

## MÉTODOS E MODELOS EM ECONOMETRIA ESPACIAL. UMA REVISÃO

Alexandre Xavier de CARVALHO YWATA<sup>1</sup>  
Pedro Henrique de Melo ALBUQUERQUE<sup>2</sup>

- RESUMO: Este texto apresenta uma discussão sobre diversos modelos econométricos para estimação de modelos paramétricos na presença de dependência espacial, com dados cross-section. O foco inicial são modelos de dependência espacial com lags espaciais da variável resposta ou lags espaciais do resíduo, com estimação dos parâmetros feita via máxima verossimilhança. Uma análise crítica destes modelos é apresentada em seguida, além de se discutirem testes para detectar presença de dependência espacial. Finalmente, discutem-se métodos de estimação mais robustos, os quais permitem a contabilização de endogeneidade em algumas das variáveis explicativas.
- PALAVRAS-CHAVE: Econometria espacial; dependência espacial; dados cross-section.

### 1 Introdução

Nas últimas décadas, um conjunto cada vez maior de ferramentas analíticas para tratamento de dados espaciais tem surgido na literatura especializada. Estas ferramentas têm auxiliado pesquisadores em diferentes campos da ciência a lidar com a crescente disponibilidade de bases de dados georreferenciados. De fato, diferentemente de séries temporais macroeconômicas, por exemplo, uma base de dados totalmente nova e detalhada, com dados *cross-section* espaciais, pode surgir de um ano para o outro. Além disso, o crescente desenvolvimento de dispositivos de coleta e armazenamento de dados geográficos tem contribuído para a construção de inúmeras bases de dados com componentes espaciais.

Apesar de todo o avanço ocorrido nas décadas recentes, ainda há um grande terreno a ser explorado em termos de ferramentas para dados geograficamente localizados. Os avanços esperados para os próximos anos têm a ver tanto com a formalização de resultados matemáticos, quanto com avanços mais conceituais sobre a aplicação dos modelos que vêm sendo utilizados até o presente momento. Uma discussão sobre tópicos de natureza mais conceitual pode ser encontrada, por exemplo, em Holmes (2010), McMillen (2010) e Pinkse e Slade (2010).

---

<sup>1</sup> Instituto de Pesquisa Econômica Aplicada – IPEA, Diretoria de Estudos Regionais, Urbanos e Fiscais – DIRUR, SBS, Quadra 1, Bloco 1, Edifício BNDES, CEP: 70076-900, Brasília, DF, Brasil. E-mail: alexandre.ywata@ipea.gov.br

<sup>2</sup> Universidade de Brasília – UnB, Departamento de Administração, CEP: 70910-900, Brasília, DF, Brasil. E-mail: pedroa@unb.br

Holmes (2010) apresenta uma discussão interessante sobre os três tipos básicos de abordagem para estudos empíricos em análise de dados espaciais. As três abordagens discutidas são: *i*) abordagem estruturalista; *ii*) abordagem experimentalista; e *iii*) abordagem descritiva. Um entendimento destas três abordagens é importante, para que os pesquisadores possam identificar em quais das três um determinado trabalho empírico se situa, de forma que as vantagens e as limitações do trabalho fiquem mais claras.

Na abordagem estruturalista, o exercício empírico parte de um modelo econômico totalmente especificado, com base em uma teoria geralmente microfundamentada. O objetivo do exercício é estimar parâmetros estruturais do modelo (*deep model parameters*), relativos a preferências e/ou tecnologias. A partir do modelo estimado, é possível simular impactos de políticas (inclusive políticas que ainda não foram implementadas). Na literatura de organização industrial mais recente,<sup>3</sup> os modelos microfundamentados estimados permitem, por exemplo, avaliar *a priori* o impacto da fusão de duas empresas. Apesar de a abordagem estruturalista estar mais desenvolvida para pesquisas em organização industrial, pesquisadores em economia política (Epple e Sieg, 1999) e economia do trabalho (Keane e Wolpin, 1997; Eckstein e Wolpin, 1999) já começaram a utilizá-la.

A abordagem experimentalista surgiu inicialmente na literatura de economia do trabalho. Nesta abordagem, o interesse principal é a identificação do efeito causal de uma determinada política (efeito tratamento). Ao invés de se preocupar com a especificação de um modelo teórico, a ideia básica é encontrar experimentos naturais ou instrumentos válidos para a identificação de causalidade de políticas que já foram implementadas. Para maiores detalhes, o leitor pode recorrer a manuais como Angrist e Pischke (2009) ou Cameron e Trivedi (2005). Nesse contexto, métodos de estimação do tipo mínimos quadrados de dois estágios, ou de forma mais geral, métodos de momentos generalizados, têm um papel muito importante. Outro procedimento comumente empregado é a regressão de descontinuidade (Hahn *et al.*, 2001).

Ao contrário das duas abordagens anteriores, a abordagem descritiva não tem por objetivo quantificar o efeito causal de determinadas políticas. Em geral, os artigos que utilizam a abordagem descritiva se iniciam com uma discussão da teoria econômica, que pode estar ou não embasada em modelos matematicamente fundamentados. A partir de regressões e outros indicadores estatísticos, os autores buscam encontrar evidências nas relações entre as variáveis, que possam corroborar uma determinada teoria (possivelmente, em detrimento de teorias alternativas). As regressões em geral correspondem a formas reduzidas de equações estruturais mais completas. Uma das limitações desta abordagem é que, além de não permitir inferências causais, ela também está sujeita à crítica de Lucas. Dessa forma, alterações no regime econômico podem incorrer em alterações nos parâmetros do modelo, tornando a utilização dos modelos reduzidos menos críveis do ponto de vista de simulações *a priori* de impactos de políticas.<sup>4</sup>

A maioria dos estudos em economia regional e urbana segue a abordagem descritiva. Nos últimos anos, têm surgido estudos que utilizam a abordagem experimentalista para avaliação de políticas. Por sua vez, a utilização da abordagem estruturalista pode trazer vários benefícios para economia regional, dada a dificuldade de se encontrar bons

---

<sup>3</sup> Ver Berry *et al.* (1995 e 2004), Nevo (2001), Petrin (2002) e Akerberg *et al.* (2007).

<sup>4</sup> Ver Hendry (1995).

instrumentos ou bons experimentos naturais. Uma das dificuldades na utilização da abordagem experimentalista em economia regional é a disponibilidade de dados (comparando-se ao número de observações de estudos em economia do trabalho, por exemplo). Uma sugestão para o uso da abordagem experimentalista em economia regional e urbana é a utilização de dados em nível de firmas, por exemplo, ao invés de dados em nível de municípios.

A utilização da abordagem estruturalista para economia regional e/ou urbana deve se iniciar com a construção de um modelo teórico (o que pode não ser tão fácil como no caso de modelos de organização industrial). Por seu turno, a utilização de abordagens estruturalistas em economia regional poderia ser interessante para simulações de políticas públicas. No entanto, pouco tem sido feito neste sentido até agora.

Neste trabalho, apresenta-se uma discussão sobre alguns dos modelos econométricos comumente utilizados para modelagem de dados espaciais. De maneira geral, os modelos apresentados aqui estariam mais adequados para estudos empíricos seguindo as abordagens experimentalista e descritiva. De fato, o estimador de mínimos quadrados de dois estágios, de Kelejian e Prucha, e o estimador de método de momentos generalizado, de Conley (1999), permitem a estimação de parâmetros na presença de variáveis endógenas do lado direito da equação, contabilizando e/ou corrigindo para a presença de autocorrelação espacial nos resíduos do modelo. Mesmo não tratando diretamente a abordagem estruturalista, as ideias apresentadas neste texto fornecerão ao leitor uma noção dos procedimentos para estimação com dados com presença de dependência espacial, o que poderá ser útil para a estimação de parâmetros estruturais em modelos microfundamentados.

Dado o grande avanço pelo qual a literatura em métodos estatísticos para dados espaciais tem passado nos últimos anos, não há interesse aqui em ser exaustivo em termos de metodologias discutidas. Pelo contrário, optou-se por apresentar apenas alguns dos métodos mais comumente utilizados, de forma a transmitir ao leitor uma ideia básica, mas elucidativa, sobre os fundamentos da estimação de modelos econométricos com dependência espacial. Nesse sentido, não serão tratados, por exemplo, dados de painel (vejam-se, por exemplo, Elhorst, 2003; Druska e Horrow, 2004; e Egger *et al.*, 2005), mas apenas dados *cross-section*. Além disso, a abordagem será predominantemente frequentista. Apesar da simpatia em relação aos métodos bayesianos, principalmente no contexto de dados espaciais, para não se estenderem demasiado os autores preferiram ater-se aos procedimentos frequentistas. O leitor poderá encontrar boas exposições em Banerjee *et al.* (2004) e Schabenberger e Gotway (2009), entre outros.

Finalmente, o texto apresenta uma discussão sobre um tópico comumente empregado na literatura: o tratamento de heterogeneidade espacial. Mais especificamente apresenta-se o método de expansão de Casetti (1972) e o método de regressão geograficamente ponderada (*geographically weighted regression*), estes permitem que haja uma variação suave nos parâmetros estimados ao longo do espaço (vide, por exemplo, Fotheringham *et al.*, 2000 e 2002). A regressão geograficamente ponderada é uma metodologia que pode ser uma alternativa útil para a modelagem de processos econômicos regionais no Brasil, dada a grande heterogeneidade entre as unidades da federação brasileiras. Além de discutir a metodologia de regressão geograficamente ponderada comumente encontrada na literatura, discute-se também uma extensão dessa metodologia para estimação utilizando-se método de momentos generalizados. Essa

extensão, conforme abordado em Camargo *et al.* (2010), é denominada método de momentos generalizados geograficamente ponderados.

Além desta introdução, este texto contém mais seis seções. Na seção 2, apresenta-se uma discussão sobre os modelos econométricos espaciais para dados *cross-section* provavelmente mais utilizados na literatura. Na seção 3, discutem-se algumas das críticas mais comuns aos modelos espaciais apresentados na seção 2. Na seção 4, são apresentados alguns dos testes mais utilizados para verificação da presença ou não de dependência espacial. As seções 5 e 6 discutem procedimentos de estimação para contabilizar para a presença de variáveis endógenas no lado direito da equação: a seção 5 apresenta o estimador espacial de mínimos quadrados de dois estágios, e a seção 6 apresenta o estimador de método de momentos generalizados, com correção para a presença de autocorrelação espacial. A heterogeneidade espacial dos dados é discutida na seção 7 e comentários finais encontram-se na seção 8.

## 2 Modelos paramétricos para dependência espacial

Nesta seção, será feita uma discussão de alguns dos modelos paramétricos comumente utilizados em econometria espacial. A discussão se limitará a regressões com dados *cross-section*.<sup>5</sup> Para modelos envolvendo dados de painel espacial, o leitor pode recorrer a Elhorst (2003), Druska e Horrace (2004), Egger *et al.* (2005).

### 2.1 Modelos SAR

Um dos modelos mais comumente utilizados para modelagem de correlação espacial é o modelo autorregressivo espacial (*spatial autoregressive model*), ou simplesmente modelo SAR. A ideia dos modelos SAR é utilizar a mesma ideia dos modelos AR (autorregressivos) em séries temporais, por meio da incorporação de um termo de *lag* entre os regressores da equação. Na sua forma mais simples, o modelo SAR tem expressão:

$$y = \rho W y + \epsilon, \quad (1)$$

onde  $y$  é um vetor coluna, contendo  $n$  observações na amostra para a variável resposta  $y_i$ , o coeficiente escalar  $\rho$  corresponde ao parâmetro autorregressivo, esse parâmetro possui como interpretação o efeito médio da variável dependente relativo à vizinhança espacial na região em questão, já o termo  $\epsilon$  corresponde a um vetor coluna contendo os resíduos  $\epsilon_i$  da equação. Por enquanto, assume-se que os resíduos  $\epsilon_i$  são independentes e identicamente distribuídos, com distribuição normal, com média zero e variância homogênea  $\sigma^2$ . Um dos componentes presentes em uma grande quantidade de modelos espaciais é a matriz  $W$ . Esta matriz é conhecida como matriz de vizinhança, e pode ser definida de diversas formas, o que traz críticas aos modelos espaciais utilizando  $W$  (muitos autores consideram as definições para  $W$  deveras arbitrárias; a este respeito, ver Pinkse e Slade, 2010).

---

<sup>5</sup> Ver Anselin (1988), Lesage (1997 e 1999), Pace e Barry (1997 e 1998), Anselin e Florax (2000), Anselin *et al.* (2004), Lesage e Pace (2009).

Uma das formas mais comumente empregadas de definição da matriz  $W$  se dá por meio da identificação de vizinhos de primeira ordem. Considere-se que cada observação no vetor  $y$  esteja associada a um polígono e um sistema georreferenciado. Por exemplo, o vetor  $y$  pode corresponder a observações de uma determinada variável observada para cada município brasileiro, ou corresponder a observações de uma variável para cada setor censitário na cidade de São Paulo. Neste caso, o elemento  $W_{i,j}$  da matriz  $W$  assume valor  $W_{i,j} = 1$ , caso os polígonos  $i$  e  $j$  sejam vizinhos, e  $W_{i,j} = 0$ , caso  $i$  e  $j$  não sejam vizinhos. A diagonal principal de  $W$  possui todos os elementos iguais a zero, por definição.

