Essa simulação foi feita utilizando o software Matlab e a rotina base está incluída no apêndice ao final dessa dissertação.

Por questões computacionais, não foi possível realizar o mesmo número de iterações para todos os tamanhos de amostra. Desse modo, para  $n=27$ , foram realizadas 1000 iterações; para  $n=1000$ , 500 iterações. No caso da amostra grande ( $n=4267$ ), a simulação de estimação para 500 extrações foi feita apenas com a matrizes  $W_{contig}$ . Como os dados eram gerados utilizando a matriz  $W_{dist}$ , no máximo 1 ou 2 estimações foram feitas com a própria matriz, servindo apenas de referência.

## 15. Resultados e Conclusões

A seguir são apresentadas as tabelas com os resultados das simulações realizadas. São apresentadas 5 tabelas para cada tamanho de amostra, cada uma com um *set* de parâmetros, totalizando 15 tabelas. Vale ressaltar que os dados são gerados utilizando-se uma matriz de distâncias (Dist) e estimados com a mesma matriz e com uma matriz de contigüidade (Contig).

Em cada célula são apresentados a média dos estimadores encontrados, o desvio padrão desses valores e a porcentagem de rejeição da hipótese nula (coeficiente ser igual 0). Na primeira coluna são reproduzidos os parâmetros originais.

Desse modo, o primeiro número de cada célula nos permite observar se há ou não viés no estimador e o tamanho desse viés. O segundo número permite observar a eficiência do estimador, ao comparar os diferentes modelos. Utilizando o primeiro número em conjunção com o segundo nos diferentes tamanhos de amostra pode-se observar se um determinado estimador diminui sua variância e se o viés é reduzido, ou seja, se é consistente.

O terceiro número de cada célula permite avaliar o tamanho ou o poder dos testes, ou, de foram equivalente, o erro do Tipo I (quando um parâmetro real é igual a 0) ou o erro do Tipo II (quando o parâmetro real é diferente de 0).

Com as tabelas a seguir é possível ver diferenças entre escolhas de modelos, tamanhos de amostra e matrizes de ponderação espacial.

**Modelo: Nenhuma Correlação Espacial**

**n=27**

**repetições: 1000**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	19.9460	20.7690	20.4170	20.2930	20.1900	19.9660	19.9560	19.9630	19.9610	20.4220	20.2940	25.8430	22.8150
		2.0542	7.4091	5.3025	7.5342	5.3354	2.1739	2.1274	2.1526	2.1107	7.8276	5.3972	12.9450	8.0692
		1.0000	0.8510	0.9700	0.8010	0.9540	1.0000	1.0000	1.0000	1.0000	0.8580	0.9670	0.8270	0.9600
1.5	b1	1.5025	1.5020	1.5025	1.5022	1.5026	1.5024	1.5026	1.5024	1.5026	1.5019	1.5028	1.5021	1.5027
		0.0291	0.0306	0.0308	0.0307	0.0307	0.0303	0.0299	0.0301	0.0298	0.0316	0.0311	0.0310	0.0310
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.7986	0.7982	0.7987	0.7983	0.7987	0.7984	0.7984	0.7984	0.7983	0.7980	0.7982	0.7982	0.7983
		0.0288	0.0298	0.0296	0.0298	0.0296	0.0305	0.0302	0.0302	0.0297	0.0321	0.0310	0.0310	0.0305
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	rho		-0.0057	-0.0035	-0.0023	-0.0018					-0.0030	-0.0025	-0.0020	-0.0020
			0.0507	0.0346	0.0515	0.0348					0.0542	0.0353	0.0529	0.0353
			0.0980	0.0870	0.0740	0.0760					0.1400	0.1090	0.1040	0.0890
0	lambda						-0.2048	-0.1024	-0.2462	-0.1184	-0.3323	-0.1203	-0.2987	-0.1351
							0.3541	0.2460	0.4210	0.2681	0.6108	0.2769	0.4532	0.2744
							0.1500	0.1080	0.1230	0.0570	0.1820	0.0900	0.7570	0.6900
0.7	Desvio Médio do Erro	0.6979	0.6433	0.6433	0.6436	0.6435	0.6343	0.6375	0.6355	0.6391	0.6141	0.6220	0.6164	0.6217
	R2 ajustado médio	0.9926	0.9929	0.9929	0.9929	0.9929	0.9931	0.9931	0.9931	0.9930	0.9936	0.9934	0.9935	0.9934
	likelihood médio		-16.4350	-16.4290			-16.450	-16.533			-0.1066	-0.3939		
	Tempo Médio		0.0592	0.0440	0.0012	0.0008	1.0299	0.9121	0.0371	0.0333	0.1950	0.1493	0.0311	0.0290
	Tempo Total		59.2400	43.9910	1.1860	0.8305	1,029.90	912.100	37.0580	33.2635	194.9650	149.2500	31.0500	28.9955

**Tabela 15. 1 - Resultados das Simulações, Sem Correlação Espacial, n = 27**



**Modelo: SAR**

**n=27**

**repetições: 1000**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	152.6900	21.2300	78.6280	19.9210	79.8220	164.0400	163.4100	160.2800	159.4900	20.6320	84.2220	25.2110	67.7120
		7.9802	11.8660	21.0970	12.0220	21.6630	6.5092	5.7352	6.2907	6.6164	12.6840	23.0650	18.5830	26.0590
		1.0000	0.5670	0.9850	0.4780	0.9810	1.0000	1.0000	1.0000	1.0000	0.6180	0.9890	0.5510	0.9810
1.5	b1	1.5292	1.5004	1.5020	1.5004	1.5018	1.3829	1.3903	1.4314	1.4412	1.5009	1.4950	1.5006	1.4971
		0.1086	0.0302	0.0612	0.0302	0.0612	0.0669	0.0705	0.0823	0.0871	0.0317	0.0608	0.0306	0.0607
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8172	0.8006	0.8021	0.8006	0.8021	0.7367	0.7421	0.7630	0.7692	0.8003	0.7978	0.8006	0.7990
		0.0919	0.0294	0.0564	0.0294	0.0567	0.0572	0.0646	0.0684	0.0732	0.0317	0.0568	0.0305	0.0562
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5	rho		0.4952	0.2820	0.5001	0.2776					0.4974	0.2634	0.5004	0.2712
			0.0434	0.0723	0.0440	0.0747					0.0469	0.0798	0.0461	0.0777
			1.0000	0.9920	1.0000	0.9720					1.0000	0.9640	1.0000	0.9650
0	lambda						0.7887	0.6661	0.4998	0.4004	-0.3529	0.1728	-0.3080	0.1492
							0.2000	0.2112	0.1941	0.1599	0.6280	0.3413	0.4375	0.2888
							0.9560	0.9130	0.7570	0.5780	0.1900	0.2770	0.7430	0.7620
0.7	Desvio Médio do Erro	2.0908	0.6468	1.2368	0.6468	1.2399	1.3384	1.3802	1.7075	1.7366	0.6179	1.1531	0.6204	1.1636
	R2 ajustado médio	0.9460	0.9930	0.9737	0.9934	0.9777	0.9737	0.9723	0.9592	0.9580	0.9940	0.9806	0.9939	0.9803
	likelihood médio		-17.3260	-33.7940			-38.6340	-39.5610			-0.9974	-17.1930		
	Tempo Médio		0.0653	0.0469	0.0006	0.0007	0.3737	0.7376	0.0375	0.0419	0.2044	0.1391	0.0193	0.0179
	Tempo Total		65.2640	46.9100	0.6040	0.6880	373.7250	737.5970	37.5380	41.8640	204.3920	139.0750	19.3310	17.8940

**Tabela 15. 2 - Resultados das Simulações, SAR, n = 27**

**Modelo: SEM**

**n=27**

**repetições: 1000**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	20.0250	19.2550	19.5760	19.7320	19.9100	20.0360	20.0370	20.0600	20.0370	19.8850	19.9990	15.3770	16.7620
		2.1492	9.3618	6.1935	9.4507	6.2239	2.0834	2.1355	2.2146	2.1335	9.4302	6.1344	11.8420	7.7528
		1.0000	0.7570	0.9250	0.7340	0.9180	1.0000	1.0000	0.9990	1.0000	0.7650	0.9280	0.7180	0.9050
1.5	b1	1.4996	1.4993	1.4993	1.4990	1.4991	1.4994	1.4994	1.4994	1.4993	1.4985	1.4992	1.4989	1.4993
		0.0301	0.0322	0.0314	0.0323	0.0314	0.0288	0.0296	0.0288	0.0295	0.0321	0.0309	0.0318	0.0310
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.7997	0.7998	0.7994	0.7996	0.7992	0.7997	0.7996	0.7997	0.7996	0.7996	0.7994	0.7998	0.7995
		0.0312	0.0324	0.0321	0.0324	0.0320	0.0302	0.0305	0.0302	0.0304	0.0322	0.0315	0.0318	0.0315
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	rho		0.0058	0.0036	0.0025	0.0012					0.0015	0.0005	0.0020	0.0011
			0.0602	0.0378	0.0608	0.0380					0.0610	0.0379	0.0603	0.0382
			0.1590	0.1250	0.1230	0.0920					0.1390	0.1270	0.1090	0.0920
0.5	lambda						0.2633	0.1832	0.2552	0.1763	0.1919	0.1647	0.2021	0.1534
							0.3354	0.2551	0.3910	0.2642	0.5099	0.2677	0.4365	0.2805
							0.3540	0.2550	0.2870	0.1700	0.3190	0.2160	0.7700	0.7560
0.7	Desvio Médio do Erro	0.7407	0.6774	0.6802	0.6777	0.6805	0.6550	0.6600	0.6585	0.6641	0.6367	0.6447	0.6382	0.6454
	R2 ajustado médio	0.9920	0.9924	0.9924	0.9924	0.9924	0.9929	0.9928	0.9929	0.9927	0.9933	0.9932	0.9933	0.9931
	likelihood médio		-17.7820	-17.9060			-17.4810	-17.6310			-1.2322	-1.5070		
	Tempo Médio		0.0628	0.0464	0.0007	0.0003	0.7501	0.7888	0.0340	0.0402	0.1914	0.1360	0.0186	0.0179
	Tempo Total		62.8060	46.4250	0.6550	0.2680	750.0510	788.7560	34.0490	40.1800	191.3890	136.0140	18.6380	17.8610

**Tabela 15.3** - Resultados das Simulações, SEM, n = 27

**Modelo: SAC 1**  
**n=27**  
**repetições: 1000**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	-31.4780	19.4280	-6.1206	19.9450	-5.6394	-29.5760	-33.2890	-31.5210	-32.3120	19.7280	-6.2161	15.3040	-5.8808
		6.7378	6.6785	8.6608	6.7925	9.0354	10.8710	8.9926	6.8713	7.0594	6.8410	8.7351	9.9607	9.1368
		1.0000	0.9070	0.2560	0.8980	0.2080	0.9870	0.9970	1.0000	1.0000	0.9230	0.2530	0.8980	0.2160
1.5	b1	1.5840	1.5005	1.5401	1.4996	1.5393	1.5594	1.6077	1.5843	1.5949	1.5000	1.5440	1.4997	1.5432
		0.0877	0.0354	0.0652	0.0356	0.0660	0.1447	0.1173	0.0903	0.0918	0.0348	0.0640	0.0348	0.0643
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8454	0.8002	0.8220	0.7997	0.8215	0.8321	0.8578	0.8460	0.8513	0.7999	0.8246	0.7999	0.8241
		0.0815	0.0317	0.0584	0.0319	0.0585	0.0986	0.0914	0.0810	0.0831	0.0315	0.0590	0.0311	0.0585
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
-0.5	rho		-0.4940	-0.2445	-0.4990	-0.2491					-0.4969	-0.2470	-0.4991	-0.2487
			0.0552	0.0671	0.0560	0.0693					0.0588	0.0680	0.0566	0.0689
			1.0000	0.9760	1.0000	0.9700					1.0000	0.9780	1.0000	0.9730
0.5	lambda						-0.1545	0.0951	-0.0036	0.0467	0.2032	0.0688	0.2081	0.0608
							0.7097	0.4412	0.2512	0.1904	0.4901	0.2960	0.4320	0.2871
							0.7130	0.4770	0.1900	0.0340	0.3200	0.1660	0.7670	0.7420
0.7	Desvio Médio do Erro	1.7829	0.6739	1.2193	0.6740	1.2200	1.4790	1.5549	1.6470	1.6520	0.6317	1.1691	0.6331	1.1711
	R2 ajustado médio	0.9618	0.9933	0.9792	0.9934	0.9796	0.9702	0.9672	0.9633	0.9631	0.9942	0.9813	0.9942	0.9813
	likelihood médio		-18.1750	-33.4520			-39.8920	-40.7810			-1.6375	-17.2470		
	Tempo Médio		0.0625	0.0466	0.0006	0.0004	0.3007	0.7755	0.0341	0.0405	0.1965	0.1377	0.0188	0.0177
	Tempo Total		62.5360	46.6160	0.6040	0.4420	300.6800	775.4620	34.1360	40.4710	196.5290	137.7300	18.7920	17.6520

**Tabela 15. 4 - Resultados das Simulações, SAC1, n = 27**

**Modelo: SAC 2**  
**n=27**  
**repetições: 1000**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	152.6800	19.5680	77.2640	19.8610	79.2710	164.2500	163.5300	160.7400	159.8300	20.7540	85.3660	15.3600	63.3950
		8.1345	15.7410	22.0460	15.9270	22.4740	6.2868	5.6436	6.2141	6.5558	15.2320	23.2110	15.4570	26.4950
		1.0000	0.5030	0.9770	0.4530	0.9800	1.0000	1.0000	1.0000	1.0000	0.4960	0.9860	0.4180	0.9770
1.5	b1	1.5317	1.4972	1.5031	1.4971	1.5033	1.3802	1.3899	1.4271	1.4388	1.4975	1.4947	1.4977	1.4978
		0.1112	0.0316	0.0658	0.0315	0.0660	0.0624	0.0726	0.0818	0.0882	0.0316	0.0650	0.0313	0.0648
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8156	0.7998	0.8000	0.7998	0.8001	0.7356	0.7403	0.7596	0.7653	0.7996	0.7962	0.7999	0.7980
		0.0943	0.0304	0.0593	0.0304	0.0595	0.0550	0.0640	0.0670	0.0722	0.0302	0.0590	0.0299	0.0587
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5	rho		0.5022	0.2874	0.5011	0.2799					0.4978	0.2596	0.5014	0.2695
			0.0554	0.0745	0.0560	0.0765					0.0535	0.0789	0.0542	0.0772
			1.0000	0.9890	1.0000	0.9650					1.0000	0.9550	1.0000	0.9610
0.5	lambda						0.8165	0.6856	0.5403	0.4239	0.1925	0.2534	0.1825	0.2129
							0.1769	0.2003	0.2013	0.1629	0.4716	0.3050	0.4686	0.2797
							0.9640	0.9210	0.8090	0.6370	0.2880	0.3650	0.7480	0.7590
0.7	Desvio Médio do Erro	2.1345	0.6733	1.2671	0.6739	1.2717	1.3213	1.3749	1.7065	1.7417	0.6363	1.1622	0.6368	1.1757
	R2 ajustado médio	0.9440	0.9906	0.9717	0.9929	0.9767	0.9747	0.9728	0.9594	0.9579	0.9937	0.9804	0.9937	0.9800
	likelihood médio		-18.3000	-34.4040			-38.4710	-39.5940			-1.8838	-17.5600		
	Tempo Médio		0.0636	0.0470	0.0007	0.0005	0.3073	0.7299	0.0339	0.0403	0.2039	0.1387	0.0190	0.0178
	Tempo Total		63.5770	46.9820	0.6710	0.5430	307.3060	729.9040	33.8950	40.3220	203.9200	138.7400	19.0280	17.7680

**Tabela 15. 5** - Resultados das Simulações, SAC2, n = 27

**Modelo: Nenhuma Correlação Espacial****n=1000****repetições: 1000****Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	20.0000	20.0280	20.0110	20.0180	19.9960	20.0000	19.9990	20.0000	19.9990	20.0190	20.0010	20.1500	20.1390
		0.3272	1.0359	0.8940	1.0391	0.8981	0.3284	0.3279	0.3284	0.3279	1.0405	0.8983	1.5679	1.2890
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.5	b1	1.4999	1.4999	1.4999	1.4999	1.4999	1.4999	1.5000	1.4999	1.5000	1.4999	1.4999	1.4999	1.4999
		0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8001	0.8001	0.8001	0.8001	0.8001	0.8001	0.8001	0.8001	0.8001	0.8001	0.8001	0.8001	0.8001
		0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	rho		-0.0002	-0.0001	-0.0001	0.0000					-0.0001	0.0000	-0.0001	0.0000
			0.0073	0.0062	0.0073	0.0063					0.0073	0.0063	0.0073	0.0063
			0.2200	0.1770	0.0490	0.0450					0.2120	0.1740	0.0470	0.0470
0	lambda						-0.0058	-0.0066	-0.0065	-0.0062	-0.0055	-0.0066	-0.0071	-0.0075
							0.0628	0.0510	0.0648	0.0522	0.0642	0.0515	0.0652	0.0528
							0.6130	0.6310	0.1540	0.1320	0.5990	0.6390	0.7240	0.7750
0.7	Desvio Médio do Erro	0.6994	0.6980	0.6980	0.6981	0.6980	0.6979	0.6979	0.6979	0.6979	0.6975	0.6975	0.6975	0.6975
	R2 ajustado médio	0.9933	0.9933	0.9933	0.9933	0.9933	0.9933	0.9933	0.9933	0.9933	0.9933	0.9933	0.9933	0.9933
	likelihood médio	0.0000	-712.39	-712.36			-712.38	-712.40			-139.55	-139.53		
	Tempo Médio	0.0000	10.4850	0.2782	0.1658	0.0030	13.5080	0.6312	19.9630	0.1480	33.0360	0.8057	0.9293	0.0289
	Tempo Total	0.00	10,485.00	278.15	165.75	3.00	13,508.00	631.24	19,963.00	148.01	33,036.00	805.69	929.31	28.87

**Tabela 15. 6 - Resultados das Simulações, Sem Correlação Espacial, n = 1000**

**Modelo: SAR**  
**n=1000**  
**repetições: 500**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	150.9200	20.0220	77.7150	20.0280	83.2340	163.2800	157.4800	159.4500	156.5500	20.0660	86.2460	20.2270	60.2890
		1.2474	1.4969	4.6996	1.5120	4.7810	1.1159	1.0240	0.8063	1.0201	1.5032	4.8933	1.9303	4.9746
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.5	b1	1.5535	1.5002	1.5320	1.5002	1.5336	1.3920	1.4683	1.4426	1.4803	1.5002	1.5321	1.5002	1.5323
		0.0167	0.0045	0.0125	0.0045	0.0126	0.0111	0.0132	0.0105	0.0132	0.0045	0.0124	0.0045	0.0124
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8284	0.7999	0.8168	0.7999	0.8176	0.7415	0.7824	0.7687	0.7889	0.7999	0.8166	0.7999	0.8168
		0.0146	0.0044	0.0113	0.0044	0.0115	0.0100	0.0115	0.0102	0.0116	0.0044	0.0111	0.0044	0.0111
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5	rho		0.4999	0.2773	0.4999	0.2564					0.4997	0.2457	0.4999	0.2471
			0.0055	0.0166	0.0056	0.0171					0.0056	0.0174	0.0056	0.0172
			1.0000	1.0000	1.0000	1.0000					1.0000	1.0000	1.0000	1.0000
0	lambda						0.9420	0.6962	0.5686	0.5648	-0.0088	0.3166	-0.0099	0.2968
							0.0207	0.0376	0.0422	0.0315	0.0629	0.0609	0.0628	0.0569
							1.0000	1.0000	1.0000	1.0000	0.6400	1.0000	0.7020	1.0000
0.7	Desvio Médio do Erro	2.0465	0.6998	1.6454	0.6998	1.6497	1.3532	1.6697	1.7052	1.7747	0.6993	1.6003	0.6993	1.6024
	R2 ajustado médio	0.9488	0.9932	0.9626	0.9940	0.9666	0.9775	0.9658	0.9643	0.9614	0.9940	0.9686	0.9940	0.9685
	likelihood médio		-732.37	-1,575.30			-1,457.20	-1,639.20			-159.52	-983.55		
	Tempo Médio		10.8	0.3	0.2	0.0	107.7	0.9	19.8	0.1	40.2	1.0	0.9	0.0
	Tempo Total		5,418.8	139.7	82.9	1.1	53,846.0	467.2	9,920.9	72.0	20,096.0	518.3	469.0	13.5

