UNIVERSIDADE FEDERAL DE JUIZ DE FORA - *CAMPUS* Governador Valadares
ECONOMETRIA ESPACIAL
Prof. Vinícius de Azevedo Couto Firme

Econometria Espacial:

Capítulo 1 – O que é Econometria Espacial?

Estrutura da Apresentação:

1. Introdução

2. Hipóteses de Gauss-Markov

3. Efeitos Espaciais

4. Os efeitos Espaciais e as hipóteses de Gauss-Markov

1.Introdução:

- Os trabalhos aplicados à economia têm se preocupado cada vez mais com a interação dos agentes no espaço. Esta seria uma crítica aos modelos que assumem que os agentes são atomísticos e agem de forma independente uns dos outros. Assim, ao considerar o espaço admite-se que a interação entre agentes heterogêneos possa gerar uma espécie de externalidade espacial (Anselin, 1999).

- "*Spatial externalities play a central role in the recent emergence of "spatial thinking" in the mainstream social sciences (Goodchild et al. 2000). For example, in economics, greatly increased attention is being paid to models of social interaction, which introduce a dependence among actors in a system* (Anselin, 2003, p.153)."

1.Introdução:

- A importância do espaço já era discutida por Isard (1956) e, posteriormente, voltou à tona através da *Nova Geografia Econômica* (NGE), resgatando conceitos como as "externalidades Marshallianas" e as "Economias de Aglomeração" (ARTHUR, 1989; KRUGMAN, 1991a, 1991b, 1998; GLAESER *et al*, 1992).

- Contudo, a possibilidade de testar os efeitos espaciais ocorreu apenas na década de 90, após o avanço do Sistema de Informações Geográficas (*Geographic Information Systems* - GIS) e da disponibilidade de dados econômicos e sociais geo-codificados.

- Deste modo, a econometria espacial surgiu como uma forma aplicada de tratar os efeitos da interação entre os agentes no espaço (auto-correlação espacial) e as diferenças socioeconômicas existentes entre localidades distintas (heterogeneidade espacial) (ALMEIDA *et al*, 2005).

1.Introdução:

- Note que o "fator espaço" pode afetar os resultados de estudos empíricos. Considere o caso onde y é uma variável dependente qualquer, X é uma matriz de variáveis explicativas e W é uma matriz de pesos espaciais que capta o efeito associado ao espaço. Assim, é possível estimar a seguinte equação:

$$y = \beta_1 Wy + \beta_2 X + \beta_3 WX + \varepsilon, \qquad \text{onde: } \varepsilon = \beta_4 W\varepsilon + u \tag{1}$$

Logo, $\beta_1 \neq 0$ indica autocorrelação espacial associada à variável dependente.

- Este tipo de dependência tem alcance global e, caso seja ignorada (ausência de Wy), pode gerar estimativas enviesadas, em pequenas amostras, ou inconsistentes, em grandes amostras. (ANSELIN E BERA, 1998, p.246)

1.Introdução:

$\beta_3 \neq 0$, indica que as variáveis explicativas geram um "efeito transbordamento" que afeta y.