Para identificar polígonos (municípios, setores censitários etc.) vizinhos, pode-se considerar uma vizinhança do tipo *queen*, quando além das fronteiras com extensão diferente de zero, puderem ser considerados os vértices como contíguos, na visualização de um mapa. Esse padrão equivale ao movimento da “rainha” no xadrez. Similarmente, uma matriz do tipo *rook* equivale ao movimento das “torres” em um jogo de xadrez. Em outras palavras, a matriz do tipo *rook* ocorre quando apenas as fronteiras com extensão diferente de zero são consideradas, não se levando em conta os vértices na visualização do mapa. Note-se que a vizinhança do tipo *queen* é menos restritiva do que a vizinhança do tipo *rook*. Além da vizinhança de primeira ordem, podem-se utilizar vizinhanças de ordem maior. Na definição de vizinhança de segunda ordem, por exemplo, os polígonos  $i$  e  $j$  são vizinhos caso exista outro polígono  $k$ , para o qual  $i$  e  $k$  sejam vizinhos de primeira ordem, e  $j$  e  $k$  também sejam vizinhos de primeira ordem.<sup>6</sup>

A matriz  $W$ , com elementos 0 ou 1, é conhecida como matriz de vizinhança não normalizada, em contraposição à matriz  $W^*$  normalizada. A matriz  $W^*$  normalizada é construída a partir da matriz  $W$  original (não normalizada), dividindo-se todos os elementos de cada linha de  $W$  pela soma da linha. Portanto, a matriz  $W^*$  possui todas as linhas com soma igual a 1. Por sua vez, a matriz  $W$  original é simétrica, o que não vale para a matriz  $W^*$ . O vetor  $y_W = Wy$  é conhecido como *lag* espacial. No caso de se utilizar a matriz de contiguidade normalizada, o vetor  $y_W = W^*y$  corresponde a um vetor de médias simples das observações para a variável  $y$  dos vizinhos. A partir de agora, a matriz de contiguidade será referida simplesmente como  $W$ , independentemente de ser uma matriz normalizada ou não normalizada.

Apesar da aparente arbitrariedade na escolha da matriz  $W$  (Pinkse e Slade, 2010) alguns autores propõem abordagens que auxiliam na escolha da matriz de contiguidade. Baumont (2004) propõem que a matriz de vizinhança seja construída por meio do método dos  $k$  vizinhos mais próximos, a qual é definida da seguinte forma:

$$\begin{cases} w_{ij}(k) = 0, se & i = k, \forall k \\ w_{ij}(k) = 1, se & d_{ij} \leq d_i(k), \\ w_{ij}(k) = 0, se & d_{ij} > d_i(k) \end{cases}$$

onde  $w_{ij}(k)$  é um elemento da matriz normalizada,  $d_{ij}$  é uma distância de corte definida para cada unidade  $i$ , ou seja,  $d_{ij}$  é a menor distância entre a região  $i$  e todas as outras unidades de modo que cada unidade  $i$  tem exatamente  $k$  vizinhos. Para escolher o número de vizinhos mais próximos, Baumont (2004) sugere o seguinte procedimento:

1. Estime o modelo de interesse.

---

<sup>6</sup> Ver Lesage e Pace (2009).

2. Em seguida teste a presença de autocorrelação espacial por meio do  $I$  de Moran para diversas matrizes contendo  $k$  vizinhos mais próximos onde  $k$  deve variar, por exemplo, de 1 a 20.
3. Finalmente, escolha-se a matriz de vizinhança que produza o maior valor do índice  $I$  de Moran.

Similarmente, outras abordagens podem ser utilizadas para a escolha da matriz  $W$ , como por exemplo, escolher a matriz de contiguidade que produza a maior log-verossimilhança ou menor critério de informação como o AIC de Akaike ou o BIC de Schwarz.

O modelo paramétrico em (1) contém, como parâmetros desconhecidos, o coeficiente  $\rho$  e a variância  $\sigma^2$ . A estimação do parâmetro  $\rho$  permite, por exemplo, inferir o grau de correlação espacial entre as observações  $y_i$ . Além disso, testando-se a significância do parâmetro  $\rho$ , tem-se um procedimento para inferir a presença ou não de dependência espacial entre as observações. A seguir, se discutirá o processo de inferência dos parâmetros do modelo em (1).

Uma das primeiras sugestões para a estimação do coeficiente  $\rho$  é a utilização do estimador de mínimos quadrados ordinários. No entanto, quando o vetor de covariáveis (variáveis do lado direito da equação) é correlacionado com o resíduo da regressão, sabe-se que o estimador de mínimos quadrados ordinários é inconsistente. Esta correlação entre os resíduos e o regressor é observada no modelo em (1).<sup>7</sup> Portanto, estimação via mínimos quadrados ordinários resultaria em uma estimativa inconsistente para o coeficiente  $\rho$ .

Para entendermos melhor a existência da correlação entre os resíduos da regressão e o regressor  $Wy$ , note que podemos reescrever o modelo em (1) como  $y = (I_n - \rho W)^{-1} \epsilon$ , onde  $I_n$  é uma matriz identidade com dimensão  $n$ . Sob certas condições de regularidade (restrições sobre  $\rho$  e sobre os autovalores da matriz  $W$ ), podemos expandir o termo  $(I_n - \rho W)^{-1}$  em uma série infinita da forma:  $(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 + \dots$

Portanto,  $Wy = W\epsilon + \rho W^2\epsilon + \rho^2 W^3\epsilon + \rho^3 W^4\epsilon + \dots$ , o que implica que cada regressor  $Wy_i$  é função também do resíduo  $\epsilon_i$ . Ressaltamos que, apesar de a diagonal principal de  $W$  ser nula, as diagonais principais das matrizes  $W^2$ ,  $W^3$ ,  $W^4$ , ..., podem possuir valores não nulos, o que explica a presença de correlação entre o regressor  $Wy_i$  e o resíduo  $\epsilon_i$ .

Como alternativa, o analista pode utilizar estimação via máxima verossimilhança, que não sofre do problema de inconsistência do estimador de mínimos quadrados ordinários, devido à endogeneidade do regressor  $Wy$ . Em linhas gerais, a estimação via máxima verossimilhança dos parâmetros  $\rho$  e  $\sigma^2$  parte da distribuição normal multivariada para o vetor de resíduos  $\epsilon$ . A partir de (1), pode-se escrever

$$y = (I - \rho W)^{-1} \epsilon, \quad (2)$$

onde  $I$  é uma matriz identidade com dimensão  $n$ . Dado que  $\epsilon$  possui distribuição normal multivariada, com média nula e covariância  $\sigma^2 I$ , então o vetor observado  $y$  possui distribuição normal multivariada com média nula e covariância  $\Sigma_y = \sigma^2 (I - \rho W)^{-1} [I - \rho W - 1] T$ . A partir desta matriz de covariância, pode-se escrever a função de log-verossimilhança  $l(\rho, \sigma^2) = \log L(\rho, \sigma^2)$ . Maximizando-se  $\log L(\rho, \sigma^2)$ , obtêm-se os estimadores de máxima verossimilhança dos parâmetros do modelo.

<sup>7</sup> Ver Anselin (1988) e Lesage e Pace (2009).

Uma das dificuldades na estimação de modelos SAR (mesmo no caso mais simples, no qual não há covariáveis exógenas) é a necessidade de se realizarem operações com matrizes de grandes dimensões. No processo iterativo para obtenção do máximo da função  $\log L(\rho, \sigma^2)$ , é preciso calcular o logaritmo do determinante da matriz  $(I - \rho W)$ , que possui dimensão  $n$ . Se o analista estiver fazendo uma aplicação com observações de setores censitários da cidade de São Paulo, por exemplo, o valor de  $n$  está em torno de 18 mil; portanto, a matriz  $(I - \rho W)$  possui dimensão 18 mil por 18 mil. Felizmente, pela própria definição da matriz de contiguidade  $W$ , pode-se tratá-la como matriz esparsa; ou seja, a grande maioria dos elementos de  $W$  são nulos. Para matrizes esparsas, existe uma literatura bem desenvolvida sobre algoritmos que tornam o processo computacional mais eficiente.<sup>8</sup> Portanto, apesar de a codificação do estimador de máxima verossimilhança não ser trivial (é preciso programar algumas rotinas para matrizes esparsas), o esforço computacional pode ser bastante reduzido.

Uma vez dentro do arcabouço de estimação via máxima verossimilhança, pode-se recorrer a vários dos resultados para este tipo de estimador. Pode-se, então, testar a significância do parâmetro  $\rho$ , utilizando-se o teste de Wald, o teste da razão de verossimilhança ou o teste dos multiplicadores de Lagrange. Testando-se a significância do parâmetro  $\rho$ , se está implicitamente testando a presença de dependência espacial das observações para a variável  $y_i$ .

O modelo SAR em (1) pode ser estendido, para incorporar variáveis exógenas no lado direito da equação, obtendo-se:

$$y = \rho W y + X \beta + \epsilon, \quad (3)$$

onde a matriz  $X$  é uma matriz contendo as observações das variáveis exógenas. A dimensão de  $X$  é  $n \times k$ , sendo  $k$  o número de regressores. Cada linha da matriz  $X$  corresponde a uma observação na base de dados (um polígono, em um sistema georreferenciado). No caso de a regressão incluir um intercepto, a primeira coluna da matriz  $X$  possui apenas valores 1. O vetor  $\beta$  é um vetor coluna de coeficientes para as variáveis exógenas, e possui dimensão  $k \times 1$ . O modelo em (3) é conhecido como modelo SAR misto.

Da mesma forma que no SAR simples (Equação (1)), a estimação dos parâmetros no modelo SAR misto via mínimos quadrados ordinários também produz estimativas inconsistentes, uma vez que o vetor de *lags* espaciais  $W y$  é correlacionado com o vetor de resíduos  $\epsilon$ . Novamente, pode-se utilizar máxima verossimilhança, a partir da hipótese de que o vetor de resíduos  $\epsilon$  possui distribuição normal multivariada com média nula e covariância  $\sigma^2 I$ . Pode-se então escrever

$$y = (I - \rho W)^{-1} X \beta + (I - \rho W)^{-1} \epsilon, \quad (4)$$

e o vetor de variáveis observadas  $y$  possui distribuição (condicional a  $X$ ) normal multivariada, com média condicional

$$E[y|X] = (I - \rho W)^{-1} X \beta, \quad (5)$$

e matriz de variância condicional

---

<sup>8</sup> Ver Davis (2006).

$$\Sigma_{y|x} = \sigma^2(I - \rho W)^{-1}[(I - \rho W)^{-1}]^T. \quad (6)$$

A partir da distribuição de  $y$ , obtém-se a função de log-verossimilhança condicional  $\log L(\rho, \beta, \sigma^2)$ . Maximizando-se a função de log-verossimilhança em relação aos parâmetros do modelo, encontram-se as estimativas para os coeficientes e para a variância dos resíduos. Para uma discussão sobre o processo iterativo para estimação dos parâmetros do modelo SAR misto, podem-se consultar Anselin (1988) e Lesage e Pace (2009).

## 2.2 Modelos SEM

Da mesma forma que os modelos SAR partem da especificação de modelos AR para séries temporais, outra classe de modelos espaciais parte da especificação de modelos MA (médias móveis) para observações no tempo. Estes modelos espaciais são denominados modelos de erros espaciais (*spatial error models*), ou simplesmente SEM. Os modelos SEM possuem a seguinte especificação:

$$y = X\beta + u. \quad (7)$$

No caso, os resíduos da equação observada possuem uma estrutura autorregressiva, da forma

$$u = \lambda Wu + \epsilon. \quad (8)$$

O vetor de resíduos  $\epsilon$  possui distribuição normal multivariada, com média nula e matriz de covariância  $\sigma^2 I$ . O coeficiente escalar  $\lambda$  indica a intensidade da autocorrelação espacial entre os resíduos da equação observada. Mais especificamente, esse parâmetro mensura o efeito médio dos erros dos vizinhos em relação ao resíduo da região em questão. Note-se que, ao contrário dos modelos SAR, os modelos SEM não apresentam a variável resposta como uma função direta dos seus *lags* espaciais. A autocorrelação espacial nos modelos SEM aparece nos termos de erro.

Outra diferença dos modelos SEM em relação aos modelos SAR é que os coeficientes no vetor  $\beta$  podem ser estimados consistentemente via mínimos quadrados ordinários. De fato, a regressão em (7) pode ser vista como uma regressão linear com resíduos correlacionados. O estimador de mínimos quadrados ordinários produz estimativas consistentes, mas a matriz de covariância das estimativas  $\hat{\beta}_{ols}$  não será mais  $\sigma^2[X'X]^{-1}$ . Devido aos erros correlacionados, a matriz de covariância de  $\hat{\beta}_{ols}$  é dada por<sup>9</sup>

$$Var[\hat{\beta}_{ols}] = [X'X][X'\Omega^{-1}X]^{-1}[X'X], \quad (9)$$

onde  $\Omega = Var[u] = \sigma^2(I - \lambda W)^{-1}[(I - \lambda W)^{-1}]^T$ . Note-se que a matriz  $\Omega$  depende do coeficiente  $\lambda$  e da variância  $\sigma^2$ . A estimativa destes dois parâmetros pode ser obtida consistentemente a partir da estimação de um modelo SAR via máxima verossimilhança, conforme discutido no item anterior, para os resíduos  $\hat{u} = y - X\hat{\beta}_{ols}$ . Uma vez estimados os escalares  $\lambda$  e  $\sigma^2$ , pode-se obter uma estimativa para a matriz de covariância de  $\hat{\beta}_{ols}$

<sup>9</sup> Ao longo deste texto, a expressão da forma  $A^T$  denotará o transposto do elemento em  $A$ , onde  $A$  é uma matriz, um vetor coluna, um vetor linha, ou mesmo um escalar.



$$Var[\widehat{\beta}_{ols}] = [X'X] [X'\widehat{\Omega}^{-1}X]^{-1} [X'X], \quad (10)$$

onde  $\widehat{\Omega} = \hat{\sigma}^2(I - \hat{\lambda}W)^{-1}[(I - \hat{\lambda}W)^{-1}]^T$ .

