ECONOMETRIA Espacial APLICADA

Eduardo Almeida

Alínea
EDITORA

DIRETOR GERAL
Wilon Mazalla Jr.

COORDENAÇÃO EDITORIAL
Willian F. Mighton

COORDENAÇÃO DE REVISÃO E COPYDESK
Catarina C. Costa

REVISÃO DE TEXTOS
Bianca Quiento Fernandes

EDITORAÇÃO ELETRÔNICA
Fabio Diego da Silva
Tatiane de Lima

CAPA
Paloma Leslie

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Almeida, Eduardo
Econometria espacial aplicada / Eduardo Almeida. -- Campinas, SP : Editora Alínea, 2012.

Bibliografia

1. Econometria 2. Espaço na economia I. Título.

12-07510 CDD-330.015195

Índices para Catálogo Sistemático

1. Econometria espacial aplicada 330.015195

ISBN 978-85-7516-601-7

Todos os direitos reservados à

Editora Alínea

Rua Tiradentes, 1053 - Guanabara - Campinas-SP
CEP 13023-191 - PABX: (19) 3232.9340 e 3232.0047
www.atomoealinea.com.br

Impresso no Brasil

Apoio cultural

FAPEMIG
Fundação de Amparo à Pesquisa do Estado de Minas Gerais

3.1 Introdução

A autocorrelação espacial significa que o valor de uma variável de interesse numa certa região i tende a estar associado ao valor dessa variável nas regiões vizinhas j. Em dados na forma de corte transversal, a autocorrelação espacial pode ser expressa pela matriz de variância-covariância:

$$Cov(y_i, y_j) = E(y_i y_j) + E(y_i)E(y_j) \neq 0 \quad \text{para} \quad i \neq j \tag{3.1}$$

Como essa matriz tem dimensão n por n, contendo n^2–n elementos de covariância, existem duas abordagens possíveis aqui para mensurar a dependência espacial. Uma solução é a abordagem direta da Geoestatística em que todos os pares de observações são classificados a partir da distância que os separa, usando o variograma, que expressa a força da autocorrelação como função contínua da distância.[56] Na abordagem paramétrica, por outro lado, a ideia de estimar os termos de covariância da matriz não se mostra possível. Isto implica um número alto de interações para que se possa levar em conta, uma vez que o número de parâmetros a ser estimado crescem mais que o tamanho da amostra. A isso se dá o nome de problema do parâmetro incidental, inviabilizando a estimação de tantos parâmetros (Anselin; Bera, 1998). Para resolver esse problema, tornando-o tratável na abordagem paramétrica, é preciso impor um arranjo para a ocorrência das interações espaciais entre as regiões. Havendo tal arranjo, o objetivo é reduzir a quantidade de parâmetros a ser estimada. Na verdade, o que se pretende é ter de estimar parcimoniosamente alguns poucos parâmetros – de preferência, apenas um único – que forneçam o grau de interação espacial.[57]

[56] Variograma é definido como sendo a variância da diferença entre uma variável observada em duas regiões diferentes, separadas por uma determinada distância. É suposto que o variograma seja função crescente dessa distância separadora. Existem várias funções crescentes em que o variograma pode se basear, tais como as funções exponencial, gaussiana, esférica, entre outras (Fotheringham et al., 2000; Johnston et al., 2001).

[57] Existem ainda duas outras abordagens para lidar com a autocorrelação espacial presente nos dados: a filtragem espacial e a abordagem das variáveis latentes com base no modelo de equações estruturais. Essas duas abordagens voltarão a ser mencionadas no resumo do capítulo.

Com tal intuito, especifica-se uma matriz de ponderação espacial (W) que procura refletir um determinado arranjo espacial das interações resultantes do fenômeno a ser estudado. Regiões mais conectadas entre si interagem mais do que regiões menos conectadas. Cada conexão entre duas regiões é representada numa célula desta matriz, sendo denominada de peso espacial. A questão crucial é saber o que usar para mensurar o grau de conexão entre as regiões. Uma medida deste grau de conexão costuma ser a proximidade entre as regiões. Convém notar que esta proximidade não precisa seguir um critério apenas geográfico, podendo seguir um critério socioeconômico ou ainda outro critério qualquer relevante para o fenômeno sendo estudado.[58]

Na abordagem paramétrica, há dois problemas recorrentes na determinação da matriz de ponderação espacial a ser adotada no estudo. Em primeiro lugar, muitas vezes a escolha da matriz W é arbitrária, dado que não existe um teste formal para defini-la (Plaigin, 2009). Em segundo lugar, existe o problema da sensibilidade dos resultados à escolha desta matriz. Por exemplo, os parâmetros espaciais estimados numa regressão podem ser sensíveis à escolha da matriz de pesos espaciais. Portanto, a discussão a respeito da tipologia das matrizes, compondo um leque de matrizes W disponíveis ao pesquisador – e qual delas escolher com propriedade – são questões importantes e espinhosas na literatura da econometria espacial.

Este capítulo está organizado da seguinte forma: na próxima seção, é apresentada a tipologia de matrizes de ponderação espacial; na terceira seção, é discutida a normalização de uma matriz W; na quarta seção, é introduzido o conceito de operador de defasagem espacial, importante para se calcular as variáveis defasadas espacialmente a serem colocadas posteriormente na regressão para se controlar a dependência espacial; na quinta seção, discutem-se as propriedades desejáveis que uma matriz W deveria ostentar; a sexta seção discute procedimentos para se escolher uma matriz de ponderação espacial de forma não arbitrária. Na derradeira seção são tecidas as considerações finais.

[58] Plaigin (2009) adota pesos culturais e pesos institucionais no contexto da análise de convergência de renda entre países, inovando em tentar encontrar um grau de conexão baseado na proximidade cultural ou institucional. Na elaboração da matriz de pesos culturais, levam-se em conta fatores como religião, línguas, sistema legal, entre outros. Na preparação da matriz de pesos institucionais, considera-se o índice de governança, composto por aspectos como estabilidade política e ausência de violência, qualidade regulatória, cumprimento das leis, controle da corrupção, liberdade de expressão, liberdade de imprensa e liberdade de associação, entre outros. Plaigin (2009) faz a comparação destas matrizes com matrizes de pesos espaciais mais convencionais, tais como matrizes de contiguidade e de proximidade geográfica, usando a análise exploratória de dados espaciais.

3.2 Tipologia das matrizes de ponderação espacial (W)

Uma matriz de ponderação espacial é uma matriz quadrada de dimensão n por n. Os pesos espaciais w_{ij} representam o grau de conexão entre as regiões segundo algum critério de proximidade, mostrando a influência da região j sobre a região i. Assim, a matriz W é útil por realizar uma espécie de ponderação da influência que as regiões exercem entre si.

O grau de conexão expresso nas matrizes de pesos espaciais[59] pode ser classificado de acordo com um critério geográfico ou socioeconômico. O critério de conexão geográfica da matriz de pesos espaciais apoia-se na ideia de proximidade, que, por sua vez, pode ser definida de acordo com a contiguidade e/ou com a distância geográfica segundo uma determinada métrica.

