



Universidade de Brasília
Instituto de Ciências Exatas

Departamento de Estatística

Dissertação de Mestrado

**Análise da dependência espacial no contexto de
dados em painel: O caso espaço-temporal**

por

Calebe de Oliveira Figueiredo

Orientador: Prof. Dr. Alan Ricardo da Silva

Janeiro de 2011

Calebe de Oliveira Figueiredo

Análise da dependência espacial no contexto de dados em painel: O caso espaço-temporal

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Brasília, Janeiro de 2011

It is not selfish to think for oneself.

A man who doesn't think for himself doesn't think at all.

Oscar Wilde

Agradecimentos

À minha família pelo apoio em todo o tempo e por ter me mostrado a importância do conhecimento.

Ao meu orientador, Alan Ricardo da Silva, cuja paciência e boa vontade foram decisivas para a realização deste trabalho.

Agradeço, também, ao Professor George von Borries por ter me ensinado os primeiros e valiosos passos na pesquisa.

Agradeço, por fim, à Lígia Freitas, pelo apoio e incentivo que fizeram esta jornada bem mais interessante por ter sido compartilhada.

Sumário

Lista de Figuras	iv
Lista de Tabelas	v
Resumo	1
Abstract	2
Introdução	3
1 Dados em Painel	5
1.1 Introdução	5
1.2 O Modelo Combinado	7
1.3 O Modelo de Efeitos Fixos	8
1.3.1 Modelo com 1 efeito fixo	9
1.3.2 O procedimento <i>Within Transformation</i>	10
1.3.3 Modelo com 2 efeitos fixos	11
1.3.4 Estimação de parâmetros via Máxima Verossimilhança	12
1.4 O Modelo de Efeitos Aleatórios	14
1.4.1 O modelo com 1 efeito aleatório	14
1.4.2 O modelo de 2 efeitos aleatórios	17
1.4.3 Estimando parâmetros via máxima verossimilhança	18
1.5 Testes de Hipóteses	19
1.5.1 Testando a adequabilidade do modelo combinado	20
1.5.2 Testando a existência de efeitos aleatórios	22
1.5.3 Testando a existência de efeitos fixos	23

1.5.4	O teste de especificação de Hausman	24
1.6	O Modelo de Hausman-Taylor	25
2	Econometria Espacial	28
2.1	A matriz de proximidades \mathbf{W}	29
2.2	Indicadores globais de dependência espacial	29
2.2.1	O índice I de Moran	29
2.2.2	O índice C de Geary	30
2.3	O modelo espacial autoregressivo de primeira ordem	30
2.4	O modelo espacial regressivo-autoregressivo	32
2.4.1	O modelo SAR : aspectos probabilísticos	32
2.4.2	O modelo SAR : aspectos computacionais	34
2.5	O modelo espacial de erros autoregressivos	34
2.5.1	O modelo SEM : aspectos computacionais	36
2.6	O modelo espacial Durbin	37
2.7	O modelo $MESS$	38
2.7.1	Estimando parâmetros do modelo $MESS$ via máxima verossimilhança	39
3	Painel Espacial	41
3.1	Painel espacial de efeitos fixos	42
3.1.1	O modelo espacial autoregressivo de efeitos fixos	42
3.1.2	O modelo espacial autoregressivo de efeitos aleatórios	45
3.2	O modelo $MESS$ em painel	45
4	Painel Espacial - Aplicações	50
4.1	Estudo de caso 1: Municípios de Rondônia	50
4.2	Estudo de caso 2: Municípios de São Paulo	53
4.3	Estudo de caso 3: Consumo de cigarros nos Estados Unidos	54
4.4	Estudo de caso 4: Municípios do Brasil	59
5	Conclusões e Recomendações	61
5.1	Conclusões	61

5.2	Recomendações para Trabalhos Futuros	62
	Appendices	66
	A	67
A.1	Within Transformation	67
A.2	Matriz de Covariâncias - Modelo <i>MESPS</i>	70
A.3	Procedimento computacional - <i>MESPS</i>	71
A.4	Procedimento computacional - <i>MESPS</i> - Grandes conjuntos de dados	75

Lista de Figuras

4.1	Dependência espacial dos municípios do Estado de Rondônia	52
4.2	Dependência espacial entre municípios de São Paulo	55
4.3	Tempo de processamento (em segundos) em função da ordem do polinômio de α	56
4.4	Estimativas de α em função da ordem de q	56
4.5	Estimativas de α em função da ordem do polinômio de α	57

Lista de Tabelas

4.1	Resultados dos Modelos - Municípios de Rondônia	51
4.2	Resultados dos Modelos - Municípios de São Paulo	54
4.3	Resultados dos Modelos - Consumo de cigarros nos Estados Unidos .	58
4.4	Resultados dos Modelos - Municípios do Brasil	60

Resumo

Amplamente utilizados em aplicações econômicas, os modelos de dados em painel desempenham papel imprescindível por sua natureza que incorpora observações de corte transversal e ao longo do tempo. Todavia, o desenvolvimento desses modelos se deu no sentido de corrigir a dependência temporal entre as observações. Os modelos espaciais buscam preencher essa lacuna no sentido de incorporar a localização (vizinhança, mais especificamente) das observações no processo de estimação. O fato de que negligenciar a dependência espacial e utilizar técnicas clássicas (que pressupõem independência das observações) produz modelos inconsistentes e de especificação questionável.

Este trabalho apresenta, de forma geral, algumas das técnicas de modelos em painel e de regressão espacial mais amplamente utilizadas e, posteriormente, propõe adaptações nos modelos em painel, no caso o modelo MESPS (*Matrix Exponential Spatial Panel Specification*), de forma a incorporar a dependência espacial em suas variadas formas. Além disso, foi construído um procedimento computacional em linguagem *SAS - IML*[©] para estimar o painel espacial de efeitos fixos. Este procedimento foi utilizado em quatro conjuntos de dados com características distintas e se provou uma excelente alternativa aos procedimentos tradicionais quando a dependência espacial não é excessivamente alta.

Palavras Chave: *Econometria Espacial, Modelos em Painel, Efeitos Fixos, Efeitos Aleatórios, Matriz de Proximidades, MESS, MESPS, Painel Espacial.*

Abstract

Panel data play an important role in Econometrics due to its cross-section and time-series features. However, the developments in this subject are usually made in order to embody dependence and trends related to the time-series framework. Spatial models are intended to take the position or neighbourhood of an individual into consideration while modeling. It should be noticed that neglecting the spatial structure of data would generate bias making the estimates inconsistent. In addition, the advantage of estimating the spatial spill-over effect would not be taken.

This thesis reviews the most widely used panel data model, such as pooled, fixed and random effects. It also approaches a mixed fixed effects panel data and spatial model named *MESPS - Matrix Exponential Spatial Panel Specification*. A *SAS - IML* code was implemented to estimate the *MESPS*. Four cases were analysed in different perspectives, comprehending a wide pool of situations.

key words: *panel data, spatial models, MESPS, spatial weight matrix, MESS, fixed effects, random effects.*

Introdução

Modelos para dados em painel desempenham um importante papel na econometria em virtude de sua natureza, que aborda de forma integrada observações de corte transversal (*cross-section*) e série temporal. Em decorrência dessa abordagem mais ampla, as técnicas convencionais de regressão de corte transversal não podem ser aplicadas diretamente. Em contrapartida, dados em painel permitem modelagens mais robustas e que captam efeitos não identificáveis no caso de corte transversal.

Uma das limitações dos modelos em painel é a grande exigência computacional, principalmente em aplicações econômicas, onde geralmente o número de observações é grande. Ao integrar técnicas de painel e regressão espacial, os procedimentos se tornam ainda mais computacionalmente intensivos¹. Com o intuito de dar um passo em direção à maior tratabilidade computacional desses modelos, foi construído um procedimento em linguagem *SAS - IML* - cuja conversão para linguagem *S* é direta e sem maiores complicações - para estimar modelos espaciais com estrutura de painel.

Nesta dissertação, serão apresentados diferentes abordagens de dados em painel, sendo que o foco será nas técnicas mais tradicionais, como modelagem de efeitos fixos e de efeitos aleatórios. Além disso, esses modelos serão apresentados num contexto em que a organização espacial das observações é levada em consideração. Os ganhos dessa iniciativa são claros quando há, de fato, uma dependência espacial nos dados. Caso contrário, os resultados serão similares aos dos modelos convencionais.

Note, também, que nem sempre é razoável assumir que a média e a heterogeneidade (variância) - dados os níveis das covariáveis envolvidas - é constante ao longo do espaço (LeSage, 1999). Num contexto, em que se esteja estudando a renda *per ca-*

¹Alguns modelos não são estimáveis mesmo para amostras intermediárias ($N \geq 500$) porque superam o limite da capacidade de processamento dos computadores atuais.

pita da população em diferentes cidades do Brasil, por exemplo, além das diferenças no nível da variável renda, seria natural observar diferentes variâncias em regiões distintas.

Os modelos do tipo painel espacial foram projetados de modo a captar diferentes médias e heterogeneidades ao longo do espaço e, em adição, englobar todas as vantagens de um estudo em painel, como identificação de tendências e grande disponibilidade de graus de liberdade (Elhorst, 2010).