Sabe-se que, no caso de modelos lineares com regressores exógenos (o que é o caso nos modelos SEM), com resíduos correlacionados, o estimador de mínimos quadrados ordinários é consistente, mas não é eficiente, havendo outros estimadores lineares que produzem variâncias menores.<sup>10</sup> Especificamente para o modelo SEM, o estimador linear com variância mínima é o estimador de mínimos quadrados generalizados (*generalized least squares* – GLS), dado por

$$\hat{\beta}_{gls} = [X'\Omega^{-1}X]^{-1} [X'\Omega^{-1}y]. \quad (11)$$

Na prática, não se conhece a matriz  $\Omega$ , uma vez que esta depende dos parâmetros desconhecidos  $\lambda$  e  $\sigma^2$ . Utiliza-se então o estimador de mínimos quadrados generalizados exequíveis (*feasible generalized least squares* – FGLS), com expressão

$$\hat{\beta}_{fgls} = [X'\widehat{\Omega}^{-1}X]^{-1} [X'\widehat{\Omega}^{-1}y], \quad (12)$$

onde  $\widehat{\Omega} = \hat{\sigma}^2(I - \hat{\lambda}W)^{-1}[(I - \hat{\lambda}W)^{-1}]^T$ , com  $\hat{\sigma}^2$  e  $\hat{\lambda}$  estimativas via máxima verossimilhança do modelo SAR simples, a partir dos resíduos  $\hat{u} = y - X\hat{\beta}_{ols}$ . Portanto, uma alternativa para a estimação dos parâmetros do modelo SEM é dada pelos passos:

- i) Obter a estimativa de mínimos quadrados ordinários  $\hat{\beta}_{ols} = [X'X]^{-1}[X'y]$ .
- ii) Calcular os resíduos  $\hat{u} = y - X\hat{\beta}_{ols}$ .
- iii) Estimar os parâmetros  $\lambda$  e  $\sigma^2$ , via máxima verossimilhança, para o modelo SAR em  $\hat{u}$ ,  $\hat{u} = \lambda W\hat{u} + \epsilon$ .
- iv) Calcular a estimativa  $\widehat{\Omega} = \hat{\sigma}^2(I - \hat{\lambda}W)^{-1}[(I - \hat{\lambda}W)^{-1}]^T$ .
- v) Obter a estimativa  $\hat{\beta}_{fgls} = [X'\widehat{\Omega}^{-1}X]^{-1} [X'\widehat{\Omega}^{-1}y]$ .
- vi) Obter a estimativa para a covariância  $\hat{\beta}_{fgls}$ ,  $Var[\widehat{\beta}_{fgls}] = [X'\widehat{\Omega}^{-1}X]^{-1}$ .

Inferência para os coeficientes em  $\beta$  pode ser efetuada a partir da matriz  $[X'\widehat{\Omega}^{-1}X]^{-1}$ . Note-se que a estimativa final para o vetor  $\beta$  não precisa parar no passo (v) acima. De fato, uma vez obtida uma estimativa  $\hat{\beta}_{fgls}$ , pode-se obter um novo vetor  $\hat{u} = y - X\hat{\beta}_{fgls}$ . Para este novo vetor  $\hat{u}$ , estimam-se novamente os parâmetros  $\lambda$  e  $\sigma^2$ , repetindo-se em seguida os passos (iv) e (v). Este processo pode ser efetuado repetidamente até que os valores no vetor  $\hat{\beta}_{fgls}$  atinjam a convergência. Finalizam-se então as estimações com o passo (vi).

Além das estimativas via mínimos quadrados ordinários (com correção da matriz de covariância das estimativas dos coeficientes) e das estimativas via mínimos quadrados generalizados exequíveis (FGLS), a literatura apresenta uma discussão sobre estimação

<sup>10</sup>Quando os autores se referem a *variâncias menores*, na verdade referem-se ao fato de que a diferença  $Var[\hat{\beta}_{ols}] - Var[\hat{\beta}]$  é uma matriz positiva definida, onde  $\hat{\beta}$  é um estimador linear mais eficiente do que o estimador de mínimos quadrados ordinários.

dos parâmetros do modelo SEM via máxima verossimilhança. Combinando as expressões (7) e (8), obtém-se

$$y = X\beta + (I - \lambda W)^{-1}\epsilon, \quad (13)$$

onde  $\epsilon$  possui distribuição normal multivariada com média nula e covariância  $\sigma^2 I$ . Portanto, o vetor de variável resposta  $y$  possui distribuição normal multivariada com média condicional

$$E[y|X] = X\beta, \quad (14)$$

e matriz de variância condicional

$$\Sigma_{y|X} = \sigma^2(I - \lambda W)^{-1}[(I - \lambda W)^{-1}]^T. \quad (15)$$

A partir da distribuição de  $y$ , obtém-se a função de log-verossimilhança condicional  $\log L(\lambda, \beta, \sigma^2)$ . Maximizando-se a função de log-verossimilhança em relação aos parâmetros do modelo, encontram-se as estimativas para os coeficientes e para a variância dos resíduos. Para uma discussão sobre o processo iterativo para estimação dos parâmetros do modelo SEM, consultem-se Anselin (1988) e Lesage e Pace (2009). Similarmente às estimações no caso de modelos SAR, a estimação de modelos SEM também envolvem operações com matrizes esparsas. Novamente, utilizando-se rotinas mais eficientes para matrizes esparsas, o esforço computacional pode ser bem menor. Lesage e Pace (2009) apresentam uma extensão dos modelos SAR e SEM denominados respectivamente de modelos de *Durbin* espacial e modelo de *Durbin* do erro espacial, nos quais, além da matriz de delineamento  $X$ , há também uma matriz de variáveis explicativas defasadas espacialmente. Assim, os modelos de *Durbin* espacial e de *Durbin* do erro espacial são representados respectivamente por  $y = \rho W y + X\beta + WX\theta + \epsilon$  e  $y = X\beta + WX\theta + u$ , onde  $u = \lambda W u + \epsilon$ .

### 2.3 Modelos SARMA

Finalmente, os modelos SEM e SAR podem ser combinados em uma especificação mais geral, seguindo a ideia nos modelos ARMA (*autorregressive and moving average*) para séries temporais. Os modelos SARMA (*spatial autorregressive and moving average*) têm uma especificação da forma:

$$y = \rho W_1 y + X\beta + u, \quad (16)$$

na qual os resíduos da equação observada possuem uma estrutura autorregressiva, da forma:

$$u = \lambda W_2 u + \epsilon. \quad (17)$$

As matrizes  $W_1$  e  $W_2$  são matrizes de contiguidade não necessariamente iguais. De fato, quando  $W_1 = W_2$ , o modelo é não identificado, e as estimativas para os coeficientes  $\lambda$  e  $\rho$  podem resultar bastante instáveis,<sup>11</sup> a menos que a matriz de delineamento  $X$  contenha pelo menos uma variável exógena além do intercepto. Uma das críticas em relação à utilização dos modelos SARMA é justamente o fato de eles exigirem, em alguns casos, a

<sup>11</sup> Ver Anselin (1988), e Lesage e Pace (2009).

especificação de duas matrizes de contiguidade diferentes. Em geral, a escolha de uma matriz de contiguidade é arbitrária; a escolha de duas matrizes diferentes implica em um grau de arbitrariedade ainda mais criticável.

Estimação dos parâmetros do modelo SARMA pode ser feita via máxima verossimilhança. A partir das expressões (16) e (17), pode-se escrever

$$\begin{aligned}(I - \rho W_1)y &= X\beta + (I - \lambda W_2)^{-1}\epsilon \\ \Rightarrow y &= (I - \rho W_1)^{-1}X\beta + (I - \rho W_1)^{-1}(I - \lambda W_2)^{-1}\epsilon.\end{aligned}$$

Assumindo-se que  $\epsilon$  possui distribuição normal multivariada, com média zero e covariância  $\sigma^2 I$ , conclui-se que o vetor de observações para a variável resposta  $y$  possui distribuição normal multivariada com média condicional

$$E[y|X] = (I - \rho W_1)^{-1}X\beta, \quad (18)$$

e matriz de variância condicional

$$\Sigma_{y|X} = \sigma^2(I - \rho W_1)^{-1}(I - \lambda W_2)^{-1}[(I - \rho W_1)^{-1}(I - \lambda W_2)^{-1}]^T. \quad (19)$$

Utilizando-se a fórmula para a distribuição normal multivariada, pode-se chegar à função de log-verossimilhança  $\log L(\lambda, \rho, \beta, \sigma^2)$ , como função dos parâmetros desconhecidos do modelo. Similarmente aos modelos SAR e SEM, as estimativas de máxima verossimilhança não possuem fórmula fechada, necessitando de um processo iterativo para maximização da função  $\log L(\lambda, \rho, \beta, \sigma^2)$ . Uma discussão sobre os passos no processo iterativo para estimação dos parâmetros no modelo SARMA pode ser encontrada em Anselin (1988) e Lesage e Pace (2009).

### 3 Críticas aos modelos de dependência espacial

Apesar do seu uso bastante disseminado, os modelos paramétricos para tratamento de dependência espacial (exemplos: SAR, SEM e SARMA) vêm recebendo várias críticas na literatura. Estas críticas não necessariamente retiram destes modelos quaisquer utilidades em pesquisas empíricas. No entanto, alguns dos pontos levantados pelos críticos são importantes para: *i*) antecipar aos usuários alguns cuidados e limitações acerca dos quais eles devem estar cientes; *ii*) fornecer um certo balizamento para pesquisas futuras para os modelos espaciais, de maneira a corrigir/amenizar algumas das limitações. Nesta seção, será feita uma discussão sobre algumas das críticas aos modelos apresentados na seção 3 (e seus equivalentes para dados de painel). Estas críticas se aplicam mais fortemente ao problema de especificação paramétrica (ou não) para capturar corretamente a dependência espacial. No caso de testes de hipótese para a presença ou não de dependência espacial, os testes atualmente disponíveis (conforme seção 4) se comportam de forma bastante satisfatória em situações regulares. Nos casos mais irregulares, utilizam-se frequentemente os testes LM Robustos. Maiores detalhes podem ser encontrados, por exemplo, em Pinkse e Slade (2010).

De maneira geral, o embasamento teórico para a modelagem em econometria espacial ainda se encontra em um estágio inicial. Dessa forma, uma das dificuldades é encontrar um modelo que se adeque a todos os tipos de situação. Nesse sentido, alguns autores defendem que os pesquisadores se concentrem no desenvolvimento de teorias

específicas para classes particulares de aplicações, ao invés de seguirem na busca de extensões para técnicas já existentes.

Entre as limitações para os modelos de SAR e outros modelos da forma ARMA espaciais (incluindo extensões para dados de painel), podem-se citar os itens a seguir:

i) Hipótese improvável e desnecessária de normalidade dos resíduos.

ii) O fato de  $y_i$  depender dos seus próprios *lags* espaciais pode implicar que  $y_i$  também dependa dos *lags* espaciais do vetor de covariáveis  $x_i$ , incorrendo no problema de reflexão (*reflexion problem*), apontado por Manski (1993). A consequência prática é que a inclusão de *lags* espaciais de  $x_i$  pode ocasionar uma matriz de design  $X$  com altíssimo grau de multicolinearidade.

iii) Os modelos SAR e demais modelos ARMA assumem linearidade nos parâmetros  $y_i$ . Este fato nem sempre é verdade na prática, e pode haver a necessidade de especificações não lineares da relação entre o vetor de regressões  $x_i$  e a variável  $y_i$  por meio dos parâmetros.

iv) Os modelos SAR e correlatos não levam em consideração a presença de dependência entre o vetor de regressores  $x_i$  e os resíduos  $u_i$ , causada pela presença de regressores endógenos em  $x_i$  e/ou pela presença de heteroscedasticidade condicionada aos regressores. Entretanto, outras propostas como o estimador de Kelejian e Prucha e o estimador HAC, ambos apresentados posteriormente, visam a corrigir esses problemas.

v) Há fortes críticas à representação excessivamente simplista de toda a dependência espacial em um único coeficiente  $\rho$ .

vi) A matriz de contiguidade  $W$  implica um alto grau de arbitrariedade na sua especificação, principalmente levando-se em consideração a irregularidade dos mapas de municípios e de setores censitários, por exemplo.