3.2.1 Matrizes de proximidade geográfica

3.2.1.1 Contiguidade

A matriz de pesos espaciais binários pode ser construída em consonância com a ideia de vizinhança baseada na contiguidade, em que duas regiões são vizinhas, caso elas partilhem de uma fronteira física comum.[60] A ideia é que duas regiões contíguas possuem uma maior interação espacial. Com base neste conceito de contiguidade, é atribuído um valor unitário na matriz a duas regiões vizinhas; caso contrário, atribui-se um valor nulo.[61] Formalmente:

$$w_{ij} = \begin{cases} 1 & \text{se } i \text{ e } j \text{ são contíguos} \\ 0 & \text{se } i \text{ e } j \text{ não são contíguos} \end{cases} \tag{3.2}$$

Convencionalmente, é presumido que $w_{ii} = 0$, ou seja, a região não é considerada como vizinha de si própria, implicando que a matriz de contiguidade possua a sua diagonal principal composta por valores nulos.[62] Apesar da aparente simplicidade deste conceito, escondem-se várias possibilidades para definir contiguidade, conforme distintas convenções. O problema reside em como se define o conceito de fronteira geográfica por intermédio da observação de um mapa. O mapa é uma mera representação abstrata da real configuração geográfica. Por isso, contém erros de medida. Levando em conta esses erros, e em alusão ao movimento de peças num tabuleiro de xadrez, a convenção de contiguidade é dita ser *rainha* (*queen*), caso, além das fronteiras com extensão diferente de zero, puderem ser considerados os vértices como contíguos, na visualização de um mapa. Caso apenas as fronteiras físicas com extensão diferente de zero entre as regiões sejam levadas em conta, a convenção de contiguidade é considerada como *torre* (*rook*). Se apenas os vértices forem considerados para definir a contiguidade, a convenção é denominada *bispo* (*bishop*). As duas primeiras convenções são as mais utilizadas na literatura.[63] As diferentes convenções para a matriz de pesos espaciais binários são mostradas nas Figuras 3.1 até 3.3, em que os vizinhos das regiões A, B e C, respectivamente, estão hachurados.

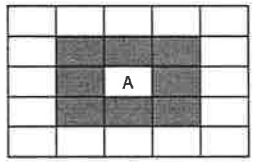

Figura 3.1. Convenção "rainha" de contiguidade.

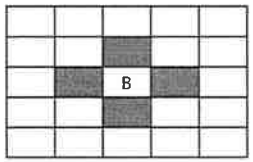

Figura 3.2. Convenção "torre" de contiguidade.

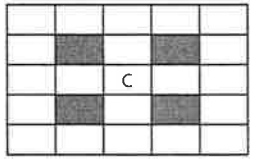

Figura 3.3. Convenção "bispo" de contiguidade.

[59] Matriz de ponderação espacial e matriz de pesos espaciais são usadas como sinônimos.

[60] Pinkse e Slade (1998) apresentam uma interessante matriz W binária no seu estudo sobre postos de combustíveis nas ruas de Vancouver no Canadá. Os pesos foram definidos de acordo com o seguinte critério: se os postos de combustíveis i e j estão na mesma rua recebem o valor unitário; e zero, caso contrário. Os autores chamaram esses pesos de *street weights*.

[61] A matriz de contiguidade foi a primeira matriz de ponderação espacial da literatura, e foi proposta nos estudos de Moran (1948) e Geary (1954).

[62] Apesar de ser pouco comum, na literatura admite-se que a região i seja vizinha de si própria ao se elaborar uma matriz de ponderação espacial em que a diagonal principal seja composta por valores unitários.

[63] Para outras convenções de contiguidade para matrizes de pesos espaciais binários, ver LeSage (1999).

A título de ilustração, a Tabela 3.1 mostra a matriz de pesos espaciais binários, cinco por cinco, das macrorregiões brasileiras segundo a convenção rainha ou torre,[64] composta por 25 elementos:

Tabela 3.1. Matriz binária de pesos espaciais para as macrorregiões brasileiras (Convenção Rainha ou Torre).

	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

Nota-se que, como a região Sudeste (SE) faz fronteira com a região Nordeste (NE), ou em outros termos, são contíguas, as respectivas células (quarto elemento da segunda linha ou o quarto elemento da segunda coluna) assumem o valor unitário. Por seu turno, a região Norte (N) não faz fronteira com a região Sul (S), logo as respectivas células (quinto elemento da primeira linha ou o quinto elemento da primeira coluna) assumem o valor nulo, representando a ausência de contiguidade. Cabe destacar que nenhuma região é contígua a si própria, assim, os elementos da diagonal principal são valores nulos.[65] Convém, ainda, observar que a matriz de contiguidade binária é simétrica no sentido de que a influência que a região i exerce sobre a região j é a mesma que esta exerce sobre aquela.

O conceito da multidirecionalidade do espaço está também representado na matriz de ponderação espacial, uma vez que as interações ocorrem em todas as direções. Se a interação espacial fosse unidirecional, de tal sorte que a interação ocorresse somente na direção Sul-Norte, a região Sul influenciaria o Sudeste, mas esta última região não exerceria nenhuma influência sobre a primeira. A consequência disso seria que a matriz de ponderação espacial seria triangular, composta por elementos não negativos apenas nas células acima da diagonal principal.

A desvantagem da matriz binária de contiguidade reside no fato de que não é garantida uma conectividade balanceada, uma vez que pode haver regiões com muitos vizinhos, ao passo que podem existir regiões com poucos vizinhos. Por exemplo, pela observação das Figuras 3.1, 3.2 e 3.3, percebe-se que uma região numa matriz de pesos

[64] Nesse caso particular, a convenção rainha coincide com a convenção torre.

[65] Como será visto mais adiante, esta convenção pode ser modificada a fim de permitir que as regiões sejam contíguas de si próprias, fazendo com que $w_{ii} = 1$.

binários para um mesmo gride regular pode ter de um a oito vizinhos, dependendo da convenção de contiguidade adotada e de onde esta região se localiza no gride.[66]

Uma vantagem comum a todas as matrizes de pesos espaciais de contiguidade é a possibilidade de definir, com mais clareza, contiguidades de ordens superiores. Uma matriz de contiguidade de primeira ordem é composta dos vizinhos diretos das regiões. Uma matriz de contiguidade de segunda ordem, por sua vez, é composta dos vizinhos dos vizinhos das regiões, isto é, os vizinhos de segunda ordem ou vizinhos indiretos, sem que esses vizinhos de segunda ordem sejam, eles próprios, vizinhos de primeira ordem. Para entender o conceito de vizinhança de segunda ordem, com a devida exclusão das redundâncias ou circularidades, considera-se a matriz binária de pesos espaciais conforme a convenção rainha dos estados do Brasil. Os vizinhos de primeira ordem do estado de São Paulo são Minas Gerais, Rio de Janeiro, Paraná e Mato Grosso do Sul. Os vizinhos de segunda ordem do estado de São Paulo são Santa Catarina (vizinho do Paraná), Bahia (vizinho de Minas Gerais), Espírito Santo (vizinho do Rio de Janeiro e Minas Gerais) e Goiás (vizinho de Minas Gerais). Porém, o estado do Mato Grosso do Sul, que é vizinho de primeira ordem de Minas Gerais – e, lembre-se, este é um vizinho de primeira ordem de São Paulo –, não é vizinho de segunda ordem de São Paulo porque já é vizinho de primeira ordem de São Paulo, sendo, portanto, redundante. Uma matriz de vizinhos de segunda ordem deve ter apenas vizinhos de segunda ordem e não vizinhos de primeira ordem, como é o caso de Mato Grosso do Sul para São Paulo. Os vizinhos de primeira ordem numa matriz de contiguidade de segunda ordem são chamados de redundantes.

Além de matrizes de segunda ordem, é possível construir matrizes de terceira ordem, quarta ordem e assim por diante. Na construção das matrizes binárias de contiguidade de ordem superior, é recomendável eliminar as redundâncias antes de usá-las para o teste de hipótese ou a estimação de modelos econométrico-espaciais, uma vez que a existência de redundâncias pode afetar as propriedades dos estimadores (Anselin, 2006).

3.2.1.2 Distância Geográfica

Outro critério de proximidade na definição dos pesos espaciais é a distância geográfica. A ideia por trás é que duas regiões próximas geograficamente têm uma maior interação espacial. Uma matriz W muito adotada na literatura é a matriz dos k vizinhos mais próximos, $w_{ij}(k)$. Trata-se de uma matriz binária cuja convenção de

[66] Uma região em qualquer canto do gride, usando a convenção bispo para a matriz de pesos binários, tem apenas um vizinho.

proximidade é baseada na distância geográfica, medida em quilômetros ou milhas. Formalmente:

$$w_{ij}(k) = \begin{cases} 1 & \text{se } d_{ij} \leq d_i(k) \\ 0 & \text{se } d_{ij} > d_i(k) \end{cases} \tag{3.3}$$

em que $d_i(k)$ é a distância de corte para a região i especificamente, a fim de que esta região i tenha k vizinhos. De novo, é assumido que $w_{ii}(k) = 0$, por convenção. Mais precisamente, $d_i(k)$ é a menor distância para a região i a fim de que ela possua exatamente k vizinhos. E esta distância de corte varia de região para região, por isso, o subscrito i em $d_i(k)$. Assim, a Expressão 3.3 informa que a proximidade é baseada num critério de distância de tal sorte que duas regiões são consideradas vizinhas, caso encontrem-se dentro de uma distância de corte necessária para que se tenha o número predeterminado de vizinhos.