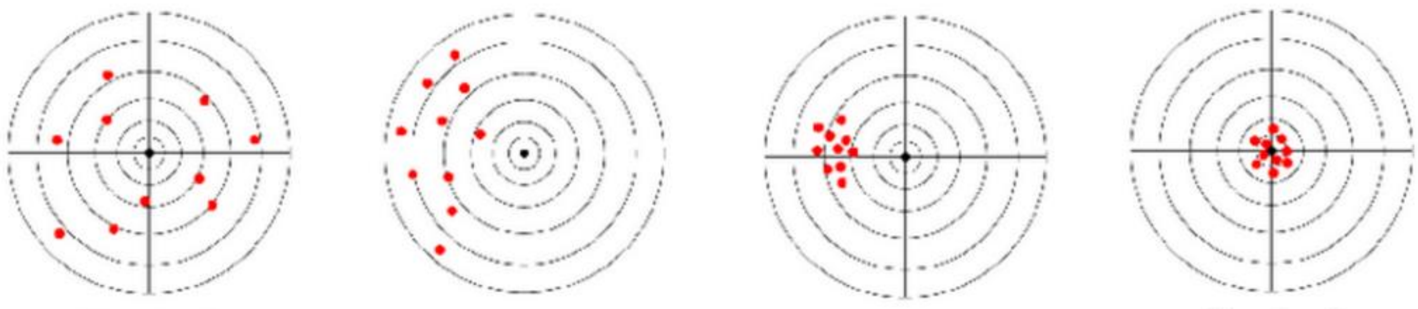
- Este efeito poderia ser estimado via MQO. Contudo, a ausência do termo defasado espacialmente (WX) causaria um problema semelhante à omissão de variável relevante, gerando viés nos coeficientes estimados (Rey e Montouri, 1999).

$\beta_4 \neq 0$ indica que os erros são espacialmente correlacionados.

- Deste modo, um choque ocorrido em uma região se espalha não só para os seus vizinhos e sim por todas as outras unidades (alcance global). Ignorar $W\varepsilon$, quando $\beta_4 \neq 0$, violaria o pressuposto da homocedasticidade, tornando o modelo ineficiente. Ainda assim, este efeito não gera viés nem inconsistência nos estimadores (ANSELIN, 1988; ANSELIN E BERA, 1998; ANSELIN, 2003).

1.Introdução:

- Relembrando - Propriedades de um Estimador:

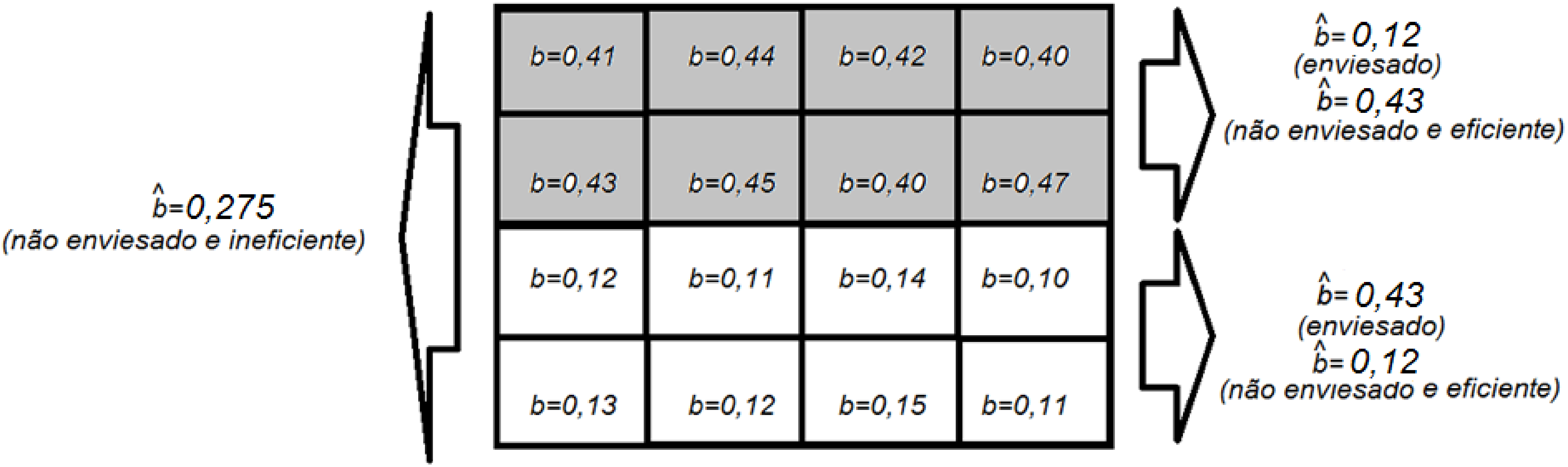


Não Enviesado Não Eficiente | Enviesado Não Eficiente | Enviesado Eficiente | Não Enviesado Eficiente

- Consistência: Um estimador é consistente quando o parâmetro estimado, $\hat{\beta}$, fica mais próximo do verdadeiro valor de β, a medida em que o tamanho da amostra cresce ($n \rightarrow \infty$).