De maneira geral, os modelos SAR e correlatos foram inicialmente propostos como possíveis extensões dos modelos para dependência em séries temporais. No entanto, há uma série de críticas à analogia dos procedimentos para dependência espacial com os procedimentos para dependência temporal. Algumas destas críticas estão listadas a seguir:

a) Os dados não são igualmente espaçados.

b) A presença de observações ausentes (*missing values*) pode incorrer na presença de endogeneidade, ocasionando vieses nos estimadores de máxima verossimilhança.

c) Observações espaciais, em muitos casos, são agregações de observações (por polígono, por exemplo) do comportamento de vários agentes. Portanto, modelos baseados no comportamento de agentes individuais podem não ser mais válidos.

d) Nos modelos para séries temporais, os procedimentos são teoricamente validados a partir de proposições sobre o comportamento assintótico dos estimadores, quando o número de observações  $T$  (intervalo total da série histórica) assume valores cada vez maiores ( $T \rightarrow \infty$ ). Para modelos para dados espaciais, não é claro se a expansão assintótica ocorre com o aumento da densidade de observações dentro do mapa (*infill asymptotics*), ocorre com o aumento das fronteiras (*increasing domain asymptotics*), ou ocorre com as suas expansões simultaneamente.

e) O item anterior é particularmente importante, porque não há garantia de que as relações de dependência espacial se alteram quando mais observações são adicionadas aos dados. Por exemplo, no caso de *infill asymptotics*, a adição de novas observações pode ocasionar um aumento da dependência espacial, uma vez que as observações estarão cada mais próximas em média.

f) Diferentemente dos modelos para séries temporais, a estimação dos modelos com dados espaciais pode sofrer do grave problema de endogeneidade das decisões locais das unidades observadas na amostra. Uma consequência da endogeneidade das localizações é que as distâncias entre os agentes, bem como as estruturas de vizinhança, também são endógenas. Este problema tem se mostrado de difícil solução até o momento, e vem sendo desprezado na maioria das aplicações.

Diversos artigos recentes têm focalizado alguns dos problemas discutidos anteriormente. Para adicionar maior flexibilidade à modelagem da vizinhança, por exemplo, algumas extensões do modelo SAR tradicional consistem em substituir a matriz de contiguidade  $W$  por uma expansão de funções base, da forma

$$y = \left[ \sum_{k=0}^{\infty} \rho_k W_k y \right] + X\beta + u. \quad (20)$$

Na prática, é necessário truncar o número de elementos no somatório da expressão (20), até um número  $K(n)$ . Como é típico em estimações com expansões de funções base, faz-se  $K(n)$  aumentar para o infinito, quando o tamanho  $n$  da amostra aumenta. Neste caso, a expressão torna-se

$$y = \left[ \sum_{k=0}^{K(n)} \rho_k W_k y \right] + X\beta + u, \quad (21)$$

e o problema de rigidez em relação à forma funcional da dependência espacial pode ser amenizado (para maiores detalhes, ver Pinkse *et al.*, 2002; Pinkse e Slade, 2004; e Pofahl, 2007).

Boa parte dos problemas de endogeneidade pode ser tratada com a utilização de variáveis instrumentais apropriadas, conforme discutido nas seções 5 e 6. Para o problema de observações ausentes (*missing data*), no qual o processo gerador das observações ausentes é exógeno, podem-se utilizar procedimentos de mínimos quadrados de dois estágios (Lee, 2007). Para situações nas quais a geração das observações ausentes é endógena, não há solução conhecida na literatura. De maneira geral, ainda existe um grande caminho a ser trilhado em termos de procedimentos e tratamentos teóricos, para lidar com os problemas nos modelos para dados espaciais.

#### 4 Testes para dependência espacial

Na seção anterior, foram discutidos alguns modelos mais comumente utilizados para contabilizar para a presença de dependência espacial nos resíduos (ou na própria variável resposta) do modelo de regressão. Nesta seção, será apresentada uma discussão sobre testes para dependência espacial. De maneira geral, os modelos paramétricos apresentados na seção 2 têm sofrido diversas críticas, conforme será visto na seção 4. Por seu turno, os testes para a presença de dependência espacial não sofrem o mesmo ataque, e são relativamente bem aceitos na literatura.

#### 4.1 Estatística de Moran

Uma das estatísticas para testes de dependência espacial mais disseminada é a estatística  $I$  de Moran. Esta estatística pode ser aplicada à variável  $y_i$  diretamente, ou aos resíduos da regressão de  $y_i$  versus um conjunto de variáveis explicativas. Considere-se então um modelo de regressão linear, da forma

$$y = X\beta + u, \quad (22)$$

onde  $y$  é um vetor coluna ( $n \times 1$ ) de variáveis resposta,  $X$  é uma matriz com cada linha contendo as observações para as variáveis explicativas, além de uma coluna unitária associada ao intercepto do modelo,  $\beta$  é um vetor de coeficientes e  $u$  é um vetor coluna contendo os resíduos da regressão. A partir da estimativa de mínimos quadrados ordinários para o vetor de coeficientes, obtém-se a seguinte expressão para os resíduos

$$\hat{u} = y - X[X'X]^{-1}[X'y]. \quad (23)$$

A estatística  $I$  de Moran para a autocorrelação espacial pode ser aplicada nos resíduos do modelo de regressão de maneira direta. Formalmente, a estatística  $I$  é dada por

$$I = \frac{n}{s} \left[ \frac{\hat{u}'W\hat{u}}{\hat{u}'\hat{u}} \right], \quad (24)$$

onde  $\hat{u}$  é o vetor de resíduos da regressão por mínimos quadrados ordinários,  $W$  é a matriz de contiguidade espacial,  $n$  é o número de observações da amostra e  $s$  é um fator de padronização igual à soma de todos os elementos da matriz  $W$ . A partir da estatística  $I$ , pode-se construir um teste para a hipótese nula de presença de independência espacial. Por sua vez, a especificação da hipótese alternativa não é tão simples.

A distribuição assintótica para a estatística  $I$  foi derivada por Cliff e Ord (1972). Dessa forma, considere-se

$$z_I = \frac{I - E(I)}{\sqrt{V(I)}}, \quad (25)$$

onde  $E(I)$  e  $\sqrt{V(I)}$  são, respectivamente, a média e a variância assintótica da estatística  $I$  de Moran. Sob a hipótese nula, a distribuição da estatística  $z_I$  pode ser estimada via simulações de Monte Carlo. Quando a estatística  $I$  é construída a partir dos resíduos  $\hat{u}$ , a rejeição da hipótese nula implica em evidências de que há autocorrelação espacial no modelo de regressão. Esse teste é afetado pela ausência de normalidade e pela presença de heterocedasticidade, o que pode invalidar as conclusões inferenciais resultantes das estimações. Entretanto, a utilização do método de Monte Carlo (via *bootstrap*) e a utilização da transformação de Box-Cox na variável dependente podem auxiliar na correção desses problemas. A partir daí, o analista pode recorrer a um dos modelos paramétricos discutidos na seção 2, na seção 4 ou na seção 5.

#### 4.2 Teste de Kelejian-Robinson

Kelejian e Robinson (1992) propuseram um teste com o mesmo objetivo do teste  $I$  de Moran. No entanto, diferentemente do teste  $I$  de Moran, o teste de Kelejian-Robinson

não pressupõe normalidade da variável sendo testada (a variável observada  $y_i$  ou os resíduos  $\hat{u}_i$  da regressão). Portanto, o teste de Kelejian-Robinson é mais robusto à não normalidade dos resíduos ou da variável observada, sendo mais apropriado quando a hipótese similaridade ao padrão gaussiano seja questionável.

O teste de Kelejian-Robinson tem como pressuposto inicial

$$\text{Cov}(\epsilon_i, \epsilon_j) = \sigma_{ij} = Z_{ij}\alpha, \quad (26)$$

onde  $Z_{ij}$  é um vetor  $1 \times q$  de covariáveis, tipicamente tomadas como funções das variáveis explicativas originais para  $i$  e  $j$ , com  $i$  e  $j$  sendo localidades "contíguas" em um espaço geral de observações ordenadas. Por exemplo,  $Z_{ij}$  pode ser construído a partir de produtos cruzados dos elementos de  $X_i$  e  $X_j$ . O vetor  $Z_{ij}$  não necessariamente possui a mesma dimensão de  $X_i$  (ou  $X_j$ ). O elemento  $\alpha$  é um vetor  $q \times 1$  de parâmetros, indicando o quanto os componentes de  $Z_{ij}$  podem explicar a covariância entre os resíduos. Intuitivamente, a ausência de autocorrelação espacial poderá não produzir relações significativas entre  $\text{Cov}(\epsilon_i, \epsilon_j)$  e  $Z_{ij}$ , resultando em estimativas não significantes para os coeficientes no vetor  $\alpha$ .

A hipótese nula é então construída como  $H_0: \alpha = 0$  em (24). Dada uma amostra de tamanho  $n$ , seja  $c$  um vetor de dimensões  $h_n \times 1$ , contendo as covariâncias  $\sigma_{ij}$ 's não nulas<sup>12</sup> (por construção) para todo  $i < j$ . O teste é implementado regredindo-se os  $h_n$  produtos cruzados  $\hat{c}_{ij} = \hat{u}_i \hat{u}_j$  dos resíduos versus os vetores  $Z_{ij}$ , para todo  $i < j$ , com  $i$  e  $j$  polígonos vizinhos. Seja então a matriz  $Z$ , com dimensão  $h_n \times q$ , construída a partir do empilhamento dos vetores linha  $Z_{ij}$ , e seja  $\hat{c}$  um vetor coluna, com dimensão  $h_n \times 1$ , construído a partir do empilhamento dos valores de  $\hat{c}_{ij} = \hat{u}_i \hat{u}_j$ . Uma estimativa para  $\alpha$  pode ser obtida via mínimos quadrados ordinários, resultando em

$$\hat{\alpha} = (Z'Z)^{-1}Z'\hat{c}.$$

A partir da estimativa  $\hat{\alpha}$ , pode-se construir a estatística teste de Kelejian-Robinson, dada pela expressão

$$KR = \frac{\hat{\alpha}'Z'Z\hat{\alpha}}{\hat{\sigma}^4}, \quad (27)$$

onde  $\hat{\sigma}^4$  é um estimador consistente de  $\sigma^4$ , e  $\sigma^2$  é a variância para o resíduo da regressão de  $\hat{c}_{ij} = \hat{u}_i \hat{u}_j$  versus  $Z_{ij}$ . Uma estimativa para  $\sigma^4$  pode ser dada, por exemplo, por

$$\hat{\sigma}^4 = \left[ \frac{(\hat{c} - Z\hat{\alpha})'(\hat{c} - Z\hat{\alpha})}{h_n} \right]^2.$$

Sob a hipótese nula, temos que  $\frac{\hat{c}'\hat{c}}{h_n}$  converge em probabilidade para  $\sigma^2$ . Pode-se mostrar então que uma forma alternativa para a estatística teste é dada por

---

<sup>12</sup> Nesse caso, as covariâncias não nulas são aquelas para as quais os polígonos  $i$  e  $j$  são vizinhos, de acordo com a definição de vizinhança utilizada para a análise.

$$KR = h_n^2 \frac{\hat{c}'Z(Z'Z)^{-1}Z'\hat{c}}{[\hat{c}'\hat{c}]^2}. \quad (28)$$

Sob a hipótese nula de ausência de dependência espacial, a estatística KR possui distribuição assintótica qui-quadrada, com  $q$  graus de liberdade. Este teste, no entanto, é baseado em uma estrutura espacial arbitrária, a qual admite apenas contigüidade de primeira ordem na definição da iteração entre as unidades espaciais.

### 4.3 Testes assintóticos a partir de especificações paramétricas

Nas seções 3.1 e 3.2, foram discutidos dois procedimentos de testes estatísticos para presença de dependência espacial, os quais não dependem de uma especificação paramétrica para a forma de autocorrelação no espaço. Nesta seção, serão revisitados os modelos discutidos na seção 2, para se construírem outros procedimentos de testes, a partir de especificações paramétricas. De forma geral, os procedimentos aqui discutidos são obtidos a partir de três metodologias tradicionais, empregadas para testes de hipóteses em geral. Estas metodologias são:

- i) teste de Wald;
- ii) teste da razão de verossimilhança (*likelihood ratio – LR*);
- iii) teste dos multiplicadores de Lagrange (*Lagrange multipliers – LM*).

#### 4.3.1 Princípios gerais

Os testes de Wald, LR e LM são baseados nas propriedades dos estimadores de máxima verossimilhança.<sup>13</sup> Mais especificamente, estas propriedades seguem do pressuposto de normalidade assintótica dos estimadores. Formalmente, seja  $\theta$  um vetor de parâmetros e  $\hat{\theta}$  suas respectivas estimativas por máxima verossimilhança, satisfazendo a convergência em distribuição

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{L} N(0, [I_1(\theta_0)]^{-1}),$$

onde  $\theta_0$  é o valor real do parâmetro no modelo (assumindo um modelo corretamente especificado), o elemento  $I_1(\theta_0)$  é a matriz de informação de Fisher para uma observação, e  $n$  é o número de observações na amostra. Considere-se então que o conjunto de hipóteses, sobre os parâmetros do modelo a serem testadas, pode ser escrito da forma:

$$\begin{cases} H_0: g(\theta_0) = 0 \\ H_a: g(\theta_0) \neq 0 \end{cases}$$

onde  $g, g: \mathfrak{R}^k \mapsto \mathfrak{R}^q$ , é uma função linear ou não linear do vetor de parâmetros  $\theta \in \Theta \subset \mathfrak{R}^k$ . Considerem-se, por exemplo, os modelos SAR ou SEM, vistos na seção 2. Como casos especiais de testes de hipóteses para os modelos paramétricos, tem-se os testes individuais dos parâmetros de autocorrelação espacial:  $H_0: \rho = 0$  no modelo SAR, ou  $H_0: \lambda = 0$  no modelo SEM.

<sup>13</sup> O teste de Wald pode ser utilizado também em outros contextos, que não o de estimação via máxima verossimilhança.