**Tabela 15. 7 - Resultados das Simulações, SAR, n = 1000**

**Modelo: SEM**  
**n=1000**  
**repetições: 500**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	19.9970	18.7120	19.3140	19.9830	19.9530	19.9970	19.9990	19.9970	19.9990	19.9520	19.9460	10.1010	14.5290
		0.3287	1.4537	1.1715	1.4780	1.1703	0.2959	0.3141	0.2958	0.3141	1.2241	1.1461	1.1224	1.2170
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.5	b1	1.5001	1.5001	1.5001	1.5001	1.5001	1.5001	1.5001	1.5001	1.5001	1.5001	1.5001	1.5001	1.5001
		0.0046	0.0046	0.0046	0.0046	0.0046	0.0043	0.0044	0.0043	0.0044	0.0045	0.0046	0.0044	0.0046
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.7999	0.7999	0.7999	0.7999	0.7999	0.7999	0.7999	0.7999	0.7999	0.7999	0.7999	0.7999	0.7999
		0.0047	0.0047	0.0047	0.0047	0.0047	0.0043	0.0045	0.0043	0.0045	0.0044	0.0046	0.0044	0.0046
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	rho		0.0095	0.0051	0.0001	0.0003					0.0003	0.0004	0.0002	0.0004
			0.0096	0.0078	0.0097	0.0077					0.0079	0.0075	0.0079	0.0075
			0.4880	0.2920	0.1180	0.0860					0.1820	0.1720	0.0420	0.0520
0.5	lambda						0.4907	0.2690	0.4953	0.2725	0.4899	0.2691	0.4936	0.2712
							0.0505	0.0477	0.0509	0.0480	0.0512	0.0488	0.0519	0.0485
							1.0000	0.9960	1.0000	1.0000	1.0000	0.9900	1.0000	1.0000
0.7	Desvio Médio do Erro	0.7419	0.7397	0.7401	0.7403	0.7403	0.6983	0.7237	0.6982	0.7236	0.6980	0.7233	0.6978	0.7232
	R2 ajustado médio	0.9924	0.9924	0.9924	0.9924	0.9924	0.9933	0.9928	0.9933	0.9928	0.9933	0.9928	0.9933	0.9928
	likelihood médio		-770.33	-770.84			-729.75	-755.55			-156.88	-182.64		
	Tempo Médio		10.8280	0.2752	0.1649	0.0030	84.7480	2.2636	19.8520	0.1452	37.8580	0.8664	0.9331	0.0273
	Tempo Total		5,414.00	137.62	82.43	1.50	42,374.00	1,131.80	9,926.10	72.61	18,929	433.20	466.55	13.64

**Tabela 15. 8** - Resultados das Simulações, SEM, n = 1000

**Modelo: SAC 1**  
**n=1000**  
**repetições: 500**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	-28.6660	19.3300	-7.5237	20.0320	-7.1585	-18.808	-29.8390	-25.1500	-29.5720	20.0310	-7.4047	10.1500	-6.9866
		1.1150	0.9042	1.6664	0.8977	1.7083	1.4603	1.0492	0.9991	1.0235	0.7694	1.6455	1.0693	1.6946
		1.0000	1.0000	0.9980	1.0000	0.9940	1.0000	1.0000	1.0000	1.0000	1.0000	0.9940	1.0000	0.9940
1.5	b1	1.5474	1.5005	1.5340	1.4998	1.5338	1.4186	1.5628	1.5015	1.5593	1.4999	1.5345	1.4999	1.5344
		0.0147	0.0048	0.0124	0.0048	0.0124	0.0194	0.0139	0.0136	0.0136	0.0046	0.0122	0.0046	0.0122
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8260	0.8005	0.8185	0.8001	0.8183	0.7576	0.8340	0.8016	0.8322	0.8001	0.8186	0.8001	0.8186
		0.0128	0.0049	0.0112	0.0049	0.0112	0.0141	0.0124	0.0116	0.0123	0.0047	0.0111	0.0047	0.0111
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
-0.5	rho		-0.4931	-0.2233	-0.5003	-0.2272					-0.5003	-0.2250	-0.5003	-0.2254
			0.0076	0.0170	0.0075	0.0164					0.0064	0.0162	0.0063	0.0164
			1.0000	1.0000	1.0000	1.0000					1.0000	1.0000	1.0000	1.0000
0.5	lambda						-0.7181	0.1337	-0.2395	0.1024	0.4896	0.0569	0.4932	0.0531
							0.1095	0.0787	0.0695	0.0526	0.0502	0.0639	0.0508	0.0642
							0.9980	0.9400	0.9960	0.7320	1.0000	0.7440	1.0000	0.7580
0.7	Desvio Médio do Erro	1.7994	0.7408	1.6089	0.7408	1.6088	1.5931	1.7869	1.7375	1.7907	0.6986	1.6057	0.6984	1.6058
	R2 ajustado médio	0.9596	0.9934	0.9671	0.9931	0.9676	0.9682	0.9600	0.9622	0.9598	0.9939	0.9677	0.9939	0.9677
	likelihood médio		-787.22	-1,549.50			-1,572.0	-1,652.0			-173.76	-975.80		
	Tempo Médio		10.8160	0.2759	0.1651	0.0028	48.6450	2.8351	19.8770	0.1461	44.5660	0.9460	0.9327	0.0276
	Tempo Total		5,407.80	137.96	82.55	1.41	24,323.	1,417.60	9,938.30	73.04	22,283	472.99	466.36	13.82

**Tabela 15. 9** - Resultados das Simulações, SAC1, n = 1000



**Modelo: SAC 2**

**n=1000**

**repetições: 500**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	150.82	17.9840	75.8890	20.0920	83.0750	163.4100	157.51	159.7900	156.6700	20.0770	87.1010	10.1920	55.7660
		1.3048	2.3003	4.8696	2.3113	4.9812	1.4609	1.0015	0.8151	1.0110	2.0992	5.0120	1.3837	4.5290
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.5	b1	1.5540	1.4992	1.5317	1.5001	1.5338	1.3905	1.4669	1.4372	1.4779	1.5001	1.5315	1.5001	1.5319
		0.0179	0.0045	0.0134	0.0045	0.0136	0.0108	0.0135	0.0103	0.0137	0.0044	0.0134	0.0044	0.0134
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8297	0.8000	0.8175	0.8005	0.8187	0.7422	0.7829	0.7672	0.7889	0.8004	0.8174	0.8004	0.8176
		0.0145	0.0047	0.0115	0.0047	0.0115	0.0092	0.0112	0.0093	0.0112	0.0045	0.0111	0.0045	0.0111
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5	rho		0.5076	0.2839	0.4996	0.2567					0.4996	0.2424	0.4996	0.2444
			0.0084	0.0169	0.0083	0.0174					0.0075	0.0174	0.0076	0.0170
			1.0000	1.0000	1.0000	1.0000					1.0000	1.0000	1.0000	1.0000
0.5	lambda						0.9567	0.7174	0.6031	0.5932	0.4890	0.3793	0.4921	0.3545
							0.0169	0.0345	0.0402	0.0295	0.0506	0.0568	0.0509	0.0522
							1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.7	Desvio Médio do Erro	2.1078	0.7392	1.6950	0.7403	1.7016	1.3361	1.6844	1.7070	1.7940	0.6980	1.6274	0.6979	1.6307
	R2 ajustado médio	0.9460	0.9906	0.9597	0.9933	0.9647	0.9782	0.9654	0.9644	0.9607	0.9940	0.9677	0.9940	0.9675
	likelihood médio		-787.48	-1,605.30			-1,448.9	-1,652.3			-174.28	-1,004.6		
	Tempo Médio		10.8180	0.2765	0.1642	0.0031	111.6400	1.1278	19.8480	0.1440	44.4240	1.0443	0.9328	0.0269
	Tempo Total		5,408.80	138.27	82.08	1.53	55,820.00	563.90	9,924.10	71.99	22,212	522.15	466.38	13.44

**Tabela 15. 10** - Resultados das Simulações, SAC2, n = 1000

**Modelo: Nenhuma Correlação Espacial**  
**n=4267**  
**repetições: 1000 Para Contig 1 para Dist**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	19.9950	20.3830	20.0100	20.3920	20.0100	19.8110	19.9950	19.8130	19.9950	20.3870	20.0110	20.3130	20.0120
		0.1468 1.0000	1.0000	0.3336 1.0000	1.0000	0.3348 1.0000	1.0000	0.1469 1.0000	0.5000	0.1469 1.0000	1.0000	0.3341 1.0000	1.0000	0.4489 1.0000
1.5	b1	1.5001	1.5020	1.5001	1.5020	1.5001	1.5020	1.5001	1.5015	1.5001	1.5020	1.5001	1.5019	1.5001
		0.0021 1.0000	1.0000	0.0021 1.0000	1.0000	0.0021 1.0000	1.0000	0.0021 1.0000	0.5000	0.0021 1.0000	1.0000	0.0021 1.0000	1.0000	0.0021 1.0000
0.8	b2	0.8000	0.8016	0.8000	0.8016	0.8000	0.8017	0.8000	0.8023	0.8000	0.8017	0.8000	0.8017	0.8000
		0.0021 1.0000	1.0000	0.0021 1.0000	1.0000	0.0021 1.0000	1.0000	0.0021 1.0000	0.0000	0.0021 1.0000	1.0000	0.0021 1.0000	1.0000	0.0021 1.0000
0.5	rho		-0.0046	-0.0001	-0.0047	-0.0001					-0.0046	-0.0001	-0.0077	-0.0001
			1.0000	0.0022 0.1330	0.5000	0.0022 0.0490					1.0000	0.0022 0.1320	0.5000	0.0027 0.1050
0	lambda						0.0185	-0.0011	-0.0118	-0.0004	0.0195	0.0000	0.0228	-0.0004
							0.5000	0.0242 0.5850	0.0000	0.0257 0.0960	0.5000	0.0246 0.5980	1.0000	0.0259 0.7590
0.7	Desvio Médio do Erro	0.7002	0.7046	0.6998	0.7046	0.6998	0.7049	0.6998	0.6995	0.6998	0.7045	0.6997	0.7045	0.6998
	R2 ajustado médio	0.9933	0.9932	0.9933	0.9932	0.9933	0.9932	0.9933	0.9932	0.9933	0.9932	0.9933	0.9932	0.9933
	likelihood médio		-3,082	-3,052			-3,083	-3,052			-639	-610		
	Tempo Médio		182.91	1.10	0.52	0.01	171.96	1.43	4,645.10	0.50	726.73	3.76	67.84	0.0603
	Tempo Total		366	1,097	1	7	344	1,431	4,645	500	1,454	3,760	136	60

**Tabela 15. 11 - Resultados das Simulações, Sem Correlação Espacial, n = 4267**

**Modelo: SAR**

**n=4267**

**repetições: 500 para Contig e 2 para Dist**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	133.27	20.01	19.80	20.03	111.16	124.26	138.71	133.01	138.61	20.00	134.25	20.17	15.53
		1.4816		2.8666		8.6063		0.5935		0.5890		1.6934		1.7692
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.5	b1	1.5268	1.4997	1.5009	1.4997	1.5212	1.4396	1.4565	1.5326	1.4581	1.4997	1.4622	1.4996	1.5655
		0.0211		0.0157		0.0182		0.0083		0.0083		0.0084		0.0138
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8142	0.8038	0.8003	0.8038	0.8112	0.7698	0.7768	0.8149	0.7776	0.8039	0.7798	0.8038	0.8708
		0.0208		0.0155		0.0187		0.0080		0.0080		0.0081		0.0119
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5	rho		0.4992	0.4615	0.4991	0.0901					0.4992	0.0171	0.5000	0.0933
				0.0089		0.0304						0.0060		0.0230
			1.0000	1.0000	1.0000	0.9670					1.0000	0.9390	1.0000	0.9870
0	lambda						0.9900	0.9144	1.0000	0.8733	-0.0180	0.9129	-0.0192	0.8533
								0.0053		0.0051		0.0054		0.0096
							1.0000	1.0000	1.0000	1.0000	0.0000	1.0000	0.0000	1.0000
0.7	Desvio Médio do Erro	6.7984	0.7016	5.0519	0.7016	6.2691	2.0984	2.7148	0.0000	2.8141	0.7016	2.7146	0.7015	3.7199
	R2 ajustado médio	0.6183	0.9958	0.5558	0.9960	0.6749	0.9640	0.9391	1.0000	0.9346	0.9960	0.9391	0.9960	0.8856
	likelihood médio		-3,099	-11,586			-7,938	-9,374			-657	-6,928		
	Tempo Médio		173.10	2.50	1.10	0.04	136.83	7.08	48.39	2.54	523.21	5.4349	33.3700	0.0626
	Tempo Total		346	2,499	2	36	274	7,085	97	2,541	1,046	5,435	67	63

**Tabela 15. 12 - Resultados das Simulações, SAR, n = 4267**

**Modelo: SEM**

**n=4267**

**repetições: 500 para Contig e 1 para Dist**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	20.0010	20.5390	19.7490	20.7470	20.0010	19.8200	20.0010	19.8200	20.0010	20.6860	19.9890	14.8170	17.2120
		0.1538		0.3871		0.3942		0.1516		0.1516		0.3806		0.4839
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.5	b1	1.5000	1.5013	1.5000	1.5013	1.5000	1.5016	1.5000	1.5016	1.5000	1.5011	1.5000	1.5058	1.4994
		0.0021		0.0021		0.0021		0.0021		0.0021		0.0021		0.0021
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8000	0.8024	0.8000	0.8024	0.8000	0.8022	0.8000	0.8022	0.8000	0.8019	0.8000	0.8099	0.7999
		0.0022		0.0022		0.0022		0.0021		0.0021		0.0021		0.0022
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	rho		-0.0058	0.0019	-0.0075	0.0000					-0.0067	0.0001	-0.0630	-0.0109
				0.0025		0.0025						0.0025		0.0032
			1.0000	0.2880	1.0000	0.0650					1.0000	0.1320	1.0000	0.9760
0.5	lambda						0.4870	0.1894	0.4920	0.1998	0.4950	0.1889	0.4916	0.1996
								0.0248		0.0253		0.0249		0.0255
							1.0000	0.9780	1.0000	1.0000	1.0000	0.9740	1.0000	1.0000
0.7	Desvio Médio do Erro	0.7206	0.7198	0.7202	0.7198	0.7202	0.6998	0.7121	0.6997	0.7118	0.6992	0.7120	0.7680	0.7154
	R2 ajustado médio	0.9929	0.9928	0.9929	0.9928	0.9929	0.9932	0.9930	0.9932	0.9930	0.9932	0.9930	0.9918	0.9930
	likelihood médio		-3,173	-3,175			-3,087	-3,142			-642	-699		
	Tempo Médio		183.9100	2.6941	0.5620	0.0390	181.5000	3.6135	4,800	2.6763	836.7300	4.7463	67.9220	0.0626
	Tempo Total		184	2,694	1	39	182	3,614	4,800	2,676	837	4,746	67.9220	62.6290

**Tabela 15. 13 - Resultados das Simulações, SEM, n = 4267**

**Modelo: SAC1**

**n=4267**

**repetições: 500 para Contig e 1 para Dist**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	-24.1980	19.9620	-19.1960	20.2870	-12.6920	-27.4040	-27.5000	-25.8730	-27.2510	20.5650	-15.7490	8.9852	-5.6630
		0.4553		0.9460		1.1669		0.3762		0.3747		0.8699		0.5552
		2.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.5	b1	1.5235	1.4964	1.5212	1.4963	1.5183	1.5695	1.5665	1.5473	1.5633	1.4964	1.5365	1.5110	1.5251
		0.0064		0.0063		0.0063		0.0051		0.0051		0.0050		0.0056
		2.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8124	0.8020	0.8112	0.8020	0.8097	0.8368	0.8352	0.8251	0.8335	0.8018	0.8191	0.8152	0.8089
		0.0058		0.0057		0.0057		0.0049		0.0049		0.0047		0.0049
		2.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
-0.5	rho		-0.4990	-0.0522	-0.5027	-0.1200					-0.5060	-0.1024	-0.4890	-0.1090
				0.0082		0.0099						0.0079		0.0083
			1.0000	1.0000	1.0000	1.0000					1.0000	1.0000	1.0000	1.0000
0.5	lambda						0.5500	0.5926	0.2940	0.5359	0.5420	0.6215	0.5378	0.5971
								0.0273		0.0212		0.0227		0.0201
							1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.7	Desvio Médio do Erro	1.7870	0.7277	1.7714	0.7278	1.7939	1.6556	1.5007	1.6998	1.5337	0.7024	1.4475	0.7858	1.4896
	R2 ajustado médio	0.9590	0.9934	0.9612	0.9932	0.9586	0.9648	0.9710	0.9629	0.9698	0.9937	0.9731	0.9921	0.9715
	likelihood médio		-3,252	-7,015			-6,771	-6,476			-703	-3,904		
	Tempo Médio		204.2200	1.6623	0.6560	0.0100	4,434	11.6510	4,842	0.6111	937.7300	6.5392	67.9380	0.0780
	Tempo Total		204	831	1	5	4,434	5,826	4,842	306	938	3,270	67.9380	38.9930

**Tabela 15. 14 - Resultados das Simulações, SAC1, n = 4267**

**Modelo: SAC2**

**n=4267**

**repetições: 500 para Contig e 1 para Dist**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
20	Const	133.3300	20.4030	19.7120	20.4910	111.4300	123.8300	138.7500	144.8900	138.6400	20.5160	134.3400	18.6840	15.3300
		1.4253		2.7716		8.1340		0.5985		0.5943		1.8505		1.7245
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.5	b1	1.5263	1.5015	1.5004	1.5015	1.5208	1.4450	1.4559	1.4354	1.4575	1.5015	1.4615	1.5011	1.5660
		0.0206		0.0150		0.0178		0.0084		0.0084		0.0087		0.0135
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	b2	0.8134	0.8025	0.7997	0.8025	0.8106	0.7731	0.7767	0.7687	0.7776	0.8021	0.7797	0.8033	0.8716
		0.0209		0.0156		0.0190		0.0084		0.0084		0.0085		0.0115
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5	rho		0.4975	0.4621	0.4971	0.0892				0.0000	0.4970	0.0169	0.4385	0.0954
				0.0087		0.0287				0.0000		0.0066		0.0227
			1.0000	1.0000	1.0000	0.9760				0.0000	1.0000	0.9340	1.0000	0.9840
0.5	lambda						0.9900	0.9146	1.1642	0.8742	0.4820	0.9134	0.4887	0.8544
								0.0055		0.0052		0.0071		0.0095
							1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.7	Desvio Médio do Erro	6.8072	0.7195	5.0585	0.7195	6.2828	2.0772	2.7164	2.3402	2.8146	0.6995	2.7160	0.7145	3.7236
	R2 ajustado médio	0.6173	0.9948	0.5540	0.9956	0.6734	0.9635	0.9390	0.9536	0.9345	0.9959	0.9390	0.9957	0.8854
	likelihood médio		-3,207	-11,591			-7,895	-9,377			-677	-6,932		
	Tempo Médio		183.6100	1.6900	0.5160	0.0106	175.6400	8.1357	5,023	0.6225	987.0100	7.1176	68.0470	0.0850
	Tempo Total		184	845	1	5	176	4,068	5,023	311	987	3,559	68.0470	42.4900

**Tabela 15. 15** - Resultados das Simulações, SAC1, n = 4267

Um aspecto geral observado em todas as tabelas é que, quando há a presença de  $\rho$ , uma estimação SEM sempre aponta para a existência de componente espacial, no caso  $\lambda$ . Apesar de isso ser incorreto, o problema observado é o mesmo da estatística  $I$  de Moran.

A existência do componente de correlação espacial da variável dependente também multiplica o termo de erro na forma reduzida da equação, como pode ser observado em (14.1) e (14.2). Se não for controlada a correlação espacial na variável dependente, a estrutura de dispersão do erro irá refletir essa correlação e provocará, incorretamente, a inferência da existência de um componente de correlação espacial no erro.

Com isso, é possível antever que todo modelo que apresente  $\rho$  diferente de 0 terá  $\lambda$  incorretamente estimado se for utilizado um modelo SEM. Portanto, se um modelo SAR apontar para a existência de  $\rho$  deve-se esperar que o modelo SEM aponte para a existência de  $\lambda$ .

O contrário, no entanto, não ocorre. Pela própria natureza dos componentes espaciais, quando existe apenas a presença do componente de correlação espacial no erro, isso não provoca uma estimação incorreta de  $\rho$ . Portanto, o modelo SAR é capaz de dizer corretamente se  $\rho$  é ou não diferente de 0, enquanto que o modelo SEM sozinho não.

Um uso importante de uma simulação é indicar estratégias de análises para as situações em que a geração de dados não está sobre o controle do pesquisador, ou seja, a situação usual de uma análise econométrica. Isso é possível pois a simulação permite conhecer a verdade, enquanto que na análise usual, temos apenas as inferências.

Dessa forma podemos concluir que, caso o modelo SAR aponte para a existência de  $\rho$ , a estimação por SEM não adicionará muita informação. O ideal, então, seria estimar por um modelo completo (SAC), que continua controlando a presença de  $\rho$  e testa a presença de  $\lambda$ . Caso o modelo SAR não aponte para a existência de  $\rho$ , pode-se utilizar uma estimativa SEM

Em contraposição a esse tipo de busca de especificação, que seria uma análise *forward*, poderíamos partir de uma *backward*, ou seja, iniciar por SAC e se um dos componentes espaciais não for encontrado, passar para um modelo mais simples.