O trabalho está estruturado como a seguir: O Capítulo 1 apresenta os principais conceitos de modelos de dados em painel, com destaque para os modelos de efeitos fixos e aleatórios. No Capítulo 2 são abordadas as técnicas de econometria espacial, iniciando com alguns conceitos de estatística espacial como matriz de proximidades \mathbf{W} e testes de dependência espacial e finalizando com os modelos de regressão espacial. O Capítulo 3 trata dos modelos de dados em painel com estrutura espacial, fazendo exatamente uma compilação dos dois capítulos anteriores, além de abordar as principais consequências computacionais dessa junção. No Capítulo 4 são feitos quatro estudos de caso, a fim de verificar as particularidades dos modelos, a saber:

1. Municípios de Rondônia - baixa dependência espacial;
2. Municípios de São Paulo - média dependência espacial;
3. Consumo de cigarros nos Estados Unidos - alta dependência espacial;
4. Municípios do Brasil - base com muitos registros e baixa dependência espacial².

Por fim, o Capítulo 5 apresenta as principais conclusões e recomendações para trabalhos futuros.

²Base municipal contendo 5.560 municípios distribuídos em quatro anos (Silva and Alves, 2007)

Capítulo 1

Dados em Painel

1.1 Introdução

De modo geral, é possível classificar dados de três formas: corte transversal (*cross section*), série temporal e dados em painel (genericamente tratados como dados longitudinais). Dados de corte transversal são mensurações de várias unidades observacionais (geralmente independentes) em um único instante no tempo. Uma série temporal consiste em uma única observação medida ao longo de vários períodos de tempo. Usualmente essas mensurações são correlacionadas, demandando estrutura de covariâncias mais estruturada que dados de corte transversal. Conjuntos de dados em painel, consistem de várias observações de corte transversal (*cross section*) mensuradas em diferentes pontos no tempo (não necessariamente equiespaçados). A modelagem desse tipo de dado envolve algumas especificidades, como a dependência ao longo do tempo, uma vez que a mesma unidade observacional é medida em diferentes pontos no tempo. Algumas das grandes vantagens do uso de dados em painel quando comparado com dados de corte transversal são:

- Menor variabilidade intra-observação que entre observações, isto é, o mesmo indivíduo apresenta menor heterogeneidade em diferentes pontos no tempo que diferentes indivíduos no mesmo instante;
- As mensurações ao longo do tempo dão uma idéia de tendência, além de captar relações dinâmicas entre indivíduos;

- O mesmo indivíduo pode ser controle e submetido a diferentes tratamentos, basta que o seja em diferentes momentos.

A literatura estatística é bastante extensa no âmbito de dados longitudinais. Há que se ressaltar, todavia, que a modelagem tradicional de dados longitudinais é um pouco mais geral que a de dados em painel. Dados em painel tem sido tratados genericamente como observações de corte transversal empilhadas ao longo do tempo (Frees, 2004). Em estudos longitudinais, tipicamente se tem um experimento controlado com mais pontos no tempo que num painel. Este, por sua vez, é formado por dados observacionais que envolvem uma série de fatores exógenos de influência desconhecida e que são fonte de heterogeneidade. Uma ressalva que deve ser feita, é que dados longitudinais implicam em repetições na mesma unidade observacional, mesmo que o fator temporal não desempenhe papel importante na modelagem. No painel, as repetições são feitas ao longo do tempo e o fator temporal é imprescindível para captar relações dinâmicas entre indivíduos. Mesmo assim, o uso de medidas repetidas é extremamente válido, uma vez que gera mais graus de liberdade e controla para influências exógenas que são constantes ao longo do tempo. Estudos longitudinais possuem, em geral, repetições suficientes para estimar parâmetros específicos de cada indivíduo¹ (Hedeker and Gibbons, 2006).

Um exemplo de estudo longitudinal apresentado em McCulloch and Searle (2001) é a avaliação do impacto de um medicamento para controle da pressão arterial. Este medicamento é ministrado em duas doses mais um controle (placebo) em cada indivíduo em diferentes pontos no tempo. Evidentemente, pessoas diferentes tem pressão média - sem serem submetidas a nenhum tratamento - diferentes, bem como respondem de forma distintas a cada dosagem do medicamento. Portanto, um modelo que seja flexível para comportar essas particularidades deve permitir um intercepto diferente para cada indivíduo, assim como diferentes impactos do tratamento (doses) na resposta (pressão arterial) de cada pessoa. No decorrer deste capítulo, serão apresentados alguns modelos que se adaptam a situações parecidas num contexto econômico. Para maiores detalhes em modelos para dados longitudinais, veja McCulloch and Searle (2001), Verbeke and Molenberghs (2000) e Hedeker and Gibbons (2006).

¹Numa abordagem de corte transversal, esses parâmetros seriam não identificáveis.

No caso de dados em painel, basicamente, o vetor resposta \mathbf{y} tem como elemento típico y_{it} o valor da variável Y na i -ésima observação e no t -ésimo instante de tempo. Num contexto em que se tem K variáveis explicativas², pode-se, de modo geral, representar matricialmente o modelo de dados em painel como:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{NT} \end{bmatrix} = \begin{bmatrix} x_{111} & x_{121} & \dots & x_{1k1} \\ x_{112} & x_{122} & \dots & x_{1k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1T} & x_{N2T} & \dots & x_{NkT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{NT} \end{bmatrix}$$

onde c_i é o efeito individual de cada observação e, portanto, constante ao longo do tempo. É válido ressaltar que o efeito individual não é um parâmetro a ser estimado e sim uma variável aleatória que carrega efeitos individuais não observáveis. Note que uma vez que esse efeito carrega características intrínsecas do indivíduo, faz sentido que ele seja mantido constante ao longo do tempo. Há diversos tipos de modelagem de dados de painel, sendo os mais conhecidos:

1. Modelo Combinado;
2. Modelo de Efeitos Fixos;
3. Modelo de Efeitos Aleatórios.

A escolha do modelo depende, primordialmente, da natureza do efeito individual c_i , i.e, se c_i puder ser dissociado das variáveis explicativas X . A verificação deste fato não é trivial e este tema será abordado com mais detalhes nas seções seguintes.

1.2 O Modelo Combinado

Num contexto em que $c_i = c \forall i$, o modelo combinado pode ser utilizado sem problemas e o estimador de mínimos quadrados é consistente e eficiente (Greene, 2006). Note que este modelo³ é extremamente simples e remete ao modelo de regressão

²Elemento típico da matriz X : x_{ijt} , onde $i = 1, 2, \dots, N$ unidades observacionais, $t = 1, 2, \dots, T$ instantes no tempo e $j = 1, 2, \dots, K$ variáveis explicativas.

³A matriz X não contém o vetor de 1's para estimação do intercepto, uma vez que este está sendo representado explicitamente por c

linear para dados de corte transversal, a saber:

$$y_{it} = x'_{it}\boldsymbol{\beta} + c + \epsilon_{it} \quad (1.1)$$

em que $x'_{it} = [x_1, x_2, \dots, x_K]$ para a i -ésima observação no período de tempo t . Este modelo funciona bem se todos os pressupostos do modelo de regressão clássico forem atendidos, a saber: $E(\epsilon_{it}|x'_{it}) = 0$, $\boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{I})$, observações independentes e exogeneidade estrita de x'_{it} . O estimador de mínimos quadrados $\hat{\boldsymbol{\beta}}$ e sua respectiva variância são os mesmos da regressão linear (MQO) sendo dado por:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \text{Var}(\hat{\boldsymbol{\beta}}) &= \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Algumas desvantagens deste modelo são facilmente reconhecíveis. De imediato, observa-se que em nenhum momento da modelagem se leva em consideração a estrutura temporal presente nos dados. Em outras palavras, assume-se que todas as observações são independentes. Todavia, é evidente que há mensurações no mesmo indivíduo em instantes de tempo distintos. Utilizar o modelo combinado é assumir que essas mensurações são independentes. De fato, há situações em que este modelo é bastante razoável. Na maioria dos casos - e principalmente em dados econômicos - este tipo de modelagem é pouco realista. Caso haja a suspeita de dependência entre as observações, vale a pena utilizar uma modelagem mais robusta, como por exemplo, um modelo de efeitos fixos ou aleatórios.

Evidentemente, a estrutura de covariâncias assumida por $\boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{I})$ é condição para construir intervalos de confiança e realizar testes de hipóteses. Caso a amostra em estudo seja razoavelmente grande, então essa estrutura é alcançada assintoticamente. Caso contrário, os estimadores das variâncias são viesados e os intervalos de confiança inúteis.