1.Introdução:

Exemplo: Coeficientes estimados ($\hat{b}$) para as regiões Norte, Sul e área total da região abaixo. Sendo "b" os parâmetros individuais verdadeiros.



Nota: a heterogeneidade espacial (Norte *vs* Sul) produz erros espacialmente correlacionados, reduzindo a eficiência do estimador.

Solução: regimes espaciais.

1.Introdução:

- O modelo da Equação 1 visa controlar a dependência espacial. Todavia, caso as unidades geográficas (ex: municípios) apresentem uma dinâmica distinta no espaço (devido às diferenças de cultura, localização, recursos naturais, clima, tecnologia ou outro fator inerente à região), haverá uma *hererogeneidade espacial*, que possivelmente comprometerá o controle da auto-correlação espacial.

- Caso y represente o crescimento econômico, tem-se que: "*Spatial heterogeneity means in turn that economic behavior is not stable across space and may generate characteristic spatial patterns of economic development under the form of spatial regimes: a cluster of rich regions being distinguished from a cluster of poor regions*" (Ertur e Gallo, 2003, p. 176).

1.Introdução:

- Embora parte desta heterogeneidade possa ser captada através da inclusão de variáveis no modelo, nem sempre é possível controlá-la (em sua totalidade) desta forma. Geralmente, utiliza-se algum tipo de agrupamento ou regime espacial a fim minimizar a heterogeneidade espacial no caso de dados *cross-section* (Monastério e Ávila, 2004, p.291; Firme e Simão Filho, 2014, p.681; Gonçalves *et all*, 2011, p.310).

2. Hipóteses de Gauss-Markov:

- Por quê a Econometria convencional falha ao desconsiderar os efeitos espaciais?

- Inicialmente, considere o seguinte MGD: $y_{nx1} = X_{nxk}\beta_{kx1} + \varepsilon_{nx1}$ (2)

Onde:
y = variável dependente (vetor); X = matriz de variáveis explicativas;
β = vetor de coeficientes estimados; ε = termos de erro (vetor)
n = número de observações (ex: 5561 municípios do Brasil)
k = número de variáveis explicativas (incluindo a constante).

Logo:

$$\underset{n\times 1}{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times k}{X} = \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & \cdots & X_{kn} \end{bmatrix} \qquad \underset{k\times 1}{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix} \qquad \underset{n\times 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

2. Hipóteses de Gauss-Markov:

- Para que o MQO seja considerado "BLUE" (MELNE em português), os 5 pressupostos de Gauss-Markov precisam ser satisfeitos:

1) Hipótese da linearidade dos parâmetros:

- y é uma função linear de $x_2 \ldots x_k$;

Questão: Como estimar $\hat{\alpha}$ numa função do tipo *Cobb-Douglas* ($Y = AK^{\alpha}L^{1-\alpha}$)?

2) Hipótese da não colinearidade perfeita:

- x_j não é uma função linear das demais variáveis x_s ($j \neq s, \forall\, j, s = 1 \ldots k$.

Nota: só assim é possível inverter a matriz quadrada $(X'X)$. Lembrar que EMQO é: $\hat{\beta} = (X'X)^{-1}X'Y$

Caso contrário: Impossível obter valor único para $\hat{\beta}$.

Nota: A presença de multicolinearidade forte não viola o pressuposto 2. Contudo os testes t tendem a aceitar a hipótese nula de significância individual das variáveis.

2. Hipóteses de Gauss-Markov:

3) Hipótese da média condicional zero (FORTE!):

- $E(\varepsilon_i|X) = 0$

Para garantir Hipótese 3, é preciso que:

3.1. y esteja bem especificada pelo conjunto de variáveis X.
Caso contrário: Omissão de variável relevante - Estimador enviesado.

