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

Econometria Espacial:

Capítulo 3 – Matrizes de Ponderação Espacial

Estrutura da Apresentação:

1. Introdução

- A autocorrelação espacial indica que a variável de interesse da região *i* (y_i) está associada ao valor desta mesma variável nas regiões vizinhas à *i* (Wy).

- Logo, a $Cov(y_i, y_j) \neq 0 \; para \; i \neq j$.

- Como a matriz *Var-Cov* tem dimensão nxn, contendo $(n^2 - n)$ elementos de covariância, seria impossível estimar todos os parâmetros de covariância.

Nota: números de parâmetros estimados crescem mais que o aumento da amostra (*problema do parâmetro incidental*).

- Solução: envolve o uso de uma matriz de pesos/ponderação espaciais (W) que permite a estimação de um único parâmetro representativo da "vizinhança".

- Problema: A escolha de W pode afetar consideravelmente os resultados. Logo, como definir W? Ou seja, que regiões serão consideradas "vizinhas"?

2. Tipologia das Matrizes de Ponderação Espacial (W)

- Uma Matriz de Ponderação Espacial (W) possui dimensão nxn, onde cada elemento w_{ij} representa o grau de conexão (*i.e.*: proximidade geográfica, econômica, política, cultural, social, etc.) entre as regiões *i* e *j*.

a) <u>**Matrizes de proximidade Geográfica**</u>: Geralmente são baseadas na contiguidade e/ou distância.

a.1) <u>Contiguidade</u>: matriz binária definida com base na fronteira entre *i* e *j*.

$$w_{ij} = \begin{cases} 1 \; se \; i \; e \; j \; são \; contíguos \\ 0 \; se \; i \; e \; j \; não \; são \; contíguos \end{cases} \tag{1}$$

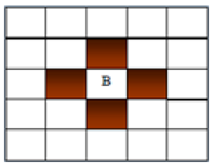
Convencionalmente, assume-se que $w_{ii} = 0$.

<u>Desvantagens</u>: a) conectividade desbalanceada (poucos vizinhos *x* muitos vizinhos); b) problema de “ilhas” (sem vizinhos).

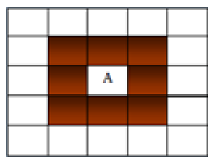
2. Tipologia das Matrizes de Ponderação Espacial (W)

- Exemplos de matrizes contíguas de primeira ordem:

A. Torre (*rook*)



B. Rainha (*queen*)



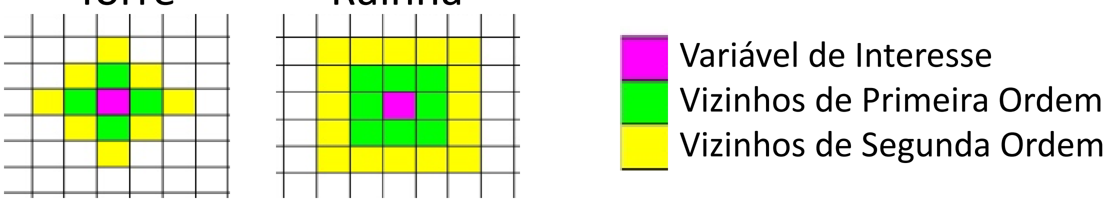
C. Matriz de Contigüidade para as Regiões do Brasil

	N	*NE*	*CO*	*SE*	*S*
N	0	1	1	0	0
NE	1	0	1	1	0
CO	1	1	0	1	1
SE	0	1	1	0	1
S	0	0	1	1	0

Regiões Brasil

- Exemplos de matrizes contíguas de segunda ordem:

Torre Rainha



<u>Nota</u>: **a)** Matriz C mostra que há multidirecionalidade de efeitos (ex: $\mathrm{N} \leftrightarrow SE$). Caso contrário seria uma matriz triangular (ex: apenas parte de cima da diagonal principal); **b)** é possível considerar ordens superiores (> 2).

2. Tipologia das Matrizes de Ponderação Espacial (W)

a.2) Distância Geográfica: matriz baseada na ideia de que regiões mais próximas geograficamente tem maior interação espacial.

- **Versão Binária** (k vizinhos mais próximos):

$$w_{ij}(k) = \begin{cases} 1 \; se \; d_{ij} \leq d_i(k) \\ 0 \; se \; d_{ij} > d_i(k) \end{cases} \qquad (2)$$

Onde: $d_i(k)$ representa a menor distância possível para que a região *i* possua *k* vizinhos.