1.3 O Modelo de Efeitos Fixos

O modelo de efeitos fixos é utilizado em situações em que não é possível dissociar o efeito individual c_i das variáveis explicativas, sendo que a forma da dependência entre c_i e \mathbf{X} é geralmente desconhecida. Este modelo, no entanto, é robusto o suficiente para

permitir que essa dependência não tenha forma conhecida. O nome “efeito fixo” não significa que c_i seja tratado como não estocástico, mas que haja uma correlação arbitrária entre c_i e \mathbf{X} (Wooldridge, 2001). Na maioria dos casos, essa correlação é tida como arbitrária pelo simples fato de c_i não ser mensurável (fonte de heterogeneidade desconhecida). Numa aplicação em que se deseja explicar a evolução da renda de certos indivíduos ao longo do tempo com base no tempo de estudo deles, o efeito fixo c_i capturaria características como aptidão e/ou postura profissional (além de outros fatores influentes e não mensurados ou não mensuráveis). Veja que essas podem ter influenciado o tempo de estudo do indivíduo, uma vez que um indivíduo com postura profissional de grande comprometimento tende a estudar com mais seriedade, ao passo que indivíduos mais aptos tendem a se sentir incentivados com bons resultados na escola/universidade advindos de sua aptidão. Tentar mensurar aptidão e postura profissional é um tanto complicado por vários motivos, como estabelecer uma unidade de medida, estabelecer níveis comparáveis, e mesmo pelo simples fato de que as pessoas podem ter escolhido carreiras para as quais não são mais aptas (e isso não significa que elas sejam inaptas de modo geral).

1.3.1 Modelo com 1 efeito fixo

Sendo assim, pode-se representar o modelo de 1 efeito fixo⁴ (*One-Way Fixed Effects*) da seguinte forma:

$$y_{it} = x'_{it}\boldsymbol{\beta} + c_i + \epsilon_{it} \quad (1.2)$$

pois leva em consideração apenas o efeito individual de cada unidade observacional.

O processo de estimação mais comum é conhecido como Mínimos Quadrados de Variáveis Indicadoras (*LSDV - Least Squares Dummy Variables*). Este procedimento é, basicamente, mínimos quadrados ordinários aplicados num modelo cuja matriz de variáveis explicativas contém uma variável indicadora para cada indivíduo⁵ (unidade observacional). Representando na forma matricial, tem-se:

⁴Para conhecer o modelo de 1 efeito fixo com maior nível de aprofundamento, veja Baltagi (2008), Wooldridge (2001) e Johnston and Dinardo (1996).

⁵A matriz X não deve conter o vetor de 1's para que não haja colinearidade perfeita entre o intercepto e $\sum_{i=1}^N c_i$. Caso o intercepto seja de relevância imprescindível para a interpretação do modelo, é necessário impor a restrição $\sum_{i=1}^N c_i = 1$.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} j & 0 & \dots & 0 \\ 0 & j & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & j \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}$$

onde $\underbrace{X_i}_{\text{TxK}} = \begin{bmatrix} x_{i11} & x_{i21} & \dots & x_{ik1} \\ x_{i12} & x_{i22} & \dots & x_{ik2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i1T} & x_{i2T} & \dots & x_{ikT} \end{bmatrix}$, $\underbrace{j}_{\text{Tx1}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ e $\epsilon_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{iT} \end{bmatrix}$.

Os erros ϵ_{it} são comumente chamados erros **idiosincráticos** (ou perturbações idiosincráticas) por variarem tanto ao longo do tempo quanto ao longo das observações, ou seja, são perturbações próprias da natureza estocástica de cada indivíduo em cada instante no tempo. Uma vez obtido \hat{c}_i , cabe lembrar que a interpretação desse é a alteração na média de y_i . Em outras palavras, não faz sentido avaliar o impacto marginal de c_i em y_i uma vez que este não possui unidade de medida quantificável. Lembre-se de que c_i reúne peculiaridades não mensuráveis (latentes) do indivíduo i . Um problema crítico do método LSDV é a grande exigência computacional. Num conjunto de dados com N observações de corte transversal a matriz de variáveis explicativas teria dimensão $NT \times (N + K)$. Numa aplicação em que N é grande, realizar operações nessa matriz é impraticável. Para contornar esse problema foi proposto um método menos computacionalmente intensivo chamado *Within Transformation*.

1.3.2 O procedimento *Within Transformation*

O procedimento *Within Transformation* consiste em estimar o modelo

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\boldsymbol{\beta} + \epsilon_{it} - \bar{\epsilon}_i. \quad (1.3)$$

A prova de que este modelo é equivalente a (1.2) está no Apêndice A1. Note que $\boldsymbol{\beta}$ mede o impacto da variação de x_{it} em relação a \bar{x}_i no desvio $(y_{it} - \bar{y}_i)$. Sendo assim, não é possível utilizar variáveis constantes ao longo do tempo (em todas as observações) na modelagem de y_{it} . Esse fato pode ser verificado facilmente no modelo

1.2, uma vez que haveria colinearidade perfeita entre c_i e uma variável constante ao longo do tempo. Seja, então, a matriz

$$\mathbf{D} = \begin{bmatrix} \mathbf{j} & 0 & \dots & 0 \\ 0 & \mathbf{j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{j} \end{bmatrix}$$

Se $\mathbf{M} = (\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}')$, então o estimador de $\boldsymbol{\beta}$ é similar ao de mínimos quadrados generalizados dado por:

$$\hat{\boldsymbol{\beta}} = [(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{M}\mathbf{y})] \quad (1.4)$$

Matricialmente, é mais simples representar a matriz \mathbf{M} como

$$\mathbf{M} = \mathbf{I}_N \otimes \mathbf{I}_T - \frac{1}{T}\mathbf{j}_T\mathbf{j}'_T \otimes \mathbf{I}_N \quad (1.5)$$

1.3.3 Modelo com 2 efeitos fixos

O procedimento *LSDV* também é útil para estimar os efeitos específicos de cada período de tempo (Greene, 2006). Sendo assim, um modelo que leva em consideração os efeitos específicos de cada unidade observacional e de cada instante de tempo é chamado de 2 efeitos fixos (*Two-Way Fixed Effects*) e tem a forma:

$$y_{it} = x'_{it}\boldsymbol{\beta} + c_i + \gamma_t + \epsilon_{it} \quad (1.6)$$

onde γ_t é o efeito específico do t -ésimo instante de tempo. Num contexto *LSDV*,

$$\gamma_t = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

É importante ressaltar que o modelo deve ser estimado com $T - 1$ variáveis indicadoras, uma vez que o primeiro efeito fixo (efeito indivíduo) já soma 1 e, portanto, se todos os períodos de tempo fossem levados em consideração, haveria colinearidade

perfeita entre os efeitos indivíduo e tempo. Neste caso, o período de tempo que foi suprimido funcionaria como um período base, e os γ_t 's seriam contrastes em relação a ele. Há, ainda, outra forma de estimar o modelo com 2 efeitos fixos sem ter de recorrer à interpretação de contrastes. Trata-se de um modelo com restrições nos parâmetro c e γ do tipo $\sum c = \sum \gamma = 0$. Este modelo é equivalente a regredir y_{it}^* em x_{it}^* , onde

$$\begin{aligned}
y_{it}^* &= y_{it} - \bar{y}_i. - \bar{y}_{.t} + \bar{\bar{y}}.. \\
x_{it}^* &= x_{it} - \bar{x}_i. - \bar{x}_{.t} + \bar{\bar{x}}.. \\
\bar{y}_{.t} &= \frac{1}{n} \sum_{i=1}^N y_{it} \\
\bar{y}_i. &= \frac{1}{T} \sum_{t=1}^T y_{it} \\
\bar{\bar{y}}.. &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{it} \\
\hat{c}_i &= (\bar{y}_i. - \bar{\bar{y}}..) - (\bar{x}_i. - \bar{\bar{x}}..)'\beta \\
\hat{\gamma}_t &= (\bar{y}_{.t} - \bar{\bar{y}}..) - (\bar{x}_{.t} - \bar{\bar{x}}..)'\beta
\end{aligned} \tag{1.7}$$

Assim como na seção anterior, há uma representação matricial que simplifica o procedimento *Within Transformation*, a saber:

$$\mathbf{M} = \mathbf{I}_N \otimes \mathbf{I}_T - \frac{1}{T} \mathbf{j}_T \mathbf{j}'_T \otimes \mathbf{I}_N - \frac{1}{N} \mathbf{I}_N \otimes \mathbf{j}_N \mathbf{j}'_N + \frac{1}{NT} \mathbf{j}_{NT} \mathbf{j}'_{NT} \tag{1.8}$$

1.3.4 Estimação de parâmetros via Máxima Verossimilhança

Estimar parâmetros num contexto de máxima verossimilhança é interessante porque permite incorporar informações *a priori* e guiar o processo de estimação para uma perspectiva Bayesiana. Há, todavia, três abordagens diferentes que geram, sob certas circunstâncias, os mesmos estimadores dos efeitos fixos. São elas: verossimilhança conjunta, verossimilhança condicional e verossimilhança marginal. Neste texto, serão abordadas as duas primeiras. Para mais informações nesse sentido, consulte Arellano (2003).