3.2. Não exista dependência linear ou não linear entre ε e X, seja ela intra-regional (ε_i e X_i) ou **inter-regional** (ε_i e X_j).
Caso contrário: Problema de simultaneidade – Estimador enviesado.

3.3. $x_2 \ldots x_k$ sejam não estocásticas (exógenas/determinísticas).
Caso contrário: Endogeneidade - Estimador enviesado.

2. Hipóteses de Gauss-Markov:

4) Hipótese de Homocedasticidade:

- $Var(\varepsilon_i|X) = \sigma^2$

Nota: erros estimados apresentam a mesma variância para as *n* regiões.
Caso contrário: Enviesa erros-padrão das estimativas, afetando credibilidade das Estatística *t* e F.

5) Hipótese da Independência dos erros:

- $Corr(\varepsilon_i \varepsilon_j | X) \cong Corr(\varepsilon_i \varepsilon_j) = 0, para\ i \neq j$

Nota: erro da região i não está linearmente correlacionado com o erro da região j. Isto é, não existe interação social entre as regiões (Hipótese Forte!).

2. Hipóteses de Gauss-Markov:

- A combinação das hipóteses 1, 2 e 3, garantem que o MQO é não viesado.

- A combinação das hipóteses 4 e 5 garantem que o MQO é o mais eficiente dentre os estimadores lineares.

- Neste caso, a matriz Var-Cov da estimação via MQO torna-se:

- $Var(\varepsilon|X) = \sigma^2 I_n$ ou seja, $Var(\varepsilon|X)=\begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix}$

- A combinação das hipótese 1 a 5 garante que o MQO é "BLUE".

2. Hipóteses de Gauss-Markov:

- Caso <u>adicionemos</u> a Hipótese de distribuição normal para y e ε:

6) $\varepsilon \sim \text{Normal}(0, \sigma^2 I_n)$; $\text{y} \sim \text{Normal}(\text{X}\beta, \sigma^2 I_n)$

O MQO se torna o melhor estimador dentre todos os estimadores, lineares e não lineares.

- Nota: A hipótese 6 garante a acurácia das estatísticas t e F em amostras limitadas ou pequenas. No casos das grandes amostras, vale-se o "**Teorema Central do Limite**".

3. Efeitos Espaciais:

3.1 Dependência Espacial (Auto-correlação Espacial)

- Ocorre quando os indivíduos analisados (ex: famílias, empresas, regiões, etc) não são independentes entre si.

- Este tipo de dependência está sujeita à Lei de Tobler (1ª Lei da Geografia).

"*Tudo está relacionado com tudo o resto, mas coisas próximas estão mais relacionadas do que coisas distantes*."

- Contudo, a "distância" não deve ser pensada apenas em termos geográficos (ex: distância política, comercial, socioeconômica, etc).

- Logo, a variável de interesse y, na região i, poderá depender desta mesma variável em seus j vizinhos (y_j). Formalmente: $y_i = f\left(y_j, X\right), com\ i \neq j$.

3. Efeitos Espaciais:

3.1 Dependência Espacial (Auto-correlação Espacial)

- Exemplo para 2 regiões:

$$\begin{cases} y_1 = \rho_2 y_2 + \beta X_1 + \varepsilon_1 \\ y_2 = \rho_1 y_1 + \beta X_2 + \varepsilon_2 \end{cases}$$

Logo: y_1 afeta y_2 e vice-versa (problema de simultaneidade) $\rightarrow$ $\boxed{y_1 \longleftrightarrow y_2}$

- Exemplo para *n* regiões:

$$\begin{cases} y_1 = \rho_a y_2 + \rho_b y_3 \ldots + \rho_c y_n + \beta X_1 + \varepsilon_1 \\ y_2 = \rho_d y_1 + \rho_e y_3 \ldots + \rho_f y_n + \beta X_2 + \varepsilon_2 \\ y_3 = \rho_g y_1 + \rho_h y_2 \ldots + \rho_i y_n + \beta X_3 + \varepsilon_3 \\ \vdots \\ y_n = \rho_j y_1 + \rho_k y_2 \ldots + \rho_l y_{n-1} + \beta X_n + \varepsilon_n \end{cases} \quad (3)$$

- <u>Questão</u>: Como estimar tantos parâmetros no sistema de equações 3?