- **Versão não Binária** (também baseada nos k vizinhos mais próximos):

$$w_{ij}(k) = \begin{cases} {}^{1}/_{d_{ij}} & se \; d_{ij} \leq d_i(k) \\ 0 & se \; d_{ij} > d_i(k) \end{cases} \qquad (3)$$

Na especificação (3), a influência de cada k vizinho é ponderada pela distância em relação à região *i*.

2. Tipologia das Matrizes de Ponderação Espacial (W)

Vantagem: Evita o desequilíbrio de conectividade (para qualquer região *i* haverá o mesmo número de vizinhos, *k*).

Desvantagem: Escolha de k é arbitrária.

Questão: Como definir k? Procedimento de Baumont (2004) – defasado!.

Passo 1: estima-se modelo via MQO.
Passo 2: testa-se a autocorrelação espacial nos resíduos via I de Moran usando as matrizes k=1, ... ,20.
Passo 3: usa-se a matriz com o maior I de Moran.

- **Distância Inversa e exponencial** (baseadas exclusivamente na distância geográfica – não há necessidade de definir k).

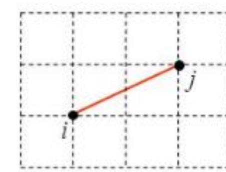
$$w_{ij} = d_{ij}^{-b} \tag{4}$$
$$w_{ij} = \exp(-bd_{ij}) \tag{5}$$

Problema 1: definição arbitrária do parâmetro de sensibilidade da distância (b). Geralmente adota-se b igual a 1 ou 2.

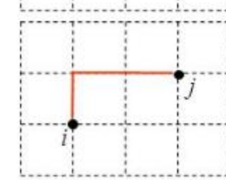
2. Tipologia das Matrizes de Ponderação Espacial (W)

Problema 2: quando $d_{ij} \rightarrow 0, w_{ij} \rightarrow \infty$.

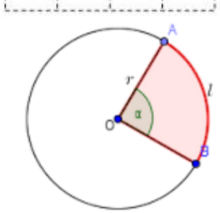
Problema 3: definição da métrica de distância.



a) Distância Euclidiana: $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$



b) Distância Manhattan: $d_{ij} = |x_i - x_j| + |y_i - y_j|$



c) Distância no Arco: $d_{ij} = Rcos^{-1}[sen\phi_i sen\phi_j + cos\phi_i cos\phi_j \cos(k_i - k_j)]$

Onde: x e y representam as coordenadas e abscissas em um plano cartesiano; R é o raio da Terra em torno do Equador (6378Km); ϕ e k são a latitude a longitude.

2. Tipologia das Matrizes de Ponderação Espacial (W)

- **Matriz de Distância Inversa com Corte de Distância Fixo:**

a) Não binária:

$$w_{ij}(d) = \begin{cases} {}^{1}/_{d_{ij}^{b}} & se\ d_{ij} \leq \bar{d} \\ 0 & se\ d_{ij} > \bar{d} \end{cases} \tag{6}$$

b) Binária:

$$w_{ij}(d) = \begin{cases} 1 & se\ d_{ij} \leq \bar{d} \\ 0 & se\ d_{ij} > \bar{d} \end{cases} \tag{7}$$

Nota: As matrizes “a” e “b” apresentam conectividade desbalanceada.

2. Tipologia das Matrizes de Ponderação Espacial (W)

- Matrizes de Distância Inversa:

Matriz de Distância Regional (Km)

	N	NE	CO	SE	S
N	0	2074	2017	2933	3854
NE	2074	0	2332	2660	3779
CO	2017	2332	0	926	1847
SE	2933	2660	926	0	1109
S	3854	3779	1847	1109	0

Referência: São Paulo (SE), Porto Alegre (S), Goiânia (CO), Recife (NE) e Belém (N).

Matriz de Distância Inversa – Não Binária

	N	NE	CO	SE	S
N	0	0.00048	0.00050	0.00034	0.00026
NE	0.00048	0	0.00043	0.00038	0.00026
CO	0.00050	0.00043	0	0.00108	0.00054
SE	0.00034	0.00038	0.00108	0	0.00090
S	0.00026	0.00026	0.00054	0.00090	0

Distância Inversa c/ Corte Dist. Fixo – não binária $(\bar{d} = 3000\ Km)$

	N	NE	CO	SE	S
N	0	0.00048	0.00050	0.00034	0.00000
NE	0.00048	0	0.00043	0.00038	0.00000
CO	0.00050	0.00043	0	0.00108	0.00054
SE	0.00034	0.00038	0.00108	0	0.00090
S	0.00000	0.00000	0.00054	0.00090	0