Um exemplo de uma matriz de dois vizinhos mais próximos para as macrorregiões brasileiras é dado pela Tabela 3.2. A distância de corte para definir os dois vizinhos mais próximos da região Norte foi diferente que a distância crítica adotada para se determinar os dois vizinhos mais próximos da região Sudeste e assim por diante.

Tabela 3.2. Matriz de dois vizinhos mais próximos para as regiões brasileiras.

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

A vantagem desta convenção é combater o desequilíbrio da conectividade de uma matriz, pois todas as unidades espaciais terão o mesmo número de vizinhos. Ademais, essa matriz garante que não haverá "ilhas", ou seja, regiões com nenhum vizinho. É possível ainda construir uma matriz de k vizinhos sem ser binária, mas baseada na distância inversa (Tyszler, 2006):

$$w_{ij}(k) = \begin{cases} 1/d_{ij} & \text{se} \quad d_{ij} \leq d_i(k) \\ 0 & \text{se} \quad d_{ij} > d_i(k) \end{cases} \tag{3.4}$$

De novo, $d_i(k)$ é a distância de corte para a região i especificamente, a fim de esta região i tenha k vizinhos. Todavia, agora, a influência de algum desses k vizinhos não é a mesma sobre a região i, sendo, aliás, diferente dependendo do inverso da distância que os separa.

Uma questão espinhosa a ser enfrentada na construção deste tipo de matriz é como definir k, ou seja, quantos vizinhos mais próximos considerar? Com o intuito de tornar menos arbitrária esta definição, e com o objetivo de capturar o máximo de autocorrelação espacial contida no fenômeno sendo estudado, Baumont (2004) propõe o seguinte procedimento em três passos:

1. Roda-se o modelo clássico de regressão linear por mínimos quadrados ordinários (MQO);
2. Testam-se os resíduos para autocorrelação espacial por intermédio do valor da estatística de I de Moran,[67] usando L matrizes de k vizinhos mais próximos e variando L de $k = 1$ a $k = 20$;
3. Define-se k que tenha gerado o maior valor do I de Moran, significativo estatisticamente.

Outra matriz de ponderação espacial de natureza geográfica é aquela que define proximidade com base na distância inversa a partir dos pontos centroides dos polígonos representando as regiões ou baseada na distância da principal cidade da região. A ideia intuitiva pode ser estabelecida como sendo a seguinte: quanto mais distantes duas regiões estiverem, menor será a interação entre elas. Genérica e formalmente:

$$w_{ij}(d_{ij}) = f(d_{ij}) \qquad f' < 0 \tag{3.5}$$

Os pesos espaciais, $w_{ij}(d_{ij})$, são uma função da distância entre as regiões i e j. Vale destacar que a função f pode assumir algumas especificações, tais como:

a) Função de distância inversa:

$$w_{ij} = d_{ij}^{-b} \tag{3.6}$$

b) Função de distância exponencial:

$$w_{ij} = \exp(-bd_{ij}) \tag{3.7}$$

sendo que b é um parâmetro de amortecimento da influência da distância sobre a força de interação. Quanto maior b, menor a força de interação entre duas regiões.

[67] O I de Moran é uma estatística que mensura a autocorrelação espacial contida numa variável, como será visto com mais detalhes no próximo capítulo.

Um problema com esse tipo de matriz é que o parâmetro b é, muitas vezes, determinado arbitrariamente e *a priori*, independentemente do restante da análise.[68] Todavia, o principal problema da convenção da distância surge quando d_{ij} aproxima-se de zero, fazendo que $w_{ij}(d_{ij})$ torne-se muito grande, aproximando-se do infinito.

Outra questão refere-se à métrica adotada para a definição da distância. A métrica mais usada é a distância euclidiana. Outras métricas são a distância Manhattan e a distância do arco, também chamada de "grande círculo".

A distância euclidiana na reta é definida como:

$$d_{ij} = \sqrt{[(u_i - v_j)^2 + (u_i - v_j)^2]} \tag{3.8}$$

em que, neste caso, u_i é a ordenada e v_i é a abscissa, projetadas no mapa.

A distância Manhattan é expressa como:

$$d_{ij} = |u_i - u_j| + |v_i - v_j| \tag{3.9}$$

A distância no arco ou grande círculo é a menor distância entre qualquer par de pontos sobre um caminho na superfície de uma esfera (Plaigin, 2009). Por sua vez, a distância no arco é computada como:

$$d_{ij} = R\cos^{-1}[sen\phi_1 sen\phi_2 + \cos\phi_1\cos\phi_2\cos(\kappa_1 - \kappa_2)] \tag{3.10}$$

em que R é o raio da Terra em torno do Equador (6.378 Km), ϕ é a latitude e κ é a longitude.

É válido tecer algumas observações sobre as diferentes métricas. A distância euclidiana é uma forma aproximada de se calcular as distâncias entre duas regiões em linha reta, não se levando em conta a curvatura da Terra e, com isso, subestimando a verdadeira distância que separa as regiões. Contudo, a distorção é pequena. Na prática, adota-se a métrica euclidiana quando se tem as distâncias baseadas nas coordenadas. Por outro lado, adota-se a métrica do arco quando se tem a informação de latitude e longitude para definir a distância que separa as regiões sobre a superfície de uma esfera como o globo terrestre.

A matriz de pesos espaciais baseada na distância inversa pode ser a melhor representação do arranjo da interação espacial por representar o amortecimento do efeito da distância sobre o fenômeno em estudo, fazendo com o que está mais próximo dependa mais do que aquilo que está mais distante. Ou seja, observações mais próximas revelam uma maior autocorrelação espacial do que observações que se encontram mais distantes. Para se construir a matriz de distância inversa para as regiões brasileiras, em primeiro lugar, é preciso ter em mãos a matriz de distâncias,

[68] É comum calibrar o parâmetro como sendo $b = 1$ ou $b = 2$ na Equação 3.6.

em quilômetros, entre essas regiões, com base na principal cidade de cada região,[69] conforme Tabela 3.3:

Tabela 3.3. Matriz de distâncias (em Km) entre as regiões brasileiras.

	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

Usando a especificação da distância inversa, como expressa pela Equação 3.6, com o parâmetro b sendo estipulado em um e com a métrica euclidiana, obtém-se a matriz, representada pela Tabela 3.4. Em primeiro lugar, cabe observar que os pesos são pequenos nesta matriz W. A maior interação é verificada entre as regiões Sudeste e Centro-Oeste, enquanto a menor interação é observada entre as regiões Norte e Sul. Também, nesta matriz, considera-se que uma região não influencia a ela mesma, e, com isso, a diagonal principal é composta por valores nulos.

Tabela 3.4. Matriz de distância inversa entre as regiões brasileiras.