Outro aspecto geral que pode ser observado é que independentemente do modelo escolhido, matriz e do tamanho de amostra, não há muita diferença em termos de esperança dos demais parâmetros, exceto em relação à constante. Isso é um pouco inesperado, pois ignorar o componente espacial na variável dependente,  $\rho$ , poderia levar a um viés significativo de omissão de variável.

Observando as estimações OLS, pode-se ver que o viés verificado é muito pequeno nos modelos que possuem  $\rho$  diferente de 0. O tamanho do viés chega a ser desprezível, e em intervalo de confiança, em geral, inclui o valor real do parâmetro. O que se tem claramente é um ganho de eficiência ao se incluir a os parâmetros espaciais, tanto na estimação dos coeficientes quanto na estimação da variância do erro.

O fato não haver tanta diferença na esperança do estimador, assim com o problema da estatística de Moran e o problema do modelo SEM incorretamente estimar  $\lambda$  quando há a presença de  $\rho$ , é simples de ser compreendido. Haverá viés de omissão de variável quando esta variável omitida, além de estar correlacionada com  $y$ , ou seja, tiver seu parâmetro diferente de 0, também for correlacionada com ao menos uma das variáveis explicativas. Para a variável dependente defasada espacialmente provocar viés nos estimadores das demais variáveis é preciso que:

$$\rho \neq 0 \quad (15.1)$$

$$\text{Corr}(X, Wy) \neq 0 \quad (15.2)$$

Se (15.1) ou (15.2) não forem verificadas de forma significativa, não haverá viés, ou este será muito pequeno, e praticamente desprezível. Mais especificamente, Anselin (1999) explicita e decompõe o estimador obtido por máxima verossimilhança (ML) de um modelo SAR:

$$\hat{\beta}_{ml} = (X'X)^{-1} X'(y - \rho Wy) \quad (15.3)$$



Decompondo (15.3), pode-se chegar a:

$$\begin{aligned}\hat{\beta}_{ml} &= (X'X)^{-1}X'y - \rho(X'X)^{-1}X'(Wy) \\ \hat{\beta}_{ml} &= \hat{\beta}_{ols} - \rho\hat{\beta}_{wy}\end{aligned}\quad (15.4)$$

Pela expressão (15.4) é possível concluir o mesmo que se conclui com as expressões (15.1) e (15.2). Se  $\rho=0$  ou se a correlação de  $Wy$  com  $X$  for baixa, não haverá muita diferença entre os parâmetros estimados por OLS e por ML, exceto para os parâmetros espaciais e constante.

Para verificar isso, fez-se o cálculo em 500 iterações para estimar  $\beta_{wy}$ , ou seja, a seguinte regressão.

$$Wy = X\beta_{wy} + \xi \quad (15.5)$$

A matriz  $W$  utilizada foi a matriz de distâncias invertida, tanto para a geração de dados quanto para a estimação da equação (15.5).

A tabela a seguir apresenta os resultados médios, desvios e R2 ajustado médio:

	n=27	n=1000	n=4267
<b>Constante</b>	265.7778 14.9074	262.3836 2.3558	226.2991 0.0317
<b>b1</b>	0.0611 0.1932	0.1007 0.0313	0.0549 0.0423
<b>b2</b>	0.0310 0.1784	0.0525 0.0251	0.0317 0.0413
<b>R2 ajustado médio</b>	0.0229	0.0234	0.0005

**Tabela 15. 16** - Poder de explicação de X sobre Wy,  
valores médios e desvios em 500 simulações

Pela tabela (15.16) é possível perceber que os coeficientes de  $b1$  e  $b2$  são próximos de zero, sendo que para estimar o impacto no estimador OLS ainda deverão ser multiplicados pela estimativa de  $\rho$ . O R2 ajustado também é muito baixo, caindo com o aumento de  $n$ . Com isso é de se esperar pouca diferença nos coeficientes de  $b1$  e  $b2$ .

Fica pendente a questão teórica do que significaria uma alta correlação de  $X$ , exceto a constante, com  $Wy$ , fato esse que acarretaria um grande viés nos coeficientes. Em termos teóricos seria necessário discutir o que causaria um grupo de variáveis explicativas no local estarem correlacionado com a variável explicada defasada espacialmente.

A mudança no valor da constante, no entanto, é esperada e facilmente compreendida. A constante, em um modelo econométrico, tem justamente o papel de captar a esperança das variáveis omitidas. Por esse motivo, nos modelos em que  $\rho$  é diferente de 0, a constante oscila bastante no seu valor, em especial quando são contrastados modelos que estimam  $\rho$  ou não.

Por esta razão, olhar a variação da constante entre um modelo e outro pode dar indicações sobre a especificação adequada. Evidentemente uma análise não poderia apenas observar a constante, pois de acordo com a construção das variáveis e nível de controle incluído ela pode ter um valor ou uma variação pequena por construção, o que dificultaria tal análise.

O fato do valor dos parâmetros para as outras variáveis não sofrer alteração substancial não é uma crítica direta aos modelos espaciais e nem invalida seu uso. A crítica fica ao porquê do uso.

Considerando como correta a verificação experimental de que há pouca mudança no valor dos parâmetros estimados, e a série de restrições econométricas em termos teóricos e operacionais que o uso de modelos espaciais implica, pode-se levantar a questão sobre quando resolver utilizá-los.

A inferência que pode ser feita é que o uso das ferramentas de econometria espacial deve ser aplicado quando o interesse da análise é na estimação dos parâmetros de correlação espacial. A simples presença do componente de correlação espacial não parece provocar desvios da estimativa dos parâmetros que justifiquem a sua inclusão apenas como controle. Claramente há um ganho de eficiência, porém isso ocorreria com a inclusão de qualquer variável relevante omitida que fosse fonte de variação no erro.

Um teste possível para avaliar o benefício potencial do uso de estimações espaciais seria justamente estimar os parâmetros da equação (15.5). Os parâmetros estimados ainda seriam ponderados por  $\rho$ , que por sua vez está entre  $(-1)$  e  $(+1)$ , no caso da matriz normalizada.

Pode-se considerar a matriz de contigüidade (Contig) como uma simplificação de uma medida mais complexa de distância, o quadrado da distância euclidiana invertida (Dist). Essa última matriz, como já indicado, foi utilizada na geração de dados.

Comparando as colunas das estimativas feitas com a matriz original (Dist) com a estimação feita com a matriz de contigüidade (Contig), é possível verificar a adequabilidade dessa simplificação.

Observando as estimativas dentro de cada tamanho de amostra é possível notar que a matriz de contigüidade não é, de forma geral, uma boa substituição para a matriz de distâncias. Ela capta parcialmente o efeito de  $\rho$  e  $\lambda$ , além de captar erroneamente  $\lambda$ , quando este inexistente.

O fato de uma medida simplificada captar parcialmente o efeito de  $\rho$  é esperado, já que ela contém apenas parte da informação da matriz completa com as distâncias. O mesmo vale para a captação parcial de  $\lambda$ . Quanto a captar erroneamente  $\lambda$  quando inexistente, a explicação seria a mesma do problema com os modelos SEM. Como, mesmo estimando por SAC, a matriz de contigüidade não consegue captar todo efeito de  $\rho$ , o efeito restante é incluído no erro. Com isso, o resíduo ainda possui características espaciais que acabam sendo, de forma incorreta, estimado como  $\lambda$ .

Outros aspectos importantes chamam a atenção. Na amostra pequena, mesmo utilizando o modelo e a matriz corretos,  $\lambda$  é estimado com viés. Ao se aumentar o tamanho da amostra, o viés é eliminado com a matriz correta, e com a matriz simplificada a captação é parcial. Dessa forma não parece ser recomendado o uso de modelos espaciais em amostras muito pequenas quando o objetivo for estimar o parâmetro de correlação espacial no resíduo.

O aspecto mais importante, no entanto, é que há, pelas simulações, um efeito prejudicial no aumento do tamanho da amostra. Quanto maior o número de observações, menor o poder de captação da matriz de contigüidade. Uma possível explicação para isso é que quanto maior a malha original, o conjunto de pontos e relações captados por uma matriz de contigüidade será cada vez menor em termos percentuais. Isso, no entanto, depende das matrizes e da variabilidade de  $y$ .

Com isso podemos inferir que numa amostra pequena há viés na estimação de  $\lambda$ , mesmo com o modelo correto. Em particular isso parece ser verificado com a matriz que reflete as distâncias e vizinhanças dos Estados do Brasil. Na amostra grande, a escolha incorreta da matriz pode gerar viés e subestimar os parâmetros.

Para comparar o quanto  $Wdist*y$  pode ser substituído por  $Wcontig*y$ , foram feitos dois testes, também em 500 iterações. A primeira medida é a correlação entre  $Wdist*y$  e  $Wcontig*y$ . A segunda é uma regressão, segundo a equação a seguir:

$$Wcontig * y = \text{constante} + Wdist * y + \xi \quad (15.6)$$

A tabela a seguir apresenta os resultados médios, desvios e  $R^2$  ajustado médio:

$Y = Wcontig * y$	$n=27$	$n=1000$	$n=4267$
<b>Constante</b>	-32.3643 68.7105	72.4576 14.6499	144.4 2.734
$Wdist * y$	1.1197 0.2544	0.7316 0.0543	0.45862 0.01183
<b>R2 ajustado médio</b>	0.6226	0.4066	0.45384
<b>Correlação entre <math>Wdist*y</math> e <math>Wcontig*y</math>.</b>	0.7918	0.6373	0.67363

**Tabela 15. 17** - Correlação e ajuste entre  $Wdist*y$  e  $Wcontig*y$ , valores médios e desvios em 500 simulações

Note que a correlação cai quando passamos de  $n=27$  para  $n=1000$ . Além disso, o coeficiente de  $Wdist * y$  cai com o aumento do tamanho da amostra e a constante aumenta, indicando que com o crescimento da amostra a associações entre as duas variáveis defasadas espacialmente se torna menor.

Pelas tabelas (15.11) a (15.15), com  $n=4267$ , exceto no modelo SAR-ML quando  $\rho = 0.5$ , a matriz de contigüidade praticamente não capta nenhuma correlação espacial e ocasiona um viés grande na estimativa de  $\lambda$ .

Uma questão relevante é se o inverso seria válido, ou seja, se dados gerados por uma relação de contigüidade pode ser estimado corretamente por uma matriz mais complexa, como a de distância. Apesar de ser difícil de considerar teoricamente uma relação de dependência espacial efetiva que se dá de forma binária, esta simplificação pode refletir externalidades de redes mais locais e uma estimação correta por uma matriz mais complexa também auxiliaria na robustez do modelo. Em um teste de SAR com  $n=1000$ , foram gerados dados com a matriz de contigüidade e estimados com a matriz de distâncias.

As estimativas foram boas, porém subestimaram o parâmetro espacial. Pode-se afirmar que, mesmo assim, foram capazes de captar bem a relação. No entanto, mesmo esse resultado parece reforçar que a adequação da matriz de ponderação é essencial para uma estimativa correta do parâmetro espacial. A tabela a seguir mostra os resultados para um modelo SAR estimado por GMM, com dados gerados utilizando a matriz de contigüidade. Vale ressaltar, no entanto, que este seria um aspecto que poderia ser mais explorado em futuros estudos.

REAL		OLS	SAR GMM	
			Dist	Contig
20	Const	148.7000	41.1800	20.0410
		1.6100	7.5402	1.2233
		1.0000	1.0000	1.0000
1.5	b1	1.5816	1.5465	1.4999
		0.0218	0.0149	0.0045
		1.0000	1.0000	1.0000
0.8	b2	0.8437	0.8254	0.7998
		0.0174	0.0128	0.0045
		1.0000	1.0000	1.0000
0.5	$\rho$		0.4082	0.4999
			0.0267	0.0047
			1.0000	1.0000
0.7	Desvio Médio do Erro	2.5536	1.9047	0.6976
	R2 ajustado médio	0.9250	0.9581	0.9944
	Tempo Médio		0.2	0.0
	Tempo Total		119.6	1.2

**Tabela 15. 18** - Simulação de Dados Gerados com a Matriz de Contigüidade.  
SAR, 500 iterações,  $n = 1000$

Um aspecto operacional relevante de ser analisado é o tempo médio de cada regressão e a exigência da capacidade do computador. As simulações foram realizadas em computadores com 512Mb de RAM e 1,4Ghz e 2,0 Ghz.

A matriz de contigüidade reduz de maneira significativa o tempo de cada regressão, em especial quando o tamanho da amostra grande. Esse resultado também é esperado, visto que os cálculos necessários para as estimações muitas vezes envolvem a inversão e demais operações matriciais que consomem tempo. No caso das simulações feitas, apenas como referência, a estimativa de um modelo SEM com amostra grande chega a durar cerca de 4500 segundos, ou 1h25min. A estimação por SEM-GMM é o processo que mais consome recursos do computador, sendo que foi necessário ajustar o uso de memória da rotina original de LeSage (1999) para que o computador fosse capaz de concluir os cálculos.

Apesar de atrativo o ganho de tempo e recurso da matriz de contigüidade, tendo em consideração as desvantagens já indicadas sobre seu poder de explicação, o ideal seria buscar outra alternativa para o melhor uso de recursos. Neste aspecto vale a comparação entre os modelos de estimação por ML e por GMM.

De um modo geral os modelos GMM parecem ser mais atrativos que os modelos ML. Em termos operacionais, exceto no caso do modelo SEM, são mais rápidos e consomem menos memória. Em termos econométricos, são mais robustos pois não precisam assumir a normalidade dos resíduos. Quanto à esperança dos estimadores, eficiência e erros do tipo I e II são muito próximos aos modelos ML na simulação. Vale ressaltar que a simulação experimental gera, pela sua definição operacional, erros normalmente distribuídos e homoscedásticos, e em alguns casos os modelos ML são ligeiramente superiores.

Um resultado não usual importante de destacar, tanto em relação aos modelos GMM e ML e quanto no uso das diferentes matrizes é a estimativa por SAR-ML com a matriz de contigüidade para dados gerados por SAR ou por SAC2 na amostra grande,  $n=4267$ . Nesses dois casos específicos a estimação SAR-ML com a matriz de contigüidade não só é bem diferente da SAR-GMM como consegue captar praticamente toda a correlação espacial na variável dependente.

Não se conseguiu compreender por completo esse resultado que se mostra como exceção. Aparentemente pode-se entender esse resultado como uma particularidade da matriz gerada a partir das AMCs brasileiras. De forma a comprovar isso foram geradas matrizes de distância e contigüidade a partir de coordenadas aleatórias de tamanho 4267. Nessa nova simulação a estimação SAR-ML e SAR-GMM com a matriz de contigüidade foram próximas e não captaram bem a correlação espacial, o resultado mais usual observado com nessa simulação. A tabela a seguir resume os resultados.

REAL		OLS	SAR	
			ML	GMM
20	<b>Const</b>	151.3900	85.5940	91.0350
		0.5634	2.2537	2.2700
		1.0000	1.0000	1.0000
1.5	<b>b1</b>	1.5468	1.5298	1.5312
		0.0076	0.0060	0.0061
		1.0000	1.0000	1.0000
0.8	<b>b2</b>	0.8251	0.8161	0.8168
		0.0064	0.0053	0.0053
		1.0000	1.0000	1.0000
0.5	<b>rho</b>		0.2485	0.2280
			0.0078	0.0080
			1.0000	1.0000
0	<b>lambda</b>			
0.7	<b>Desvio Médio do Erro</b>	1.9017	1.5805	1.5842
	<b>R2 ajustado médio</b>	0.9550	0.9656	0.9688
	<b>likelihood médio</b>		-6,550.40	
	<b>Tempo Médio</b>		1.7	0.0
	<b>Tempo Total</b>		833.8	5.4

**Tabela 15. 19** - Simulação de Dados Estimados com a Matriz de Contigüidade.  
500 iterações, n= 4267, mapa aleatório

Essa tabela confirma assim, que a diferença observada é uma particularidade da matriz gerada a partir da malha municipal brasileira. Como já indicado, cabem maiores investigações sobre o tema.

Em termos operacionais, devido às rotinas de cálculo em Matlab de LeSage (1999), a estimação por SAR-ML pode ser considerada com a mais “independente”, pois as demais ML partem de estimativas GMM para a sua otimização. Com isso, por construção operacional, os demais coeficientes deveriam estar mais próximos entre si do que com a estimação feita por SAR-ML. Sabe-se que esta explicação não é completa, porém aponta mais uma vez para a particularidade da estrutura de determinadas matrizes em conjunto com a dispersão das variáveis, o que incentiva futuros estudos particularizando as especificidades das matrizes espaciais.



## PARTE V: APLICAÇÃO EMPÍRICA

### 16. Contextualização da análise empírica

De modo a complementar a análise por meio de simulações e aprofundar a discussão sobre a escolha de modelos e matrizes espaciais, esta parte da dissertação explora a análise espacial com dados empíricos, dando continuidade e aprofundando a análise exploratória de dados espaciais apresentada no capítulo 3.

No contexto de análise de políticas públicas será testado quais decisões de políticas, refletidas na composição do gasto local estão relacionadas com variação em indicadores de performance. Para essa análise empírica, foi escolhida a taxa de variação no emprego formal.

Vale ressaltar que o objeto da análise da dissertação não é o caso particular aqui apresentado. Ele serve para ilustrar uma análise de dados espaciais e verificar o impacto das mudanças nas escolhas de matrizes e modelos. Sendo assim, serão feitas as devidas observações sobre a especificação do modelo e sua fundamentação teórica de forma a embasar e contextualizar a análise empírica. A análise crítica pertinente à essa dissertação, porém, terá foco no uso da econometria espacial em seus diferentes modelos e no uso de diferentes matrizes com esses dados.

O próximo capítulo traz elementos que apresentam e fundamentam o modelo escolhido, seguido da apresentação das variáveis utilizadas. Em seguida os resultados econométricos iniciais são apresentados e discutidos. Por fim, o quinto capítulo dessa parte se concentra na discussão crítica da escolha dos modelos espaciais, das matrizes e no papel da convergência assintótica.

## **17. Fundamentos e Especificação do Modelo da Análise Empírica**

No contexto da avaliação das políticas públicas, há uma preocupação em saber quais escolhas promovem um melhor avanço em termos de algum ou alguns indicadores de performance. Para essa dissertação, há um interesse em saber quais decisões sobre políticas públicas, refletidas na composição do gasto, promovem um melhor crescimento de um indicador escolhido, a taxa de crescimento de emprego formal.

A especificação do modelo aqui utilizada é baseada em um modelo de crescimento de Solow-Swan convencional descrito a seguir. Outros fatores de bem-estar e igualdade são também relevantes. Os modelos de convergência são adequados para este tipo de análise, pois podem fornecer uma boa indicação de como testar a efetividade de políticas orientadas ao desenvolvimento econômico local.

A hipótese de convergência da renda per capita pode ser sintetizada como uma tendência de diminuição contínua, no tempo, das diferenças de renda entre as economias mais avançadas e as menos avançadas. A convergência é uma das principais previsões do modelo neoclássico de crescimento desenvolvido por Solow (1956) e Swan (1956), sendo uma implicação da hipótese de existência de retornos decrescentes para os fatores de produção. Posto que a produtividade do capital é maior em economias relativamente menos avançadas, a taxa de crescimento seria mais elevada nos países com menor estoque de capital, com que haveria uma progressiva redução nos diferenciais de crescimento entre países mais e menos avançados e uma convergência para mesmos níveis de taxa de crescimento e de renda.

Nos estudos empíricos dos condicionantes do crescimento da renda per capita (ou da convergência/divergência de renda) fatores institucionais podem ser considerados. Gallup et. al. (1995) tentam relacionar fatores políticos e geográficos ao desempenho econômico, selecionando regiões específicas de países em desenvolvimento. Em geral, regiões localizadas próximas ao litoral ou próximas a grandes rios navegáveis apresentaram ao longo do tempo taxas de crescimento maiores que as demais. Por outro lado, países com maior instabilidade política (maiores taxas de inflação, menor participação da população nos processos de decisão etc.) apresentaram taxas de crescimento menores.

Para o Brasil, em geral, os fatores que mais aparecem na literatura referem-se ao capital humano, infra-estrutura local e especialização da atividade econômica (Azzoni, 1998; Azzoni et. al., 2000; Andrade & Serra, 1998; Chagas & Toneto Jr., 2002).

Há vários trabalhos que tratam da questão da convergência no caso brasileiro. A maioria desses estudos aponta a existência de convergência entre os estados brasileiros, principalmente no tocante à renda per capita, mas não há um consenso com relação à velocidade dessa convergência. Ferreira & Diniz (1995), Schwartzman (1996) e Zini Jr. (1998) não rejeitam a hipótese de convergência absoluta entre as rendas per capita dos estados brasileiros para o período entre 1970 e 1985. Azzoni (1997, 2001), empregando uma série mais longa (1939-1996), igualmente encontrou indicações de convergência absoluta da renda, porém com uma velocidade muito menor.