Distância Inversa c/ Corte Dist. Fixo – binária $(\bar{d} = 3000\ Km)$

	N	NE	CO	SE	S
N	0	1	1	1	0
NE	1	0	1	1	0
CO	1	1	0	1	1
SE	1	1	1	0	1
S	0	0	1	1	0

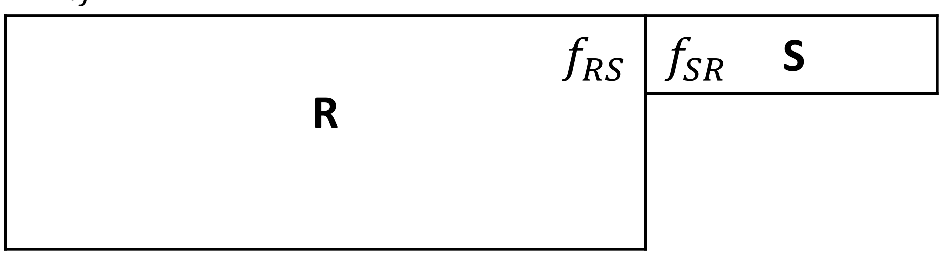
2. Tipologia das Matrizes de Ponderação Espacial (W)

- Analogamente às “matrizes de distância inversa”, também seria possível criar matrizes de “tempo de viagem inversa” ou de “custos monetários da viagem inversa”.

a.3) Matriz de Pesos Espaciais Gerais de Cliff e Ord: consideram tanto a distância quanto o tamanho da fronteira entre i e j.

$$w_{ij} = \frac{f_{ij}^{\phi}}{d_{ij}^{\delta}} \tag{8}$$

Ex:



Nota: $f_{RS} < f_{SR}$.

Problemas: a) Matriz não simétrica ($f_{RS} < f_{SR}$);
b) Definição arbitrária de ϕ e δ;
c) Conectividade desbalanceada (R tende a ter mais vizinhos que S).

2. Tipologia das Matrizes de Ponderação Espacial (W)

b) <u>Matrizes de proximidade Socioeconômica</u>: geralmente são baseadas em critérios de similaridade, dissimilaridade ou algum tipo de fluxo.

b.1) <u>Matriz socioeconômica (SE) baseada na similaridade</u>:

$$w_{ij}(d_{SE}) = \frac{1}{\left(|x_i - x_j|\right)^b} \tag{9}$$

<u>Ex</u>: Regiões com IDHs semelhantes apresentariam maior conectividade, independente da distância entre elas.

<u>Problemas</u>: a) Definição arbitrária de b;
b) $w_{ij} \to \infty$ quando $x_i \cong x_j$; Se $x_i = x_j$, w_{ij} fica indefinido.
c) Matriz pode se tornar endógena ao fenômeno analisado

Nota: a matriz sempre deve ser exógena. Caso contrário: possível viés e inconsistência.

- <u>Exemplo do problema “c”</u>: Usar matriz com base no IDH para analisar o PIB.

2. Tipologia das Matrizes de Ponderação Espacial (W)

b.2) Matriz socioeconômica (SE) baseada na dissimilaridade:

$$w_{ij}(d_{SE}) = \left|x_i - x_j\right| \qquad (10)$$

Ex: Guerra Fiscal – quanto maior a diferença entre o ICMS de i e j, maior será o fluxo de empresas entre estas regiões (fugindo do imposto elevado).

b.3) Matriz socioeconômica (SE) baseada em fluxos: baseada em fluxos migratórios, comerciais, entre outros.

→ Intensidade Comércio Bilateral

$$w_{ij}(CB) = \left(E_{ij} + M_{ij}\right)/(E_i + M_i) \qquad (11)$$

→ Intensidade Importação Bilateral

$$w_{ij}(IB) = M_{ij}/M_i \qquad (12)$$

3. Normalização da Matriz de Pesos Espaciais

- Geralmente, a normalização é realizada nas linhas da matriz. Formalmente:

$$w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}}, \text{ onde } \sum_j w_{ij}^* = 1 \quad (13)$$

<u>Desvantagem</u>: a) matriz se torna assimétrica $\left(w_{ij}^* \neq w_{ji}^*\right)$.
b) afeta o decaimento de influência em matrizes de distância inversa. **Nota**: evitar normalização nestes casos.

<u>Vantagens</u>: a) soma dos w_{ij}^* elementos da matriz normalizada é igual a *n*.
b) permite analisar o valor médio da vizinhança.

Exemplo de Normalização: Matriz Binária *Queen*					
	N	NE	CO	SE	S
N	0,00	0,50	0,50	0,00	0,00
NE	0,33	0,00	0,33	0,33	0,00
CO	0,25	0,25	0,00	0,25	0,25
SE	0,00	0,33	0,33	0,00	0,33
S	0,00	0,00	0,50	0,50	0,00

Nota: NE representa 50% dos vizinhos de N.