	N	NE	CO	SE	S
N	0,00000	0,00048	0,00050	0,00034	0,00026
NE	0,00048	0,00000	0,00043	0,00038	0,00026
CO	0,00050	0,00043	0,00000	0,00108	0,00054
SE	0,00034	0,00038	0,00108	0,00000	0,00090
S	0,00026	0,00026	0,00054	0,00090	0,00000

Uma variante deste tipo de ponderação espacial é representada por uma matriz também baseada na distância inversa, mas com uma especificação particular dos pesos espaciais:

$$w_{ij}(d) = \begin{cases} 1/d_{ij}^b & \text{se} \quad d_{ij} \leq \bar{d} \\ 0 & \text{se} \quad d_{ij} > \bar{d} \end{cases} \tag{3.11}$$

[69] Foram consideradas arbitrariamente as seguintes cidades para representar cada região: São Paulo (Sudeste), Porto Alegre (Sul), Goiânia (Centro-Oeste), Recife (Nordeste) e Belém (Pará).

em que $\overline{d}$ é a distância de corte fixada e o restante da notação permanece como antes. Até a distância de corte $\overline{d}$, a influência da região j sobre a região i se dá pela distância inversa. A partir de $\overline{d}$, considera-se que a influência torna-se nula. Como se depreende da ausência do subscrito i, a distância de corte $\overline{d}$ não varia para a região i individualmente, sendo, portanto, a mesma para todas as regiões da amostra, ao contrário do que ocorre com a matriz de k vizinhos mais próximos. A desvantagem deste tipo de matriz para regiões com grande área é que a distância de corte fixada pode deixar estas regiões com poucos vizinhos. Por outro lado, para regiões de pequena área, esta distância limite fixa pode determinar muitos vizinhos. Isso leva ao problema da ausência de balanceamento da conectividade, já discutido anteriormente.

A título de exemplo, a Tabela 3.5 reporta os pesos espaciais com essa especificação para as regiões brasileiras, estipulando como a distância de corte fixada em três mil quilômetros:

Tabela 3.5. Matriz da distância de corte fixa das regiões brasileiras.

	N	NE	CO	SE	S
N	0,00000	0,00048	0,00050	0,00034	0,00000
NE	0,00048	0,00000	0,00043	0,00038	0,00000
CO	0,00050	0,00043	0,00000	0,00108	0,00054
SE	0,00034	0,00038	0,00108	0,00000	0,00090
S	0,00000	0,00000	0,00054	0,00090	0,00000

Pode-se, ainda, construir uma matriz binária de pesos espaciais baseada na distância de corte fixa.

$$w_{ij}(d) = \begin{cases} 1 & \text{se} \quad d_{ij} \leq \overline{d} \\ 0 & \text{se} \quad d_{ij} > \overline{d} \end{cases} \tag{3.12}$$

Novamente, $\overline{d}$ é a distância de corte fixa e o restante da notação permanecendo com o mesmo sentido anteriormente apresentado. A Tabela 3.6 reporta os pesos espaciais binários com base na distância de corte fixa de três mil quilômetros. As regiões vizinhas que se localizam abaixo desta distância de corte fixa recebem o valor unitário como peso espacial. Caso contrário, o peso espacial é estabelecido como zero, indicando ausência de interação espacial.

Tabela 3.6. Matriz binária da distância de corte fixa das regiões brasileiras.

	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

Nota-se que como a região Sudeste está dentro do raio de três mil quilômetros em torno da cidade de Belém, que representa a região Norte, portanto, na célula correspondente, o peso espacial assume o valor um. Já a região Sul está fora deste raio e, consequentemente, o peso espacial que denota a interação entre essas duas regiões assume o valor nulo.

Uma alternativa à matriz de distância inversa é a matriz de tempo de viagem inversa.[70] O tempo de viagem entre pares de regiões pode ser uma medida mais refinada da dificuldade de acessibilidade inter-regional. Por exemplo, para um estudo sobre a localização de atividades econômicas, o tempo de viagem de uma determinada região em relação ao principal mercado consumidor é uma variável relevante. Isso é conseguido construindo uma matriz de tempos de viagem. Com base nessa matriz, e analogamente ao que foi feito nas Tabelas 3.4, 3.5 e 3.6, é possível calcular a matriz de tempo de viagem inversa, a matriz de pesos espaciais baseados no tempo de viagem de corte fixo e a matriz binária de pesos espaciais baseados no tempo de viagem de corte fixo.

3.2.1.3 Matriz de pesos espaciais gerais de Cliff e Ord

Intuitivamente, é razoável supor que regiões que compartilham maior extensão de fronteira entre si tenham uma interação maior. Do mesmo modo, quanto mais próximas duas regiões se encontram, maior a interação entre si. Estas duas forças geográficas indutoras de interação estão condensadas na matriz de pesos espaciais gerais, proposta por Cliff e Ord (1981). Os pesos espaciais gerais ou pesos Cliff-Ord consistem no comprimento relativo da fronteira comum, ajustado pela distância inversa entre as duas observações. Formalmente, os pesos Cliff-Ord podem ser expressos como:

[70] Um outra alternativa é construir uma matriz de custos monetários. Com base nessa matriz, é possível construir posteriormente a matriz inversa de custos monetários ou uma matriz de custo monetário de corte, como foi feito com a matriz de distância de corte fixo (Ribeiro, 2011).

$$w_{ij} = \frac{f_{ij}^{\phi}}{d_{ij}^{\delta}} \tag{3.13}$$

em que f_{ij} é a proporção da fronteira comum entre as observações i e j no perímetro de i, e ϕ e δ são parâmetros a serem definidos. Convém notar que f_{ij} não é necessariamente igual a f_{ji}, como pode ser observado na Figura 3.4, supondo que haja duas regiões, uma região maior R e outra menor S:

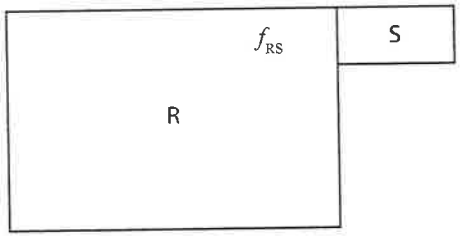

Figura 3.4. Representação dos pesos espaciais de Cliff-Ord.

Claramente, tem-se que $f_{RS}<f_{SR}$. Isto é, a proporção da fronteira comum entre as unidades espaciais R e S com relação ao perímetro de R (f_{RS}) é menor que a proporção desta fronteira comum no perímetro de S (f_{SR}). Isso obviamente acarreta que a matriz W dos pesos Cliff-Ord não seja simétrica. Se não forem vizinhos, tem-se que $f_{ij}= 0$ e, portanto, $w_{ij}= 0$. Uma desvantagem desta matriz repousa no fato de que é necessário obter valores para dois parâmetros ϕ e δ, e não apenas para um. Ademais, os valores destes parâmetros são, frequentemente, determinados arbitrariamente. Outra desvantagem é que, com esse tipo de matriz W, não se garante a conectividade balanceada. Assim, regiões com uma fronteira mais extensa tendem a ter mais vizinhos do que outras regiões com fronteira menor.

3.2.2 Matrizes de proximidade socioeconômica

Em alguns casos, o arranjo para a interação espacial pode ser mais bem representado por forças socioeconômicas do que por forças geográficas, tais como contiguidade ou distância. Existem matrizes socioeconômicas de três naturezas:

1. As matrizes baseadas no conceito da similaridade;
2. As matrizes com base no conceito da dissimilaridade; e
3. Matrizes baseadas em fluxos.

No que tange ao primeiro tipo de matrizes socioeconômicas, o conceito que define os seus pesos espaciais é o de similaridade, ou seja, quem é mais similar, influencia mais do que os mais diferentes. As distâncias neste tipo de matriz são dadas por valores denotando a diferença de similaridade. Nas matrizes socioeconômicas, é possível expressar a similaridade entre regiões pela diferença da composição demográfica (por exemplo, porcentagem de mulheres na população), diferenças de indicadores sociais ou econômicos (renda *per capita*, taxa de desemprego, IDH, proporção de pobres etc.) ou uma potência dessa diferença entre duas regiões i e j (Tyszler, 2006). Por exemplo, se usar a renda *per capita* como distância econômica, então se duas regiões tiverem rendas *per capita* muito semelhantes, terão uma interação elevada, independentemente de estarem próximas ou distantes geograficamente.