Nota: aumento de “*n*” aumenta parâmetros a serem estimados (*problema do parâmetro incidental*).

3. Efeitos Espaciais:

3.1 Dependência Espacial (Auto-correlação Espacial)

- Solução: Usa-se uma matriz de ponderação espacial (W) para se obter uma "média" dos n-1 vizinhos de i.

- Cabe destacar que a dependência espacial também poderia estar associada às variáveis explicativas (WX) e ao termo de erro ($W\varepsilon$)

- Como é difícil visualizar esta dependência na prática, iremos utilizar testes a fim de verificar sua existência (ex: *I de Moran, G de Getis-Ord*, *C de Geary*).

Fraqueza: Os testes captam apenas a dependência espacial linear.

3. Efeitos Espaciais:

3.1 Dependência Espacial (Auto-correlação Espacial)

Fontes primárias de Dependência Espacial:

a) Interação Espacial

Conceito: causada pela movimentação de bens, pessoas e informações através do espaço

Fontes de Interação Espacial

a.1) Difusão
EX: transbordamento tecnológico.

a.2) Troca de bens, serviços e rendas
EX: aluno de outra cidade morando em GV.

a.3) Comportamento Estratégico:
EX: Dilema do sorveteiro; Disposição de franquias.

a.4) Espraiamento
EX: migração de pessoas, êxodo rural.

b) Erros de Medida

Conceito: causado pela agregação espacial de dados.

Ex: Análise das regiões independentes A, B, C.

Problema: Agregação em regiões 1 e 2.

A	B	C

1	2

Neste caso:

$$y_1 = y_A + \theta y_B$$
$$y_2 = y_C + (1-\theta) y_B$$

c) Má especificação do modelo

Conceito: causada pela omissão de variável relevante ou *outliers* espacial.

EX: Ignorar o nível educacional numa análise do PIB poderia gerar resíduos com dependência Espacial.

3. Efeitos Espaciais:

3.2 Heterogeneidade Espacial

- Fenômeno decorrente da instabilidade estrutural (social, econômica, cultural, etc.) no espaço.

- Neste caso, pode haver diferentes respostas dependendo do lugar.

Ex: a) Coeficientes distintos em diferentes regiões (mudança estrutural)
b) Variância do erro não constante (heterocedasticidade)
c) Formas funcionais diferentes para subconjuntos de dados.

Formalmente, a Heterogeneidade pode ser representada com:

$y_i = f_i(X_i, \beta_i, \xi_i)$ onde $\xi_i \sim (0, \Omega)$ (4)

Onde o subscrito i indica que pode haver um β, um ξ e uma f específicos para cada região analisada. O termo Ω representa uma matriz heterocedástica.

3. Efeitos Espaciais:

3.2 Heterogeneidade Espacial

Nota: Este fenômeno pode ser controlado através da subdivisão amostral, isto é, via criação de regimes espaciais (ex: Norte x Sul, Urbano x Rural, Centro x Periferia, etc.). Contudo, quando há *heterogeneidade espacial extrema*, haveria a necessidade de estimar um coeficiente (β_i) ou Equação específicos para cada região analisada (f_i).

Problema: Faltaria graus de liberdade para efetuar estimações no caso extremo.

Solução: Regressões Ponderadas Geograficamente – RPG (FOTHERINGHAN *ET AL*, 2000, 2002).