4. Operador de Defasagem Espacial

- Em séries de tempo, a defasagem temporal de uma variável ocorre em apenas uma direção.

Exemplo: Se Y representa o PIB de 2010, Y(-1) será o PIB de 2009.

- Contudo, na análise espacial a defasagem é multidirecional.

Exemplo (matriz Torre): Y(-1) seriam os vizinhos de primeira ordem de uma região nas regiões Norte, Sul, Leste, Oeste. Isto é: $Y_{i-1,j}$, $Y_{i,j+1}$, $Y_{i+1,j}$, $Y_{i,j-1}$.

Exemplo (matriz Rainha): Y(-1) seriam os vizinhos de primeira ordem de uma região em todas as direções.

$Y_{i-1,j-1}$	$Y_{i-1,j}$	$Y_{i-1,j+1}$
$Y_{i,j-1}$	Y_{ij}	$Y_{i,j+1}$
$Y_{i+1,j-1}$	$Y_{i+1,j}$	$Y_{i+1,j+1}$

4. Operador de Defasagem Espacial

- O operador de defasagem tem a capacidade de criar novas variáveis a partir de outras já existentes.

$$\boldsymbol{W_{nxn}} * \boldsymbol{y_{nx1}} = \boldsymbol{Wy_{nx1}}$$

$$\begin{bmatrix} 0{,}000 & 0{,}500 & 0{,}500 & 0{,}000 & 0{,}000 \\ 0{,}333 & 0{,}000 & 0{,}333 & 0{,}333 & 0{,}000 \\ 0{,}250 & 0{,}250 & 0{,}000 & 0{,}250 & 0{,}250 \\ 0{,}000 & 0{,}333 & 0{,}333 & 0{,}000 & 0{,}333 \\ 0{,}000 & 0{,}000 & 0{,}500 & 0{,}500 & 0{,}000 \end{bmatrix} * \begin{bmatrix} 50{,}6 \\ 144{,}1 \\ 76{,}5 \\ 636{,}4 \\ 193{,}5 \end{bmatrix} = \begin{bmatrix} 110{,}3 \\ 254{,}5 \\ 256{,}2 \\ 138{,}0 \\ 356{,}5 \end{bmatrix} \qquad \begin{bmatrix} N \\ NE \\ CO \\ SE \\ S \end{bmatrix} \tag{14}$$

Onde: $\boldsymbol{W}$ é a matriz rainha das regiões brasileiras, normalizada nas linhas; $\boldsymbol{y}$ representa o PIB (em R$ Bilhões) de cada região.

<u>Nota</u>: seria possível considerar que um região é vizinha de si mesma (incomum). Neste caso, $w_{ii} \neq 0$.

<u>Exemplo</u>: Na Equação 14, se $w_{ii} \neq 0$, tem-se que $w_{11} = w_{12} = w_{13} = 0{,}333$.

5. Propriedades das Matrizes de Ponderação Espacial

- As características desejáveis de W são:

a) $0 \leq w_{ij} < \infty$

- Não existe distância negativa nem infinita.

b) $\sum_j w_{ij}^* \neq 0$

- Matriz não deve conter "ilhas", principalmente quando a ilha em questão interage com as demais regiões, *e.g.*: Japão (omissão de variável relevante).

c) $w_{ii} = 0$

- Uma região não pode ser vizinha de si mesma. Esta hipótese ainda evita maiores problemas de endogeneidade e multicolinearidade.

d) $E(w_{ij}, \varepsilon) = 0$

- Matriz deve ser exógena ao fenômeno estudado (Ex.: Distância entre i e j afeta PIB_i, mas PIB_i não afeta distância entre i e j $\Rightarrow$ matriz exógena). Caso contrário: Inconsistência.

6. Escolha da Matriz de Ponderação Espacial

- Não há um consenso sobre a escolha de W. Geralmente, opta-se por:

a) Procedimento de Baumont (2004) - sujeito à críticas:

a.1) Estima-se o modelo clássico de regressão linear;
a.2) Testa-se a autocorrelação espacial nos resíduos, via I de Moran, com várias matrizes W.
a.3) Utiliza-se a matriz W que obteve o maior I de Moran.

b) Outra possibilidade (bastante trabalhosa) consiste em estimar diversos modelos considerando a maior quantidade de matrizes possíveis. Feito isto, define-se o melhor modelo com base nos critérios AIC e/ou BIC.

- Sugestão: Usar Mix de (a) e (b). Defina algumas matrizes via procedimento (a) e siga para o procedimento (b) com um número reduzido de matrizes.

Referência

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