Os testes de Wald, LR e LM são baseados nas distâncias das estimativas para o modelo irrestrito e as estimativas satisfazendo às restrições impostas pela hipótese nula. Por exemplo, se o vetor de parâmetros  $\theta$  é particionado em dois vetores distintos, da forma  $\theta' = [\theta'_1, \theta'_2]$ , e a hipótese nula pode ser escrita da forma  $H_0: \theta_1 = 0$ , a estimativa  $\hat{\theta}_r$  de  $\theta$  no modelo restrito consistirá das estimativas para  $\theta_2$  concatenada com todos os elementos de  $\theta_1$  iguais a zero. A estimativa irrestrita  $\hat{\theta}_u$  é a estimativa do vetor completo  $\theta$ . Os testes serão então baseados na medida da diferença entre as estimativas do modelo completo  $\hat{\theta}_u$  e o vetor restrito  $\hat{\theta}_r$ . Intuitivamente, se a distância entre os dois resultados é muito grande, a hipótese nula é rejeitada. Para a realização dos testes é necessário estimar:

- i) Wald: apenas o modelo completo (irrestrito);
- ii) RV: o modelo completo (irrestrito) e o modelo restrito (sob a hipótese nula); e
- iii) LM: apenas o modelo restrito (sob a hipótese nula).

A seguir se fará uma discussão um pouco mais detalhada dos três tipos de testes. Dadas certas condições de regularidade, e assumindo-se que a hipótese nula é verdadeira, as estatísticas testes comumente empregadas para os três procedimentos possuem distribuição assintótica qui-quadrada  $\chi^2_q$ , com número de graus de liberdade iguais a  $q$  (dimensão da função vetorial  $g(\cdot)$ ).

#### 4.3.2 Teste de Wald

O teste de Wald pode ser expresso na forma geral

$$W = g(\hat{\theta}_u)' [G\hat{\Sigma}G']^{-1} g(\hat{\theta}_u), \quad (29)$$

com  $g(\cdot)$  um vetor  $q \times 1$  das estimativas obtidas por máxima verossimilhança dos parâmetros irrestritos,  $G$  uma matriz de derivadas da função  $g(\theta)$  e  $\hat{\Sigma}$  uma estimativa consistente da matriz de variâncias e covariâncias do estimador do vetor de parâmetros  $\hat{\theta}_u$ .

Considere-se, por exemplo, o modelo espacial SARMA, com resíduos homocedásticos, com um parâmetro de autocorrelação igual a  $\rho$ , e suponha-se que há interesse em testar se este parâmetro é igual a zero. Para isso, pode-se escrever a hipótese nula como:

$$H_0: [1, 0'] [\rho, \beta', \lambda, \sigma^2]' = \rho = 0.$$

A derivada  $G = \frac{\partial}{\partial \rho} [1, 0'] [\rho, \beta', \lambda, \sigma^2]' = [1, 0']$ , e chega-se então a

$$W = \hat{\rho}' \{ [1, 0'] \hat{\Sigma} [1, 0'] \}^{-1} \hat{\rho} = \frac{\hat{\rho}^2}{\hat{\Sigma}_{11}} \xrightarrow{L} \chi^2_1,$$

onde  $\hat{\Sigma}_{11}$  é o primeiro elemento da diagonal principal da estimativa  $\hat{\Sigma}$ .

#### 4.3.3 Teste da razão de verossimilhança

Considere-se o modelo paramétrico indexado pelo parâmetro  $\theta \in \Omega$ . A partir de uma amostra de tamanho  $n$ , constrói-se a função de log-verossimilhança, como função de  $\theta$ . Seja  $l(\hat{\theta}_r)$  o valor da função de log-verossimilhança, computada no ponto  $\theta = \hat{\theta}_r$ , e seja

$l(\hat{\theta}_r)$  o valor da função de log-verossimilhança, computada no ponto  $\theta = \hat{\theta}_r$ . Conforme discutido anteriormente,  $\hat{\theta}_u$  é a estimativa irrestrita do parâmetro  $\theta$  tal que  $\hat{\theta}_u = \arg \max_{\theta \in \Omega} l(\theta)$ , e  $\hat{\theta}_r$  é a estimativa do parâmetro  $\theta$ , impondo-se a restrição correspondente à hipótese nula, de forma que  $g(\theta) = 0$ . Ou seja,

$$\hat{\theta}_r = \arg \max_{\theta \in \Omega, \text{ s.t. } g(\theta)=0} l(\theta).$$

A estatística do teste da razão de verossimilhança é dada por

$$LR = 2[l(\hat{\theta}_u) - l(\hat{\theta}_r)]. \quad (30)$$

Sob a hipótese nula, e assumindo certas condições de regularidade, tem-se  $LR \xrightarrow{L} \chi_q^2$ . Considerando-se novamente o modelo SARMA, pretende-se testar a hipótese nula  $H_0: \rho = 0$ . A função de log-verossimilhança do modelo irrestrito tem expressão

$$\begin{aligned} l(\rho, \sigma^2, \lambda, \beta) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) + \log|I - \lambda W_2| + \log|I - \rho W_1| \\ &= -\frac{1}{2\sigma^2} \{(I - \lambda W_2)[(I - \rho W_1)y - X\beta]\}^T \{(I - \lambda W_2)[(I - \rho W_1)y - X\beta]\}, \end{aligned}$$

enquanto a função de log-verossimilhança do modelo restrito é dada por

$$\begin{aligned} l(\sigma^2, \lambda, \beta) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) + \log|I - \lambda W_2| + \log|I| \\ &= -\frac{1}{2\sigma^2} \{(I - \lambda W_2)[y - X\beta]\}^T \{(I - \lambda W_2)[y - X\beta]\}. \end{aligned}$$

A estatística teste é dada por  $LR = l(\rho, \sigma^2, \lambda, \beta) - l(\sigma^2, \lambda, \beta)$ , e tem distribuição assintótica  $\chi_1^2$ .

#### 4.3.4 Teste dos multiplicadores de Lagrange

O teste dos multiplicadores de Lagrange, também conhecido como teste do escore, é baseado na abordagem de otimização, mais precisamente, nas condições de primeira ordem da função lagrangiana da função de log-verossimilhança

$$f(\theta) = l(\theta) + \eta g(\theta),$$

onde  $\eta$  é o vetor dos multiplicadores de Lagrange correspondendo às  $q$  restrições em  $g(\theta) = 0$ . A estatística deste é dada por

$$LM = g(\hat{\theta}_r)' [I(\hat{\theta}_r)]^{-1} g(\hat{\theta}_r),$$

onde  $g(\hat{\theta}_r)$  é o vetor escore do modelo restrito calculado sob a hipótese nula.  $[I(\hat{\theta}_r)]$  é a matriz de informação de Fisher calculada sob a hipótese nula. A estatística LM terá distribuição  $\chi_q^2$ .

### 4.3.5 Teste dos multiplicadores de Lagrange no modelo SEM

No caso do modelo de erros espaciais (SEM), os resíduos são modelados na forma  $u = \lambda Wu + \epsilon$ , e, para se testar a hipótese de ausência de autocorrelação espacial, o interesse reside em se testar a hipótese nula de que  $\lambda = 0$ . Das três abordagens de testes (Wald, razão de verossimilhança e multiplicadores de Lagrange), a mais conveniente é a abordagem dos multiplicadores de Lagrange, uma vez que ela requer apenas a estimação do modelo restrito. Neste caso, a partir da estimação dos coeficientes da regressão via mínimos quadrados ordinários, e das estimativas para os erros da regressão, dados por  $\hat{u} = y - X(X'X)^{-1}X'y$ , pode-se mostrar que a estatística teste tem expressão

$$LM = \frac{[\hat{u}'W\hat{u}]}{T\hat{\sigma}^4}, \quad (31)$$

onde  $T = \text{traço}[(W' + W)W]$ . Caso a matriz  $W$  seja simétrica (*i.e.*,  $W = W'$ ), obtém-se  $T = n - 1$ . Computacionalmente, os testes de Wald e da razão de verossimilhança são mais complexos, uma vez que é necessário o cálculo das estimativas de máxima verossimilhança sem a restrição sobre o parâmetro  $\lambda$ . A estatística teste em (29) converge assintoticamente para uma distribuição qui-quadrada com um grau de liberdade. Note-se que o teste dos multiplicadores de Lagrange constitui-se em um procedimento simples para se testar a hipótese de ausência de dependência espacial nos erros da regressão.

### 4.3.6 Testes LM e LR robustos

O teste dos multiplicadores de Lagrange (LM) apresentado anteriormente segue uma distribuição qui-quadrada, com um grau de liberdade, sob a hipótese nula de que o parâmetro de autocorrelação espacial é igual a zero. Entretanto, no caso de má especificação, a distribuição do teste será uma distribuição qui-quadrada não centrada, implicando assim na rejeição da hipótese nula mais frequentemente, do que especificado no nível do teste. Nesse contexto, surgem os testes LM e LR robustos sugeridos por Anselin *et al.* (1996) e Anselin e Bera (1998), apresentados a seguir. Considere o modelo SARMA apresentado em (16) e (17). O teste para  $H_0: \lambda = 0$ , na presença do parâmetro  $\rho$ , é dado por:

$$LM_{\lambda}^* = \frac{[\bar{d}_{\lambda} - T_{12}\hat{\sigma}^2\hat{D}^{-1}\bar{d}_{\rho}]^2}{T_{22} - (T_{12})^2\hat{\sigma}^2\hat{D}}. \quad (32)$$

Sobre  $H_0: \lambda = 0$  e  $\rho = \delta/\sqrt{n}$ , a estatística  $LM_{\lambda}^*$  converge para uma distribuição qui-quadrada com um grau de liberdade, onde  $T_{ij} = \text{traço}[W_i W_j + W_i^T W_j]$  com  $i = 1, 2$  e  $j = 1, 2$ .  $\hat{D} = (W_1 X \hat{\beta})^T M (W_1 X \hat{\beta}) + T_{11} \hat{\sigma}^2$ ,  $M = I - X(X^T X)^{-1} X^T$  e  $\bar{d}_{\lambda}$ ,  $\bar{d}_{\rho}$  são as estatísticas escore dos parâmetros  $\lambda$  e  $\rho$ , respectivamente, do modelo SARMA. Similarmente, o teste robusto para  $H_0: \rho = 0$  na presença do parâmetro  $\lambda$  é dado por:

$$LM_{\rho}^* = \frac{[\bar{d}_{\rho} - T_{12} T_{22}^{-1} \bar{d}_{\lambda}]^2}{\hat{\sigma}^2 \hat{D} - (T_{12})^2 T_{22}^{-1}}. \quad (33)$$

No entanto, Anselin e Bera (1998) realçam que há um preço a ser pago na robustificação do teste. Por exemplo, no caso em que  $\boldsymbol{\rho} = \mathbf{0}$ , e na presença do parâmetro de dependência do erro espacial, por meio de  $\boldsymbol{\lambda} = \boldsymbol{\tau}/\sqrt{\mathbf{n}}$ , sobre esse formato, os parâmetros de não centralidade dos testes  $\mathbf{LM}_\lambda^*$  e  $\mathbf{LM}_\rho^*$  são respectivamente  $\mathbf{T}_{22}\boldsymbol{\tau}^2/\sqrt{\mathbf{n}}$  e  $\boldsymbol{\tau}^2(\mathbf{T}_{22} - \mathbf{T}_{12}^2\boldsymbol{\sigma}^2\mathbf{D}^{-1})/\mathbf{n}$ . Desde que  $\boldsymbol{\tau}^2\mathbf{T}_{12}^2\boldsymbol{\sigma}^2\mathbf{D}^{-1}/\mathbf{n} \geq \mathbf{0}$  o poder assintótico de  $\mathbf{LM}_\lambda^*$  será menor do que  $\mathbf{LM}_\lambda$  quando não há o parâmetro de defasagem  $\boldsymbol{\rho}$ . Similarmente a mesma discussão se enquadra para o teste  $\mathbf{LM}_\rho^*$ . A idéia por trás dos testes robustos é ajustar a estatística teste LM levando em consideração a não centralidade da distribuição do teste, no caso de má especificação do modelo. Assim, uma forma direta para especificar qual modelo utilizar pode ser dado pelos seguintes passos:

1. Estimativa OLS - teste LM:
  - a. Ausência de significância:
    - i. Assuma o modelo OLS.
  - b. LM-erro significativo e LM-lag não significativo:
    - i. Assuma o modelo de erro espacial.
  - c. LM-lag significativo e erro-LM não significativo:
    - i. Assuma o modelo de defasagem espacial.
  - d. LM do erro e LM-lag significativos:
    - i. Utilizar testes robustos e selecionar aquele com maior significância como alternativa.

## 5 Estimação via mínimos quadrados de dois estágios

Os modelos apresentados na seção 2 tratam de situações nas quais não há variáveis explicativas endógenas no lado direito da equação, de forma que a estimação via máxima verossimilhança fornece estimativas consistentes para os parâmetros do modelo. No entanto, em muitas situações, principalmente quando se tem o objetivo de identificar relações de causalidade entre determinadas políticas, o problema de endogeneidade aparece nos modelos espaciais, surgindo a necessidade de se utilizarem abordagens que estendam, por exemplo, os estimadores de variáveis instrumentais para situações com dependência espacial. Kelejian e Prucha, em diversos artigos,<sup>14</sup> exploraram este problema, e propuseram o estimador espacial de mínimos quadrados de dois estágios (S2SLS).