Verossimilhança Conjunta

Sendo assim, sob normalidade

$$\mathbf{y}_i | \mathbf{X}_i, c_i \sim N(\mathbf{X}'_i \boldsymbol{\beta}, \sigma_c^2 \mathbf{I}_T) \tag{1.9}$$

e a log-verossimilhança conjunta de N observações de corte transversal é dada por

$$l(\mathbf{y}|\mathbf{X}, \mathbf{c}) = \sum_{i=1}^N \log f(\mathbf{y}_i|\mathbf{X}_i, c_i) \quad (1.10)$$

onde

$$\log f(\mathbf{y}_i|\mathbf{X}_i, c_i) \propto -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{c}_i)' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{c}_i) \quad (1.11)$$

O estimador de $\boldsymbol{\beta}$ e os resíduos gerados via máxima verossimilhança conjunta (*Joint Likelihood*) são idênticos aos do procedimento *Within Transformation*. O estimador $\hat{\sigma}_{JL}^2 = \frac{1}{NT} e'e$, todavia, é inconsistente porque é função de $\hat{\boldsymbol{\beta}}$ e \mathbf{c} , mas a dimensão de \mathbf{c} cresce conforme o número de observações de corte transversal ($N \rightarrow \infty$). Nesse caso, como o interesse é estimar $\hat{\boldsymbol{\beta}}$, o estimador de máxima verossimilhança é, ainda, uma excelente alternativa por causa de suas propriedades.

Outra abordagem bastante relevante e que fornece estimadores consistentes é a Verossimilhança Condicional.

Verossimilhança Condicional

Seja a densidade (em logaritmo)

$$\log f(\bar{\mathbf{y}}_i|\mathbf{X}_i, c_i) = K - \frac{1}{2} \log \sigma^2 - \frac{T}{2\sigma^2} (\bar{\mathbf{y}}_i - \bar{\mathbf{X}}_i \boldsymbol{\beta} - \mathbf{c}_i)' (\bar{\mathbf{y}}_i - \bar{\mathbf{X}}_i \boldsymbol{\beta} - \mathbf{c}_i) \quad (1.12)$$

onde K é uma constante.

Então,

$$\begin{aligned} f(\mathbf{y}_i|\mathbf{X}_i, c_i, \bar{\mathbf{y}}_i) &= \frac{f(\mathbf{y}_i|\mathbf{X}_i, c_i)}{f(\bar{\mathbf{y}}_i|\mathbf{X}_i, c_i)} \\ \log f(\mathbf{y}_i|\mathbf{X}_i, c_i, \bar{\mathbf{y}}_i) &= \log f(\mathbf{y}_i|\mathbf{X}_i, c_i) - \log f(\bar{\mathbf{y}}_i|\mathbf{X}_i, c_i) \\ \log f(\mathbf{y}_i|\mathbf{X}_i, c_i, \bar{\mathbf{y}}_i) &\propto -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{c}_i)' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{c}_i) + \frac{1}{2} \log \sigma^2 + \\ &\quad \frac{T}{2\sigma^2} (\bar{\mathbf{y}}_i - \bar{\mathbf{X}}_i \boldsymbol{\beta} - \mathbf{c}_i)' (\bar{\mathbf{y}}_i - \bar{\mathbf{X}}_i \boldsymbol{\beta} - \mathbf{c}_i) \\ \log f(\mathbf{y}_i|\mathbf{X}_i, c_i, \bar{\mathbf{y}}_i) &\propto -\frac{T-1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (\epsilon_{it} - \bar{\epsilon}_i)^2 \end{aligned}$$

Note que $\log f(\mathbf{y}_i|\mathbf{X}_i, c_i, \bar{\mathbf{y}}_i)$ não depende de c_i . Portanto, $\bar{\mathbf{y}}_i$ é estatística suficiente para c_i (Casella and Berger, 2001).

Assim, a verossimilhança condicional

$$\mathcal{L}(\beta, \sigma^2; \mathbf{y}, \mathbf{X}) = \sum_{i=1}^N \log f(y_i|\mathbf{X}_i, \bar{\mathbf{y}}_i) \quad (1.13)$$

De modo semelhante à seção anterior, o estimador de máxima verossimilhança condicional de β é o mesmo gerado pela *Within Transformation*, mas

$$\hat{\sigma}_{CL}^2 = \frac{1}{N(T-1)} \sum_{i=1}^N \bar{\epsilon}_i^2, \text{ onde } \bar{\epsilon}_i = \sum_{t=1}^T \epsilon_{it}.$$

Sendo assim, essas duas abordagens são interessantes caso se deseje estudar a fundo o impacto de *prioris* informativas. Um excelente exemplo de modelagem em painel sob a perspectiva Bayesiana está em Moral-Benito (2009). Todavia, o escopo deste trabalho se restringe a metodologias frequentistas.

1.4 O Modelo de Efeitos Aleatórios

A principal diferença entre o modelo de efeitos fixos e o modelo de efeitos aleatórios é que neste último os efeitos individuais são ortogonais às variáveis exógenas (explicativas). Além disso, o modelo de efeitos aleatórios é ideal para aplicações em que os indivíduos foram amostrados de uma população maior ou janela de tempo mais extensa (Greene, 2006).

O modelo de efeitos aleatórios é útil quando há suspeita de heterogeneidade ao longo as observações, isto é, o efeito individual leva em consideração uma fonte de variação comum aquele indivíduo (além da variação idiossincrática).

Outras grandes vantagens desse tipo de modelagem são a correção, de forma natural, da dependência intra observação (Baltagi, 2008) (no mesmo indivíduo em pontos diferentes no tempo) e a flexibilidade no espaçamento dos pontos no tempo (Verbeke and Molenberghs, 2000).

De modo geral, a estrutura de covariâncias assumida em contextos econômicos é a de simetria composta (*Compound Symmetry*). Algumas outras estruturas seriam possíveis, como a autoregressiva, de *Toeplitz* e não estruturada (Hedeker and Gibbons, 2006). Todavia, estas demandam estudos com muitos pontos no tempo, característica raramente encontrada em dados em painel. O debate acerca da estrutura de covariâncias mais indicada para cada caso é bastante extenso e envolve algumas questões, como o ganho prático de se utilizar um modelo mais complicado (mensurado, por exemplo, como a diferença nos estimadores das dispersões) e a viabilidade do modelo dadas as condições da amostra (número de observações e pontos no tempo). Neste texto, o enfoque será na estrutura de simetria composta por ela ser

amplamente abordada.

1.4.1 O modelo com 1 efeito aleatório

Seja o modelo

$$y_{it} = x'_{it}\beta + c_i + \epsilon_{it} \quad (1.14)$$

onde c_i é o efeito individual, tal que $E(c_i|x_{it}) = E(c_i) = 0$ e $E(c_i^2) = \sigma_c^2$. Conforme dito anteriormente, a condição de ortogonalidade entre o efeito aleatório e as variáveis exógenas pode ser escrito na forma $E(x'_{it}c_i)$.

Uma vez que os termos c_i e ϵ_{it} são aleatórios, pode-se escrever o modelo 1.14 da seguinte forma:

$$y_{it} = x'_{it}\beta + v_{it} \quad (1.15)$$

em que o termo $v_{it} = c_i + \epsilon_{it}$ capta a heterogeneidade presente na mensuração y_{it} , seja essa heterogeneidade própria do i -ésimo indivíduo (c_i), seja ela puramente aleatória⁶ (ϵ_{it}). Então, do termo $v_{it} = c_i + \epsilon_{it}$, tem-se que:

$$\begin{aligned} E(v_{it}) &= E(c_i) + E(\epsilon_{it}) = 0 \\ Var(v_{it}) &= E(v_{it}^2) = E(c_i + \epsilon_{it})^2 \\ &= E(c_i^2 + 2c_i\epsilon_{it} + \epsilon_{it}^2) \\ &= E(c_i^2) + E(\epsilon_{it}^2) \\ &= \sigma_c^2 + \sigma_\epsilon^2 \end{aligned}$$

$$\begin{aligned} Cov(v_{it}, v_{is}) &= E[(c_i + \epsilon_{it})(c_i + \epsilon_{is})] \\ &= E(c_i^2) + E(c_i\epsilon_{is}) + E(c_i\epsilon_{it}) + E(\epsilon_{it}\epsilon_{is}) \\ &= E(c_i^2) = \sigma_c^2, t \neq s \end{aligned}$$

Seja $v_i = [v_{i1}, \dots, v_{iT}]'$. Então, se $\Omega_i = E(v_i v_i')$, tem-se que

$$\Omega_i = \begin{bmatrix} \sigma_c^2 + \sigma_\epsilon^2 & \sigma_c^2 & \dots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_\epsilon^2 & \dots & \sigma_c^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_c^2 & \sigma_c^2 & \dots & \sigma_c^2 + \sigma_\epsilon^2 \end{bmatrix}$$

⁶Idiossincrática.

Seja $\mathbf{j}_{T \times 1} = [1, \dots, 1]'$. Então, é possível escrever a matriz $\mathbf{\Omega}_i$ como $\mathbf{\Omega}_i = \sigma_c^2 I_{T \times T} + \sigma_\epsilon^2 \mathbf{j}\mathbf{j}'$. Esta forma é usualmente conhecida como estrutura de efeitos aleatórios (Wooldridge, 2001). Note que independentemente do tamanho de T , $\mathbf{\Omega}$ dependerá apenas dos parâmetros σ_c^2 e σ_ϵ^2 . Esta é uma vantagem que este tipo de modelo tem em relação aos de efeitos fixos que, em virtude da existência de um parâmetro para cada indivíduo, são inevitavelmente superparametrizados em grandes amostras. É interessante ressaltar, também, que para qualquer t e s , ($t \neq s$), $Corr(v_{it}, v_{is}) = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_\epsilon^2}$. Esta correlação informa a respeito do nível de heterogeneidade explicado pelo efeito c_i .