A ideia de se apoiar em critérios socioeconômicos para a construção das matrizes de pesos espaciais pode se valer da noção de distância inversa, já introduzida na subseção anterior. Contudo, na elaboração da matriz W, os pesos espaciais, $w_{ij}(d_{SE})$, são definidos conforme a interação baseada na distância socioeconômica (d_{SE}). Formalmente:

$$w_{ij}(d_{SE}) = \frac{1}{|x_i - x_j|} \tag{3.14a}$$

ou

$$w_{ij}(d_{SE}) = \frac{1}{(x_i - x_j)^b} \tag{3.14b}$$

em que, agora, x representa uma variável que mede a distância socioeconômica e b é um número positivo par.

Na Equação 3.14a, usa-se o módulo da diferença a fim de garantir que os pesos não sejam negativos. Em 3.14b, nota-se que b denota a potência à qual a diferença é elevada. Um exemplo desse tipo de matriz W, usando como medida de distância econômica o Índice de Desenvolvimento Humano (IDH) de 2000, em módulo, foi elaborada para as regiões brasileiras, conforme a Tabela 3.7.

Tabela 3.7. Matriz de distância socioeconômica de similaridade para as regiões brasileiras.

	N	NE	CO	SE	S
N	0,000	20,199	14,823	15,160	12,139
NE	20,199	0,000	8,549	8,660	7,582
CO	14,823	8,549	0,000	666,667	67,039
SE	15,160	8,660	666,667	0,000	60,914
S	12,139	7,582	67,039	60,914	0,000

A maior interação é encontrada entre as regiões Sudeste e Centro-Oeste, ao passo que a menor interação é entre as regiões Nordeste e Sul. A matriz é simétrica, significando que a influência que o Sudeste exerce sobre o Centro-Oeste é a mesma que o Centro-Oeste impõe sobre o Nordeste, medida em 666,7.

Apesar do grande apelo de se considerar como medida da força da interação algum critério que não sejam fatores como a distância geográfica e a contiguidade, é preciso ter pleno conhecimento dos problemas que podem surgir desta decisão. Em primeiro lugar, é importante evitar o problema derivado da situação em que a mesma variável que define a distância socioeconômica na matriz de pesos espaciais seja inserida no modelo da regressão como variável dependente (Anselin, 1988). Os pesos espaciais são presumidamente exógenos, a não ser que exista um motivo teórico para o contrário. É preciso evitar situações em que os pesos sejam endógenos na regressão (Anselin; Bera, 1998). Outro problema diz respeito à distância zero, quando $x_i = x_j$. Por exemplo, se o critério de distância for a proporção de pobres, pode ocorrer que tanto a região i quanto a região j tenham a mesma proporção de pobres. Logo, nesta célula da matriz, tem-se uma divisão por zero, que não é definida matematicamente. Mesmo que não se tenha exatamente $x_i = x_j$, mas que os valores sejam muito próximos, os pesos podem ser muito elevados, caracterizando pesos explosivos, com consequências deletérias na estimação dos parâmetros e na computação de estatísticas de autocorrelação espacial. A checagem visual da matriz e a computação de algumas características da matriz, tais como o número de elementos não nulos em relação ao número total de elementos da matriz e o autovalor são medidas recomendáveis para se evitar a utilização de matrizes inapropriadas nas etapas de especificação, estimação e de testes de hipóteses do trabalho econométrico.

Agora no que concerne às matrizes socioeconômicas de dissimilaridade, o princípio que guia a sua construção é o seguinte: regiões dissimilares possuem maior conectividade entre si. Um exemplo de aplicação deste tipo de matriz é a guerra fiscal em que regiões estão disputando alíquotas de imposto (Tyszler, 2006). A especificação dos pesos espaciais é:

$$w_{ij}(d_{SE}) = |x_i - x_j| \tag{3.15}$$

Assim, um maior peso para a influência da região j sobre i é dado para a maior dissimilaridade na medida de distância socioeconômica. Um exemplo de matriz de distância socioeconômica, baseada na dissimilaridade, usando como medida o IDH para as regiões brasileiras, conforme Tabela 3.8.

Tabela 3.8. Matriz de distância socioeconômica de dissimilaridade para as regiões brasileiras.

	N	NE	CO	SE	S
N	0,0000	0,0495	0,0675	0,0660	0,0824
NE	0,0495	0,0000	0,1170	0,1155	0,1319
CO	0,0675	0,1170	0,0000	0,0015	0,0149
SE	0,0660	0,1155	0,0015	0,0000	0,0164
S	0,0824	0,1319	0,0149	0,0164	0,0000

Outro tipo de matriz socioeconômica é aquela que usa fluxos em vez de diferença de indicadores sociais ou econômicos. Uma matriz deste tipo pode usar, por exemplo, fluxos econômicos para definir os pesos espaciais (w_{ij}), tais como correntes de comércio (exportações mais importações) ou fluxos migratórios. Assim, quanto mais comércio entre duas regiões, maior deve ser sua interação espacial. Por exemplo, Cabrer-Borrás e Serrano-Domingo (2007), analisando os fatores explicativos da atividade inovativa entre as regiões da Espanha, definiram a proximidade entre regiões com base em suas relações comerciais. Deste modo, a matriz de ponderação espacial foi composta pelos seguintes pesos: $w_{ij} = 1$ se a intensidade em suas relações comerciais é maior que a média, caso contrário, $w_{ij} = 0$. Os autores construíram duas medidas para tal intensidade. A primeira medida é a intensidade de fluxos de comércio bilateral (IF_{CB}):

$$IF_{CB} = (E_{ij} + M_{ij})/(E_i + M_i) \tag{3.16}$$

em que E_{ij} são as exportações da região i para a região j, M_{ij} representam as importações da região i, enviadas pela região j; E_i são as exportações totais da região i, ao passo que M_i são as importações totais da região i.

A segunda medida foi a intensidade de parcelas de importação (IF_{PI}), expressa como:

$$IF_{PI} = (M_{ij}/M_i) \tag{3.17}$$

Com base nestas medidas de intensidade, são calculadas matrizes binárias de interação comercial.[71]

[71] Outro exemplo de uso de matriz de fluxos, mas agora migratórios, é provido pelo trabalho de De la Mata et al. (2009).

3.3 Normalização da matriz de pesos espaciais

Um conceito importante na econometria espacial é a matriz de ponderação espacial normalizada na linha.[72] Em termos formais, a normalização da matriz de pesos espaciais pode ser expressa como:

$$w_{ij}^{*} = \frac{w_{ij}}{\sum_{j} w_{ij}} \tag{3.18}$$

$$\sum_{j} w_{ij}^{*} = 1 \tag{3.19}$$

As Equações 3.18 e 3.19 informam que, para normalizar uma matriz de ponderação espacial, é necessário dividir as células representando os pesos espaciais de cada linha de uma matriz W pela somatória dos pesos das respectivas linhas. A normalização na linha da matriz de pesos espaciais produz, com frequência, assimetria, significando, por exemplo, que a região i exerce uma influência distinta sobre a região j do que esta sobre a primeira região, ou seja, $w_{ij} \neq w_{ji}$. A Tabela 3.9 reporta as fórmulas de normalização para alguns pesos espaciais vistos neste capítulo.

Tabela 3.9. Fórmulas de normalização para matrizes W.

Tipo de Matriz W	Fórmula de normalização
Binária (rainha, torre ou bispo)	$w_{ij}^{*} = w_{ij} \Big/ \sum_{j} w_{ij}$
k vizinhos mais próximos	$w_{ij}^{*}(k) = w_{ij}(k) \Big/ \sum_{j} w_{ij}(k)$
Distância inversa	$w_{ij}^{*}(d) = w_{ij}(d) \Big/ \sum_{j} w_{ij}(d)$
Distância socioeconômica similar	$w_{ij}^{*}(d_{SE}) = w_{ij}(d_{SE}) \Big/ \sum_{j} w_{ij}(d_{SE})$

Em que pese isso, a relevância da normalização da matriz W reside em dois aspectos. Em primeiro lugar, uma matriz normalizada na linha significa que a soma de suas linhas e colunas é limitada a um número finito n, a saber, o tamanho da

[72] Apesar de ser pouco adotada em trabalhos aplicados, há ainda a normalização na coluna, em que cada célula da matriz W é dividida pelo total da coluna, e a normalização pelo autovalor (*eigenvalue*) da matriz, em que o valor de cada célula é dividido pelo autovalor da matriz W (Ribeiro, 2011).

amostra. Em segundo lugar, a matriz de ponderação espacial normalizada permite a interpretação de média dos valores da variável nos vizinhos para a defasagem espacial. A interpretação de média dos valores vizinhos é a chave para se definir posteriormente o conceito de defasagem espacial tanto para a variável dependente (y_i) quanto para as variáveis explicativas (X) e os termos de erros (ξ e ε). Neste sentido, a normalização presta-se a uma espécie de suavização espacial dos dados. Além disso, segundo Tyszler (2006), a normalização da matriz W facilita a interpretação do peso espacial w_{ij}, pois indica qual é a porcentagem da influência total que é exercida pela região j sobre a região i.