Entre as características da abordagem de mínimos quadrados espaciais de dois estágios de Kelejian e Prucha, podem-se citar: *i)* visa à estimação de modelos de regressão linear, com um termo de *lag* espacial da variável resposta do lado direito da equação; *ii)* permite a estimação de modelos com regressores endógenos; *iii)* os coeficientes (inclusive o coeficiente do termo de *lag* espacial da variável resposta) são todos estimados por intermédio do procedimento de mínimos quadrados de dois estágios; *iv)* o coeficiente de *lag* espacial da variável resposta tem como instrumento, para resolver o problema de endogeneidade, os *lags* espaciais dos regressores exógenos; *v)* o procedimento permite a incorporação de correções para a presença de heteroscedasticidade e autocorrelação espacial residual nos termos de erro da regressão estimada.

<sup>14</sup> Ver Kelejian e Prucha (1997, 1998); Kelejian e Robinson (2002, 2007 e 2009); e Kelejian *et al.* (2004).

Para fazer a exposição de metodologia de mínimos quadrados espacial de dois estágios, considere-se a equação geral a seguir:

$$y = \rho W y + Y v + X \beta + u, \quad (34)$$

onde  $y$  é um vetor coluna contendo as  $n$  observações empilhadas para a variável resposta,  $\rho$  é o coeficiente do *lag* espacial da variável resposta,  $W$  é uma matriz de vizinhança,  $Y$  é uma matriz com regressores endógenos, o vetor  $v$  é um vetor de coeficientes dos regressores endógenos,  $X$  é uma matriz com os regressores exógenos, o vetor  $\beta$  é o vetor com coeficientes dos regressores exógenos, o vetor  $u$  é um vetor coluna, de dimensão  $n \times 1$  com os resíduos do modelo. Escrevendo-se a equação acima de forma mais concisa, com  $Z = [W y, Y, W]$ ,  $\gamma = [\rho, v', \beta']'$ , tem-se

$$y = Z \gamma + u.$$

Seja  $H$  uma matriz com observações das variáveis instrumentais para os regressores endógenos em  $Y$ . Os instrumentos para a variável endógena  $W y$  são dados pelos *lags* espaciais dos regressores exógenos  $W X$ . A matriz com todas as variáveis instrumentais pode ser então representada como:

$$Q = [X, W X, H].$$

O estimador de mínimos quadrados espacial de dois estágios (*spatial two stage least squares* – S2SLS) tem expressão

$$\hat{\gamma}_{S2SLS} = [Z' Q (Q' Q)^{-1} Q' Z]^{-1} Z' Q (Q' Q)^{-1} Q y. \quad (35)$$

Na ausência de heteroscedasticidade e autocorrelação espacial dos resíduos, um estimador para a variância assintótica dos estimadores é dada por:

$$\hat{\Sigma}_{\hat{\gamma}_{S2SLS}} = \hat{\sigma}^2 [Z' Q (Q' Q)^{-1} Q' Z]^{-1}, \quad (36)$$

com  $\hat{\sigma}^2 = (y - Z \hat{\gamma}_{S2SLS})' (y - Z \hat{\gamma}_{S2SLS}) / n$ .

Na presença de heteroscedasticidade dos resíduos, uma estimativa robusta para a matriz de variância assintótica tem expressão:

$$\hat{\Sigma}_{\hat{\gamma}_{S2SLS}} = [Z' Q (\widehat{Q' \Omega Q})^{-1} Q' Z]^{-1}, \quad (37)$$

onde  $\widehat{Q' \Omega Q} = Q' S Q$  e  $S$  é uma matriz diagonal contendo o quadrado dos resíduos da equação estimada via S2SLS. Na presença de heteroscedasticidade e autocorrelação espacial, pode-se utilizar um estimador robusto (HAC). Para isso, é preciso estimar  $\Psi = Q' \Omega Q$ . Uma forma para esta estimativa é dada por

$$\hat{\psi}_{r,s} = \frac{1}{n} \sum_i \sum_j q_{ir} q_{is} \hat{u}_i \hat{u}_j K\left(\frac{d_{ij}}{d}\right),$$

onde  $q_{ir}$  são elementos da matriz  $Q$ , e  $\hat{u}$  é o vetor de resíduos da equação estimada via S2SLS. O termo  $K\left(\frac{d_{ij}}{d}\right)$  é uma função *kernel* (que é uma função de densidade, com integral igual a 1). Algumas alternativas para as funções *kernel* estão apresentadas na Tabela 1.

Tabela 1- Alguns tipos de *kernel* a serem utilizados no estimador HAC para a matriz de covariância assintótica do estimador S2SLS

Tipo de <i>kernel</i>	Expressão
<i>Kernel</i> triangular ou de Barlett	$K\left(\frac{d_{ij}}{d}\right) = \left[1 - \left(\frac{d_{ij}}{d}\right)\right] \times I_{[d_{ij} \leq d]}$
<i>Kernel</i> de Epanechnikov	$K\left(\frac{d_{ij}}{d}\right) = \left[1 - \left(\frac{d_{ij}}{d}\right)^2\right] \times I_{[d_{ij} \leq d]}$
<i>Kernel</i> biquadrado (bi-squared <i>kernel</i> )	$K\left(\frac{d_{ij}}{d}\right) = \left[1 - \left(\frac{d_{ij}}{d}\right)^2\right]^2 \times I_{[d_{ij} \leq d]}$

Elaboração dos autores.

Na expressão na segunda coluna da tabela 1, o valor  $d_{ij}$  corresponde à distância entre os polígonos (ou demais entidades localizadas em um espaço de coordenadas)  $i$  e  $j$ .

A distância  $d$  é uma distância máxima de corte. Pode-se escolher  $d$  com um valor fixo para todas as observações, ou  $d$  variável, de forma a considerar um número fixo de vizinhos mais próximos de cada observação  $i$  (podem-se escolher distâncias variáveis, de forma a incluir os 40 vizinhos mais próximos, por exemplo, de cada observação). A partir da equação anterior para  $\Psi = Q'\Sigma Q$ , pode-se escrever a variância assintótica, robusta à heteroscedasticidade e à autocorrelação espacial nos resíduos, para os estimadores S2SLS, com a expressão

$$\hat{\Sigma}_{\hat{\gamma}_{S2SLS}} = \left[ (Z'_q Z_q)^{-1} Z' Q (Q' Q)^{-1} \hat{\Psi} (Q' Q)^{-1} Q' Z (Z'_q Z_q)^{-1} \right], \quad (38)$$

onde  $Z'_q Z_q = Z' Q (Q' Q)^{-1} Q' Z$ .

A correção dada pela expressão (36), para contabilizar para desvios em relação à hipótese de homocedasticidade e ausência de correlação entre os resíduos da regressão, baseia-se no trabalho de Conley (1999), que propõe um estimador robusto para correção da *matrix* de variância assintótica no contexto de método de momentos generalizados. Na próxima seção, faz-se uma discussão especificamente sobre a abordagem de Conley, a qual se mostra bastante flexível, permitindo estimar modelos com especificações não lineares. Nesse contexto, será discutido, por exemplo, como a abordagem GMM de Conley pode ser utilizada para estimar modelos *probit*, *logit* etc., quando há correlação espacial entre as observações.

## 6 Método de momentos generalizado com correção para dependência espacial

Nesta seção, apresenta-se uma discussão sobre o procedimento de Conley (1999), por meio do qual se permite a estimação de modelos gerais via método de momentos generalizados, na presença de autocorrelação espacial nas observações. Entre as vantagens deste procedimento, podem-se citar: *i*) conta com a toda a flexibilidade da estimação via

GMM; *ii*) possibilita a estimação de modelos com especificações não lineares; *iii*) apresenta uma extensão, para o caso espacial, da estimação não paramétrica da matriz de variância, inicialmente proposta, para dados com dependência temporal, por Newey e West (1987); e *iv*) possibilita a estimação de sistemas de equações.

Para simplificar a exposição, serão considerados apenas modelos uniequacionais. Considere-se então a forma geral do modelo de regressão (linear ou não linear)

$$y_i = m(x_i, \beta) + u_i. \quad (39)$$

O termo  $u_i$  é um termo de erro que possui média zero. O vetor  $x_i$  é um vetor de variáveis explicativas, e  $\beta$  corresponde a um vetor de parâmetros desconhecidos do modelo. Assume-se que pode haver endogeneidade em algumas das variáveis do lado direito da equação. Considere-se então um vetor de instrumentos  $z_i$ . No caso de não haver endogeneidade, o vetor de instrumentos é exatamente o vetor de covariáveis; ou seja,  $z_i = x_i$ .

A partir do vetor de variáveis instrumentais, podem-se então escrever as condições de momento (momentos populacionais)

$$E[u_i \times z_i] = E[(y_i - m(x_i, \beta)) \times z_i] = 0. \quad (40)$$

Para prosseguir a estratégia de estimação, substituem-se os momentos populacionais por seus equivalentes amostrais, obtendo-se

$$\frac{1}{n} \sum_{i=1}^n [y_i - m(x_i, \beta)] \times z_i = 0. \quad (41)$$

Assumindo-se algumas condições de regularidade, quando o número de coeficientes é exatamente igual ao número de instrumentos, diz-se que o modelo é exatamente identificado e é possível encontrar um vetor  $\hat{\beta}$  de coeficientes para o qual a igualdade acima é satisfeita.<sup>15</sup>

No entanto, quando a dimensão de  $z_i$  é maior do que o número de coeficientes, a probabilidade de se obter uma amostra para a qual a igualdade seja exatamente satisfeita é zero (conjunto de medida nula). Uma alternativa então é encontrar o vetor  $\hat{\beta}$  que minimiza a forma quadrática

$$J(\beta) = \left[ \frac{1}{n} \sum_{i=1}^n [y_i - m(x_i, \beta)] \times z_i \right]' \Psi \left[ \frac{1}{n} \sum_{i=1}^n [y_i - m(x_i, \beta)] \times z_i \right].$$

A matriz  $\Psi$  é uma matriz positiva definida qualquer. O estimador GMM é definido como

$$\hat{\beta}_{GMM} = \arg \min_{\beta \in \Theta} J(\beta).$$

---

<sup>15</sup> Ver Hamilton (1994) e Matyas (2008).

Pode-se mostrar que o estimador GMM é consistente (assumindo que as devidas condições de regularidade são satisfeitas). Eficiência é obtida utilizando-se a matriz ótima  $\Psi = \Omega^{-1}$ , onde

$$\Omega = Cov[[y_i - m(x_i, \beta)] \times z_i].$$

Na prática, quando não há dependência entre as observações, pode-se estimar  $\Omega$  por intermédio da expressão

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^n [[y_i - m(x_i, \beta)] \times z_i] \times [[y_i - m(x_i, \beta)] \times z_i]'. \quad (42)$$

No entanto, quando há possíveis dependências entre as observações para os vetores correspondentes às condições de momento, o estimador supracitado para  $\Omega$  não é mais válido. No caso de as observações para  $y_i$ ,  $x_i$  e  $z_i$  acontecerem em períodos discretos de tempo igualmente espaçados, Newey e West (1987) propõem uma correção não paramétrica e robusta para o estimador  $\hat{\Omega}$ . Este estimador foi revisitado em Andrews (1991) e Andrews e Monahan (1992).

Conley (1999) propôs um estimador robusto tanto a heteroscedasticidade quanto autocorrelação espacial, no caso de dados *cross-section*, espacialmente distribuídos, seguindo os mesmos princípios que Newey e West (1987). De maneira geral, o estimador proposto por Conley tem expressão

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n K(i, j) \times [[y_i - m(x_i, \beta)] \times z_i] \times [[y_j - m(x_j, \beta)] \times z_j]', \quad (43)$$

onde  $K(i, j) = \left[1 - \frac{D_H(i, j)}{L_H}\right] \times \left[1 - \frac{D_V(i, j)}{L_V}\right]$ , para  $D_H(i, j) < L_H$  e  $D_V(i, j) < L_V$ , e  $K(i, j) = 0$ , caso contrário. O valor  $D_H(i, j)$  corresponde à distância horizontal entre unidades  $i$  e  $j$ , o valor  $D_V(i, j)$  corresponde à distância vertical entre  $i$  e  $j$ ,  $L_H$  é a distância de corte horizontal, e  $L_V$  é a distância de corte vertical. Em geral, a minimização de  $J(\beta)$  não resulta em uma solução explícita, devendo ser feita via algoritmos numéricos. Uma exceção ocorre no caso de modelos lineares; neste caso, o estimador GMM pode ser escrito em forma fechada, sem haver necessidade de minimização numérica.

A flexibilidade da estimação via GMM, na formulação  $y_i = m(x_i, \beta) + \epsilon_i$ , permite o tratamento de modelos não lineares, com formulações paramétricas comumente encontradas na literatura. A Tabela 2 apresenta alguns exemplos de modelos que podem ser incorporados na formulação GMM. Pode-se então proceder com a abordagem de estimação, corrigindo, por exemplo, para problemas de dependência espacial.



Tabela 2 - Exemplos de modelos paramétricos enquadrados na formulação GMM, que podem ser estimados corrigindo-se para dependência espacial

Modelos paramétricos	Formulação
Modelos lineares	$m(x_i, \beta) = x_i^T \beta$
Modelos <i>logit</i>	$m(x_i, \beta) = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}$
Modelos <i>probit</i>	$m(x_i, \beta) = \Phi(x_i^T \beta)$
Modelos <i>complementary log-log</i>	$m(x_i, \beta) = e^{x_i^T \beta}$
Modelos exponenciais	$m(x_i, \beta) = 1 - \exp(-\exp(x_i^T \beta))$

Elaboração dos autores.