O modelo de efeitos aleatórios corrige um problema bastante crítico que é a autocorrelação entre mensurações no mesmo indivíduo. Assumir que y_{it} é independente de y_{is} é pouco realista em aplicações econômicas. O modelo (1.14) leva, naturalmente, este fato em consideração.

Para estimar o modelo (1.14) pelo método de mínimos quadrados generalizados (*GLS - Generalized Least Squares*) basta, portanto, estimar de forma consistente σ_c^2 e σ_ϵ^2 . Note, porém, que é muito mais fácil, inicialmente, estimar de forma consistente σ_v^2 , uma vez que é suficiente estimar o modelo (1.14) via mínimos quadrados ordinários⁷ e guardar os resíduos desta regressão v_{it}^* .

Logo,

$$\hat{\sigma}_v^2 = \frac{1}{NT - k} \sum_{i=1}^N \sum_{t=1}^T v_{it}^{*2} \quad (1.16)$$

Para obter σ_c^2 , é interessante utilizar o fato de que $Cov(v_{it}, v_{is}) = \sigma_c^2$, $t \neq s$. Sendo assim tem-se que

$$\begin{aligned} E \left(\sum_{t=1}^T \sum_{s=t+1}^{T-1} v_{it}^* v_{is}^* \right) &= \sum_{t=1}^T \sum_{s=t+1}^{T-1} \sigma_c^2 \\ &= [(T-1) + (T-2) + \dots + 1] \sigma_c^2 = \frac{T(T-1)}{2} \sigma_c^2 \end{aligned}$$

Um estimador consistente pelo método dos momentos para σ_c^2 é tal que

$$\hat{\sigma}_c^2 = \left[\frac{NT(T-1)}{2} - k \right]^{-1} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=t+1}^{T-1} v_{it}^* v_{is}^* \quad (1.17)$$

⁷O estimador de máxima verossimilhança de σ_v^2 seria $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T v_{it}^2$, i.e, não leva em consideração a correção de graus de liberdade induzida pelo número de parâmetros estimados.

Por fim, uma vez estimados $\hat{\sigma}_c^2$ e $\hat{\sigma}_v^2$, é direto obter $\hat{\sigma}_\epsilon^2 = \hat{\sigma}_v^2 - \hat{\sigma}_c^2$.

Uma vez obtidos estimadores consistentes para $\hat{\sigma}_c^2$ e $\hat{\sigma}_\epsilon^2$, basta estimar β por mínimos quadrados generalizados⁸. Este estimador é dado por:

$$\hat{\beta} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y} \quad (1.18)$$

Este estimador leva em consideração a heterogeneidade entre os indivíduos. Em alguns casos, entretanto, é necessário, também, modelar a heterogeneidade entre os períodos de tempo. Nesse caso, o modelo mais robusto é o de dois efeitos aleatórios.

1.4.2 O modelo de 2 efeitos aleatórios

Bastante similar ao modelo de 1 efeito aleatório, o modelo de 2 efeitos aleatórios incorpora a heterogeneidade ao longo do tempo. Sendo assim, o modelo tem a forma

$$y_{it} = x'_{it}\beta + v_{it} \quad (1.19)$$

onde $v_{it} = c_i + \lambda_t + \epsilon_{it}$ e λ_t é o efeito temporal.

A matriz Ω é construída de modo que

$$\begin{aligned} Cov(v_{it}, v_{js}) &= \sigma_c^2, i = j, \text{ se } t \neq s \\ Cov(v_{it}, v_{js}) &= \sigma_\lambda^2, \text{ se } i \neq j, t = s \\ Cov(v_{it}, v_{js}) &= \sigma_c^2 + \sigma_\lambda^2 + \sigma_\epsilon^2, \text{ se } i = j, t = s \end{aligned} \quad (1.20)$$

Para ilustrar melhor Ω , foi construído um exemplo, com base numa situação hipotética em que há 4 observações mensuradas em 2 períodos de tempo. O resultado é a matriz:

$$\begin{bmatrix} \sigma_c^2 + \sigma_\lambda^2 + \sigma_\epsilon^2 & \sigma_c^2 & \sigma_\lambda^2 & 0 & \sigma_\lambda^2 & 0 \\ \sigma_c^2 & \sigma_c^2 + \sigma_\lambda^2 + \sigma_\epsilon^2 & 0 & \sigma_\lambda^2 & 0 & \sigma_\lambda^2 \\ \sigma_\lambda^2 & 0 & \sigma_c^2 + \sigma_\lambda^2 + \sigma_\epsilon^2 & \sigma_c^2 & \sigma_\lambda^2 & 0 \\ 0 & \sigma_\lambda^2 & \sigma_c^2 & \sigma_c^2 + \sigma_\lambda^2 + \sigma_\epsilon^2 & 0 & \sigma_\lambda^2 \\ \sigma_\lambda^2 & 0 & \sigma_\lambda^2 & 0 & \sigma_c^2 + \sigma_\lambda^2 + \sigma_\epsilon^2 & \sigma_c^2 \\ 0 & 0 & 0 & \sigma_\lambda^2 & \sigma_c^2 & \sigma_c^2 + \sigma_\lambda^2 + \sigma_\epsilon^2 \end{bmatrix}$$

⁸Esse estimador é conhecido em econometria como “estimador de mínimos quadrados generalizados factíveis” (FGLS - *Feasible Generalized Least Squares*) ou, especificamente no caso de painel de efeitos aleatórios, estimador de Balestra-Nerlove.

Basta, então, obter estimadores consistentes de σ_c^2 , σ_λ^2 e σ_ϵ^2 e calcular o estimador de mínimos quadrados generalizados dado por

$$\hat{\beta} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{y} \quad (1.21)$$

Este estimador é conhecido como estimador de Balestra-Nerlove (Balestra and Nerlove, 1966).

1.4.3 Estimando parâmetros via máxima verossimilhança

Apesar de o estimador de mínimos quadrados generalizados ser consistente, é possível, também, obter estimadores consistentes por meio de uma abordagem de verossimilhança, em que todos os parâmetros são estimados simultaneamente. O processo de estimação, todavia, é um pouco mais complicado que no caso de mínimos quadrados generalizados, uma vez que os estimadores são obtidos por meio de processos iterativos (*Newton-Raphson* ou *Escore de Fisher*, por exemplo). Para estimar os efeitos individuais, isto é, o nível da variável aleatória \mathbf{C} , utiliza-se, geralmente, o método de Bayes empírico (*EB - Empirical Bayes*), enquanto os parâmetros da distribuição de probabilidade de \mathbf{C} são obtidos via máxima verossimilhança marginal ou restrita (Hedeker and Gibbons, 2006). Para mais informações, veja McCulloch and Searle (2001) e Verbeke and Molenberghs (2000).

O primeiro passo é estimar a correlação intra grupos, utilizada para calcular o nível do efeito individual. A covariância intra grupos (Cov_{IG}) é estimada por

$$C\hat{ov}_{IG} = \frac{QM_I - QM_R}{T} \quad (1.22)$$

onde $QM_I = \frac{T}{N-1} \sum_{i=1}^N (\bar{y}_i - \bar{y})^2$ e $QM_R = \frac{1}{(N-1)(T-1)} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y}_i - \bar{y}_t + \bar{y})^2$. Por sua vez, a correlação ($Corr_{IG}$) é estimada por

$$C\hat{orr}_{IG} = \frac{C\hat{ov}_{IG}}{C\hat{ov}_{IG} + QM_R} = \frac{\hat{\sigma}_c^2}{\hat{\sigma}_c^2 + \hat{\sigma}_\epsilon^2} \quad (1.23)$$

Seja \tilde{c}_i o nível do efeito aleatório no i -ésimo indivíduo estimado via *EB*. Então, tem-se que

$$\begin{aligned} \tilde{c}_i &= \frac{\hat{\rho}}{T} \mathbf{j}'_T (\mathbf{y}_i - \mathbf{X}\hat{\beta}_{MV}) \\ \tilde{\sigma}_{c|\mathbf{y}_i}^2 &= \hat{\sigma}_c^2 (1 - \hat{\rho}) \end{aligned} \quad (1.24)$$

onde $\hat{\rho} = \frac{T C\hat{orr}_{IG}}{[1+(T-1)C\hat{orr}_{IG}]}$ é o coeficiente de confiabilidade de Spearman-Brown (Guilford, 1954). Note que o nível do efeito aleatório depende diretamente da dependência intra

grupos, no sentido de que caso essa dependência seja alta (altos valores de $Corr_{IG}$), então o efeito se aproxima do estimado via mínimos quadrados, ao passo que se $Corr_{IG}$ for baixo (em torno de 0), então o nível do efeito aleatório se aproxima do valor estipulado *a priori*⁹ (Hedeker and Gibbons, 2006).