Todos esses trabalhos citados no parágrafo anterior utilizam os estados como escala de análise. Outros trabalhos procuram encontrar evidências de convergência para uma escala menor. Azzoni et. al. (2000) mostram a existência de convergência condicional da renda entre regiões metropolitanas do Brasil, enquanto que Andrade & Serra (1998) sugerem a existência de convergência entre os municípios médios do Brasil no período de 1970 a 1991. Chagas & Toneto Jr. (2002) mostram a existência de convergência condicional para os municípios brasileiros no período de 1980 a 1991. Laurini, Andrade, & Pereira (2003), analisando a evolução da distribuição da renda relativa per capita para os municípios brasileiros no período de 1970 a 1996, detectam a formação de dois clubes de convergência, um de baixa renda formado pelos municípios das regiões Norte e Nordeste, e outro de alta renda formado pelos municípios das regiões Centro-Oeste, Sudeste e Sul.

O modelo de crescimento neo-clássico, desenvolvido paralelamente por Solow (1956) e Swan (1956), de maneira simplificada<sup>4</sup>, parte de uma dinâmica para a variação do estoque de capital no tempo dada por:

$$\dot{K} = I - \delta K = sY - \delta K \quad (17.1)$$

---

<sup>4</sup> Para maiores detalhes vide Barro e Sala-i-Martin (1995).

Onde  $K$  é o estoque de capital,  $I$  o total de investimento,  $\delta$  a taxa de depreciação do capital,  $s$  a taxa de poupança e  $Y$  o produto total da economia e consideramos a variação de estoques como investimento,  $I \equiv sY$ . O ponto acima da variável representa a derivada da mesma em relação ao tempo. Excepcionalmente a letra maiúscula não indica matriz. Ela é usada nesse trecho apenas para diferenciar a variável agregada da variável per capita.

O modelo assume uma taxa de crescimento constante da população dada por:

$$\frac{\frac{\partial L}{\partial t}}{L} = \frac{\dot{L}}{L} = n \quad (17.2)$$

Assumindo que a função de produção apresenta retornos constantes de escala, é possível escrever a função de produção como função apenas do capital per capita:  $y = f(k)$ , onde as variáveis minúsculas representam as variáveis brutas em relação ao total da força de trabalho, ou seja,  $y = \frac{Y}{L}$  e  $k = \frac{K}{L}$ .

A partir da definição de  $k$  e utilizando (18.2), é possível verificar que:

$$\dot{k} = \frac{\partial \left( \frac{K}{L} \right)}{\partial t} = \frac{\partial K}{\partial t} \frac{1}{L} + K(-1) \frac{1}{L^2} nL = \frac{\dot{K}}{L} - \frac{nK}{L} = \frac{\dot{K}}{L} - nk \quad (17.3)$$

Dividindo (18.1) por  $L$ , obtemos:

$$\frac{\dot{K}}{L} = \frac{sY}{L} - \frac{\delta K}{L} = sf(k) - \delta k \quad (17.4)$$

Substituindo (17.4) em (17.3), chega-se a:

$$\dot{k} = sf(k) - \delta k - nk = sf(k) - (n + \delta)k \quad (17.5)$$

O estado estacionário pode ser definido como o ponto no tempo em que a variação do estoque de capital seja constante. Em particular, no modelo de Solow-Swan o estado estacionário ocorre quando a variação do estoque de capital é zero, ou seja:

$$sf(k^*) = (n + \delta)k^* \quad (17.6)$$

Onde as variáveis com asterisco representam o seu valor em estado estacionário.

Dividindo os dois lados da equação (17.5) por  $k$  temos a variação percentual do capital:

$$\gamma_k \equiv \frac{\dot{k}}{k} = \frac{sf(k)}{k} - (n + \delta) \quad (17.7)$$

Onde  $\gamma_k$  representa a variação percentual do estoque de capital no tempo. Se substituirmos a propensão a poupar  $s$  implícita em (17.6) na equação (17.7) temos que:

$$\gamma_k = \frac{(n + \delta)k^*}{f(k^*)} \frac{f(k)}{k} - (n + \delta) = (n + \delta) \left( \frac{\frac{f(k)}{k}}{\frac{f(k^*)}{k^*}} - 1 \right) \quad (17.8)$$

A equação (17.8) implica dizer que  $\gamma_k$  será zero quando  $k = k^*$ , ou seja, no estado estacionário. Por simplificação, e de forma a garantir uma solução analítica, é possível assumir que a função de produção é Cobb-Douglas. Vale ressaltar que resultado vale para qualquer função de produção com retornos constantes de escala.

Nesse caso:

$$\begin{aligned} Y &= AK^\alpha L^{(1-\alpha)} \\ y &= AK^\alpha L^{(-\alpha)} = A(kL)^\alpha L^{(-\alpha)} \\ y &= Ak^\alpha \end{aligned} \quad (17.9)$$

Utilizando (17.9), a variação do estoque de capital (17.8) pode ser escrita como:

$$\gamma_k = (n + \delta) \left[ \left( \frac{k}{k^*} \right)^{(\alpha-1)} - 1 \right] \quad (17.10)$$

Adotando uma aproximação log-linear de (17.10) em torno do estado estacionário temos que:

$$\gamma_k \cong -\beta \ln\left(\frac{k}{k^*}\right) \quad (17.11)$$

Onde  $\beta = (1 - \alpha)(n + \delta)$ . É imediato para a função Cobb-Douglas que  $\ln\left(\frac{y}{y^*}\right) = \beta \ln\left(\frac{k}{k^*}\right)$ . Logo, é possível realizar a mesma aproximação para a variação percentual do produto no tempo,  $\gamma_y$ :

$$\gamma_y \cong -\beta \ln\left(\frac{y}{y^*}\right) \quad (17.12)$$

A equação (17.12) é uma equação diferencial com solução:

$$\ln(y_t) = (1 - e^{-\beta t}) \ln(y^*) + e^{-\beta t} \ln(y_{t-1}) \quad (17.13)$$

É possível derivar da equação (17.13) que  $\beta$  é uma boa medida da velocidade de convergência. Se  $\beta$  é muito alto o segundo termo do lado direito de (17.13) vai para zero e  $y_t = y^*$ . Na realidade se  $\beta < \infty$  a convergência ocorre apenas quanto  $t$  tende a infinito. Este é um resultado conhecido para qualquer decaimento exponencial. Dada a impossibilidade de se encontrar o tempo necessário para convergência absoluta, uma medida usualmente utilizada é o que se chama “meia vida”, ou seja, o tempo necessário para se alcançar metade do caminho necessário para convergência. Denominado o tempo de meia vida por  $\tau$ , temos que  $e^{-\beta\tau} = 0,5 \Rightarrow \tau = \frac{\ln(2)}{\beta}$ . Subtraindo  $\ln(y_{t-1})$  dos

dois lados de (17.13), é possível reescrevê-la como:

$$\ln(y_t) - \ln(y_{t-1}) = (1 - e^{-\beta t}) [\ln(y^*) - \ln(y_{t-1})] \quad (17.14)$$

A equação (17.14) dá uma boa indicação de uma especificação econométrica que permite inferir  $\beta$  e, portanto, o tempo necessário para convergência.

Uma aplicação imediata da equação (17.14) para a estimativa da taxa de convergência pode ser escrita como:

$$\ln(y_{i,t}) - \ln(y_{i,t-1}) = \alpha - \delta \ln(y_{i,t-1}) + \varepsilon_{i,t} \quad (17.15)$$

Onde  $\alpha = (1 - e^{-\beta t}) \ln(y^*)$ ,  $\delta = (1 - e^{-\beta t})$ , o subscrito  $i$  representa a região e  $\varepsilon_{i,t}$  é um erro aleatório i.i.d. com distribuição  $N(0, \sigma^2)$ . Assim, é possível estimar o parâmetro da velocidade de convergência implicitamente a partir de uma regressão utilizando a especificação (17.15) para um determinado grupo de regiões. Este é o conceito de convergência absoluta e variações desta especificação foram utilizadas por diversos autores desde os anos 80 como, por exemplo, Baumol (1986) entre muitos outros.

Ocorre, no entanto, que o resultado do modelo neo-clássico não implica obrigatoriamente em convergência absoluta. Se as regiões apresentarem preferências, tecnologia ou instituições distintas o produto de estado estacionário de cada economia pode não ser o mesmo. Isto equivale a considerar um modelo com efeitos fixos na especificação (17.15), ou seja, permitir que o intercepto seja diferente para cada região considerada.

Por outro lado, se educação, infra-estrutura pública etc influenciarem a produtividade do capital e/ou do trabalho entrando multiplicativamente na função de produção a especificação (17.15) pode ser modificada para refletir este fato. Em particular, para os interesses dessa dissertação, as políticas públicas locais podem alterar a taxa de convergência. Isso significa que podemos re-especificar (17.15) como:

$$\ln(y_{i,t}) - \ln(y_{i,t-1}) = \alpha - \delta \ln(y_{i,t-1}) + \mathbf{p}'_{i,t-1} \boldsymbol{\pi} + \mathbf{x}'_{i,t-1} \boldsymbol{\gamma} + \varepsilon_{i,t} \quad (17.16)$$

Na especificação (17.16) o vetor de parâmetros a ser estimado  $\pi$  está associado ao vetor de variáveis de políticas  $\mathbf{p}'_{i,t-1}$  (implementadas no período anterior) enquanto o vetor  $\gamma$  está associado ao vetor de variáveis de controle  $\mathbf{x}'_{i,t-1}$  também medidas no período anterior. A hipótese a ser testada é se as políticas públicas locais podem ou não influenciar a velocidade de convergência da localidade. Um problema usual dessa abordagem na grande maioria dos artigos citados anteriormente é trabalhar com variáveis de controle contemporâneas. Esse é um problema clássico de endogeneidade. Nessa especificação evita-se esse problema usual utilizando tanto as políticas como as variáveis de controle defasadas temporalmente.

O modelo de Solow-Swan fornece então uma fundamentação teórica bastante robusta para estudar o impacto de políticas locais sobre o crescimento da renda nas localidades indicando uma especificação coerente com tal micro-fundamento. A extensão dessa análise para a produtividade (salário médio) ou mesmo do emprego é trivial. Na realidade, se tivéssemos eliminado  $K$  e não  $L$  na passagem para (17.5) a equação diferencial (17.12) seria escrita em termos do emprego. De fato, a hipótese para justificar uma equação de convergência para a taxa de desemprego é mais leve do que a necessária para justificar a convergência de renda. Basta assumir retornos constantes, não é necessário impor uma Cobb-Douglass para a função de produção.

É importante ressaltar que apesar do modelo neo-classico de Solow ter essa ênfase no investimento de capital, e focar sua análise em convergência de renda, pesquisas mais recentes na área de crescimento têm colocado foco nas externalidades e nos aspectos locais para o desenvolvimento. Modelos da chamada nova economia geográfica, indicam caminhos para a dinâmica de crescimento urbano, como, por exemplo, trabalhos de Krugman (1991).

Assim, para essa dissertação a especificação (17.16) será utilizada com o emprego como  $y$ . O coeficiente da variável dependente defasada indica se está ocorrendo convergência ou não. O coeficiente das políticas públicas indica se a política aumenta ou diminui a taxa de convergência.



Em princípio, o estudo de convergência seja mais adequado para governos sub-nacionais visto que esses são muito mais abertos que os países e, portanto, tanto o capital como o fator trabalho são muito mais móveis que entre países. Isso implicaria em uma maior homogeneidade nos parâmetros tecnológicos e comportamentais e, assim, de desempenho como argumentam Glaeser et. al. (1995).

A literatura de economia espacial coloca justamente a o papel da aglomeração e clusters de atividades econômicas como causas fundamentais para o desempenho sócio-econômico, gerando externalidades que permitem condições de crescimento.

No entanto, há duas maneiras de um município aumentar seu emprego, a variável de interesse nesse estudo empírico. Ele pode implementar políticas que aumentem a sua produtividade ou de assistência bem estruturadas ou podem “roubar” ou “perder” empregos para outros municípios. Ou seja, a interação entre os vizinhos é fundamental para determinar a característica da política de desenvolvimento.

Para lidar com a interação entre os municípios iremos utilizar os métodos espaciais aqui já mencionados. Definida a matriz  $W$ , podemos alterar a nossa especificação inicial para incorporar a inter-relação entre os municípios:

$$\ln(\mathbf{y}_t) - \ln(\mathbf{y}_{t-1}) = \alpha - \delta \ln(\mathbf{y}_{t-1}) + \pi \mathbf{P}_{t-1} + \gamma \mathbf{X}_{t-1} + \rho \mathbf{W}[\ln(\mathbf{y}_t) - \ln(\mathbf{y}_{t-1})] + \mathbf{u} \quad (17.17)$$

$$\mathbf{u} = \lambda \mathbf{W}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

Esse é o modelo mais geral de especificação com auto-correlação espacial, SAC (5.1) Assumindo valores 0 para  $\rho$  ou  $\lambda$  é possível estimar variações dessa especificação. Com ambos 0, retorna-se à (17.16), ou seja, sem a nenhuma correlação espacial. Pode-se ainda obter os modelos SAR(5.3) e SEM (5.4) assumindo  $\rho = 0$  e  $\lambda = 0$ , respectivamente.

O modelo SAR nos indica possível autocorrelação espacial, positiva ou negativa, na variável de análise observada. Conforme indicado pelas simulações, quando SAR apontar para a presença do componente espacial, pouca informação é adicionada pelo modelo SEM. A estratégia seguinte, então, seria estimar pelo modelo completo. Se o modelo SAR não apontar a presença, pode ser feita a estimação por SEM. O modelo SEM nos indica correlação espacial em variáveis não observadas, como, por exemplo, determinantes estruturais para um crescimento. A presença de ambos os componentes indicaria uma ação concomitante de ambos os efeitos.

## 18. Variáveis Utilizadas

Uma dificuldade relevante na construção da base de dados é definir a unidade de análise. Para os fins desse estudo empírico, o ideal seria trabalhar na escala do município, a única escala efetivamente “local”. No entanto, entre 1991 e 2000 o número de municípios no Brasil aumentou de 4.491 para 5.507. Para criar um painel de distritos consistentes para esse período optou-se pela definição do IPEA de “Áreas Mínimas Comparáveis” (AMC) como a nossa unidade de análise. Dada a criação de 1.106 municípios no período o IPEA conseguiu criar 4.267 áreas comparáveis entre os dois períodos. Essa é a escala mais detalhada que é possível trabalhar no Brasil na década de 1990. De fato, como se pode observar na Tabela 1, 85% das AMC tinham apenas um município enquanto 10% tinham 2 municípios. Ou seja, a quantidade de AMC com 3 ou mais municípios é ínfima.

Número de Municípios	Porcentagem
1	84.71%
2	10.27%
3	2.30%
4	1.24%
5	0.56%
Mais de 5	0.91%

**Tabela 18. 1** - Número de Municípios em cada AMC

Uma alternativa para o problema da criação de municípios seria trabalhar apenas com os 85% que não sofreram alterações e constituem as AMC com 1 município. No entanto, com há o controle da correlação espacial, a matriz  $W$  e toda o uso das técnicas espaciais seriam prejudicados caso fosse feita a eliminação.

Para construir a base de dados para essas análises combinou-se uma série de fontes. Os dados das empresas e de emprego foram obtidos a partir da Relação Anual de Informações Sociais (RAIS), uma informação obrigatória para as empresas recolhida anualmente pelo Ministério do Trabalho e Emprego (MTE). O acesso aos micro-dados da RAIS, foi possível graças a um acordo do CEPESP/FGV-SP com o departamento de estatística do MTE. Dados na escala municipal ou, quando disponível, na escala da AMC foram obtidos no UNDP/IPEA, IBGE e IPEA (para as contas de finanças públicas obtidas por sua vez da Secretaria do Tesouro Nacional – STN). Quando os dados estavam disponíveis apenas na escala municipal, os dados foram re-agregados por AMC garantindo a sua comparação no tempo.

A variável dependente, como já indicado é a diferença entre os logaritmos naturais do emprego formal em 2000 e 1991. As variáveis independentes podem ser divididas em alguns grupos.

Inicialmente, para controlar o nível de onde partem as taxas de crescimento medida pela variável dependente, foi incluída o logaritmo natural do emprego formal em 1991. Essa variável, em especial, tem o papel de medir a convergência no emprego entre as diferentes AMC.

Um grupo de variáveis reflete as condições sócio-econômicas de 1991 das AMC, em logaritmo natural, exceto no caso de variáveis que representam porcentagem. Inclui-se, a população ( $\ln\_pop\_91$ ) e a população abaixo da linha de pobreza de R\$ 75,00 ( $\ln\_pop\_pobr\_91$ ). Essa linha de pobreza corresponde a meio salário mínimo de agosto de 2000. As variáveis renda per capita ( $\ln\_renda\_pc\_91$ ) e salário médio ( $\ln\_sal\_medio\_91$ ), controlando o poder de compra e nível de renda das AMC. Para controlar as diferença de capital humano, inclui-se a média de anos de estudo da população acima de 25 ( $\ln\_anos\_estudo\_91$ ).

Refletindo condições estruturais e de inclusão social, inclui-se, a porcentagem da população vivendo em habitações com banheiro e água corrente (*struc\_91*). Como *proxy* de violência, inclui-se o número de homicídios (*ln\_murder\_91*). Para controlar as diferenças de empreendedorismo bem como o ciclo de vida e a atração/repulsão das empresas, criou-se variáveis que decompõe a variação do número de empresas. Como controle das condições iniciais, foram incluídas na regressão as empresas criadas entre 91 e 96 (*ln\_emp\_Nascimentos\_91\_96*) e as empresas atraídas (*ln\_emp\_Entr\_91\_96*). Por fim, para controlar a composição do emprego em cada AMC, inclui-se a porcentagem do emprego total no setor industrial (*P\_Emprego\_Ind\_C\_91*) e no setor de serviços (*P\_Emprego\_Serv\_C\_91*).

Note que a variável dependente, como visa refletir desempenho de políticas, relaciona-se com o setor formal, já que este permitiria observar melhor a adequação de uma política de aumento de emprego. Já para a distribuição do emprego optou-se pelo emprego total, formal e informal, onde podemos ter uma idéia mais concreta de onde se distribui a população empregada.

Um outro bloco inclui variáveis de condições geográficas. Assim, o logaritmo natural da área (*ln\_area\_91*) foi incluído, assim como uma *dummy* para AMC da costa brasileira. Como controle do acesso ao mercado nacional e local, inclui-se a distância até São Paulo (*dist\_sp\_100*) e a distância até a capital do Estado (*dist\_cap\_100*).

Como um dos objetivos específicos da especificação é medir o impacto de decisões de política, dados sobre a composição do gasto foram incluídos. Os dados são uma média do período de 1994-1996. A escolha se deu por disponibilidade e confiabilidade dos dados. Como uma medida de políticas públicas com ênfase em capital humano, inclui-se a porcentagem do gasto em saúde (*P\_G\_Saude\_94\_96*) e em educação (*P\_G\_Educ\_94\_96*). Para medir investimento em capital físico, inclui-se a porcentagem dos gastos em energia, transporte, habitação e desenvolvimento regional (*P\_G\_Infra\_94\_96*). Por fim, como a receita de muitas AMC depende de transferências, inclui-se a participação dessas na receita orçamentária (*P\_Transf\_94\_96*).

Também foram incluídas *dummies* para cada um dos Estados brasileiros, exceto São Paulo. Nas tabelas de regressões os coeficientes das *dummies* são omitidos.

## 19. Resultados das Regressões

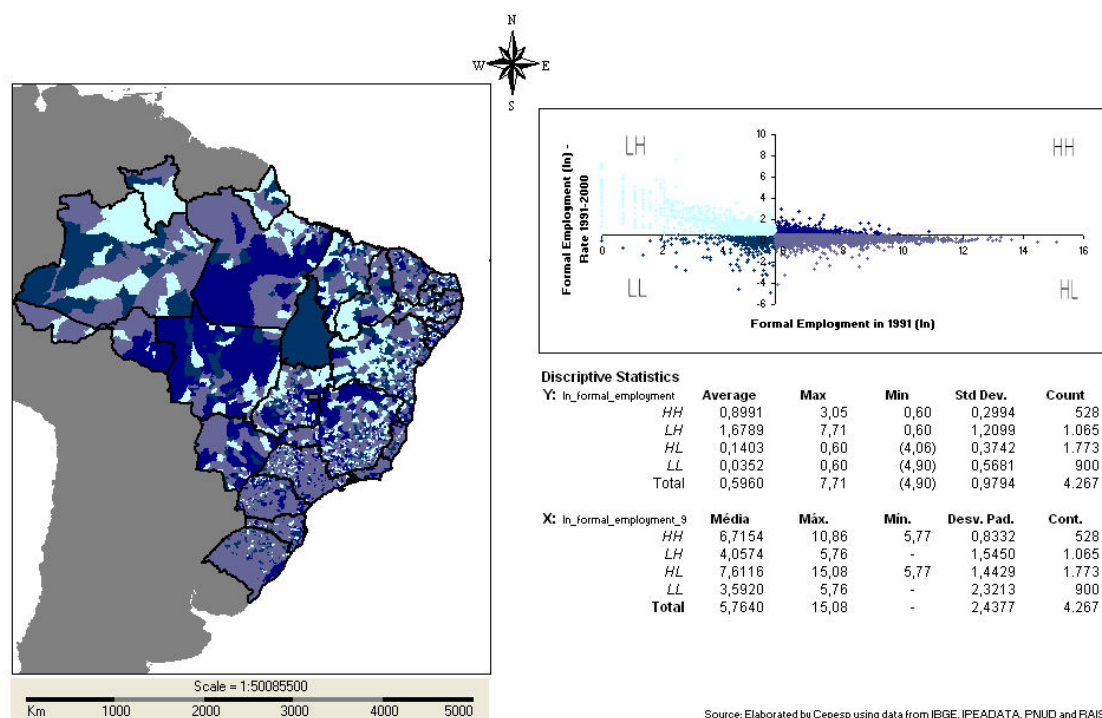
Como apresentado na seção de especificação, a variável dependente é a diferença entre os logaritmos naturais do emprego formal em 2000 e 1991. A matriz de ponderação  $W$  aqui utilizada é a matriz de distância euclidiana invertida (10.4).