Uma vez estimado o vetor de coeficientes  $\beta$ , pode-se proceder com o processo de inferência a partir da matriz de covariância dos estimadores, estimável a partir da expressão

$$\widehat{Var}(\hat{\beta}) = \left\{ n \times \left[ \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta} ([y_i - m(x_i, \beta)] \times z_i) \right]' \hat{\Omega}^{-1} \left[ \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta} ([y_i - m(x_i, \beta)] \times z_i) \right] \right\}^{-1}.$$

Quando o modelo é exatamente identificado, com número de instrumentos igual ao número de parâmetros, a minimização da forma quadrática  $J(\beta)$  resulta em  $J(\beta) = 0$ . Quando o modelo é sobreidentificado, pode ser testada a validade das condições de momento, utilizando-se a estatística de Hansen (1982).

$$J = n \times \left[ \frac{1}{n} \sum_{i=1}^n ([y_i - m(x_i, \beta)] \times z_i) \right]^T \hat{\Omega}^{-1} \left[ \frac{1}{n} \sum_{i=1}^n ([y_i - m(x_i, \beta)] \times z_i) \right]. \quad (44)$$

Sob a hipótese nula de validade dos instrumentos, pode-se mostrar que a estatística  $J$  em (42) tem distribuição assintótica qui-quadrada, com  $k - q$  graus de liberdade, sendo  $q$  o número de coeficientes e  $k$  o número de condições de momento.

## 7 Heterogeneidade espacial

Muitos fenômenos estudados nas ciências que envolvem dados regionais tratam com a instabilidade da estrutura sobre o espaço na forma de diferentes funções respostas ou através da variação sistemática dos parâmetros no espaço.

Esses efeitos são conhecidos como heterogeneidade e englobam fatores como a heterocedasticidade, coeficientes aleatórios e mudança estrutural. Em termos globais há dois tipos de distintos de heterogeneidade espacial:

- a) Instabilidade estrutural: Expressa pela mudança funcional ou pela variação dos parâmetros.

b) Heterocedasticidade: Acontece pela falta de variáveis ou outro tipo de falta de ajuste que leva o erro a apresentar variância inconstante.

O tratamento da heterogeneidade espacial inicialmente é realizado por meio do aprofundamento dos métodos de expansão de coeficientes, como Casetti (1972). No entanto, a maior parte das aplicações recentes envolve o método de regressão geograficamente ponderado como uma forma de modelar a variabilidade dos parâmetros através do espaço (Fotheringham *et al.*, 2000; Fotheringham *et al.*, 2002).

Dessa forma, a heterogeneidade é usualmente tratada por meio de dois métodos: Método de expansão dos coeficientes de Casetti (1972) e a Regressão geograficamente ponderada- Fotheringham *et al.* (2002). Do ponto de vista econométrico, o método de expansão espacial pode ser considerado como um caso especial da alteração sistematicamente dos coeficientes em um modelo de regressão.

A heterogeneidade no fenômeno sob estudo é refletida nos valores dos parâmetros os quais diferem para cada observação. Essa diferenciação é assumida ser expressa como uma função de um número de variáveis auxiliares.

Na terminologia do método de expansão de Casetti, a especificação homogênea simples original é chamada de *modelo inicial*, ao passo que a formulação heterogênea é chamada de *modelo final* (ou *modelo terminal*).

Nas aplicações iniciais do método de expansão, as variáveis auxiliares consistiam de tendências de superfícies (através de polinômios) usando as coordenadas das localizações das observações, caracterizando assim a expansão espacial, outra proposta é utilizar componentes principais para agregar o máximo de informação em apenas uma ou duas variáveis.

Sem perda de generalidade, as propriedades do método de expansão de Casetti podem ser mais formalmente ilustradas através de uma regressão simples com apenas uma variável explicativa e o intercepto. O modelo inicial é dado por:

$$y = \beta_0 + \beta_1 x + u, \quad (45)$$

onde  $\beta_0$  e  $\beta_1$  são os coeficientes da regressão e  $x$  é o vetor de observações da variável explicativa. A heterogeneidade é refletida na falta de estabilidade dos parâmetros sob as unidades amostrais, isso é incorporado no modelo assumido, que cada parâmetro individualmente (ou um subconjunto de parâmetros) é uma função exata de um número finito de variáveis de expansão, por exemplo,  $z_1$  e  $z_2$  então:

$$\beta_1 = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2. \quad (46)$$

Substituindo o parâmetro expandido  $\beta_1$  no modelo original, Casetti (1972) obtêm-se:

$$y = \beta_0 + (\gamma_0 + \gamma_1 z_1 + \gamma_2 z_2)x + u, \quad (47)$$

ou mais explicitamente:

$$y = \beta_0 + \gamma_0 x + \gamma_1 (z_1 x) + \gamma_2 (z_2 x) + u. \quad (48)$$

Caso o modelo final seja especificado corretamente, então as estimativas dos parâmetros no modelo inicial serão viesadas, graças à omissão das variáveis no problema. De maneira mais geral, as estimativas dos coeficientes do modelo inicial podem ser expressas como:

$$\hat{\theta} = (X^t M X)^{-1} X^t M y, \quad (49)$$

onde  $M = I - Z(Z^t Z)^{-1} Z^t$  e o vetor  $\theta$  é composto pelas estimativas do vetor de parâmetros  $[\beta_0, \gamma_0]$ ,  $X = [1, y]$  e  $Z = [z_1 x, z_2 x]$ . Em termos do valor esperado, pode-se mostrar que:

$$E(\hat{\theta}) = \beta + (X^t X)^{-1} X^t Z \gamma, \quad (50)$$

onde  $\gamma$  é o vetor de parâmetros  $[\gamma_1, \gamma_2]$ . Os elementos da matriz  $Z$  são os produtos dos elementos de  $X$  com as variáveis de expansão. Então,  $X$  e  $Z$  não poderão ser ortogonais e as estimativas do modelo inicial serão viesadas.

Apesar de em alguns casos as variáveis expansoras possam apresentar multicolinearidade, o método de expansão espacial fornece uma maneira simples e atrativa de tratar a heterogeneidade nos coeficientes do modelo de regressão. Entretanto, a sua implementação necessita de cautela, especialmente quando não existem boas razões para a escolha das variáveis expansoras.

Na prática, o pressuposto de uma relação exata entre os coeficientes e suas expansões espaciais é difícil de se manter, ou seja, é usual que ao realizar esse pressuposto algum erro seja cometido. Então, uma forma de incorporar esse erro é fazer:

$$\beta_1 = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \varepsilon, \quad (51)$$

onde  $\varepsilon$  é assumido ser um vetor de erros aleatórios com distribuição normal com vetor de média zero e matriz de variância e covariâncias igual a  $I\sigma_\varepsilon^2$ . Substituindo (49) no modelo inicial obtém-se:

$$y = \beta_0 + \gamma_0 x + \gamma_1 z_1 x + \gamma_2 z_2 x + \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + x\varepsilon + u, \quad (52)$$

onde o novo termo estocástico será  $\omega = x\varepsilon + u$ . Assim, o modelo pode ser escrito em uma forma mais compacta  $y = \beta_0 + \gamma_0 x + \gamma_1 z_1 x + \gamma_2 z_2 x + \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \omega$ . O qual é um modelo heterocedástico, uma vez que  $Var[\omega] = x\sigma_\varepsilon^2 + \sigma_u^2$ , caso o erro da expansão e o erro do modelo sejam não-correlacionados. Conseqüentemente, inferências sobre um modelo final expandido que não considere a heterocedasticidade podem ser falhas.

Uma abordagem mais geral e que produz usualmente melhores resultados é o método das *Regressões Espacialmente Ponderadas*. Modelos de regressão geograficamente ponderada (ou espacialmente ponderada) são modelos de regressão onde o esquema de estimação permite que os parâmetros do modelo variem no espaço, de maneira suave. Após a estimação, ao invés de um vetor de parâmetros, contendo estimativas para amostra como um todo, existem  $n$  vetores de parâmetros, onde  $n$  é o número de unidades geográficas. Portanto, é estimado um vetor de parâmetros para cada unidade geográfica.

Esse procedimento é interessante se deseja identificar diferenças nos coeficientes da regressão, por exemplo, em âmbito territorial. A regressão geograficamente ponderada pode ser útil, inclusive, como ferramenta para investigar heterocedasticidade nos parâmetros do modelo.

A estimação é dada como segue: para cada unidade geográfica  $k$ , os parâmetros da regressão são estimados por mínimos quadrados ordinários, máxima verossimilhança

ponderada ou GMM geograficamente ponderado. O vetor de ponderação é construído de forma que unidades geográficas mais próximas de  $k$  recebem peso maior. Por outro lado, uma unidade geográfica mais distante da unidade  $k$  recebe peso menor. O processo completo envolve a execução de  $n$  estimações, cada uma com um diferente vetor de pesos de dimensão  $n$ .

No contexto de estimação do GMM, para a escolha da unidade geográfica  $k$ , considere um conjunto de pesos,  $i = 1, \dots, n$  os quais decaem na medida em que a distância geográfica entre as unidades  $i$  e  $k$  aumenta. Uma função de ponderação comumente usada, a qual é popular na estimação da densidade de *kernel*, é baseada na função de densidade para a distribuição normal.

Mais especificamente, seja  $lat_i$  e  $lon_i$  a latitude e longitude da unidade  $i$ , de acordo com a projeção de algum sistema de informação geográfico. Para uma unidade geográfica específica  $k$ , os  $n$  pesos para todas as  $n$  unidades podem ser calculados da seguinte maneira:

$$u_i^k = \frac{1}{\sqrt{2\pi\lambda^2}} \exp\left\{\frac{1}{2\lambda^2} [(lat_i - lat_k)^2 + (lon_i - lon_k)^2]\right\}, \quad i = 1, \dots, n. \quad (53)$$

Note que, a seqüência de pesos  $u_i^k$ ,  $i = 1, \dots, n$ , depende da unidade geográfica focada,  $k$ . Obviamente, o valor máximo da seqüência, para  $i = 1, \dots, n$  é igual a  $u_k^k$ . O parâmetro  $\lambda$  é conhecido como *bandwidth*, ou parâmetro de decaimento, e controla a velocidade de decaimento na função *kernel*. Quando  $\lambda$  for muito pequeno, os pesos para unidades geográficas mais distantes tendem a zero muito rapidamente; quando  $\lambda \rightarrow 0$ , todos os pesos  $u_i^k \rightarrow 0$ , exceto  $u_k^k$ , que tende para o infinito. Por outro lado, quando  $\lambda \rightarrow \infty$ , todos os pesos assumem mesmo valor.

Para cada unidade geográfica  $k$ , pode-se estimar o vetor correspondente  $\beta$  minimizando a função objetivo do GMM modificado:

$$Q^k(\beta) = \left[ \frac{1}{n} \sum_{i=1}^n u_i^k [y_i - m(x_i, \beta)] \times z_i \right]' \Psi \left[ \frac{1}{n} \sum_{i=1}^n u_i^k [y_i - m(x_i, \beta)] \times z_i \right]. \quad (54)$$

em que os pesos podem ser construídos de acordo com (51).

Para evitar distorções para os erros padrões dos parâmetros estimados, os  $n$  pesos  $u_i^k$  são normalizados com a soma

$$\sum_{i=1}^n u_i^k = n.$$

Portanto, o GMM usual é um caso particular do GMM geograficamente ponderado, quando  $u_i^k = 1$  para todos  $k = 1, \dots, n$  e todos  $i = 1, \dots, n$ .

A escolha do  $\lambda$  é importante para determinar a suavidade das  $n$  estimativas de  $\hat{\beta}^k$ . Para valores grandes de  $\lambda$ , as estimativas tendem ser bastante similares. No limite, quando  $\lambda \rightarrow \infty$ , todas as estimativas  $\hat{\beta}^k$  tendem para  $\hat{\beta}$ , estimado por GMM não-ponderado. Quando  $\lambda \rightarrow 0$ , o estimador considerada somente a observação  $k$ . Em geral, a escolha de  $\lambda$  pode ser orientada por alguma regra de validação-cruzada (*cross-validation*).

Outra proposta para o tratamento da heterogeneidade espacial é a abordagem de Lesage (2004) denominada SALE (*Spatial Autocorrelation Local Estimation*). Esse método utiliza a proposta de Fotheringham *et al.* (1999) para justificar a abordagem GWR (*Geographic Weighted Regression*). Apesar de essa abordagem produzir estimativas dos parâmetros locais tendenciosas e inconsistentes, a magnitude do viés desses estimadores é menor do que alternativas que não abordam o problema da heterogeneidade espacial. Mesmo não sendo uma solução ideal, essa abordagem pode ser válida para diversas situações práticas. Lesage (2004) estendeu a abordagem GWR (que originalmente tratava apenas dos coeficientes dos regressores do modelo) para estimar também o parâmetro autoregressivo do modelo com dinâmica espacial do tipo SAR. A idéia central do algoritmo é fornecer uma ferramenta que resolva, para cada ponto na amostra, a estimativa com base em uma equação com estrutura espacial dinâmica.

### Modelo de regimes espaciais.

Uma abordagem usual para o tratamento da instabilidade estrutural dos parâmetros do modelo é a adaptação da proposta de Quantd (1958) para a abordagem espacial. Nesse caso, assume-se que o sistema siga dois ou mais regimes no espaço. Inicialmente, a proposta de Quantd (1958) foi sugerida para dois regimes no tempo para o caso de modelos de regressão linear. A extensão espacial dessa proposta pode ser utilizada para mais de dois regimes, com base nas dimensões no espaço ao invés da dimensão temporal.