O segundo passo consiste em obter estimadores consistentes, via máxima verossimilhança, para β , σ_c^2 e σ_ϵ^2 , por meio das equações

$$\begin{aligned}\hat{\beta} &= [(\mathbf{X}'\mathbf{X})]^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{I}_N \otimes \mathbf{j}_T \tilde{\mathbf{c}}) \\ \hat{\sigma}_c^2 &= \frac{1}{N}\mathbf{j}'_N \tilde{\mathbf{c}} + \tilde{\sigma}_{c|y} \\ \hat{\sigma}_\epsilon^2 &= NT\tilde{\sigma}_{c|y} + \frac{1}{NT}(\mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{I}_N \otimes \mathbf{j}_T \tilde{\mathbf{c}})'(\mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{I}_N \otimes \mathbf{j}_T \tilde{\mathbf{c}})\end{aligned}\tag{1.25}$$

Assim, a cada passo se realiza uma iteração nas equações (1.24) e (1.25), de modo que cada procedimento (EB e Máxima Verossimilhança) utiliza estimadores obtidos na última iteração. Diz-se que o algoritmo converge quando os estimadores não se alteram significativamente de uma iteração para outra.

Veja que se os efeitos individuais $\tilde{\mathbf{c}}$ são muito próximos de 0, então $\hat{\sigma}_c^2 \rightarrow 0$ e os estimadores de máxima verossimilhança de β e σ_ϵ^2 se aproximam dos estimadores de mínimos quadrados ordinários (*OLS - Ordinary Least Squares*) do modelo combinado

$$\begin{aligned}\hat{\beta}_{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \hat{\sigma}_\epsilon^2 &= \frac{1}{NT}(\mathbf{y} - \mathbf{X}\hat{\beta}_{OLS})'(\mathbf{y} - \mathbf{X}\hat{\beta}_{OLS})\end{aligned}$$

Uma ressalva importante a ser feita é o viés de $\hat{\sigma}_\epsilon^2$ em pequenas amostras (Verbeke and Molenberghs, 2000). Note que $\hat{\sigma}_\epsilon^2 = \frac{(\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}})}{NT}$, ao passo que o estimador - não viesado - de mínimos quadrados é dado por $\hat{\sigma}_\epsilon^2 = \frac{(\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}})}{NT-k-1}$. Apesar do viés ser negligenciável em grandes amostras, é complicado estabelecer um N mínimo a partir do qual isso acontece. A melhor alternativa nesse sentido seria estimar os parâmetros por meio da máxima verossimilhança restrita. Este método, todavia, não será apresentado aqui. Para mais detalhes, veja McCulloch and Searle (2001) e, ainda mais profundamente, Patterson and R. (1971). Uma grande vantagem do método de máxima verossimilhança é a aplicabilidade direta dos testes de razão de verossimilhança para comparar modelos com covariáveis diferentes. Nenhum método de estimação que corrija a verossimilhança pelos graus de liberdade referentes ao número de covariáveis possui essa atrativa propriedade.

⁹Diferentemente do contexto Bayesiano, em que o pesquisador elicita a a distribuição *a priori* com base em conhecimentos prévios, a distribuição do efeito aleatório tem seus parâmetros estimados por meio dos dados.

1.5 Testes de Hipóteses

Apesar de serem diferentes do ponto de vista da estrutura que gerou os dados, é complicado conhecer essa estrutura *a priori* em aplicações. Por isso, é interessante guiar a modelagem com base em testes estatísticos. Esta seção se dedica a expor os testes mais amplamente utilizados no estudo de dados em painel.

1.5.1 Testando a adequabilidade do modelo combinado

Pelo fato do modelo combinado ser o mais simples¹⁰, é relevante verificar se ele já fornece um ajuste razoável aos dados ou se a investigação de uma especificação mais robusta deve ser levada adiante.

Seja $\mathbf{y}_i = [y_{i1}, \dots, y_{iT}]'$, $\mathbf{Z}_i = [\mathbf{1}, \mathbf{X}_{1i}, \dots, \mathbf{X}_{ki}]$ e $\boldsymbol{\epsilon}_i = [\epsilon_{i1}, \dots, \epsilon_{iT}]'$. A hipótese nula dos testes de adequabilidade do modelo combinado é $H_0 : \delta_1 = \dots = \delta_N = \delta$ na especificação

$$\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\delta}_i + \boldsymbol{\epsilon}_i \quad (1.26)$$

Sendo assim, o modelo restrito (sob H_0) pode ser escrito como

$$\mathbf{y} = \mathbf{Z} \boldsymbol{\delta} + \boldsymbol{\epsilon} \quad (1.27)$$

onde $\mathbf{y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]'$, $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_N]'$ e $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N]'$.

O modelo irrestrito pode ser representado por

$$\mathbf{y} = \begin{bmatrix} \mathbf{Z}_1 & 0 & \dots & 0 \\ 0 & \mathbf{Z}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{Z}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_k \end{bmatrix} + \boldsymbol{\epsilon} = \mathbf{Z}^* \boldsymbol{\delta}^* + \boldsymbol{\epsilon}^*$$

Note que $\mathbf{Z} = \mathbf{Z}^* \mathbf{I}^*$, onde $\mathbf{I}^* = \mathbf{j}_N \otimes \mathbf{I}_{k+1}$ (Baltagi, 2008).

Adequabilidade do modelo combinado para $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}_{NT})$

O estimador de mínimos quadrados (e máxima verossimilhança) do modelo (1.27) é dado por

$$\hat{\boldsymbol{\delta}} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y} \quad (1.28)$$

De forma análoga, para a i -ésima observação,

$$\hat{\boldsymbol{\delta}}_i = (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \mathbf{y}_i \quad (1.29)$$

¹⁰Do ponto de vista do processo de estimação.

Sejam as matrizes $\mathbf{M} = (\mathbf{I} - \mathbf{H})\mathbf{y}$, onde $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{M}_i = (\mathbf{I} - \mathbf{H}_i)\mathbf{y}_i$, onde $\mathbf{H}_i = \mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'$ e $\mathbf{M}^* = \begin{bmatrix} \mathbf{M}_1 & 0 & \dots & 0 \\ 0 & \mathbf{M}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{M}_N \end{bmatrix}$. Além disso, \mathbf{M} é simétrica e idempotente.

Uma propriedade útil das matrizes \mathbf{M} e \mathbf{M}^* é $\mathbf{M}\mathbf{M}^* = \mathbf{M}^*$. A prova deste fato é apresentada a seguir:

$$\begin{aligned} \mathbf{M}\mathbf{M}^* &= (\mathbf{I}_{NT} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')(\mathbf{I}_{NT} - \mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}) \\ &= \mathbf{I}_{NT} - \mathbf{I}_{NT}\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{I}_{NT} + \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime} \\ &= \mathbf{I}_{NT} - \mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' + \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \\ &= \mathbf{I}_{NT} - \mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime} \\ &= \mathbf{M}^* \end{aligned}$$

uma vez que

$$\begin{aligned} \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime} &= \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{I}_{NT}\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime} \\ &= \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \end{aligned}$$

Então, a diferença na soma dos quadrados dos resíduos (via regressão por mínimos quadrados) entre o modelo restrito¹¹ e o modelo irrestrito é dado por:

$$\begin{aligned} \mathbf{e}'\mathbf{e} - \mathbf{e}^{*\prime}\mathbf{e}^* &= \mathbf{y}'\mathbf{M}\mathbf{M}\mathbf{y} - \mathbf{y}'\mathbf{M}^*\mathbf{M}^*\mathbf{y} \\ &= \mathbf{y}'\mathbf{M}\mathbf{y} - \mathbf{y}'\mathbf{M}^*\mathbf{y} \\ &= \mathbf{y}'(\mathbf{M} - \mathbf{M}^*)\mathbf{y} \end{aligned} \tag{1.30}$$

Pode-se, portanto, proceder com um teste F para várias restrições:

$$\begin{aligned} F_{obs} &= \frac{\frac{SQR_r - SQR_i}{gl_i - gl_r}}{\frac{SQR_i}{gl_i}} = \frac{\frac{\mathbf{y}'(\mathbf{M} - \mathbf{M}^*)\mathbf{y}}{tr(\mathbf{M}) - tr(\mathbf{M}^*)}}{\frac{\mathbf{y}'\mathbf{M}^*\mathbf{y}}{tr(\mathbf{M}^*)}} \\ &= \frac{\frac{\mathbf{y}'(\mathbf{M} - \mathbf{M}^*)\mathbf{y}}{(N-1)(k+1)}}{\frac{\mathbf{y}'\mathbf{M}^*\mathbf{y}}{N(T-(k+1))}} \end{aligned} \tag{1.31}$$

que tem distribuição $F_{(N-1)(k+1), N(T-(k+1))}$. A grande desvantagem do teste formulado em (1.31) é o pressuposto de que a estrutura de covariâncias de $\boldsymbol{\epsilon}$ é $\sigma^2\mathbf{I}_{NT}$.

Uma alternativa mais geral que (1.31) é o teste de adequabilidade do modelo combinado para $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Omega})$.

¹¹O vetor de resíduos dos modelos restrito e irrestrito estão representados por \mathbf{e} e \mathbf{e}^* , respectivamente.