Os coeficientes foram estimados utilizando, assim como na seção de simulação, OLS e SAR, SEM e SAC, estimados por máxima verossimilhança e GMM. Assim, temos 7 estimativas. Os resultados podem ser considerados robustos uma vez que se mantêm similar em valor e significância nos diferentes modelos. Além disso, o controle pela dependência espacial mostra-se significativa na maioria dos modelos. Por isso, a maior parte das observações estará concentrada nas estimativas do modelo SAC.

Como indicado na fundamentação do modelo, essa especificação corresponde a um teste de convergência para as AMC. O resultado do modelo de Solow-Swan aplicado para uma *cross-section* de regiões implica dizer que se o coeficiente da variável defasada for negativo, teremos convergência. O Mapa (19.1) apresenta a variação do emprego entre 1991 e 2000 dividindo os municípios em 4 grupos, similar a um diagrama de Moran: municípios que estavam acima da média de emprego em 1991 e que cresceram com uma taxa acima da média (HH); municípios que estavam abaixo da média de emprego em 1991 e que cresceram com uma taxa acima da média (LH); municípios que estavam acima da média de emprego em 1991 e que cresceram com uma taxa abaixo da média (HL); municípios que estavam abaixo da média de emprego em 1991 e que cresceram com uma taxa abaixo da média (LL)

A evidência de convergência absoluta (não condicional) em emprego é bastante forte. Os municípios estão de fato concentrados nos quadrantes noroeste e sudeste indicando que a maioria dos municípios inicialmente acima da média estão crescendo numa velocidade menor, ou até mesmo decrescendo, o nível de emprego enquanto os municípios abaixo da média estão ganhando postos de trabalho. Como já se esperava, os municípios com maior emprego inicial estão concentrados nas regiões sudeste e sul.

É interessante observar como os municípios da costa estão perdendo postos para os municípios do interior. O quanto essa convergência não condicional é desejável é uma questão um pouco mais complicada pois pode-se estar perdendo em termos de ganhos de aglomeração. Por outro lado, alguns custos de densidades extremamente elevadas também podem estar sendo evitados.



**Figura 19.1** - Mapa, Diagrama de Dispersão e Estatísticas Descritivas para Taxa de Variação no Emprego Formal (ln), 1991-2000

A análise de convergência não condicionada aponta essa tendência na variável de emprego. No entanto, sabemos que esse resultado é relativamente usual. Um aspecto interesse é saber qual o resultado obtido para convergência condicional, especialmente observando como as variáveis de política respondem ao grau de convergência. A Tabela (19.1) apresenta os resultados obtidos para as regressões utilizando os métodos explicados anteriormente. Como podemos observar, para todos os métodos, a hipótese de convergência condicional é confirmada já que o coeficiente da variável defasada é sempre negativo e significativo mesmo ao nível de 1%. Além disso, sua magnitude é uma das maiores.

	OLS	SAR - ML	SAR - GMM	SEM - ML	SEM - GMM	SAC - ML	SAC - GMM
Const	-1.0290 *** -2.809	-1.0445 *** -2.904	-1.0254 *** -2.797	-1.0766 *** -2.942	-1.0749 *** -2.908	-1.0237 *** -2.775	-0.6361 ** -2.399
ln_emprego_91	-0.4971 *** -57.61	-0.4953 *** -57.95	-0.4974 *** -57.30	-0.5009 *** -58.25	-0.5007 *** -58.16	-0.4995 *** -58.39	-0.4998 *** -57.89
ln_sal_medio_91	0.547 *** 50.71	0.546 *** 51.09	0.547 *** 50.60	0.554 *** 51.82	0.554 *** 51.63	0.553 *** 52.16	0.553 *** 51.29
ln_pop_pobr_91	-0.1276 ** -2.387	-0.1281 *** -3.505	-0.1276 ** -2.385	-0.1192 *** -2.913	-0.1195 ** -2.142	-0.1089 ** -2.475	-0.1285 ** -2.288
ln_pop_91	0.324 *** 5.976	0.322 *** 9.240	0.325 *** 5.980	0.314 *** 7.827	0.314 *** 5.548	0.304 *** 6.994	0.318 *** 5.566
ln_anos_estudo_91	-0.0010 -0.016	0.001 0.015	-0.0016 -0.024	-0.0015 -0.022	-0.0015 -0.022	-0.0094 -0.132	-0.0003 -0.004
ln_area_91	-0.0199 -1.410	-0.0216 -1.540	-0.0195 -1.378	-0.0152 -1.020	-0.0154 -1.036	-0.0089 -0.575	-0.0152 -1.012
ln_renda_pc_91	-0.1156 * -1.665	-0.1162 * -1.851	-0.1154 * -1.662	-0.1105 * -1.710	-0.1107 -1.578	-0.1057 -1.609	-0.1367 ** -1.969
ln_emp_Nascimentos_91_96	0.157 *** 8.720	0.158 *** 8.840	0.157 *** 8.699	0.158 *** 8.679	0.158 *** 8.682	0.156 *** 8.480	0.161 *** 8.800
ln_emp_Entr_91_96	0.042 *** 3.255	0.043 *** 3.327	0.042 *** 3.242	0.041 *** 3.122	0.041 *** 3.127	0.037 *** 2.823	0.040 *** 3.077
struc_91	-0.3283 *** -3.027	-0.3321 *** -3.087	-0.3278 *** -3.021	-0.3389 *** -3.017	-0.3384 *** -3.014	-0.3386 *** -2.941	-0.3323 *** -2.933
costeira	0.057 1.336	0.050 1.178	0.058 1.365	0.054 1.201	0.054 1.203	0.063 1.331	0.057 1.238
dist_cap_100	-0.0002 * -1.920	-0.0002 ** -2.149	-0.0002 * -1.865	-0.0002 * -1.757	-0.0002 * -1.750	-0.0001 -1.266	-0.0002 * -1.703

Tabela 19.1 - Resultados das Regressões - Parte I



	OLS	SAR - ML	SAR - GMM	SEM - ML	SEM - GMM	SAC - ML	SAC - GMM
dist_sp_100	-0.0002 *** -4.094	-0.0002 *** -5.914	-0.0003 *** -4.107	-0.0003 *** -5.420	-0.0003 *** -3.728	-0.0003 *** -5.312	-0.0003 *** -3.746
P_G_Educ_94_96	0.425 ** 2.420	0.420 ** 2.408	0.426 ** 2.425	0.393 ** 2.254	0.394 ** 2.260	0.370 ** 2.142	0.389 ** 2.221
P_G_Infra_94_96	0.171 1.315	0.163 1.267	0.172 1.325	0.135 1.048	0.136 1.059	0.121 0.946	0.141 1.085
P_G_Saude_94_96	0.084 0.602	0.081 0.587	0.084 0.605	0.117 0.848	0.115 0.837	0.135 0.992	0.109 0.786
P_Transf_94_96	-0.2301 ** -2.461	-0.2223 ** -2.396	-0.2317 ** -2.474	-0.2097 ** -2.234	-0.2104 ** -2.240	-0.1973 ** -2.102	-0.2208 ** -2.337
P_Emprego_Ind_C_91	0.401 *** 2.761	0.407 *** 2.951	0.400 *** 2.750	0.446 *** 3.053	0.444 *** 2.944	0.459 *** 3.044	0.436 *** 2.868
P_Emprego_Serv_C_91	0.164 0.837	0.168 0.863	0.164 0.834	0.179 0.893	0.179 0.890	0.201 0.988	0.188 0.927
ln_murder_91	-0.0128 -0.780	-0.0109 -0.675	-0.0132 -0.801	-0.0033 -0.204	-0.0037 -0.226	0.000 0.021	-0.0024 -0.144
rho		0.097 *** 6.863	-0.0193 -0.384			-0.2289 *** -9.127	-0.0923 * -1.743
lambda				0.278 *** 33.89	0.267 *** 10.69	0.476 *** 94.19	0.288 *** 9.282
Observações:	4267	4267	4267	4267	4267	4267	4267
R <sup>2</sup> ajustado	53.92%	53.71%	53.88%	54.66%	54.66%	55.45%	54.76%
log-likelihood		-2806		-2786		-335	
tempo (segundos)	0	680	39	666	3481	9427	1857

(t assintótico)

\*\*\* siginificante a 1%

\*\* siginificante a 5%

\* siginificante a 10%

**Tabela 19. 2 - Resultados das Regressões - Parte II**

Como sugerido pela literatura de economia espacial, dado o controle pela área, o coeficiente de população é significativo. Isso indica que a intensidade da aglomeração é relevante para processos de desenvolvimento econômico local.

Outro resultado forte é a distância a São Paulo, *proxy* de acesso ao mercado nacional, com coeficiente negativo e significativo. Esse resultado está de acordo com a teoria, que indica que proximidade com maiores mercados contribui para maiores atividades, e com isso já estariam em um nível mais alto de emprego inicial, aumentando a convergência. Um resultado similar, porém mais fraco é observado com relação à distância para a capital do Estado.

Para controlar o fluxo de empregos inclui-se o número de empresa criados entre 1991 e 1996 bem como as empresas que vieram de outras AMCs. Ambos são significativos e positivamente correlacionados com a taxa de crescimento de empregos, indicando que políticas de atração de empresas ou incentivo ao empreendedorismo devem impactar positivamente o nível de emprego e, conseqüentemente, o desempenho da AMC.

A variável de anos de estudo não se mostrou significativa para a taxa de crescimento de emprego, o que pode indicar uma eventual desconexão da oferta de trabalho com a demanda mais qualificada. No entanto, a porcentagem do gasto público em educação é estatisticamente significativa a 5% e com uma magnitude grande, em relação aos demais. Isso indica que investimento em educação, de fato, parece estar positivamente correlacionado com o crescimento do emprego.

As transferências como porcentagem da receita está negativamente correlacionada com a taxa de crescimento de emprego. Uma possível interpretação é que AMC que dependem mais de transferências não estão crescendo tanto. Este resultado, no entanto, pode ser enganoso. Ele pode estar apenas identificando AMC que são mais pobres, e que por isso não conseguem crescer tanto.

Analisando em particular os coeficientes de correlação espacial, sempre há significância estatística, sendo  $\rho$ , o coeficiente da variável dependente, negativo e  $\lambda$ , o coeficiente do termo de erro, positivo. O coeficiente negativo de  $\rho$  indica algum grau de competição entre as AMCs. O coeficiente positivo de  $\lambda$  indica ganhos de aglomeração em fatores não observados, possivelmente fatores estruturais.

Assim, a dinâmica desenhada para esses coeficientes espaciais é uma possível economia de aglomeração em aspectos estruturais, porém uma competição entre as AMCs pelo emprego gerado. De qualquer forma, a componente espacial se mostra relevante.

## **20. Análise Crítica da Mudança de Matrizes e Modelos**

As tabelas seguir reproduzem os resultados das regressões do capítulo anterior, sendo colocadas lado a lado com a mesma especificação calculada com a matriz de contigüidade.

	OLS	SAR - ML		SAR - GMM		SEM - ML		SEM - GMM		SAC - ML		SAC - GMM	
		Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
Const	-1.0290 *** -2.809	-1.0445 *** -2.904	-1.0280 *** -2.853	-1.0254 *** -2.797	-1.0713 *** -2.914	-1.0766 *** -2.942	-0.9972 *** -2.735	-1.0749 *** -2.908	-1.0001 *** -2.718	-1.0237 *** -2.775	-0.9849 *** -2.673	-0.6361 ** -2.399	-0.7909 *** -2.602
ln_emprego_91	-0.4971 *** -57.61	-0.4953 *** -57.95	-0.4970 *** -58.05	-0.4974 *** -57.30	-0.5006 *** -57.45	-0.5009 *** -58.25	-0.4995 *** -58.12	-0.5007 *** -58.16	-0.4993 *** -57.96	-0.4995 *** -58.39	-0.5012 *** -58.31	-0.4998 *** -57.89	-0.501 *** -57.91
ln_sal_medio_91	0.547 *** 50.71	0.546 *** 51.09	0.547 *** 51.14	0.547 *** 50.60	0.549 *** 50.68	0.554 *** 51.82	0.551 *** 51.37	0.554 *** 51.63	0.550 *** 51.20	0.553 *** 52.16	0.552 *** 51.67	0.553 *** 51.29	0.552 *** 51.12
ln_pop_pobr_91	-0.1276 ** -2.387	-0.1281 *** -3.505	-0.1277 *** -3.503	-0.1276 ** -2.385	-0.1257 ** -2.344	-0.1192 *** -2.913	-0.1238 *** -3.612	-0.1195 ** -2.142	-0.1242 ** -2.274	-0.1089 ** -2.475	-0.1146 ** -2.536	-0.1285 ** -2.288	-0.1129 ** -1.977
ln_pop_91	0.324 *** 5.976	0.322 *** 9.240	0.324 *** 9.368	0.325 *** 5.980	0.329 *** 6.051	0.314 *** 7.827	0.318 *** 9.937	0.314 *** 5.548	0.318 *** 5.744	0.304 *** 6.994	0.310 *** 6.951	0.318 *** 5.566	0.309 *** 5.369
ln_anos_estudo_91	-0.0010 -0.016	0.001 0.015	-0.0009 -0.014	-0.0016 -0.024	-0.0070 -0.106	-0.0015 -0.022	0.000 0.006	-0.0015 -0.022	0.000 0.004	-0.0094 -0.132	-0.0042 -0.059	-0.0003 -0.004	-0.0050 -0.072
ln_area_91	-0.0199 -1.410	-0.0216 -1.540	-0.0201 -1.428	-0.0195 -1.378	-0.0149 -1.048	-0.0152 -1.020	-0.0184 -1.255	-0.0154 -1.036	-0.0186 -1.269	-0.0089 -0.575	-0.0109 -0.710	-0.0152 -1.012	-0.0111 -0.722
ln_renda_pc_91	-0.1156 * -1.665	-0.1162 * -1.851	-0.1157 * -1.839	-0.1154 * -1.662	-0.1097 -1.576	-0.1105 * -1.710	-0.1186 * -1.876	-0.1107 -1.578	-0.1183 * -1.695	-0.1057 -1.609	-0.1143 * -1.721	-0.1367 ** -1.969	-0.1158 -1.626
ln_emp_Nascimentos_91_96	0.157 *** 8.720	0.158 *** 8.840	0.157 *** 8.771	0.157 *** 8.699	0.155 *** 8.586	0.158 *** 8.679	0.158 *** 8.708	0.158 *** 8.682	0.158 *** 8.714	0.156 *** 8.480	0.156 *** 8.523	0.161 *** 8.800	0.156 *** 8.485
ln_emp_Entr_91_96	0.042 *** 3.255	0.043 *** 3.327	0.042 *** 3.277	0.042 *** 3.242	0.040 *** 3.084	0.041 *** 3.122	0.042 *** 3.182	0.041 *** 3.127	0.042 *** 3.190	0.037 *** 2.823	0.038 *** 2.910	0.040 *** 3.077	0.039 *** 2.925
struc_91	-0.3283 *** -3.027	-0.3321 *** -3.087	-0.3285 *** -3.050	-0.3278 *** -3.021	-0.3201 *** -2.942	-0.3389 *** -3.017	-0.3404 *** -3.077	-0.3384 *** -3.014	-0.3391 *** -3.067	-0.3386 *** -2.941	-0.3485 *** -3.063	-0.3323 *** -2.933	-0.3441 *** -3.019
costeira	0.057 1.336	0.050 1.178	0.056 1.336	0.058 1.365	0.075 * 1.733	0.054 1.201	0.052 1.184	0.054 1.203	0.053 1.196	0.063 1.331	0.062 1.347	0.057 1.238	0.064 1.373
dist_cap_100	-0.0002 * -1.920	-0.0002 ** -2.149	-0.0002 * -1.958	-0.0002 * -1.865	-0.0001 -1.531	-0.0002 * -1.757	-0.0002 * -1.794	-0.0002 * -1.750	-0.0002 * -1.791	-0.0001 -1.266	-0.0001 -1.219	-0.0002 * -1.703	-0.0001 -1.211

Tabela 20. 1 - Comparação dos Resultados das Regressões com Diferentes Matrizes - Parte I

	OLS	SAR - ML		SAR - GMM		SEM - ML		SEM - GMM		SAC - ML		SAC - GMM	
		Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig	Dist	Contig
dist_sp_100	-0.0002 *** -4.094	-0.0002 *** -5.914	-0.0002 *** -5.999	-0.0003 *** -4.107	-0.0003 *** -4.437	-0.0003 *** -5.420	-0.0002 *** -6.904	-0.0003 *** -3.728	-0.0002 *** -3.858	-0.0003 *** -5.312	-0.0003 *** -5.146	-0.0003 *** -3.746	-0.0003 *** -3.893
P_G_Educ_94_96	0.425 ** 2.420	0.420 ** 2.408	0.425 ** 2.432	0.426 ** 2.425	0.442 ** 2.508	0.393 ** 2.254	0.405 ** 2.316	0.394 ** 2.260	0.407 ** 2.327	0.370 ** 2.142	0.394 ** 2.266	0.389 ** 2.221	0.400 ** 2.285
P_G_Infra_94_96	0.171 1.315	0.163 1.267	0.171 1.321	0.172 1.325	0.189 1.445	0.135 1.048	0.151 1.168	0.136 1.059	0.153 1.180	0.121 0.946	0.143 1.116	0.141 1.085	0.149 1.151
P_G_Saude_94_96	0.084 0.602	0.081 0.587	0.083 0.601	0.084 0.605	0.107 0.765	0.117 0.848	0.072 0.519	0.115 0.837	0.073 0.526	0.135 0.992	0.077 0.558	0.109 0.786	0.080 0.577
P_Transf_94_96	-0.2301 ** -2.461	-0.2223 ** -2.396	-0.2297 ** -2.472	-0.2317 ** -2.474	-0.2478 *** -2.638	-0.2097 ** -2.234	-0.2226 ** -2.382	-0.2104 ** -2.240	-0.2233 ** -2.389	-0.1973 ** -2.102	-0.2292 ** -2.451	-0.2208 ** -2.337	-0.2320 ** -2.460
P_Emprego_Ind_C_91	0.401 *** 2.761	0.407 *** 2.951	0.402 *** 2.910	0.400 *** 2.750	0.380 *** 2.602	0.446 *** 3.053	0.432 *** 3.069	0.444 *** 2.944	0.429 *** 2.898	0.459 *** 3.044	0.446 *** 2.997	0.436 *** 2.868	0.443 *** 2.908
P_Emprego_Serv_C_91	0.164 0.837	0.168 0.863	0.164 0.842	0.164 0.834	0.167 0.848	0.179 0.893	0.174 0.875	0.179 0.890	0.173 0.870	0.201 0.988	0.190 0.941	0.188 0.927	0.183 0.901
ln_murder_91	-0.0128 -0.780	-0.0109 -0.675	-0.0127 -0.783	-0.0132 -0.801	-0.0161 -0.977	-0.0033 -0.204	-0.0061 -0.373	-0.0037 -0.226	-0.0067 -0.409	0.000 0.021	-0.0005 -0.029	-0.0024 -0.144	-0.0015 -0.092
rho		0.097 *** 6.863	0.002 0.616	-0.0193 -0.384	-0.1190 *** -3.445					-0.2289 *** -9.127	-0.128 *** -13.76	-0.0923 * -1.743	-0.1346 *** -3.691
lambda						0.278 *** 33.89	0.095 *** 7.677	0.267 *** 10.69	0.086 *** 4.042	0.476 *** 94.19	0.207 *** 17.55	0.288 *** 9.282	0.188 *** 10.89
Observações:	4267	4267	4267	4267	4267	4267	4267	4267	4267	4267	4267	4267	4267
R <sup>2</sup> ajustado	53.92%	53.71%	53.91%	53.88%	53.66%	54.66%	54.17%	54.66%	54.15%	55.45%	54.80%	54.76%	54.73%
log-likelihood		-2806	-2810			-2786	-2802			-335	-351		
tempo (segundos)	0	680	7	39	1	666	45	3481	2	9427	93	1857	0

(t assintótico)

\*\*\* siginificante a 1%

\*\* siginificante a 5%

\* siginificante a 10%

**Tabela 20. 2** - Comparação dos Resultados das Regressões com Diferentes Matrizes - Parte II

Foram realizados também os testes de especificação apontados no capítulo 8. As tabelas a seguir apresentam os resultados utilizando a matriz de distâncias e a matriz de contigüidade:

<i>Wdist</i>	Estatística	Estatística Padronizada	P-valor
Moran	0.04	8.38	0.00
LM Erro		53.94	0.00
LM Erro Robusto		59.13	0.00
LM Var. Dep.		9.32	0.00
LM Var. Dep. Robusto		14.51	0.00

**Tabela 20. 3** – Testes de Especificação Espacial, *Wdist*

<i>Wcontg</i>	Estatística	Estatística Padronizada	P-valor
Moran	0.04	4.84	0.00
LM Erro		16.12	0.00
LM Erro Robusto		40.00	0.00
LM Var. Dep.		0.02	0.89
LM Var. Dep. Robusto		23.90	0.00

**Tabela 20. 4** – Testes de Especificação Espacial, *Wcontg*

Pelas tabelas (20.3) e (20.4) nota-se que de fato a matriz de contigüidade capta apenas parcialmente o efeito captado pela matriz de distância, e como observado nas simulações, não consegue captar, na versão não robusta do teste LM, a presença da componente espacial na variável dependente.