Suponha, sem perda de generalidade, um modelo espacial SARMA o qual assume dois regimes no espaço, tais que:

$$\begin{cases} y = \rho W_{11}y + X\beta_1 + u_1 \\ y = \rho W_{12}y + X\beta_2 + u_2 \end{cases}, \quad (55)$$

nos quais os resíduos das equações observadas possuem uma estrutura auto-regressiva, da forma

$$\begin{cases} u_1 = \lambda_1 W_{21}\epsilon_1 + \epsilon_1 \\ u_2 = \lambda_2 W_{22}\epsilon_2 + \epsilon_2 \end{cases}. \quad (56)$$

As matrizes  $W_{11}, W_{12}, W_{21}, W_{22}$ , e  $W_2$  são matrizes de contiguidade não necessariamente iguais, tais que  $W_{11}, W_{12}$ , e  $W_{21}, W_{22}$ , são partições da região como um todo<sup>16</sup>.

Essas partições são estimadas de maneira a maximizar a soma das log-verossimilhanças dos modelos em (55) e (56). Ou seja, para cada uma das  $k$  possíveis partições da matriz  $W$ , calcula-se a log-verossimilhança e o conjunto de partições que maximiza essa medida é considerada na análise. Em seguida, cada modelo é estimado separadamente e o teste da razão de verossimilhança entre o modelo particionado e o modelo completo pode ser utilizado para se testar a hipótese de que a mudança de regime tenha ocorrido no espaço.

<sup>16</sup> Diz-se que uma região possui uma partição quando há dois conjuntos de polígonos contíguos, disjuntos entre si, tais que a união dessas sub-regiões compõe a região espacial como um todo.

## 8 Comentários finais

Este texto apresenta uma discussão sobre alguns dos modelos econométricos comumente utilizados para modelagem de dados espaciais. Os modelos apresentados aqui estariam mais adequados para estudos empíricos seguindo as abordagens experimentalista e descritiva, nas quais o objetivo é identificar efeitos causais de uma determinada política, ou encontrar relações entre variáveis econômicas. De fato, o estimador de mínimos quadrados de dois estágios, de Kelejian e Prucha, e o estimador de método de momentos generalizado, de Conley (1999), ambos discutidos neste estudo; permitem a estimação de parâmetros na presença de variáveis endógenas do lado direito da equação, contabilizando e/ou corrigindo para a presença de autocorrelação espacial nos resíduos do modelo. Mesmo não tratando diretamente a abordagem estruturalista, as ideias apresentadas neste texto fornecerão ao leitor uma noção dos procedimentos para estimação com dados com presença de dependência espacial, o que poderá ser útil para a estimação de parâmetros estruturais em modelos microfundamentados.

Dado o grande avanço recente na literatura de análise de dados espaciais, optou-se por apresentar apenas alguns dos métodos mais comumente utilizados, de forma a transmitir ao leitor uma ideia básica, mas clara, dos fundamentos da estimação de modelos econométricos com dependência espacial. Não foram cobertos modelos para dados de painel,<sup>17</sup> mas apenas para dados *cross-section*. Outro tópico de extrema importância na análise de dados espaciais, que não foi tratado aqui, são os modelos estimados via abordagem bayesiana. O leitor poderá encontrar boas exposições em Tanner (1996), Banerjee *et al.* (2004) e Schabenberger e Gotway (2009), entre outros.

CARVALHO YWATA, A. X.; ALBUQUERQUE, P. H. M. Methods and models in spatial econometrics. A review. *Rev. Bras. Biom.*, São Paulo, v.29, n.2, p.273-306, 2011.

- *ABSTRACT: This paper presents a discussion on several econometric techniques for estimation of parametric models in the presence of spatial dependence, for cross-section data. We focus initially on spatial dependence models with spatial lags for the response variable or spatial lags for the equation residuals; estimation is done by maximum likelihood methods. We also present a critical analysis of the pitfalls that come up when using these spatial lag models. The paper also brings a discussion on tests for detection of the presence of spatial dependence. Finally, we discuss more robust estimation methods, which account for endogeneity in some of the explanatory variables.*
- *KEYWORDS: Spatial econometrics; spatial dependence; cross-section data.*

## Referências

ACKERBERG, D.; BENKARD, C. L.; BERRY, S.; PAKES, A. Econometric tools for analyzing market outcomes. In: HECKMAN, J. J.; LEAMER, E. E. (Ed.). *Handbook of econometrics*. Amsterdam: Elsevier, 2007. v.6A.

---

<sup>17</sup> Ver, por exemplo, Elhorst (2003), Druska e Horrace (2004), e Egger *et al.* (2005).

ANDREWS, D. W. K. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, Oxford, v.59, n.3, p.817-858, 1991.

ANDREWS, D. W. K.; MONAHAN, J. C. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, Oxford, v.60, n.4, p.953-966, 1992.

ANGRIST, J. D.; PISCHKE, J. S. *Mostly harmless econometrics: an empiricist's companion*. New Jersey: Princeton University Press, 2009. 373p.

ANSELIN, L.; BERA A. K.; FLORAX, R.; YOON, M. J. Simple diagnostic tests for spatial dependence. *Reg. Sci. Urban Econ.*, Amsterdam, v.26, p.77-104, 1996.

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

ANSELIN, L.; FLORAX, R. *Advances in spatial econometrics*. Heidelberg: Springer-Verlag, 2000. 513p.

ANSELIN, L.; FLORAX, R., REY, S. J. *Advances in spatial econometrics – methodology, tools and applications*. Heidelberg: Springer, 2004. 513p.

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

BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. *Hierarchical modeling and analysis for spatial data*. Florida: Chapman & Hall/CRC, 2004. 472p. (Monographs on Statistics and Applied Probability, 101)

BAUMONT, C. *Spatial effects in housing price models. Do housing prices capitalize urban development policies in the agglomeration of Dijon (1999)?* LEG - Document de travail - Economie 2004-04, LEG, Laboratoire d'Economie et de Gestion, CNRS UMR 5118, Université de Bourgogne, 2004.

BERRY, S.; LEVINSOHN, J.; PAKES, A. Automobile prices in market equilibrium. *Econometrica*, Oxford, v.63, n.4, p.841-890, 1995.

BERRY, S.; LEVINSOHN, J.; PAKES, A. Differentiated products demand systems from a combination of micro and macro data: the new car market. *J. Polit. Econ.*, Chicago, v.112, n.1, p.68-105, 2004.

CAMARGO, R. S.; CARVALHO, A. X. Y.; BOUERI, R. *Método de momentos generalizados geograficamente ponderados*. Brasília: IPEA, 2010. 76p. (Relatório Técnico).

CAMERON, A. C.; TRIVEDI, P. K. *Microeconometrics: methods and applications*. New York: Cambridge University Press, 2005.

CASETTI, E. Generating models by the expansion method: applications to geographic research. *Geogr. Anal.*, Columbus, v.4, p.81-91, 1972.

CONLEY, T. GMM estimation with cross-sectional dependence. *J. Econ.*, Oxford, v.92, p.1-45, 1999.

DAVIS, T. A. *Direct methods for sparse linear systems: fundamentals of algorithms*. Philadelphia: Society for Industrial and Applied Mathematics, 2006. 217p.

DRUSKA, V.; HERRACE, W. C. Generalized moments estimation for spatial panel data: Indonesian rice farming. *Am. J. Agric. Econ.*, Oxford, v.86, n.1, p.185-198, 2004.

ECKSTEIN, Z.; WOLPIN, K. Why youths drop out of high school: the impact of preferences, opportunities, and abilities. *Econometrica*, Oxford, v.67, p.1295-1340, 1999.

EGGER, P.; PFAFFERMAYR, M.; WINNER, H. An unbalanced spatial panel data approach to US state tax competition. *Econ. Lett.*, Amsterdam, v.88, n.3, p.329-335, 2005.

ELHORST, J. P. Specification and estimation of spatial panel data models. *Int. Reg. Sci. Rev.*, Philadelphia, v.26, n.3, p.244-268, 2003.

EPPLE, D.; SIEG, H. Estimating equilibrium models of local jurisdictions. *J. Pol. Econ.*, London, v.107, p.645-681, 1999.

FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. *Geographically weighted regression: the analysis of spatially varying relationships*. New York: John Wiley & Sons, 2002. 262p.

FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. *Quantitative geography: perspectives on spatial data analysis*. London: Sage Publications, 2000. 282p.

FOTHERINGHAM, A. S.; CHARLTON, M.; BRUNSDON, C. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ. Plann. A*, London, v.30, n.11, p.1905-1927, 1999.

HAHN, J.; TODD, P.; VAN DER KLAUW, W. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, Oxford, v.69, p.201-209. 2001.

HAMILTON, J. D. *Time series analysis*. Princeton: Princeton University Press, 1994. 799p.

HENDRY, D. F. *Dynamic econometrics*. Advanced texts in econometrics. Oxford: University Press, 1995.

HOLMES, T. J. Structural, experimentalist, and descriptive approaches to empirical work in regional economics. *J. Reg. Sci.*, Philadelphia, v.50, n.1, p.5-22, 2010.

KEANE, M.; WOLPIN, K. I. The career decisions of young men. *J. Pol. Econ.*, London, v.105, p.473-522, 1997.

KELEJIAN, H. H.; PRUCHA, I. R. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J. Real State Finance Econ.*, Dordrecht, v.17, n.1, p.99-121, 1998.



- KELEJIAN, H. H.; PRUCHA, I. R. Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem. *Int. Reg. Sci. Rev.*, Philadelphia, v.20, n.1, p.103-111, 1997.
- KELEJIAN, H. H.; PRUCHA, I. R.; YUZEFOVICH, Y. Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample results. In: LESAGE, J.; PACE, R. K. *Spatial and spatiotemporal econometrics, advances in econometrics*, New York: Elsevier, 2004. v.18, p.163-198.
- KELEJIAN, H. H.; ROBINSON, D. P. 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Reg. Sci. Urban Econ.*, Amsterdam, v.32, n.6, p.691-707, 2002.
- KELEJIAN, H. H.; ROBINSON, D. P. HAC estimation in a spatial framework. *J. Econ.*, Amsterdam, v.140, n.1, p.131-154, 2007.
- KELEJIAN, H. H.; ROBINSON, D. P. Spatial autocorrelation: a new computationally simple test with an application to per capita county police expenditures. *Reg. Sci. Urban Econ.*, Amsterdam, v.22, n.3, p.317-331, 1992.
- KELEJIAN, H. H.; ROBINSON, D. P. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *J. Econ.*, Amsterdam, 2009.
- LEE, L. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *J. Econ.*, Amsterdam, v.137, n.2, p.489-514, 2007.
- LESAGE, J. Bayesian estimation of spatial autoregressive models. *Int. Reg. Sci. Rev.*, Philadelphia, v.20, n.1-2, p.113-129, 1997.
- LESAGE, J.; PACE, R. K. *Introduction to spatial econometrics*. Boca Raton: CRC Press, 2009.
- LESAGE, J. The family of geographically weighted regression models. In: ANSELIN, L., FLORAX, R. J. G. M.; REY, S. J. *Advances in spatial econometrics*. New York: Springer, 2004.
- LESAGE, J. *The theory and practice of spatial econometrics*. Toledo: Department of Economics, University of Toledo, 1999.
- MANSKI, C. Identification of endogenous social effects: the reflection problem. *Rev. Econ. Stud.*, Oxford, v.60, n.3, p.531-542, 1993.
- MATYAS, L. *Generalized method of moments estimation: themes in modern econometrics*. Cambridge: Cambridge University Press, 2008. 332p.
- McMILLEN, D. P. Issues in spatial data analysis. *J. Reg. Sci.*, London, v.50, n.1, p.119-141, 2010.
- NEVO, A. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, Oxford, v.69, n.2, p.307-342, 2001.

- NEWKEY, W. K.; WEST, K. D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, Oxford, v.55, p.703-708, 1987.
- PACE, R. K.; BARRY, R. Simulating mixed regressive spatially autoregressive estimators. *Comput. Stat.*, Heidelberg, v.13, n.3, p.397-418, 1998.
- PACE, R. K.; BARRY, R. Sparse spatial autoregressions. *Stat. Probabil. Lett.*, Amsterdam, v.33, p.291-297, 1997.
- PETRIN, A. Quantifying the benefits of new products: the case of the minivan. *J. Political Econ.*, Chicago, v.110, n.4, 2002.
- PINKSE, J.; SLADE, M. E.; BRET, C. Spatial price competition: a semiparametric approach. *Econometrica*, Oxford, v.70, n.3, p.1111-1153, 2002.
- PINKSE, J.; SLADE, M. E. Mergers, brand competition, and the price of a pint. *Eur. Econ. Rev.*, Amsterdam, v.48, n.3, p.617-643, 2004.
- PINKSE, J.; SLADE, M. E. The future of spatial econometrics. *J. Reg. Sci.*, London, v.50, n.1, p.103-117, 2010.
- POFAHL, G. *Essays on horizontal merger simulation: the curse of dimensionality, retail price discrimination, and supply channel stage-games*. Tese (Doutorado), Texas A&M. University, College Station, 2007.
- QUANDT, R. The estimation of the parameters of a linear regression system obeying two separates regimes. *J. Am. Stat. Assoc.*, Schaumburg, v.53, p.873-880, 1958.
- SCHABENBERGER, O.; GOTWAY, C. A. *Statistical methods for spatial data analysis*. Florida: Chapman & Hall/CRC, 2009. (Texts in Statistical Science).
- TANNER, M. *Tools for statistical inference, methods for the exploration of posterior distributions and likelihood functions*. New York: Springer-Verlag, 1996. 207p. (Series in Statistics).

Recebido em 19.01.2011

Aprovado após revisão em 17.06.2011