Adequabilidade do modelo combinado para $\epsilon \sim N(0, \Omega)$

O pressuposto de que $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT})$ pode ser muito restritivo. Nesse caso, é possível adaptar o teste (1.31) para comportar estruturas de variâncias mais gerais.

Sendo assim, se $\Omega = \sigma^2 \Sigma$, tem-se o modelo equivalente a (1.27) (Baltagi, 2008):

$$\dot{\mathbf{y}} = \dot{\mathbf{Z}}\delta + \dot{\epsilon} \quad (1.32)$$

onde $\dot{\mathbf{y}} = \Sigma^{-\frac{1}{2}}\mathbf{y}$, $\dot{\mathbf{Z}} = \Sigma^{-\frac{1}{2}}\mathbf{Z}$ e $\dot{\epsilon} = \Sigma^{-\frac{1}{2}}\epsilon$.

Note que $E(\dot{\epsilon}) = E(\Sigma^{-\frac{1}{2}}\epsilon) = \Sigma^{-\frac{1}{2}}E(\epsilon) = 0$ e $Var(\dot{\epsilon}) = Var(\Sigma^{-\frac{1}{2}}\epsilon) = \Sigma^{-\frac{1}{2}}E(\epsilon\epsilon')\Sigma^{-\frac{1}{2}} = \sigma^2 \mathbf{I}_{NT}$.

O modelo irrestrito também é transformado de modo que $\dot{\mathbf{Z}}^* = \Sigma^{-\frac{1}{2}}\mathbf{Z}^*$.

$$\dot{\mathbf{y}} = \dot{\mathbf{Z}}^*\delta^* + \dot{\epsilon} \quad (1.33)$$

Portanto, o teste F ajustado para a matriz de covariâncias Ω é dado por:

$$\begin{aligned} \dot{F}_{obs} &= \frac{\frac{SQR_r - SQR_i}{gl_i - gl_r}}{\frac{SQR_i}{gl_i}} = \frac{\frac{\mathbf{y}'(\dot{\mathbf{M}} - \dot{\mathbf{M}}^*)\mathbf{y}}{tr(\dot{\mathbf{M}}) - tr(\dot{\mathbf{M}}^*)}}{\frac{\mathbf{y}'\dot{\mathbf{M}}^*\mathbf{y}}{tr(\dot{\mathbf{M}}^*)}} \\ &= \frac{\frac{\mathbf{y}'(\dot{\mathbf{M}} - \dot{\mathbf{M}}^*)\mathbf{y}}{(N-1)(k+1)}}{\frac{\mathbf{y}'\dot{\mathbf{M}}^*\mathbf{y}}{N(T-(k+1))}} \end{aligned} \quad (1.34)$$

de modo que $\dot{F} \sim F_{(N-1)(k+1), N(T-(k+1))}$.

Perceba que Σ é assumido como conhecido. Em aplicações práticas, se utiliza um estimador consistente $\hat{\Sigma}$.

1.5.2 Testando a existência de efeitos aleatórios

Nesta seção serão apresentados testes que indicam a existência de efeitos aleatórios diretamente (teste de *Breusch-Pagan*) e a ortogonalidade entre os efeitos aleatórios e as covariáveis (pressuposto do modelo de efeitos aleatórios).

O Teste de *Breusch-Pagan* para efeitos aleatórios

Este teste foi desenvolvido para testar a significância dos efeitos aleatórios do indivíduo e do tempo (conjuntamente ou um a um) (Breusch and Pagan, 1979).

Seja a função log-verossimilhança do modelo de 2 efeitos aleatórios:

$$l(y|\boldsymbol{\theta}, \mathbf{X}) = k - \frac{1}{2} \log|\Omega| - \frac{1}{2} \boldsymbol{\epsilon}'\Omega\boldsymbol{\epsilon} \quad (1.35)$$

onde $\boldsymbol{\theta} = (\sigma_c^2, \sigma_\lambda^2, \sigma_\epsilon^2)$ e $\boldsymbol{\Omega} = \sigma_c^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_\lambda^2(\mathbf{J}_N \otimes \mathbf{I}_T) + \sigma_\epsilon^2\mathbf{I}_{NT}$.

Assim, o vetor escore $U(\boldsymbol{\theta}) = \frac{\partial l}{\partial \boldsymbol{\theta}}$ é dado, de forma geral para a k -ésima componente de $\boldsymbol{\theta}$, pela expressão (Hartley and Rao, 1967):

$$\frac{\partial l}{\partial \theta_k} = \frac{1}{2} \text{tr} \left[\boldsymbol{\Omega}^{-1} \left(\frac{\partial \boldsymbol{\Omega}}{\partial \theta_k} \right) \right] + \frac{1}{2} \left[\boldsymbol{\epsilon}' \boldsymbol{\Omega}^{-1} \left(\frac{\partial \boldsymbol{\Omega}}{\partial \theta_k} \right) \boldsymbol{\Omega}^{-1} \boldsymbol{\epsilon} \right] \quad (1.36)$$

para $k=1,2$ e 3 . Assim, é possível verificar que

$$\begin{aligned} \frac{\partial \boldsymbol{\Omega}}{\partial \sigma_c^2} &= \mathbf{I}_N \otimes \mathbf{J}_T \\ \frac{\partial \boldsymbol{\Omega}}{\partial \sigma_\lambda^2} &= \mathbf{J}_N \otimes \mathbf{I}_T \\ \frac{\partial \boldsymbol{\Omega}}{\partial \sigma_\epsilon^2} &= \mathbf{I}_{NT} \end{aligned}$$

O estimador de máxima verossimilhança sob $H_0 : \sigma_c^2 = \sigma_\lambda^2 = 0$ é tal que $\tilde{\boldsymbol{\Omega}} = \sigma_\epsilon^2 \mathbf{I}_{NT}$ e $\tilde{\sigma}_\epsilon^2 = \frac{\tilde{\boldsymbol{\epsilon}}' \tilde{\boldsymbol{\epsilon}}}{NT}$. Além disso, $\tilde{\boldsymbol{\epsilon}}$ são os resíduos resultantes do modelo combinado estimado via máxima verossimilhança.

Sabendo que $\text{tr}(\mathbf{I}_N \otimes \mathbf{J}_T) = \text{tr}(\mathbf{J}_N \otimes \mathbf{I}_T) = \text{tr}(\mathbf{I}_{NT}) = NT$, é possível verificar que, sob H_0 :

$$U(\tilde{\boldsymbol{\theta}}) = -\frac{NT}{2\sigma_\epsilon^2} \left[1 - \frac{\tilde{\boldsymbol{\epsilon}}' \tilde{\boldsymbol{\epsilon}} (\mathbf{I}_N \otimes \mathbf{J}_T) \tilde{\boldsymbol{\epsilon}}}{\tilde{\boldsymbol{\epsilon}}' \tilde{\boldsymbol{\epsilon}}}, 1 - \frac{\tilde{\boldsymbol{\epsilon}}' \tilde{\boldsymbol{\epsilon}} (\mathbf{J}_N \otimes \mathbf{I}_T) \tilde{\boldsymbol{\epsilon}}}{\tilde{\boldsymbol{\epsilon}}' \tilde{\boldsymbol{\epsilon}}}, 0 \right] \quad (1.37)$$

$$\text{e } \mathcal{I}(\tilde{\boldsymbol{\theta}}) = E \left(-\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) = \frac{NT}{2\sigma_\epsilon^4} \begin{bmatrix} T & 1 & 1 \\ 1 & N & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

A estatística de teste é dada por $LM = U(\tilde{\boldsymbol{\theta}})' \mathcal{I}(\tilde{\boldsymbol{\theta}})^{-1} U(\tilde{\boldsymbol{\theta}})$ e tem distribuição χ^2 com 2 graus de liberdade.

A estatística LM é amplamente utilizada em virtude de sua facilidade de cálculo (são necessários apenas os resíduos do modelo combinado). Além disso, a natureza aditiva da distribuição χ^2 permite que LM seja particionada em $LM = LM_1 + LM_2$, a saber:

$$\begin{aligned} LM_1 &= \frac{NT}{2(T-1)} \left[1 - \frac{\tilde{\boldsymbol{\epsilon}}' (\mathbf{I}_N \otimes \mathbf{J}_T) \tilde{\boldsymbol{\epsilon}}}{\tilde{\boldsymbol{\epsilon}}' \tilde{\boldsymbol{\epsilon}}} \right] \\ LM_2 &= \frac{NT}{2(T-1)} \left[1 - \frac{\tilde{\boldsymbol{\epsilon}}' (\mathbf{J}_N \otimes \mathbf{I}_T) \tilde{\boldsymbol{\epsilon}}}{\tilde{\boldsymbol{\epsilon}}' \tilde{\boldsymbol{\epsilon}}} \right] \end{aligned}$$

Caso a hipótese de interesse seja $H_0 : \sigma_c^2 = 0$, deve-se utilizar a estatística LM_1 . Se o interesse for em $H_0 : \sigma_\lambda^2 = 0$, utiliza-se LM_2 . Tanto LM_1 quanto LM_2 têm distribuição χ^2 com 1 grau de liberdade.

1.5.3 Testando a existência de efeitos fixos

Uma vez que efeitos fixos se comportam como parâmetros a serem estimados, o procedimento para testar se eles são estatisticamente significativos é análogo ao caso de corte transversal. Veja Wooldridge (2001) e Johnston and Dinardo (1996) para maiores detalhes.