Observando as tabelas (20.1) e (20.2), é possível notar os mesmo efeitos indicados na simulação. Os coeficientes das variáveis explicativas não mudam muito entre os diferentes modelos, nem em termos de valores nem em termos de significância.

Realizando o mesmo teste especificado na equação (15.5), aqui reproduzida, seria possível obter uma indicação na mudança dos coeficientes entre OLS e um estimador ML, representando o benefício potencial da estimação dos modelos espaciais.

$$W\mathbf{y} = X\boldsymbol{\beta}_{W\mathbf{y}} + \boldsymbol{\xi} \quad (15.5)$$

A tabela a seguir apresenta os resultados para as duas matrizes:

	<b>Wdist</b>	<b>Wcontig</b>
	Coefficient	Coefficient
Const	0.157	-0.3545
ln_emprego_91	-0.0176 ***	-0.0292 ***
ln_sal_medio_91	0.013 ***	0.020 ***
ln_pop_pobr_91	0.004	0.016
ln_pop_91	0.026	0.043
ln_anos_estudo_91	-0.0214	-0.0497
ln_area_91	0.017 ***	0.042 ***
ln_renda_pc_91	0.006	0.048
ln_emp_Nascimentos_91_96	-0.0109	-0.0158
ln_emp_Entr_91_96	-0.0066	-0.0174 **
struc_91	0.038	0.069
costeira	0.073 ***	0.149 ***
dist_cap_100	0.000 ***	0.000 ***
dist_sp_100	-0.0000 **	-0.0002 ***
P_G_Educ_94_96	0.052	0.142
P_G_Infra_94_96	0.077	0.147 *
P_G_Saude_94_96	0.027	0.194 **
P_Transf_94_96	-0.0792 **	-0.1483 ***
P_Emprego_Ind_C_91	-0.0608	-0.1819 **
P_Emprego_Serv_C_91	-0.0386	0.022
ln_murder_91	-0.0189 ***	-0.0278 ***
R <sup>2</sup> ajustado	18.00%	20.72%
*** significante a 1% ** significante a 5% * significante a 10%		

**Tabela 20. 5** – Ajuste de X em W\*y

É possível notar um R<sup>2</sup> baixo entre as matrizes, e os coeficientes, mesmo os significantes são baixos. De forma não esperada, o ajuste parece ser melhor aos dados defasados pela matriz de contigüidade do que aos dados defasados espacialmente pela matriz de distâncias. Vale ressaltar que esses parâmetros ainda devem ser multiplicados pela estimativa de  $\rho$ .

Observando ainda as tabelas (20.1) e (20.2) pode-se notar que os coeficientes dos parâmetros espaciais, de um modo geral, são reduzidos na estimação pela matriz de contigüidade, conforme esperado. Algumas mudanças, no entanto, não são tão esperadas: pela estimação SAR-GMM, ao se utilizar a matriz de contigüidade,  $\rho$  aumenta em valor e em significância; o mesmo ocorre na estimação SAC-GMM.



Do mesmo modo que no capítulo 15, é feito também o teste especificado na equação (15.16) aqui reproduzida, além de testada a correlação entre as variáveis defasadas calculadas a partir das duas matrizes..

$$W_{contig} * y = \text{constante} + W_{dist} * y + \xi \quad (15.16)$$

A tabela a seguir apresenta os resultados:

<b>Y = <math>W_{contig} * y</math></b>	<b>n=27</b>
<b>Constante</b>	-0.022635 -2.226930
$W_{dist} * y$	1.096431 65.765527
<b>R2 ajustado médio</b>	0.5034
<b>Correlação entre <math>W_{dist} * y</math> e <math>W_{contig} * y</math>.</b>	0.7096

**Tabela 20. 6** - Ajuste entre  $W_{dist} * y$  e  $W_{contig} * y$

Note que para o caso empírico a substituição de uma matriz pela outra parece ser uma aproximação válida de acordo com os resultados da tabela (20.6), indicando que de fato caminham perfeitamente no mesmo sentido. O R2 de 50% no entanto alerta que efeito a ser captado deve ser em torno dessa ordem de grandeza, o que é de fato observado, em especial no modelo SAC-ML.

Em relação à escolha do melhor modelo, uma análise mais cuidadosa dos resíduos, em especial em termos de normalidade deveria ser feita para indicar qual seria mais adequado. Porém os sinais dos parâmetros espaciais são constantes em quase todas as estimações, sendo *rho* negativo e *lambda* positivo, o que dá robustez a esse resultado.

Note, porém, que de SAR-ML para SAC-ML é o único caso de mudança de sinal de *rho*, o que pode indicar algum problema com os pressupostos do modelo, já que a mudança de sinal de SAR para SAC não seria esperada. Com isso, o modelo SAC-GMM parece ser mais confiável.

## PARTE VI – TESTANDO DISTÂNCIAS ECONÔMICAS E UMA ESTRUTURA ALTERNATIVA DA DINÂMICA ESPACIAL

### 21. Influências Opostas

*Todo mundo influencia; alguns influenciam positivamente outros negativamente.*

Quando se passa do campo geográfico-físico para distâncias sócio-econômicas, não se está medindo apenas influências a partir de similaridades ou dissimilaridades. Neste campo, muitas vezes, não importam apenas as diferenças em valor absoluto, mas também o “sinal” dessa diferença. Para determinados índices socioeconômicos, de forma comparativa, uma determinada unidade está melhor que outra na questão em observação de acordo com o valor delas para esse índice. Dessa forma, se as similaridades ou dissimilaridades vão ser medidas por índices como esses, em que o “sinal” da diferença é importante, é aceitável se supor que as influências também possam se dar em sentidos opostos. Pode-se acreditar, por exemplo, que quem está melhor que uma unidade em observação irá prejudicar essa unidade de forma relativa; da mesma maneira quem estiver pior, irá ajudá-la de forma relativa.

No entanto, os modelos de econometria espacial, por apresentarem um conjunto de regressor e matriz de correlação espacial para uma determinada variável, impõem que todas as influências sejam em um único sentido, deixando-se a mudança nos sinais apenas para os valores nas medidas, dentro da matriz. Porém, como há a padronização dos valores, todas as medidas acabam se tornando positivas, sendo a única possibilidade de variação de sinal encontrada no regressor único. Visto isso, e as situações apresentadas no parágrafo anterior, viu-se a necessidade de uma modelagem alternativa que permitisse decompor as influências.

Como já colocado nas sessões anteriores, as influências espaciais são resumidas em uma matriz de pesos espaciais  $W$ . No modelo para a regressão, seja  $W$  ponderando a variável dependente, independente ou o termo de erro, o que se tem implícito é que a influência se dá sempre no mesmo sentido, positivo ou negativo, de acordo com o sinal de  $\rho$ . A inovação metodológica aqui proposta, de forma exploratória, é justamente possibilitar que a influência espacial possa ser decomposta em dois grandes grupos: uma influência supostamente negativa e outra influência supostamente positiva.

Para isso, antes de se calcular os pesos, ordenamos as unidades em ordem crescente do valor com o qual serão medidas as distâncias. Tal ordenação é necessária para garantir que seja possível separar a matriz  $W$  inicial em duas matrizes triangulares, como será explicado adiante.

A partir disso, calcula-se uma matriz de distâncias usuais. Após a normalização por linhas, a matriz possui sua diagonal central com zeros. Com isso, dada a ordenação anterior, todos os elementos da diagonal superior influenciam para um sentido e os da diagonal inferior para o outro. O passo seguinte é separar a matriz  $W$  em duas matrizes triangulares  $W^l$  e  $W^h$  de tal forma que a soma de  $W^l$  e  $W^h$  seja  $W$  original.

Dessa maneira, a normalização continua igual. Isso significa dizer que os pesos continuam comparáveis e que a soma, em absoluto, dos pesos para cada unidade é igual a 1. Uma interpretação para os pesos é que eles indicam qual a porcentagem do total de influências que a unidade  $j$  tem sobre a unidade  $i$ . A separação da matriz não muda a importância relativa, apenas permite que estas estejam em sinais contrários.

Formalmente, do modelo original (5.1):

$$\mathbf{y} = \rho W_1 \mathbf{y} + X\boldsymbol{\beta} + \mathbf{u} \quad (5.1)$$

$$\mathbf{u} = \lambda W_2 \mathbf{u} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

Passamos para:

$$\mathbf{y} = (\rho_1 W_1^l + \rho_2 W_1^h) \mathbf{y} + X\boldsymbol{\beta} + \mathbf{u} \quad (21.1)$$

$$\mathbf{u} = (\lambda_1 W_2^l + \lambda_2 W_2^h) \mathbf{u} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$$

Onde:

$\rho_k, \lambda_k$  são os coeficientes respectivamente de  $W_1^j$  e  $W_2^j$ ,  $j = l, h$

$W_i^l$  é uma matriz triangular inferior

$W_i^h$  é uma matriz triangular superior

$$W_i^l + W_i^h = W_i$$

Veja que no caso limite em que  $\rho_1 = \rho_2$  temos que  $(\rho_1 W_1^l + \rho_2 W_1^h) = \rho W_1$ .

A dificuldade para tal estimação é que não existem ainda técnicas desenvolvidas que permitam estimar essa especificação, estimando os coeficientes separados. Se forem assumidas algumas simplificações, no entanto, é possível ao menos testar a especificação.

Se considerarmos  $\rho_1 = -\rho_2$  temos que  $(\rho_1 W_1^l - \rho_1 W_1^h) = \rho(W_1^l - W_1^h)$ . Com isso, bastaria ajustar os sinais dos pesos para poder obter uma matriz  $W$  que reflita essa estrutura. É possível partir da especificação de pesos pelo critério de dissimilaridade (12.1) aqui reproduzido:

$$w_{ij}^* = \begin{cases} |d_i - d_j|, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases} \quad (12.1)$$

Onde  $d_i$  e  $d_j$  são valores da variável de interesse para  $i$  e  $j$ .

A simples alteração seria retirar o módulo da medida:

$$w_{ij}^{**} = \begin{cases} d_i - d_j, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases} \quad (21.2)$$

Para obter-se a matriz  $W$ , utiliza-se normalização baseada na (10.2) combinando o módulo no denominador, porém não no numerador:

$$w_{ij} = \frac{w_{ij}^{**}}{\sum_j w_{ij}^{**}} \quad (21.3)$$

Com isso se chega à estrutura desejada.

## 22. Influências Opostas na Simulação

De forma a testar o impacto e adequabilidade da estrutura espacial proposta, foram realizadas algumas simulações.

Para um tamanho de amostra de 1000 observações foram gerados dados SAR utilizando a definição da matriz que combina (12.1) e (21.2) resultando em (21.3). Em 50 iterações foram estimados os 7 modelos já apresentados a usando a própria matriz (chamada de  $W1$  nas tabelas) e uma matriz de dissimilaridade usual ( $W2$ ). A matriz  $W2$  foi feita utilizando (12.1) e a normalização usual (10.2). Com isso pode-se testar o impacto dessa nova matriz.

A tabela a seguir resume os resultados.

**Modelo: SAR**  
**n=1000**  
**repetições: 50**

**Em cada célula:**

Coeficientes

Desvio

Rejeição  $H_0=0$  ( $t \geq 1.96$ )

REAL		OLS	SAR ML		SAR GMM		SEM ML		SEM GMM		SAC ML		SAC GMM	
			W1	W2	W1	W2	W1	W2	W1	W2	W1	W2	W1	W2
20	Const	2.0225	19.9670	120.1500	19.4810	245.6200	2.2255	2.4381	2.2077	2.8690	19.9660	271.4100	29.4230	422.9800
		19.8770	0.3262	11.0450	9.7502	5.0497	19.8860	6.3903	19.8840	3.6640	0.3256	80.2620	156.6100	13.2610
		0.1200	1.0000	1.0000	0.8200	1.0000	0.1200	0.2200	0.1200	0.0800	1.0000	1.0000	0.8200	1.0000
1.5	b1	1.5156	1.5004	1.5011	1.5244	1.4869	1.5156	1.4997	1.5156	1.4897	1.5004	1.4943	1.4011	1.4877
		0.2319	0.0043	0.1226	0.0742	0.0502	0.2319	0.0707	0.2319	0.0414	0.0043	0.0267	1.8596	0.0443
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9400	1.0000
0.8	b2	0.7805	0.8005	0.7924	0.8074	0.8063	0.7801	0.7990	0.7802	0.8043	0.8005	0.8097	0.6675	0.8050
		0.2716	0.0050	0.1501	0.1131	0.0681	0.2717	0.0869	0.2717	0.0595	0.0050	0.0250	1.8735	0.0613
		0.8400	1.0000	1.0000	0.9600	1.0000	0.8400	1.0000	0.8400	1.0000	1.0000	1.0000	0.9000	1.0000
0.5	rho		0.5001	-1.0000	0.5297	-2.0632					0.5001	-2.2771	0.4379	-2.0259
			0.0011	0.0000	0.2705	0.0151					0.0012	0.6771	0.8431	0.0191
			1.0000	1.0000	0.9800	1.0000					1.0000	1.0000	0.9800	1.0000
0	lambda						0.0204	-0.9900	0.0186	-2.0831	-0.0216	-7.2287	0.0194	-0.7496
							0.0004	0.0000	0.0014	0.0039	0.1328	0.7615	0.0122	0.0415
							0.0000	1.0000	1.0000	1.0000	0.0200	1.0000	0.0000	0.0000
0.7	Desvio Médio do Erro	40.9180	0.6933	22.3530	21.9930	8.6491	40.8490	22.5440	40.8500	9.1138	0.6932	5.4050	348.5900	8.1603
	R2 ajustado médio	0.0413	0.9997	0.0422	0.7197	0.9570	0.0417	0.7081	0.0417	0.9523	0.9997	0.9832	-69.8000	0.9618
	likelihood médio		-690.96	-4,179.90			-4,782	-4,188			-118.37	-2,188		
	Tempo Médio		15.6	15.7	0.2	0.2	13.9	13.9	46.1	45.5	70.2	79.9	3.2	1.1
	Tempo Total		780.5	782.9	12.1	12.1	695.6	695.6	2,307.0	2,273.4	3,511.9	3,994.5	161.2	57.1

**Tabela 22. 1** - Resultados das Simulações com Influências Opostas, SAR, n = 1000

Como esperado a Matriz usada na geração de dados ( $W1$ ) estima melhor os dados do que uma matriz mal especificada ( $W2$ ). A matriz  $W2$ , inclusive, não consegue captar em nenhum modelo o coeficiente de correlação espacial, sugerindo mais uma vez a sensibilidade desses modelos à matriz adequada. Isso porém é um aspecto positivo, pois seria indesejado que qualquer matriz pudesse captar um coeficiente de correlação espacial.

Ainda aqui os demais coeficientes não sofrem muitas alterações.

Um aspecto, no entanto, chama a atenção aqui. Os modelos ML são um pouco superiores aos modelos GMM, em especial os modelos SAC.

Vale ressaltar aqui que não se sabe ainda, porém, a aplicabilidade empírica dessa estrutura alternativa. De forma a estimar, também de forma preliminar, essa possibilidade, no capítulo seguinte a mesma especificação de matriz de ponderação aqui apresentada será aplicada aos dados empíricos do usados na Parte V dessa dissertação.

### 23. Análise Empírica com Diferentes Medidas Geográficas e Sócio-Econômicas

Para complementar essa última parte exploratória, tentou-se aplicar diferentes medidas sócio-econômicas aos dados empíricos. Além das matrizes de distância e contigüidade já utilizadas, foram criadas mais 3 matrizes com medidas sócio-econômicas, todas utilizando o salário médio dos municípios em 1991.

A primeira é uma medida de similaridade, conforme (11.1), dando maior peso para AMCs com salários médios parecidos. A segunda é uma medida de dissimilaridade, conforme (12.1), dando maior peso para AMCs com salários mais diferentes. A terceira, por fim, segue (21.3), criando uma matriz com influências opostas.

Neste ultimo caso, terão pesos positivos as AMCs com salários menores que a AMC em análise. Para as AMCs com salários maiores, o peso será negativo. Dessa maneira, a estimativa de  $\rho$  observada na tabela a seguir para o último modelo deve ser compreendido como o coeficiente para as AMCs com salário menor. Para as AMCs com salário maior, o coeficiente deve ser lido com o sinal invertido.

A tabela a seguir resume os resultados dos modelos estimados. Todas as regressões foram feitas utilizando-se um modelo SAC-GMM, apenas variando a matriz.



	OLS	Dist	Contig	Inv.Sal	Sal	Influências Opostas
Const	-1.0290 ***	-0.6361 **	-0.7909 ***	-0.9777 ***	-3.5157 ***	-2.9610 ***
	-2.809	-2.399	-2.602	-2.873	-10.76	-8.096
ln_emprego_91	-0.4971 ***	-0.4998 ***	-0.501 ***	-0.4986 ***	-0.5183 ***	-0.5262 ***
	-57.61	-57.89	-57.91	-57.64	-58.63	-60.06
ln_sal_medio_91	0.547 ***	0.553 ***	0.552 ***	0.564 ***	0.670 ***	0.648 ***
	50.71	51.29	51.12	48.43	61.13	49.89
ln_pop_pobr_91	-0.1276 **	-0.1285 **	-0.1129 **	-0.1220 **	0.002	-0.1183 **
	-2.387	-2.288	-1.977	-2.274	0.042	-2.316
ln_pop_91	0.324 ***	0.318 ***	0.309 ***	0.319 ***	0.181 ***	0.391 ***
	5.976	5.566	5.369	5.845	3.263	7.373
ln_anos_estudo_91	-0.0010	-0.0003	-0.0050	-0.0011	0.009	0.013
	-0.016	-0.004	-0.072	-0.017	0.136	0.215
ln_area_91	-0.0199	-0.0152	-0.0111	-0.0201	-0.0076	-0.0142
	-1.410	-1.012	-0.722	-1.422	-0.529	-1.026
ln_renda_pc_91	-0.1156 *	-0.1367 **	-0.1158	-0.1145 *	-0.0880	0.055
	-1.665	-1.969	-1.626	-1.652	-1.236	0.843
ln_emp_Nascimentos_91_96	0.157 ***	0.161 ***	0.156 ***	0.160 ***	0.200 ***	0.130 ***
	8.720	8.800	8.485	8.831	10.85	7.403
ln_emp_Entr_91_96	0.042 ***	0.040 ***	0.039 ***	0.042 ***	0.027 **	0.047 ***
	3.255	3.077	2.925	3.203	2.065	3.707
struc_91	-0.3283 ***	-0.3323 ***	-0.3441 ***	-0.3235 ***	-0.1860 *	-0.3151 ***
	-3.027	-2.933	-3.019	-2.980	-1.672	-2.966
costeira	0.057	0.057	0.064	0.055	0.083 *	0.067
	1.336	1.238	1.373	1.295	1.907	1.615

**Tabela 23. 1** - Resultado das Regressões - Medidas Geográficas e Sócio-Econômicas - Parte I

	OLS	Dist	Contig	Inv.Sal	Sal	Influências Opostas
dist_cap_100	-0.0002 *	-0.0002 *	-0.0001	-0.0002 *	-0.0001	-0.0001
	-1.920	-1.703	-1.211	-1.948	-0.869	-0.972
dist_sp_100	-0.0002 ***	-0.0003 ***	-0.0003 ***	-0.0002 ***	-0.0002 ***	-0.0002 ***
	-4.094	-3.746	-3.893	-4.105	-3.901	-3.936
P_G_Educ_94_96	0.425 **	0.389 **	0.400 **	0.419 **	0.342 *	0.450 ***
	2.420	2.221	2.285	2.386	1.901	2.614
P_G_Infra_94_96	0.171	0.141	0.149	0.171	0.205	0.137
	1.315	1.085	1.151	1.313	1.538	1.077
P_G_Saude_94_96	0.084	0.109	0.080	0.079	0.042	0.099
	0.602	0.786	0.577	0.570	0.295	0.731
P_Transf_94_96	-0.2301 **	-0.2208 **	-0.2320 **	-0.2311 **	-0.2499 ***	-0.1417
	-2.461	-2.337	-2.460	-2.475	-2.606	-1.550
P_Emprego_Ind_C_91	0.401 ***	0.436 ***	0.443 ***	0.411 ***	0.362 **	0.547 ***
	2.761	2.868	2.908	2.828	2.427	3.825
P_Emprego_Serv_C_91	0.164	0.188	0.183	0.163	0.149	0.177
	0.837	0.927	0.901	0.831	0.739	0.923
ln_murder_91	-0.0128	-0.0024	-0.0015	-0.0128	-0.0231	-0.0176
	-0.780	-0.144	-0.092	-0.782	-1.371	-1.105
rho		-0.0923 *	-0.1346 ***	-0.0968 **	4.479 ***	-0.2776 ***
	0	-1.743	-3.691	-2.268	126.9	-9.804
lambda		0.288 ***	0.188 ***	0.072	0.131	0.042
	0	9.282	10.89	1.532	0.323	0.025
Observações:	4267	4267	4267	4267	4267	4267
R <sup>2</sup> ajustado	53.92%	54.76%	54.73%	54.03%	51.53%	55.77%
log-likelihood						
tempo (segundos)	0	1,857	0	120	135	152

(\* t assintótico)

\*\*\* siginificante a 1%

\*\* siginificante a 5%

\* siginificante a 10%

**Tabela 23. 2** - Resultado das Regressões - Medidas Geográficas e Sócio-Econômicas - Parte II

Como esperado, exceto no caso do parâmetro espacial, não há muita mudança nos coeficientes.