A despeito dessas vantagens, Anselin (1988) alerta que a noção de decaimento da influência à medida que a distância geográfica entre as regiões aumenta, contida essencialmente na matriz de distância inversa, pode ser perdida com a normalização. A normalização desta matriz na linha pode não fazer sentido, pois tal amortecimento do efeito da distância desaparece. Neste caso particular, é melhor trabalhar com a matriz não normalizada na linha.

Tabela 3.10. Matriz de pesos binários de contiguidade do tipo rainha normalizada na linha para as macrorregiões brasileiras.

	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-se na Tabela 3.10 que a matriz de pesos espaciais binários, baseados na contiguidade do tipo rainha, normalizada na linha, exibe a propriedade de que todas as suas linhas somem o valor unitário. A interpretação desses pesos normalizados pode ser feita do seguinte modo: a influência do Nordeste sobre a região Norte responde por 50% de toda a influência sofrida por esta última região e assim por diante.

3.4 Operador de defasagem espacial

O operador de defasagem espacial desempenha um papel fundamental na metodologia da econometria espacial porque o controle para a dependência espacial é realizado por intermédio da inclusão de variáveis defasadas espacialmente no lado direito do modelo de regressão.

Vale a pena traçar uma comparação com o conceito de defasagem temporal, encontrado na literatura de séries de tempo. Autocorrelação em séries de tempo significa correlação entre o valor de uma variável no período t e o período t-h, em que h é a defasagem temporal. Por exemplo, em séries de tempo, se y_{2010} é o PIB em 2010, B^2y_{2010} é o PIB dois períodos para trás, ou seja, o PIB em 2008. Aqui, B refere-se ao operador de defasagem em séries de tempo.

No domínio do espaço, por conta da multidirecionalidade, o significado do operador de defasagem é muito diferente. Isto ocorre porque não se tem uma clara definição, sem incorrer em ambiguidades, do operador de defasagem espacial que desloca h regiões no espaço da variável em análise. Como exemplo, considera-se uma variável qualquer $y_{i,j}$, localizada num gride regular, tal como nas Figuras 3.1, 3.2 ou 3.3. Esta variável pode ser "deslocada" no espaço de diversas maneiras: para o Norte, Sul, Leste, Oeste, Sudeste, Nordeste etc. Usando o conceito de convenção da torre, os deslocamentos possíveis são $y_{i-1,j}$, $y_{i,j-1}$, $y_{i+1,j}$ e $y_{i,j+1}$. Para a convenção rainha, graficamente, existem oito possíveis deslocamentos no espaço para y_{ij}, como se pode observar na Figura 3.5:

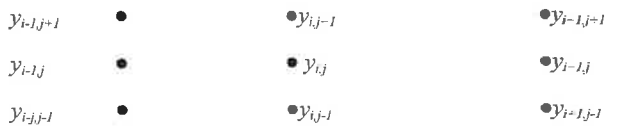

Figura 3.5. Defasagens espaciais num gride regular.
Fonte: Adaptado de Anselin (1988, p. 22).

Assim, o conceito de operador de defasagem espacial não é tão simples e direto como o operador de defasagem em séries de tempo devido à natureza multidirecional do processo de interação no espaço. Na verdade, o significado do operador de defasagem espacial de uma variável y, formalmente Wy, é a média do valor dessa variável nas regiões vizinhas. O operador de defasagem espacial tem a capacidade de transformar variáveis em outras variáveis, só que, agora, variáveis defasadas espacialmente. Como exemplo, computa-se a defasagem espacial do PIB macrorregional (Wy) como exemplificada na Figura 3.6:

$$\underset{W}{\begin{pmatrix} 0{,}000 & 0{,}500 & 0{,}500 & 0{,}000 & 0{,}000 \\ 0{,}333 & 0{,}000 & 0{,}333 & 0{,}333 & 0{,}000 \\ 0{,}250 & 0{,}250 & 0{,}000 & 0{,}250 & 0{,}250 \\ 0{,}000 & 0{,}333 & 0{,}333 & 0{,}000 & 0{,}333 \\ 0{,}000 & 0{,}000 & 0{,}500 & 0{,}500 & 0{,}000 \end{pmatrix}} \underset{y}{\begin{pmatrix} 50{,}6 \\ 144{,}1 \\ 76{,}5 \\ 636{,}4 \\ 193{,}5 \end{pmatrix}} = \underset{Wy}{\begin{pmatrix} 110{,}3 \\ 254{,}5 \\ 256{,}2 \\ 138{,}0 \\ 356{,}5 \end{pmatrix}}$$

Figura 3.6. O operador de defasagem espacial.

A primeira matriz da Figura 3.6 diz respeito à matriz normalizada na linha de pesos espaciais W da Tabela 3.10. O vetor y refere-se ao PIB, em bilhões de reais, das regiões Norte, Nordeste, Centro-Oeste, Sudeste e Sul. O último vetor na Figura 3.6, apresenta a defasagem espacial do PIB macrorregional, isto é, o PIB médio das regiões vizinhas (Wy), em bilhões de reais. O valor de 110,3 para a defasagem espacial Wy na região Norte representa a média do PIB na região vizinha do Nordeste (144,1) e do PIB da região Centro-Oeste (76,5). É importante notar que o operador de defasagem (W) quando multiplicado por uma variável (y), transforma-se numa outra variável (Wy). Como se verá a partir do quinto capítulo, utiliza-se este conceito para definir defasagens tanto na variável dependente (Wy), quanto na variável independente (WX) e defasagem nos termos de erro ($W\xi$ e $W\varepsilon$). A interpretação continua sendo a média ponderada da variável nos vizinhos.

Se a matriz de ponderação espacial considerar que as regiões são vizinhas entre si e, portanto, a diagonal principal da matriz é composta por valores unitários, a defasagem espacial continua sendo interpretada como média dos valores da variável nos vizinhos, mas agora com a região i fazendo parte desta média. A Figura 3.7 retrata a situação com uma matriz de pesos binários do tipo rainha, mas levando em conta que $w_{ii} = 1$.

$$\underset{W}{\begin{pmatrix} 0{,}333 & 0{,}333 & 0{,}333 & 0{,}000 & 0{,}000 \\ 0{,}250 & 0{,}250 & 0{,}250 & 0{,}250 & 0{,}000 \\ 0{,}200 & 0{,}200 & 0{,}200 & 0{,}200 & 0{,}200 \\ 0{,}000 & 0{,}250 & 0{,}250 & 0{,}250 & 0{,}250 \\ 0{,}000 & 0{,}000 & 0{,}333 & 0{,}333 & 0{,}333 \end{pmatrix}} \underset{y}{\begin{pmatrix} 50{,}6 \\ 144{,}1 \\ 76{,}5 \\ 636{,}4 \\ 193{,}5 \end{pmatrix}} = \underset{Wy}{\begin{pmatrix} 90{,}4 \\ 226{,}9 \\ 220{,}2 \\ 262{,}6 \\ 302{,}1 \end{pmatrix}}$$

Figura 3.7. O operador de defasagem espacial com matriz W com as regiões como vizinhas de si mesmas.