3. Efeitos Espaciais:

3.2 Heterogeneidade Espacial

Fontes de Heterogeneidade Espacial:		
a) Estrutura Espacial	***b) Erros de Medida***	***c) Má especificação do modelo***
Conceito: diferenças na estrutura espacial (observadas ou não) que podem gerar parâmetros distintos, dependendo da localidade.	**Conceito:** A heterocedasticidade, uma das manifestações da heterogeneidade, pode ser causada por erros de medida.	**Conceito:** a omissão de uma variável relevante fará parte do termo de erro, tornando-os não homocedásticos e não independentes.
Ex: Diferenças culturais, sociais, econômicas, políticas, geográficas, relacionadas às preferências locais e outras características associadas ao espaço.	**Ex:** A variância do termos de erro (σ^2) depende do número de observações incluídas em cada "g" agrupamento (σ^2/n_g).	**Ex:** Ignorar o nível educacional numa análise do PIB poderia gerar resíduos com heterocedasticidade e dependência Espacial.
Solução: boa especificação dos modelos; uso de *dummies* de região, regimes espaciais ou RPG.	**Solução:** Usar "g" homogêneos e variáveis intensivas (ex: PIB per capita, produção p/ hectare, etc).	**Solução:** Verificar literatura sobre o tema a fim de produzir boas especificações.

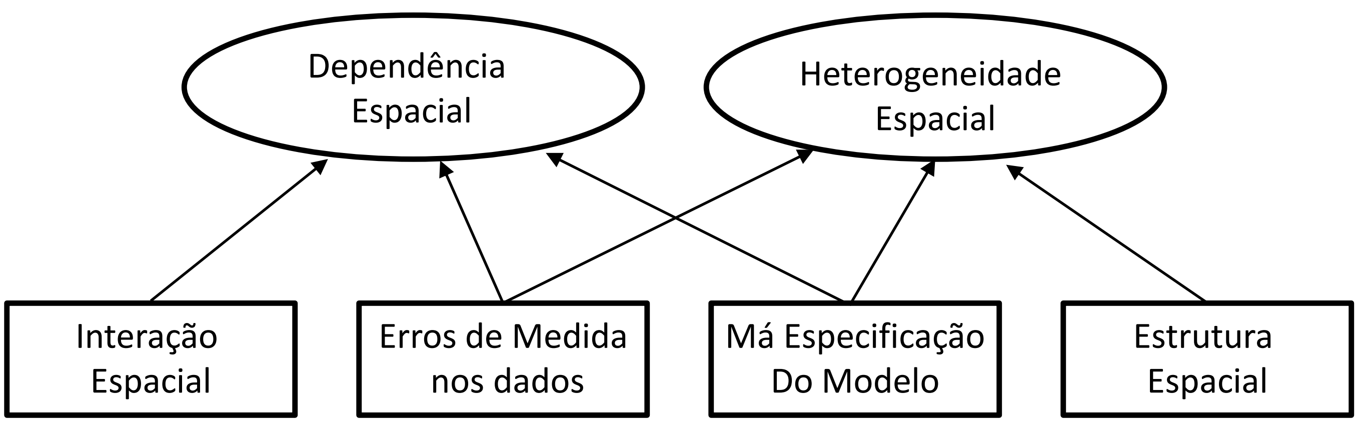
3. Efeitos Espaciais:

3.3 Imbricação dos Efeitos Espaciais

- A auto-correlação espacial gera heterogeneidade espacial e vice-versa.

Ex: Um modelo que desconsidere a heterogeneidade espacial nos parâmetros (β_i) produzirá resíduos com dependência espacial ($W\varepsilon$).

- Fontes de Imbricação:



3. Efeitos Espaciais:

3.4 Efeitos Espaciais e as Hipótese de Gauss-Markov

1) Linearidade dos parâmetros:

- A presença da heterogeneidade espacial geralmente demanda algum tipo de flexibilização associada aos coeficientes (de inclinação e/ou intercepto) a fim de que esta hipótese seja mantida.