Analisando apenas o coeficiente espacial na variável dependente,  $\rho$ , observamos que todos indicam para a mesma interpretação geral, no sentido da convergência e competição. Os parâmetros estimados pelas matrizes geográficas já foram interpretados no capítulo correspondente, mas, como referência, indicam um grau de competição espacial pelo emprego, dentro das aglomerações.

Observando o  $\rho$  estimado pela primeira matriz sócio-econômica, este é muito similar estimado pela matriz de distância. Sua interpretação poderia ser similar, considerando o espaço de salários. Assim, pode-se dizer que há um certo padrão de competição pelo emprego entre os municípios com salários médios parecidos.

Ao olhar o parâmetro espacial estimado com a matriz de dissimilaridade, este fica positivo e com alta magnitude, o que não é esperado. Em relação ao sinal, a interpretação pode caminhar no sentido da convergência: se os municípios que possuem salário médio mais diferentes crescem, o município em análise cresce. Assim, um município intermediário seria beneficiado pelo crescimento tanto do extremo de maior salário, quanto pelo extremo de menor salário. Vale ressaltar, mais uma vez, que a magnitude desse estimador levanta suspeitas sobre sua validade.

Já no caso da matriz que reflete as influências opostas, o sinal negativo indica que o crescimento do emprego das AMCs com menor salário que o de uma unidade em análise é prejudicial a ela, enquanto que o crescimento nas unidades com salário maior seria positivo. Isso indicaria um padrão de convergência e do emprego buscar salários médios mais baixos.

Por fim, é importante destacar que essa análise desse capítulo é ainda exploratória e tem o papel de testar a especificação alternativa.

## PARTE VII: LIMITAÇÕES E CONCLUSÕES

### 24. Limitações e Sugestões para Próximos Estudos

Esta dissertação procurou se aprofundar no tópico das matrizes de ponderação espacial. Ao longo do texto, foram feitas algumas indicações de possíveis pontos para aprofundamento em futuros estudos, que de certa forma refletem as limitações desse estudo.

Nas simulações foram utilizados dados homoscedásticos e normalmente distribuídos. Simples alterações nessas condições poderiam ser feitas de forma a verificar o comportamento dos modelos, em espacial dos modelos de máxima verossimilhança. Poderiam ser testados dados heteroscedásticos e distribuições não normais. Nas variáveis independentes, poderiam ser incluídas correlação entre as mesmas, além de testar outras distribuições. Ainda sobre os parâmetros espaciais, como indicado, outros valores para os parâmetros poderiam ser testados. Todas essas alterações nas simulações estariam verificando a sensibilidade e robustez dos modelos.

Um campo que certamente requer futuras investigações são os testes de especificação. Foi visto que, mesmo em suas versões robustas, os testes de LM do erro, por exemplo, é muito sensível ao tamanho da amostra. Assim, caberiam estudos não só para testar seus comportamentos, como para buscar testes que melhor auxiliem a identificação do modelo adequado.

Certamente, um campo que, apesar de aqui explorado, ainda exhibe um escopo muito grande a ser explorado, são as matrizes espaciais. Foi visto que os modelos são muito sensíveis à essas matrizes. Os testes realizados aqui buscaram, de forma mais específica, verificar se uma suposta simplificação de uma medida era capaz de captar relações geradas por uma mais complexa. Valeria, no entanto, em futuros estudos, verificar diferentes combinações de geração de dados, como tempo de deslocamento, e verificar o poder de captação por outras matrizes. Ainda, verificar o quanto uma medida supostamente mais ampla e complexa seria capaz de captar relações geradas de modo mais simples.

Cabe também um estudo mais aprofundado das relações nos espaços sócio-econômicos. Seria importante verificar quando relações sócio-econômicas seriam mais adequadas do que relações medidas do ponto de vista geográfico e vice-versa.

De um modo geral, em relação às matrizes de ponderação espacial, caberia a investigação de ferramentas que permitam compreender o quanto uma certa escolha seria robusta em relação à outras. Nessa dissertação, por exemplo, sugerem-se dois testes de associação entre variáveis defasadas espacialmente calculadas por matrizes diferentes, mas certamente deve haver formas complementares para verificar essa adequabilidade.

Ainda, questões relacionadas poderiam ser levantadas nos campos de fronteira, mais especificamente, nas aplicações em painel e nos modelos semi-paramétricos. Nesses últimos cabem discussões sobre a mudança ou não dos pesos espaciais ao longo do tempo e da endogeneização da matriz espacial.

Por fim, valeriam ser aprofundados os estudos apresentados na Parte VI dessa dissertação, a decomposição de influências espaciais em dois espaços simultâneos. Para tanto seria necessário verificar as propriedades desse modelo bem como desenvolver algoritmos capazes de estimar os parâmetros apresentados.

## 25. Conclusões

Conforme indicado no início dessa dissertação, o foco de investigação é a matriz de ponderação espacial e os impactos de sua escolha nos diferentes modelos. Como na literatura prioritária de econometria espacial ela é imposta exogenamente pelo investigador, algumas questões sobre a adequação da escolha dessa matriz precisavam ser feitas.

Uma questão específica de interesse é o quanto uma medida simplificada é capaz de captar os efeitos que seriam produzidos por uma medida mais complexa e vice-versa. Outra questão é qual a diferença obtida ao se optar por diferentes medidas para a composição da ponderação espacial. Além disso, procura-se entender se os diferentes modelos são capazes de diferenciar a forma de geração de dados, ou seja, se há uma forma para se escolher o modelo mais adequado.

Utilizando de forma prioritária um método de simulação experimental, a simulação de Monte Carlo, e também respaldados pelas evidências obtidas através de dados empíricos, algumas respostas começam a ser traçadas.

Os testes de especificação espacial parecem ser bons para identificar a presença de componentes espaciais, mas não para diferenciá-los de forma confiável. Em particular, os testes LM do erro parecem sofrer prejuízos com o crescimento da amostra. Já os modelos SAR e SAC, em conjunto com o teste LM da variável dependente parecem gerar bons indicadores de qual o melhor modelo a ser utilizado.

Dadas as indicações da literatura e as evidências mostradas nessa dissertação, a estratégia mais confiável para a busca de um modelo parece ser utilizar as estimativas dos modelos, em conjunto com as estatísticas de Moran, para identificar presença de algum tipo de correlação espacial, e do LM da variável dependente, para indicar esse tipo de correlação especificamente.

Em particular, uma abordagem *forward* poderia iniciar com um modelo SAR e então avançar para um modelo SAC, no caso de existência da correlação espacial na variável dependente ou para um modelo SEM, caso isso não seja observado. Uma abordagem *backward* poderia partir do modelo SAC e, de acordo com o observado, reduzir para um modelo SAR ou SEM.

De acordo com as inferências feitas a partir das simulações, caso o modelo SAR aponte para a existência de  $\rho$ , a estimação por SEM parece não adicionar muita informação. O ideal, então, seria estimar por um modelo completo (SAC), que continua controlando a presença de  $\rho$  e testa a presença de  $\lambda$ . Caso o modelo SAR não aponte para a existência de  $\rho$ , pode-se, então, utilizar uma estimativa SEM

Outro aspecto geral que pode ser observado é que independentemente do modelo escolhido, matriz e do tamanho de amostra, não há muita diferença em termos de esperança dos parâmetros não espaciais, exceto em relação à constante.

Com isso, a conclusão que se pode traçar é que o uso das ferramentas de econometria espacial deve ser aplicado quando o interesse da análise é na estimação dos parâmetros de correlação espacial. A simples presença do componente de correlação espacial não parece provocar desvios da estimativa dos parâmetros que justifiquem a sua inclusão apenas como controle. Claramente há um ganho de eficiência, porém isso deveria ocorrer com a inclusão de qualquer variável relevante omitida que fosse fonte de variação no erro.

Durante a dissertação foi sugerido um teste de ajuste entre as variáveis explicativas e a variável dependente defasada. Este teste é um indicador do benefício potencial (ou do viés gerado), se a opção for feita pela inclusão (ou por ignorar) a componente espacial.

Comparando o grupo de análises feitas a partir das técnicas de máxima verossimilhança com as estimativas feitas por GMM, estes últimos parecem ser mais atrativos. Em termos operacionais, exceto no caso do modelo SEM, são mais rápidos e consomem menos memória, além de serem operacionalizados de forma mais simples em pacotes econométricos já estabelecidos.

Em termos econométricos, são mais robustos pois não precisam assumir a normalidade dos resíduos,. Quanto à esperança dos estimadores, eficiência e erros do tipo I e II são muito próximos aos modelos ML na simulação. Vale ressaltar, no entanto, que em algumas estimações e simulações os modelos ML se mostraram ligeiramente superior, em termos de precisão e eficiência. Porém, isso foi feito sob condições de resíduos normais e homoscedásticos.

Comparando o uso da matriz de contigüidade no lugar da matriz de distâncias, é possível notar que a matriz de contigüidade não é, de forma geral, uma boa substituição para a matriz de distâncias. Ela capta parcialmente o efeito de *rho* e *lambda*, além de captar erroneamente *lambda*, quando este inexistente.

Na amostra pequena, porém, mesmo utilizando o modelo e a matriz corretos, *lambda* é estimado com viés. Ao se aumentar o tamanho da amostra, o viés é eliminado com a matriz correta, e com a matriz simplificada a captação é parcial. Dessa forma não parece ser recomendado o uso de modelos espaciais em amostras muito pequenas quando o objetivo for estimar o parâmetro de correlação espacial no resíduo.

Quando a amostra cresce, o poder de explicação da matriz de contigüidade sob a matriz de distâncias parece ser reduzido. Assim, quanto maior o número de observações, menor o poder de captação da matriz de contigüidade, e menos recomendado seria o seu uso.

Não parece haver, assim, algo similar à convergência assintótica, em que diferentes formas de medir relações similares convergem para resultados similares. Com isso pode-se concluir que escolha adequada da matriz de ponderação espacial é crucial para uma estimativa correta dos parâmetros espaciais. Como escolher corretamente essa matriz, continuará a ser um campo em investigação.



## REFERÊNCIAS

- ACOSTA et al. *Monte Carlo Properties of Spatial Dependence LM tests*. 2003
- ANDRADE, Eduardo, Márcio LAURINI, Regina MADALOZZO, and Pedro L. Valls PEREIRA. Convergence Clubs among Brazilian Municípios. *Economic Letters* 83 (2):179-84. 2004
- ANSELIN, Luc. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, 1988.
- \_\_\_\_\_. *Spatial Econometrics*. April 26, 1999.
- \_\_\_\_\_. Spatial econometrics In: *A companion to theoretical econometrics*, ed. B. Baltagi. Oxford, England: Blackwell. 2001
- \_\_\_\_\_. Classes Notes. 2003. Disponível em: <http://sal.uiuc.edu/default.php>
- \_\_\_\_\_. Spatial Externalities, Spatial Multipliers, And Spatial Econometrics *International Regional Science Review* 26 (2). 2003.
- ANSELIN and FLORAX. Small Sample properties of tests for spatial dependence in regression models: some further results. In: Anselin and Florax (eds) *New Directions in Spatial Econometrics*.. p 21-74. 1995
- AZZONI, Carlos R. , Naercio MENEZES-FILHO, Tatiane A. de. MENEZES, and Raúl SILVEIRA-NETO. Geografia e convêrgencia de renda entre os estados brasileiros. In: *Desigualdade e pobreza no Brasil*, ed. R. Henriques. Rio de Janeiro: IPEA. 2000
- AZZONI, Carlos R., Naercio MENEZES-FILHO, Tatiana de MENEZES, and Raúl SILVEIRA-NETO. Geography and income convergence among Brazilian states. In: *Research Network Working Paper R-395*. Washington DC: IADB. 2000
- AZZONI, Carlos Roberto. Economic Growth and Regional Income Inequality in Brazil. *The Annals of Regional Science* 35:133-52. 2001
- BOARNET and GLAZER. Federal grants and yardstick competition. *Journal of Urban Economics*. Academic Press. 52., p 53-64. 2002
- BUSSAB et al. *Estatística Básica*. Saraiva: São Paulo, 2004.
- BRUCKNER. *Strategic Interaction Among Governments*. June, 2003.
- CASE and ROSEN. Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*. 52, p 285-307. 1997
- CASE. *On the use of spatial autoregressive models in demand analysis*. Research Program in Development Studies. Discussion Paper no. 135 (Princeton University). 1987

- DA SILVA PORTO JR., Sabino, and Eduardo Pontual RIBEIRO. Dinâmica Espacial da Renda Per capita e Crescimento Entre os Municípios da Região Nordeste do Brasil - uma Análise Markoviana. Paper read at *Anais do XXXI Encontro Nacional de Economia* . 2003
- DE MELLO, Luiz R. "Public Finance, Government Spending and Economic Growth: The Case of Local Governments in Brazil." *Applied Economics* 34 (15):1871-8. 2002
- ELHORST, J. Paul. *Dynamic Models in Space and Time*. Research Report 00C16, University of Groningen, Research Institute SOM (Systems, Organisations and Management). 2000
- ELHORST, J. Paul. *Panel Data Models Extended to Spatial Error Autocorrelation or a Spatially Lagged Dependent Variable*. Research Report 01C05, University of Groningen, Research Institute SOM (Systems, Organisations and Management). 2001
- ERTUR and GALLO. An Exploraty Spatial Data Analysis of European Regional Disparities, 1980-1995. In: Fingleton, *Growth*. Springer, 2003
- FERREIRA, Afonso. Convergence in Brazil: Recent Trends and Long Run Prospects. *Applied Economics* 32 (4):479-89. 2000
- FINGLETON, Bernard. Spurious Spatial Regression: Some Monte Carlo Results with a spatial unit root and spatial cointegration. *Journal of Regional Science*, Vol. 39. 1999
- \_\_\_\_\_. *European Regional Growth*. Springer, 2003
- \_\_\_\_\_. Externalities, Economic Geography, And Spatial Econometrics: Conceptual And Modeling Developments. *International Regional Science Review* 26 (2):197-207. 2003
- GASPARINI, Carlos, and Francisco S. RAMOS. "Incentivos à Eficiência na Descentralização Fiscal Brasileira: O Caso do FPM." *Pesquisa e Planejamento Econômico* 34 (1). 2004
- GLAESER, Edward L., Guy DUMAIS, and Glenn ELLISON. Geographic Concentration as a Dynamic Process. *Review of Economics & Statistics* 84 (2):193-204. 2002
- GLAESER, Edward L., Jose A. SCHEINKMAN, and Andrei SHLEIFER. Economic Growth in a Cross-Section of Cities. *Journal of Monetary Economics* 36 (1):117-43. 1995

- HAINING. *Spatial Data Analysis: Theory and Practice*. Cambridge, 2003.
- HENDRY. Monte Carlo Experimentation in Econometrics. In: Griliches and Intriligator, *Handbook of Econometrics, Vol. II*. 1984
- HERNÁNDEZ-MURILLO. Strategic Interaction in Tax Policies Among States. *Review of The Federal Reserve Bank of St. Louis*. May/June. pp 47-56. 2003
- KELEIJAN and PRUCHA.. A generalized moments estimator for the autoregressive parameter in a Spatial Model. Working Paper 95-03. University of Maryland, 1995.
- \_\_\_\_\_. A generalized spatial two stage least squares procedures form estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real State Finance and Economics* 17.. p99-121. 1998
- KELEIJAN and ROBINSON. A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results. *Regional Science and Urban Economics* 28. p389-417. 1998
- KRUGMAN, Paul R.. *Geography and Trade*. Cambridge: MIT Press. 1991a
- \_\_\_\_\_. Increasing Returns and Economic Geography. *Journal of Political Economy* 99:483-99. 1991b.
- \_\_\_\_\_. 1999. "The Role of Geography in Development." *International Regional Science Review* 22 (2):142-61
- LESAGE. *Spatial Econometrics*. December, 1998.
- LESAGE. *The theory and practice of Spatial Econometrics*. February, 1999.
- LESAGE. *Applied Econometrics Using Matlab*. October, 1998
- MORAN. A test for the serial dependence of residuals. *Biometrika* 37. 1950. p178-181
- REVELLI. Reaction or interaction? Spatial process identification in multi-tiered government structure. *Department of Economics, University of Torino, Working Paper Series. No. 02/2005*. February, 2002.
- SLADE, Margareth. *The Role of Economic Space in Decision Making*. 2004
- UPTON and FINGLETON. *Spatial Data Analysis by Example*. Wiley, 1990.

## APÊNDICES

**Rotina base de simulação – Programação MatLab**

```

%inicializa
clear;
clc
load('Matrizes.mat');

% parametros:
b = [20 1.5 0.8]';
n = 27; %parâmetros a serem ajustados
rho = 0.5; %parâmetros a serem ajustados
lambda = 0; %parâmetros a serem ajustados

sigE = 0.7; %desvio padrão do erro
sigX = 5; %desvio padrão de X
meanX = 50;

% Var fixas:
J = ones(n,1);
Wgera = WcontignormP; %parâmetros a serem ajustados
Westima = WdistnormP; %parâmetros a serem ajustados
A = sparse(inv(eye(n)-rho*Wgera));
B = sparse(inv(eye(n)-lambda*Wgera));

%looping
for i = 1:1000;
    i
    %E = random('norm', 0, sigE, n,1);
    E = randn(n,1)*sigE;
    X = randn(n,2)*sigX + meanX;
    % Gerando
    Y = A*([J X]*b) + A*B*E;

    %Estimando OLS
    resultsOLS = ols(Y,[J X]);

    % Estimando SAR
    resultsGsar = sar(Y,[J X], Wgera);
    resultsEsar = sar(Y,[J X], Westima);

    % Estimando SAR-GMM
    resultsGsarg = sar_gmm(Y,[J X], Wgera);
    resultsEsarg = sar_gmm(Y,[J X], Westima);

    % Estimando SEM
    resultsGsem = sem(Y,[J X], Wgera);
    resultsEsem = sem(Y,[J X], Westima);

    % Estimando SEM-GMM
    resultsGsemg = sem_gmm_marcelo(Y,[J X], Wgera);
    resultsEsemg = sem_gmm_marcelo(Y,[J X], Westima);

    % Estimando SAC
    resultsGsac = sac(Y,[J X], Wgera, Wgera);
    resultsEsac = sac(Y,[J X], Westima, Westima);

```

```

% Estimando SAC-GMM
resultsGsacg = sac_gmm(Y,[J X], Wgera, Wgera);
resultsEsacg = sac_gmm(Y,[J X], Westima, Westima);

% Salvando:

%ols
Gera.betaols(:,i) = [resultsOLS.beta;0;0];
Gera.tstatols(:,i) = [resultsOLS.tstat;0;0];
Gera.sigeols(i) = [resultsOLS.sige];
Gera.rbarols(i) = [resultsOLS.rbar];

%SEM
Gera.betasem(:,i) = [resultsGsem.beta ;0; resultsGsem.rho];
Estima.betasem(:,i) = [resultsEsem.beta ;0; resultsEsem.rho];

Gera.rbarsem(i) = [resultsGsem.rbar];
Estima.rbarsem(i) = [resultsEsem.rbar];

Gera.liksem(i) = [resultsGsem.lik];
Estima.liksem(i) = [resultsEsem.lik];

Gera.sigesem(i) = [resultsGsem.sige];
Estima.sigesem(i) = [resultsEsem.sige];

Gera.tstatsem(:,i) = [resultsGsem.tstat(1:3); 0 ; resultsGsem.tstat(4)];
Estima.tstatsem(:,i) = [resultsEsem.tstat(1:3); 0 ; resultsEsem.tstat(4)];

Gera.timesem(i) = [resultsGsem.time];
Estima.timesem(i) = [resultsEsem.time];

%SEMG
Gera.betasemg(:,i) = [resultsGsemg.beta ;0; resultsGsemg.lambda];
Estima.betasemg(:,i) = [resultsEsemg.beta ;0; resultsEsemg.lambda];

Gera.rbarsemg(i) = [resultsGsemg.rbar];
Estima.rbarsemg(i) = [resultsEsemg.rbar];

Gera.sigesemg(i) = [resultsGsemg.sige];
Estima.sigesemg(i) = [resultsEsemg.sige];

Gera.tstatsemg(:,i) = [resultsGsemg.tstat; 0; resultsGsemg.lambdatstat];
Estima.tstatsemg(:,i) = [resultsEsemg.tstat; 0; resultsEsemg.lambdatstat];

Gera.timesemg(i) = [resultsGsemg.time];
Estima.timesemg(i) = [resultsEsemg.time];

%SAR
Gera.betasar(:,i) = [resultsGsar.beta ;resultsGsar.rho; 0];
Estima.betasar(:,i) = [resultsEsar.beta ;resultsEsar.rho; 0];

Gera.rbarsar(i) = [resultsGsar.rbar];
Estima.rbarsar(i) = [resultsEsar.rbar];

Gera.liksar(i) = [resultsGsar.lik];
Estima.liksar(i) = [resultsEsar.lik];

Gera.sigesar(i) = [resultsGsar.sige];
Estima.sigesar(i) = [resultsEsar.sige];