Evidentemente, os valores da defasagem espacial calculados na Figura 3.7 são diferentes daqueles computados na Figura 3.6. Por exemplo, o valor de 90,4 para a

defasagem espacial Wy na região Norte representa a média do PIB na região vizinha do Nordeste (144,1), do PIB da região Centro-Oeste (76,5) e do PIB da própria região Norte (50,6). A interpretação dos outros valores é análoga.

3.5 Propriedades das matrizes de ponderação espacial

Como já foi visto, a possibilidade de construção de matrizes de ponderação espacial é ampla. Agora, tendo em mente a classificação das diferentes matrizes W, é importante apresentar um conjunto de propriedades desejáveis que seja compartilhado por uma matriz. Este conjunto de propriedades desejáveis é o seguinte:

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

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

c) $w_{ii} = 0$

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

A primeira propriedade impõe que os pesos espaciais sejam não negativos e finitos. A condição de não negatividade é óbvia. Não há sentido econômico em se ter uma interação negativa entre duas regiões. A condição de os pesos serem finitos é importante pela necessidade de se invocar as propriedades assimptóticas dos estimadores, como a Máxima Verossimilhança, e dos testes de diagnóstico. Em relação a isso, segundo Kelejian e Prucha (1999), quando o tamanho da amostra tende ao infinito, a soma das linhas e das colunas da matriz deve ser limitada uniformemente em valor absoluto, isto é, deve se aproximar de um número finito.[73] Além das matrizes binárias de contiguidade, a matriz de distância inversa também atende a esta última condição. Essa talvez seja a principal razão pela qual matrizes de ponderação espacial de natureza geográfica sejam as mais usadas na econometria espacial.

A segunda condição impede que a matriz W possua "ilhas" na base de dados espaciais. Neste caso, "ilhas" são as regiões que ficam isoladas, ou seja, que não são contíguas a nenhuma outra. Na matriz, essa situação é representada por alguma linha ou coluna que contenha apenas zeros. A frequência disto é maior com matrizes binárias de contiguidade de acordo com as convenções rainha e torre. Em que pese isso não invalide a estimação ou a construção dos testes, a existência de ilhas representa perda de graus de liberdade, pois a região desconectada não é levada em conta no modelo espacial (Anselin; Bera, 1998). Contudo, o principal problema de se ter ilhas em

[73] Lee (2004) mostrou que a condição dos pesos finitos pode ser menos restrita ainda: basta que a soma das linhas e das colunas não divirja para o infinito a uma taxa igual ou mais rápida do que a taxa de crescimento do tamanho da amostra de dados em corte transversal.

uma matriz de pesos espaciais, significando que ela não tem interação com as outras regiões, é quando essa ilha é importante. Por exemplo, ao se estudar a convergência de renda dos países, a linha da matriz referente ao Japão seria constituída de elementos nulos, significando que não haveria interação espacial com os outros países, o que não é verossímil.

A terceira propriedade convenciona que os elementos da diagonal principal sejam nulos ($w_{ii} = 0$). Por que convencionalmente é assumido isso? A resposta repousa em duas razões. Em primeiro lugar, por uma questão de lógica, ou seja, nenhuma região pode exercer uma influência sobre si mesma. Em segundo lugar, a adoção dessa convenção implica uma facilidade computacional. Para entender isso, convém perceber que, uma vez que se calcula frequentemente o traço da matriz de ponderação espacial, e como o traço é definido como a somatória dos elementos da diagonal principal da matriz, se estes forem nulos, o traço assumirá, consequentemente, o valor nulo também, facilitando uma série de contas. Em que pese isso, para certos casos, pode ser que seja interessante que a diagonal principal da matriz W seja composta por elementos unitários ($w_{ii} = 1$), implicando que a matriz W seja somada com a matriz identidade I_n. Conforme pode ser observado na Figura 3.7, a consequência disso é que a defasagem espacial da variável dependente (Wy) é agora uma média entre os valores da variável dependente nos vizinhos e na própria região i (Haining, 2003). O operador de defasagem baseado neste tipo de matriz, usado para construir as variáveis defasadas espacialmente (Wy, WX, $W\xi$ e $W\varepsilon$), também inclui a ideia de média, mas incluindo nesta média a própria região i.[74]

A quarta condição estabelece que os pesos devam ser exógenos. Se os pesos são determinados endogenamente, isso significa que os pesos influenciam a variável dependente e, por sua vez, esta variável influencia os valores dos pesos. Essa causalidade reversa tem potencial para viesar as estimativas e torná-las inconsistentes, mesmo que o tamanho da amostra se eleve ao infinito. Esse perigo de se ter pesos endógenos é maior quando se usa matrizes socioeconômicas, pois a variável dependente também tem uma natureza socioeconômica. Esta é uma vantagem das matrizes de pesos espaciais binários baseadas no critério geográfico. Existe a certeza de poder contar com pesos exógenos, afastando essa fonte de má especificação do modelo econométrico.

Em relação à simetria dos pesos espaciais, tal propriedade numa matriz W significa que a região i exerce a mesma influência sobre j do que esta sobre a região i. No caso específico de matrizes de pesos binários isto significa que $w_{ij} = 1$, mas, por sua

[74] O *software* SpaceStat 2.0 é capaz de construir matrizes de ponderação espacial em que a diagonal principal é composta por uns, considerando a região como vizinha de si própria.

vez, $w_{ji} = 1$. Em casos mais gerais de matrizes W, a ausência de simetria envolveria que $w_{ij} = a$ e $w_{ji} \neq a$, sendo que a é um número positivo qualquer pertencente ao conjunto dos números reais positivos, denotando a força da interação. Como já foi visto, existem, no entanto, muitos tipos de matrizes que não compartilham da propriedade da simetria. Na verdade, não se pode afirmar que a propriedade da simetria seja desejável. Isso ocorre porque não é razoável supor que a influência de uma região sobre outra tenha de ser rigorosamente a mesma. Efetivamente, é esperado que tais influências sejam diferentes. Por exemplo, a cidade do Rio de Janeiro supostamente deve exercer uma influência maior sobre Nova Iguaçu, localizada dentro de sua Região Metropolitana, do que a influência de Nova Iguaçu exerce sobre a Cidade do Rio de Janeiro. Assim, ter o mesmo peso na matriz W pode não fazer muito sentido.

Antes de seguir em diante, cabe aqui uma nota de alerta. A esmagadora maioria das matrizes W encontradas na literatura tem os seus pesos construídos num contexto estritamente espacial. Porém, além da dimensão espacial, o processo gerador de dados espaciais pode ter a dimensão temporal. Ou seja, os dados espaciais são coletados em diferentes períodos de tempo na forma de cortes transversais agrupados. Neste caso, definir uma matriz de pesos somente espaciais pode desconsiderar a dinâmica temporal das interações entre as regiões. Assim sendo, Dubé e Legros (2012) desenvolveram uma matriz de pesos espacial-temporais, levando em conta a importância da dimensão do tempo para computar a dependência espacial. Para fazer isso, considera-se que a matriz W de pesos espacial-temporais de dimensão n por n pode ser decomposta em outras duas, também de mesma dimensão: a matriz S com pesos espaciais e a matriz T com pesos temporais. A matriz S é especificada com pesos espaciais (s_{ij}) no critério de distância inversa, como definido na equação (3.11). Analogamente, a matriz T de pesos temporais é especificada como:

$$t_{ij} = \begin{cases} |v_i - v_j|^{-a} & \text{se} \quad |v_i - v_j| < \bar{v} \\ 1 & \text{se} \quad v_i = v_j \quad \forall i \neq j \\ 0 & \text{se} \quad v_i = v_j \quad \forall i = j \end{cases} \tag{3.20}$$

em que v_i é o período que a observação i ocorre, ao passo que v_j é o período quando a observação j acontece; a diferença $v_i - v_j$ representa o lapso de tempo entre as observações i e j, enquanto que $\bar{v}$ é um valor de corte. O parâmetro a fornece o amortecimento ao longo do tempo. Para que os pesos temporais sejam finitos e não negativos, apura-se a diferença absoluta em módulo. Convém notar que o peso temporal t_{ij} assume o valor unitário quando as observações i e j ocorrem ao mesmo

tempo. Ademais, adota-se a convenção de se considerar a diagonal principal da matriz T composta por valores nulos.