- A econometria convencional dispõe de alguns recursos para lidar com este problema, como o uso de variáveis *dummies* ou modelos do tipo *switching regressions*.

- Apesar disso, existem modelos mais específicos, próprios da Econometria Espacial, que lidam com esta questão: Modelos com Regimes Espaciais, Modelos Multiníveis, RPG, entre outros.

3. Efeitos Espaciais:

3.4 Efeitos Espaciais e as Hipótese de Gauss-Markov

2. Não colinearidade perfeita:

- Ignorar Wy e WX, quando os mesmos forem relevantes, poderia enviesar os coeficientes estimados. Contudo, o uso destas defasagens espaciais pode gerar colinearidade imperfeita (multicolinearidade).

- Em grandes amostras, esta questão é atenuada pela elevada variação entre os dados. Todavia, em pequenas amostras, este problema pode afetar o erro padrão das estimativas, afetando as inferências e resultados obtidos.

3. Efeitos Espaciais:

3.4 Efeitos Espaciais e as Hipótese de Gauss-Markov

3. Média Condicional Zero:

- Ignorar os efeitos espaciais poderia violar esta hipótese das seguintes formas:

a) Omissão de defasagem espacial relevante

Exemplo: Omitir WX poderia gerar resíduos com auto-correlação espacial (WX passa a fazer parte de ε_i). Neste caso, ε_i apresentará dependência com X_i e X_j (viola Hip. 3.1 e 3.2).

b) Viés/endogeneidade associado ao termo Wy

Exemplo: Ignorar Wy, quando este é relevante, poderia gerar resultados enviesados. Todavia, como Wy determina y e vice-versa, a inclusão de ambas gera endogeneidade, tornando o EMQO enviesado e inconsistente (viola Hip. 3.3).

Solução: Incluir Wy e usar MQ2E.

3. Efeitos Espaciais:

3.4 Efeitos Espaciais e as Hipótese de Gauss-Markov

c. Erros de Medida

<u>Exemplo</u>: Mesmo após incluir as defasagens espaciais e contornar o problema da endogeneidade de Wy, ainda pode haver problema associado ao erro de medida. Por exemplo, a escolha errada da matriz espacial poderia comprometer sua capacidade de captar toda a auto-correlação do fenômeno estudado.

d. Adoção de apenas um nível hierárquico para um fenômeno multinível

<u>Exemplo</u>: Nota do aluno depende da sua inteligência e do nível educacional de seus pais (nível individual). Contudo, se a nota também depender da qualidade do ensino local e outras características da região do aluno (ex: vive em bairro barulhento x tranquilo), haverá omissão de variável relevante (no nível de grupo).

- Evitar os problemas “a”, “b”, “c” e “d” são condições necessárias (mas não suficientes) para que o EMQO seja não enviesado e consistente.

3. Efeitos Espaciais:

3.4 Efeitos Espaciais e as Hipótese de Gauss-Markov

4. Homocedasticidade:

- A presença e imbricação da heterogeneidade espacial e da auto-correlação espacial geralmente violam a hipótese de homocedasticidade.

5. Independência dos Erros:

- A inobservância dos efeitos espaciais podem fazer com que $Corr(\varepsilon_i \varepsilon_j) \neq 0$.

<u>Exemplo</u>: Omitir WX, por exemplo, faz com que os resíduos da região *i* se tornem: $\varepsilon_i = \beta WX + u_i$. Como WX contém informações de vizinhos comuns às regiões *i* e *j*, haverá dependência entre ε_i e ε_j (ambos dependem de WX).

<u>Nota</u>: Em séries de tempo, este problema gera perda de eficiência. Contudo, em processos espaciais, isto pode gerar viés e inconsistência.

❖ A <u>normalidade dos erros</u> não faz parte das Hip. de Gauss-Markov (ver Hip. 6).

Referência

1. ALMEIDA, E. *Econometria Espacial Aplicada.* 1ª ed. Alínea, 2012.