```

```

Gera.tstatsar(:,i) = [resultsGsar.tstat;0];
Estima.tstatsar(:,i) = [resultsEsar.tstat;0];

Gera.timesar(i) = [resultsGsar.time];
Estima.timesar(i) = [resultsEsar.time];

%SARG
Gera.betasarg(:,i) = [resultsGsarg.beta ;resultsGsarg.rho; 0];
Estima.betasarg(:,i) = [resultsEsarg.beta ;resultsEsarg.rho; 0];

Gera.rbarsarg(i) = [resultsGsarg.rbar];
Estima.rbarsarg(i) = [resultsEsarg.rbar];

Gera.sigesarg(i) = [resultsGsarg.sige];
Estima.sigesarg(i) = [resultsEsarg.sige];

Gera.tstatsarg(:,i) = [resultsGsarg.tstat; resultsGsarg.rhotstat;0];
Estima.tstatsarg(:,i) = [resultsEsarg.tstat; resultsEsarg.rhotstat;0];

Gera.timesarg(i) = [resultsGsarg.time];
Estima.timesarg(i) = [resultsEsarg.time];

%SAC
Gera.betasac(:,i) = [resultsGsac.beta ;resultsGsac.rho; resultsGsac.lam];
Estima.betasac(:,i) = [resultsEsac.beta ;resultsEsac.rho;resultsEsac.lam ];

Gera.rbarsac(i) = [resultsGsac.rbar];
Estima.rbarsac(i) = [resultsEsac.rbar];

Gera.lik sac(i) = [resultsGsac.lik];
Estima.lik sac(i) = [resultsEsac.lik];

Gera.sigesac(i) = [resultsGsac.sige];
Estima.sigesac(i) = [resultsEsac.sige];

Gera.tstatsac(:,i) = [resultsGsac.tstat];
Estima.tstatsac(:,i) = [resultsEsac.tstat];

Gera.timesac(i) = [resultsGsac.time];
Estima.timesac(i) = [resultsEsac.time];

%SACG
Gera.betasacg(:,i) = [resultsGsacg.beta ;resultsGsacg.rho; resultsGsacg.lam];
Estima.betasacg(:,i) = [resultsEsacg.beta ;resultsEsacg.rho ;resultsEsacg.lam];

Gera.rbarsacg(i) = [resultsGsacg.rbar];
Estima.rbarsacg(i) = [resultsEsacg.rbar];

Gera.sigesacg(i) = [resultsGsacg.sige];
Estima.sigesacg(i) = [resultsEsacg.sige];

Gera.tstatsacg(:,i) = [resultsGsacg.tstat; resultsGsacg.rhotstat; resultsGsacg.lambdatstat];
Estima.tstatsacg(:,i) = [resultsEsacg.tstat; resultsEsacg.rhotstat; resultsEsacg.lambdatstat];

Gera.timesacg(i) = [resultsGsacg.time2];
Estima.timesacg(i) = [resultsEsacg.time2];

```

end

```

%Salvar
clear W* A B E X Y J re*;
%save('sac_small_n')

%Apresentar resultados
[ mean(Gera.betaols(1,:)) mean(Gera.betasar(1,:)) mean(Estima.betasar(1,:)) mean(Gera.betasarg(1,:))
mean(Estima.betasarg(1,:)) mean(Gera.betasem(1,:)) mean(Estima.betasem(1,:))
mean(Gera.betasemg(1,:)) mean(Estima.betasemg(1,:)) mean(Gera.betasac(1,:))
mean(Estima.betasac(1,:)) mean(Gera.betasacg(1,:)) mean(Estima.betasacg(1,:)) ; ...
std(Gera.betaols(1,:)) std(Gera.betasar(1,:)) std(Estima.betasar(1,:)) std(Gera.betasarg(1,:))
std(Estima.betasarg(1,:)) std(Gera.betasem(1,:)) std(Estima.betasem(1,:)) std(Gera.betasemg(1,:))
std(Estima.betasemg(1,:)) std(Gera.betasac(1,:)) std(Estima.betasac(1,:)) std(Gera.betasacg(1,:))
std(Estima.betasacg(1,:)) ; ...
length(find(abs(Gera.tstatols(1,:))>=1.96))/i length(find(abs(Gera.tstatsar(1,:))>=1.96))/i
length(find(abs(Estima.tstatsar(1,:))>=1.96))/i length(find(abs(Gera.tstatsarg(1,:))>=1.96))/i
length(find(abs(Estima.tstatsarg(1,:))>=1.96))/i length(find(abs(Gera.tstatsem(1,:))>=1.96))/i
length(find(abs(Estima.tstatsem(1,:))>=1.96))/i length(find(abs(Gera.tstatsemg(1,:))>=1.96))/i
length(find(abs(Estima.tstatsemg(1,:))>=1.96))/i length(find(abs(Gera.tstatsac(1,:))>=1.96))/i
length(find(abs(Estima.tstatsac(1,:))>=1.96))/i length(find(abs(Gera.tstatsacg(1,:))>=1.96))/i
length(find(abs(Estima.tstatsacg(1,:))>=1.96))/i ; ...
mean(Gera.betaols(2,:)) mean(Gera.betasar(2,:)) mean(Estima.betasar(2,:)) mean(Gera.betasarg(2,:))
mean(Estima.betasarg(2,:)) mean(Gera.betasem(2,:)) mean(Estima.betasem(2,:))
mean(Gera.betasemg(2,:)) mean(Estima.betasemg(2,:)) mean(Gera.betasac(2,:))
mean(Estima.betasac(2,:)) mean(Gera.betasacg(2,:)) mean(Estima.betasacg(2,:)) ; ...
std(Gera.betaols(2,:)) std(Gera.betasar(2,:)) std(Estima.betasar(2,:)) std(Gera.betasarg(2,:))
std(Estima.betasarg(2,:)) std(Gera.betasem(2,:)) std(Estima.betasem(2,:)) std(Gera.betasemg(2,:))
std(Estima.betasemg(2,:)) std(Gera.betasac(2,:)) std(Estima.betasac(2,:)) std(Gera.betasacg(2,:))
std(Estima.betasacg(2,:)) ; ...
length(find(abs(Gera.tstatols(2,:))>=1.96))/i length(find(abs(Gera.tstatsar(2,:))>=1.96))/i
length(find(abs(Estima.tstatsar(2,:))>=1.96))/i length(find(abs(Gera.tstatsarg(2,:))>=1.96))/i
length(find(abs(Estima.tstatsarg(2,:))>=1.96))/i length(find(abs(Gera.tstatsem(2,:))>=1.96))/i
length(find(abs(Estima.tstatsem(2,:))>=1.96))/i length(find(abs(Gera.tstatsemg(2,:))>=1.96))/i
length(find(abs(Estima.tstatsemg(2,:))>=1.96))/i length(find(abs(Gera.tstatsac(2,:))>=1.96))/i
length(find(abs(Estima.tstatsac(2,:))>=1.96))/i length(find(abs(Gera.tstatsacg(2,:))>=1.96))/i
length(find(abs(Estima.tstatsacg(2,:))>=1.96))/i ; ...
mean(Gera.betaols(3,:)) mean(Gera.betasar(3,:)) mean(Estima.betasar(3,:)) mean(Gera.betasarg(3,:))
mean(Estima.betasarg(3,:)) mean(Gera.betasem(3,:)) mean(Estima.betasem(3,:))
mean(Gera.betasemg(3,:)) mean(Estima.betasemg(3,:)) mean(Gera.betasac(3,:))
mean(Estima.betasac(3,:)) mean(Gera.betasacg(3,:)) mean(Estima.betasacg(3,:)) ; ...
std(Gera.betaols(3,:)) std(Gera.betasar(3,:)) std(Estima.betasar(3,:)) std(Gera.betasarg(3,:))
std(Estima.betasarg(3,:)) std(Gera.betasem(3,:)) std(Estima.betasem(3,:)) std(Gera.betasemg(3,:))
std(Estima.betasemg(3,:)) std(Gera.betasac(3,:)) std(Estima.betasac(3,:)) std(Gera.betasacg(3,:))
std(Estima.betasacg(3,:)) ; ...
length(find(abs(Gera.tstatols(3,:))>=1.96))/i length(find(abs(Gera.tstatsar(3,:))>=1.96))/i
length(find(abs(Estima.tstatsar(3,:))>=1.96))/i length(find(abs(Gera.tstatsarg(3,:))>=1.96))/i
length(find(abs(Estima.tstatsarg(3,:))>=1.96))/i length(find(abs(Gera.tstatsem(3,:))>=1.96))/i
length(find(abs(Estima.tstatsem(3,:))>=1.96))/i length(find(abs(Gera.tstatsemg(3,:))>=1.96))/i
length(find(abs(Estima.tstatsemg(3,:))>=1.96))/i length(find(abs(Gera.tstatsac(3,:))>=1.96))/i
length(find(abs(Estima.tstatsac(3,:))>=1.96))/i length(find(abs(Gera.tstatsacg(3,:))>=1.96))/i
length(find(abs(Estima.tstatsacg(3,:))>=1.96))/i ; ...
mean(Gera.betaols(4,:)) mean(Gera.betasar(4,:)) mean(Estima.betasar(4,:)) mean(Gera.betasarg(4,:))
mean(Estima.betasarg(4,:)) mean(Gera.betasem(4,:)) mean(Estima.betasem(4,:))
mean(Gera.betasemg(4,:)) mean(Estima.betasemg(4,:)) mean(Gera.betasac(4,:))
mean(Estima.betasac(4,:)) mean(Gera.betasacg(4,:)) mean(Estima.betasacg(4,:)) ; ...
std(Gera.betaols(4,:)) std(Gera.betasar(4,:)) std(Estima.betasar(4,:)) std(Gera.betasarg(4,:))
std(Estima.betasarg(4,:)) std(Gera.betasem(4,:)) std(Estima.betasem(4,:)) std(Gera.betasemg(4,:))
std(Estima.betasemg(4,:)) std(Gera.betasac(4,:)) std(Estima.betasac(4,:)) std(Gera.betasacg(4,:))
std(Estima.betasacg(4,:)) ; ...

```

```

length(find(abs(Gera.tstatols(4,:))>=1.96))/i length(find(abs(Gera.tstatsar(4,:))>=1.96))/i
length(find(abs(Estima.tstatsar(4,:))>=1.96))/i length(find(abs(Gera.tstatsarg(4,:))>=1.96))/i
length(find(abs(Estima.tstatsarg(4,:))>=1.96))/i length(find(abs(Gera.tstatsem(4,:))>=1.96))/i
length(find(abs(Estima.tstatsem(4,:))>=1.96))/i length(find(abs(Gera.tstatsemg(4,:))>=1.96))/i
length(find(abs(Estima.tstatsemg(4,:))>=1.96))/i length(find(abs(Gera.tstatsac(4,:))>=1.96))/i
length(find(abs(Estima.tstatsac(4,:))>=1.96))/i length(find(abs(Gera.tstatsacg(4,:))>=1.96))/i
length(find(abs(Estima.tstatsacg(4,:))>=1.96))/i ; ...
mean(Gera.betaols(5,:)) mean(Gera.betasar(5,:)) mean(Estima.betasar(5,:)) mean(Gera.betasarg(5,:))
mean(Estima.betasarg(5,:)) mean(Gera.betasem(5,:)) mean(Estima.betasem(5,:))
mean(Gera.betasemg(5,:)) mean(Estima.betasemg(5,:)) mean(Gera.betasac(5,:))
mean(Estima.betasac(5,:)) mean(Gera.betasacg(5,:)) mean(Estima.betasacg(5,:)) ; ...
std(Gera.betaols(5,:)) std(Gera.betasar(5,:)) std(Estima.betasar(5,:)) std(Gera.betasarg(5,:))
std(Estima.betasarg(5,:)) std(Gera.betasem(5,:)) std(Estima.betasem(5,:)) std(Gera.betasemg(5,:))
std(Estima.betasemg(5,:)) std(Gera.betasac(5,:)) std(Estima.betasac(5,:)) std(Gera.betasacg(5,:))
std(Estima.betasacg(5,:)) ; ...
length(find(abs(Gera.tstatols(5,:))>=1.96))/i length(find(abs(Gera.tstatsar(5,:))>=1.96))/i
length(find(abs(Estima.tstatsar(5,:))>=1.96))/i length(find(abs(Gera.tstatsarg(5,:))>=1.96))/i
length(find(abs(Estima.tstatsarg(5,:))>=1.96))/i length(find(abs(Gera.tstatsem(5,:))>=1.96))/i
length(find(abs(Estima.tstatsem(5,:))>=1.96))/i length(find(abs(Gera.tstatsemg(5,:))>=1.96))/i
length(find(abs(Estima.tstatsemg(5,:))>=1.96))/i length(find(abs(Gera.tstatsac(5,:))>=1.96))/i
length(find(abs(Estima.tstatsac(5,:))>=1.96))/i length(find(abs(Gera.tstatsacg(5,:))>=1.96))/i
length(find(abs(Estima.tstatsacg(5,:))>=1.96))/i ; ...
sqrt(mean(Gera.sigeols)) sqrt(mean(Gera.sigesar)) sqrt(mean(Estima.sigesar)) sqrt(mean(Gera.sigesarg))
sqrt(mean(Estima.sigesarg)) sqrt(mean(Gera.sigesem)) sqrt(mean(Estima.sigesem))
sqrt(mean(Gera.sigesemg)) sqrt(mean(Estima.sigesemg)) sqrt(mean(Gera.sigesac))
sqrt(mean(Estima.sigesac)) sqrt(mean(Gera.sigesacg)) sqrt(mean(Estima.sigesacg)) ; ...
mean(Gera.rbarols) mean(Gera.rbarsar) mean(Estima.rbarsar) mean(Gera.rbarsarg)
mean(Estima.rbarsarg) mean(Gera.rbarsem) mean(Estima.rbarsem) mean(Gera.rbarsemg)
mean(Estima.rbarsemg) mean(Gera.rbarsac) mean(Estima.rbarsac) mean(Gera.rbarsacg)
mean(Estima.rbarsacg) ; ...
0 mean(Gera.likсар) mean(Estima.likсар) 0 0 mean(Gera.liksem) mean(Estima.liksem) 0 0
mean(Gera.likсар) mean(Estima.likсар) 0 0 ; ...
0 mean(Gera.timesar) mean(Estima.timesar) mean(Gera.timesarg) mean(Estima.timesarg)
mean(Gera.timesem) mean(Estima.timesem) mean(Gera.timesemg) mean(Estima.timesemg)
mean(Gera.timesac) mean(Estima.timesac) mean(Gera.timesacg) mean(Estima.timesacg) ; ...
0 sum(Gera.timesar) sum(Estima.timesar) sum(Gera.timesarg) sum(Estima.timesarg) sum(Gera.timesem)
sum(Estima.timesem) sum(Gera.timesemg) sum(Estima.timesemg) sum(Gera.timesac)
sum(Estima.timesac) sum(Gera.timesacg) sum(Estima.timesacg) ]

```



## Testes de Dependência Espacial – Programação MatLab

```

function spat_specification_tests_result = spat_tests(y,x,W);
% PURPOSE: computes five spatial specification tests that uses
%   residuals and parameters from OLS regression:
%   1) Moran's I-statistic for spatial correlation
%   2) LM error test
%   3) LM error robust test
%   4) LM lag error test
%   5) LM lag error robust test
% -----
% USAGE: spat_specification_test_result = spat_specification_test(y,x,W)
% where: y = dependent variable vector
%   x = independent variables matrix
%   W = spatial weight matrix (MUST BE NORMALIZED)
% -----
% RETURNS: a structure variable
%   spat_specification_tests_result.morani = Moran's I statistic
%   spat_specification_tests_result.moranistat =  $[i - E(i)]/\text{std}(i)$ , standardized version
%   spat_specification_tests_result.moranprob = std normal marginal probability
%   spat_specification_tests_result.lmerror = LM Error statistic
%   spat_specification_tests_result.lmerrorprob = LM Error statistic marginal prob
%   spat_specification_tests_result.lmerrorrob = LM Error robust statistic
%   spat_specification_tests_result.lmerrorrobprob = LM Error robust statistic marginal prob
%   spat_specification_tests_result.lmlag = LM Lag statistic
%   spat_specification_tests_result.lmlagprob = LM Lag statistic marginal prob
%   spat_specification_tests_result.lmlagrob = LM Lag robust statistic
%   spat_specification_tests_result.lmlagrobprob = LM Lag robust statistic marginal prob
% -----
% NOTES: Moran I standardaized: Normal Distribution
%   critical value: 1.96 (5%)
%
%   All other: chi-squared distribution, with 1df
%   critical value: 6.635 (1%)
%
%   HO:
%   Moran: no spatial correlation
%   LMerror:  $\lambda = 0$ 
%   LMLag:  $\rho = 0$ 
% -----
% See also: lmerror, walds, lratios, lmsar, moran
% -----
% Written by: Marcelo Tyszler, 01/2006
% Fundação Getúlio Vargas
% Graduate Program in Public Management and Government
% Graduate Program in Economics
% São Paulo, Brazil
% mtyszler@fgvsp.br
%
% Based on:
% Functions by
% James P. LeSage, Dept of Economics
% University of Toledo
%
% Anselin, Luc. Spatial Econometrics: Methods and Models. Kluwer
% Academic Publishers, 1988.
% Anselin, Luc. Spatial Econometrics (chapter). 1999
% Anselin, Luc. Spatial Econometrics course slides (w8). 2003
% Acosta et al. Monte Carlo Properties of Spatial Dependence LM Tests. 2003

```

```

% check for correct number of arguments
if nargin <3
    error('Arguments missing')
end

% Start procedure
[n k] = size(x);

% Get OLS parameters and residuals:
OLS = ols(y,x);

beta = OLS.beta;
e = OLS.resid;

%Parâmetros:

T = trace(W*W + W'*W);

epe=e'*e;

WXB = W*x*beta;
M = eye(n) - x*(inv(x'*x))*x';
D = WXB'*M*WXB+T*(epe/n);
D1 = inv(D);

eWe=e'*W*e;
eWy = e'*W*y;

eWy=e'*W*y;

drho = eWy/(epe/n);
dlam = eWe/(epe/n);

%Moran:
MW = M*W;
MWt = M*W';
MW2 = MW*MW;
trMW = trace(MW);

morani = eWe/epe;

Ei = trMW/(n-k);
Vari = (trace(MW*MWt) + trace(MW2)+(trMW^2)) / ((n-k)*(n-k+2)) - Ei^2;

moranistat = (morani - Ei)/sqrt(Vari);

moranprob = norm_prb(moranistat);

%LM error:
lmerror = (dlam^2)/T;
lmerrorprob = 1 - chis_prb(lmerror,1);

%LM error robust:
A = T*(epe/n)*D1;
lmerrorrob = ((dlam - A*drho)^2)/(T*(1-A));
lmerrorrobprob = 1 - chis_prb(lmerrorrob,1);

```

```

%LM lag:
lmlag = ((epe/n)*drho^2)/D;
lmlagprob = 1 - chis_prb(lmlag,1);

%LM lag robust:
lmlagrob = ((epe/n)*((drho-dlam)^2))/(D-T*(epe/n));
lmlagrobprob = 1 - chis_prb(lmlagrob,1);

% Save results:
    spat_specification_tests_result.morani = morani;
    spat_specification_tests_result.moranistat = moranistat;
    spat_specification_tests_result.moranprob = moranprob;
    spat_specification_tests_result.lmerror = lmerror;
    spat_specification_tests_result.lmerrorprob = lmerrorprob;
    spat_specification_tests_result.lmerrorrob = lmerrorrob;
    spat_specification_tests_result.lmerrorrobprob = lmerrorrobprob;
    spat_specification_tests_result.lmlag = lmlag;
    spat_specification_tests_result.lmlagprob = lmlagprob;
    spat_specification_tests_result.lmlagrob = lmlagrob;
    spat_specification_tests_result.lmlagrobprob = lmlagrobprob;

```