Agora para se obter a matriz W de pesos espacial-temporais, aplica-se o operador de produto matricial Hadamard, ou seja, $W = S \odot T$, gerando pesos do tipo $s_{ij} \cdot t_{ij}$. A utilização de matriz de ponderação espacial-temporal é adequada para dados na forma de cortes transversais agrupados, em que as observações não são as mesmas ao longo do tempo. Dubé e Legros (2012) usam esta matriz de ponderação espacial--temporal para computar a dependência espacial dentro de um modelo hedônico para investigar o mercado imobiliário de Quebec, onde as residências observadas são diferentes ao longo do tempo.

3.6 Escolha da matriz de ponderação espacial

Diante desta ampla gama de matrizes de pesos espaciais, tanto do tipo geográfico quanto do tipo socioeconômico, cabe a pergunta: qual matriz de pesos espaciais deve-se usar? Idealmente, para não ser uma fonte de má especificação do modelo econométrico, a matriz de ponderação espacial deveria ser construída com o intuito de capturar toda a autocorrelação espacial subjacente ao fenômeno em estudo. Contudo, na prática, isso raramente ocorre, o que faz com que a escolha de uma adequada matriz W envolva, às vezes, um processo de tentativa e erro.[75] Em consonância com isso, como salientado por Plaigin (2009), não há ainda testes formais para determinar qual é a matriz de ponderação espacial ótima.

A sugestão da escolha da matriz a ser utilizada deve se originar, em primeiro lugar, das características do fenômeno em estudo. Em segundo lugar, os pesos espaciais da matriz W precisam atender às propriedades apresentadas na seção anterior. Mesmo assim, é possível que o pesquisador fique em dúvida entre duas ou mais matrizes de ponderação espacial. A escolha da matriz de ponderação espacial, que fornece o grau de conectividade entre as regiões, é uma possível fonte de má especificação para o modelo. Isso ocorre quando se escolhe uma matriz W que capture pouca autocorrelação espacial existente no fenômeno estudado. Assim, quando se usa tal matriz como operador de defasagem espacial para se construir as variáveis Wy, WX, $W\xi$ e $W\varepsilon$, essas variáveis serão medidas com erro sistemático, podendo viesar a estimação e torná-la inconsistente. Além disso, uma escolha inadequada da matriz W

[75] Segundo Anselin (1995), "em princípio, essa matriz deveria ser de tal forma que capturasse toda a dependência espacial. Na prática, isso nunca é conhecido com certeza e um certo grau de tentativa e erro pode ser necessário (usando a estatística de autocorrelação espacial para várias ordens de contiguidade ou bandas de distâncias)" (p. 49).

no sentido de captar pouca autocorrelação espacial faz com que os resíduos do modelo continuem apresentando dependência espacial, mesmo depois da incorporação das variáveis defasadas espacialmente no modelo de regressão para tentar controlar o problema. Para evitar este tipo de má especificação, é importante a seleção da matriz W, que capture a maior autocorrelação espacial possível do fenômeno em estudo, fazendo com que as defasagens espaciais sejam medidas corretamente.

Para contornar uma possível arbitrariedade da escolha da matriz W, existe um procedimento de especificação de pesos espaciais, baseado naquele proposto por Baumont (2004). O procedimento tem por escopo tentar capturar o máximo da dependência espacial por intermédio de teste de diagnóstico. O procedimento possui três passos:

1. Estima-se o modelo clássico de regressão linear;
2. Testam-se os resíduos desse modelo para a autocorrelação espacial, usando o I de Moran para um conjunto de matrizes W;
3. Seleciona-se a matriz de pesos espaciais que tenha gerado o mais alto valor do teste I de Moran, e que seja significativo estatisticamente.

3.7 Resumo

Na literatura econométrica, existem algumas abordagens para lidar com a autocorrelação espacial. Neste capítulo, apresentou-se a abordagem mais adotada nos trabalhos aplicados de econometria espacial, ou seja, a abordagem da matriz de ponderação espacial W. Existe a abordagem direta da Geoestatística, baseada no variograma, e usado principalmente com variáveis contínuas. Há ainda a abordagem da filtragem espacial, cuja ideia é remover a dependência espacial das variáveis espacialmente autocorrelacionadas e aplicar posteriormente as técnicas econométricas convencionais às variáveis filtradas tanto com dados em corte transversal quanto em dados em painel (Griffith, 2010; Parent; LeSage, 2011).[76] Existe também a abordagem mais recentemente apresentada na literatura de se levar em conta a autocorrelação espacial com a construção de variáveis latentes, que são variáveis que não podem ser observadas diretamente (Folmer; Oud, 2008; Liu et al., 2010). Todavia, essas duas últimas abordagens alternativas ainda não estão consolidadas na literatura, existindo uma pequena quantidade de trabalhos que as

[76] Na exposição da teoria ao longo dos capítulos, o princípio da filtragem espacial será usado algumas vezes, sobretudo em alguns estimadores que usam uma espécie de procedimento de Cochrane-Orcutt espacial, apresentados no sexto capítulo, e na estimação do modelo dinâmico de painel de dados espacial no capítulo catorze.

utiliza com regularidade. Indubitavelmente, a abordagem mais utilizada continua sendo a construção de matrizes de ponderação espacial.

Na abordagem paramétrica, a necessidade de se construir uma matriz de ponderação espacial surge a fim de pôr uma configuração na interação espacial dos agentes. Existem vários tipos de matrizes baseadas na proximidade, tais como as matrizes binárias de contiguidade nas convenções rainha, torre, bispo e k vizinhos mais próximos ou nas matrizes de pesos espaciais com base na distância inversa. As matrizes de ponderação espacial também podem ser construídas com base no conceito socioeconômico ou em outro conceito relevante para a análise, tais como critérios políticos, culturais ou institucionais.

A escolha da matriz mais adequada deve respeitar certas propriedades desejáveis e certas particularidades do estudo em questão. A sugestão da escolha da matriz a ser usada deve vir, em primeiro lugar, das características do fenômeno em estudo. Em segundo lugar, a matriz precisa atender às propriedades apresentadas na Seção 3.5. Como as possibilidades são muitas na construção de matrizes econômicas, é recomendável que se evitem matrizes extravagantes, que não atendam às propriedades desejáveis para uma matriz W. Por atender a tais propriedades, as matrizes W geográficas (rainha, torre, k vizinhos mais próximas e de distância inversa) são as mais adotadas na econometria espacial aplicada.

É aconselhável que o pesquisador adote algum procedimento para tornar a determinação da matriz W menos arbitrária. Na literatura, existe o procedimento de Baumont (2004), que busca escolher uma matriz W que capte a maior parte da dependência espacial. De qualquer modo, às vezes, é válido estimar os modelos econométricos usando mais de uma matriz de ponderação espacial e comparar os resultados, a fim de detectar discrepâncias significativas nos resultados, expondo a sensibilidade desses à escolha da matriz, checando, assim, a sua robustez.

A importância de uma boa escolha da matriz de ponderação espacial reside no fato de que esta matriz é usada na operação de construir variáveis defasadas espacialmente, a saber, Wy, WX, $W\varepsilon$ e $W\xi$, que serão inseridas no lado direito da regressão a fim de controlar a dependência espacial. Uma matriz W que capture pouca autocorrelação espacial subjacente ao fenômeno em estudo levará a erros de mensuração nessas defasagens espaciais. Se o erro de mensuração existente nessas defasagens espaciais for correlacionado pelo menos com uma variável explicativa, o estimador usado no modelo de regressão espacial será viesado e inconsistente.

No próximo capítulo, as matrizes de ponderação espacial serão usadas para se realizar a análise exploratória de dados espaciais, cujo objetivo é fazer com que o pesquisador conheça melhor a base de dados espaciais com a qual trabalha.