Basicamente, pode-se testar a significância de cada variável indicadora de forma prática por meio de um teste de Wald, a saber:

$$\hat{W} = \frac{(\hat{c}_i)^2}{Var(\hat{c}_i)} \sim \chi_{gli}^2 \quad (1.38)$$

onde gli representa os graus de liberdade do modelo irrestrito.

No caso de um teste de significância conjunta, um procedimento similar pode ser utilizado, adaptando-se as devidas dimensões:

$$\hat{W} = \hat{\mathbf{c}}'[Var(\hat{\mathbf{c}})]^{-1}\hat{\mathbf{c}} \sim \chi_{gli}^2 \quad (1.39)$$

Outro teste muito popular e que envolve informações tanto do modelo restrito quanto do irrestrito é o teste F (Razão de Verossimilhança) da forma:

$$F_{obs} = \frac{\frac{SQR_r - SQR_i}{gl_r - gl_i}}{\frac{SQR_i}{gl_i}} \quad (1.40)$$

o qual tem distribuição F com $(gl_r - gl_i, gl_i)$ graus de liberdade no numerador e denominador, respectivamente.

Nesse contexto, pode-se formular e testar uma variedade de hipóteses, por exemplo $H_0 : c_1 = \dots = c_N = 0$ versus $H_a : \text{pelo menos um } c_i \neq 0$, $H_0 : \lambda_1 = \dots = \lambda_T = 0$ versus $H_a : \text{pelo menos um } \lambda_t \neq 0$. Essas hipóteses têm aplicações práticas extremamente relevantes como H_0 : Modelo Combinado versus H_a : Modelo de 1 efeito fixo e H_0 : Modelo de 1 efeito fixo versus H_a : Modelo de 2 efeitos fixos, respectivamente.

1.5.4 O teste de especificação de Hausman

O principal pressuposto para viabilidade do modelo de efeitos aleatórios é a ortogonalidade entre os efeitos individuais e as variáveis exógenas, isto é, $E(v_{it}|\mathbf{X}_{it}) = E(v_{it}) = 0$.

Eventualmente, características intrínsecas (constantes ao longo do tempo) de cada indivíduo podem influenciar o nível de \mathbf{X}_{it} . Nesse caso, $E(v_{it}|\mathbf{X}_{it}) \neq 0$ e, conseqüentemente, $E(\hat{\beta}_{GLS}) \neq \beta$.

Lembre-se, porém, que o estimador resultante da *Within Transformation* ($\hat{\beta}_{WT}$) desconsidera variáveis constantes ao longo do tempo, como os efeitos individuais. Sendo assim, $\hat{\beta}_{WT}$ é consistente para β independentemente de $E(v_{it}|\mathbf{X}_{it})$.

Com base nesses fatos, Hausman propôs um teste comparando $\hat{\beta}_{WT}$ e $\hat{\beta}_{GLS}$, uma vez que ambos são consistentes sob $H_0 : E(v_{it}|\mathbf{X}_{it}) = 0$, mas apenas $\hat{\beta}_{WT}$ é consistente caso $E(v_{it}|\mathbf{X}_{it}) \neq 0$ (Hausman, 1978).

Seja, então, $\hat{W} = \hat{\beta}_{GLS} - \hat{\beta}_{WT}$. Note que, sob H_0 (Arellano, 2003):

$$\begin{aligned} E(\hat{W}) &= 0 \\ Var(\hat{W}) &= Var(\hat{\beta}_{WT}) + Var(\hat{\beta}_{GLS}) - 2Cov(\hat{\beta}_{WT}, \hat{\beta}_{GLS}) \end{aligned}$$

mas,

$$\begin{aligned} Var(\hat{\beta}_{GLS}) &= (\mathbf{X}'\mathbf{\Omega}\mathbf{X})^{-1} \\ Var(\hat{\beta}_{WT}) &= \sigma^2(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1} \\ Cov(\hat{\beta}_{GLS}, \hat{\beta}_{WT}) &= E(\hat{\beta}_{GLS}\hat{\beta}'_{WT}) - E(\hat{\beta}_{GLS})E(\hat{\beta}'_{WT}) \\ &= E[(\beta + (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\epsilon)(\beta + (\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}^{-1}\epsilon)'] - \beta\beta' \\ &= E[\beta(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}\epsilon + (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\epsilon\beta' + \\ &\quad (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\epsilon\epsilon'\mathbf{M}\mathbf{X}(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}] \\ E(\epsilon\epsilon') &= \mathbf{\Omega} \\ Cov(\hat{\beta}_{GLS}, \hat{\beta}_{WT}) &= E[(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{\Omega}\mathbf{M}\mathbf{X}(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \end{aligned}$$

Portanto,

$$\begin{aligned} Var(\hat{W}) &= Var(\hat{\beta}_{GLS}) + Var(\hat{\beta}_{WT}) - 2Cov(\hat{\beta}_{WT}, \hat{\beta}_{GLS}) \\ &= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} + \sigma^2(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1} - 2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \end{aligned} \tag{1.41}$$

e $\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{NT}$ e \mathbf{e} é o vetor de resíduos do modelo combinado.

Sendo assim, o teste de Wald proposto por Hausman tem a forma:

$$W_{obs} = \hat{\mathbf{W}}'[\hat{\sigma}^2(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1} - (\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}]^{-1}\hat{\mathbf{W}} \tag{1.42}$$

A estatística W_{obs} tem distribuição assintótica χ^2 com k graus de liberdade, onde k é o número de variáveis que se alteram ao longo do tempo.

1.6 O Modelo de Hausman-Taylor

Seja o modelo

$$y_{it} = x_{it}\beta + z_i\delta + c_i + \epsilon_{it} \tag{1.43}$$

onde x_{it} são as covariáveis da observação i no tempo t , z_i é um conjunto de variáveis constantes ao longo do tempo¹², c_i é um efeito aleatório ($c_i \sim N(\mathbf{0}, \sigma_c^2 \mathbf{I}_N)$) individual e ϵ é o erro idiossincrático tal que $\epsilon_i \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_{NT})$.

Infelizmente, nenhum dos procedimentos apresentados nas seções anteriores consegue estimar o efeito δ das variáveis constantes ao longo do tempo (Arellano, 2003). O estimador β_{GLS} não é consistente, uma vez que o efeito aleatório c_i é confundido com z_i e, portanto, $E(c_i|x_{it}, z_i) \neq 0$, ao passo que proceder com o método *Within Transformation* (para estimar os efeitos fixos β e δ) desconsidera δ , já que ele consiste em estimar o efeito de $z_{it} - \bar{z}_i$, mas $z_{it} = \bar{z}_i, \forall t$. Outro problema recorrente, é a possível correlação entre \mathbf{X}_i e c_i , que faz com que β_{GLS} seja inconsistente.

Sendo assim, (Hausman and Taylor, 1981) propuseram um modelo para estimar, de forma consistente, $\theta = (\beta, \delta, \sigma_c^2, \sigma_\epsilon^2)'$. Basicamente, o método consiste em selecionar algumas colunas de \mathbf{X} que não estejam correlacionadas com c_i para aproveitá-las de duas formas:

1. porque variam tanto ao longo dos indivíduos e do tempo, essas variáveis geram estimativas consistentes de β e;
2. a média dos indivíduos funciona como instrumento para z .

Uma vez escolhidas as variáveis que não são correlacionadas com c_i , é interessante testar a hipótese de ausência de correlação para não depender de justificativas essencialmente *ad hoc*.

Sejam, portanto, as partições $\mathbf{X} = (\mathbf{X}_1|\mathbf{X}_2)$ e $\mathbf{Z} = (\mathbf{Z}_1|\mathbf{Z}_2)$, de tal forma que, se $N \rightarrow \infty$, então:

$$\begin{aligned} \frac{1}{N} \mathbf{X}'_1 \mathbf{c} &\xrightarrow{p} \mathbf{0} \\ \frac{1}{N} \mathbf{X}'_2 \mathbf{c} &\xrightarrow{p} \mathbf{0} \\ \frac{1}{N} \mathbf{Z}'_1 \mathbf{c} &\xrightarrow{p} \mathbf{h}_X \\ \frac{1}{N} \mathbf{Z}'_2 \mathbf{c} &\xrightarrow{p} \mathbf{h}_Z \end{aligned}$$

e os vetores \mathbf{h}_X e \mathbf{h}_Z são diferentes de $\mathbf{0}$.

Sejam, também, as matrizes de projeção ortogonal $\mathbf{P} = (\mathbf{I}_N \otimes \frac{1}{T} \mathbf{j}_T \mathbf{j}'_T)$ e $\mathbf{M} = \mathbf{I}_{NT} - \mathbf{P}$. Então, transformando o modelo (1.43) por meio de \mathbf{M} , tem-se que

$$\mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{X}\beta + \mathbf{M}\mathbf{Z}\delta + \mathbf{M}\mathbf{c} + \mathbf{M}\epsilon \quad (1.44)$$

¹²Numa aplicação em que pessoas são unidades observacionais, alguns exemplos de características constantes ao longo do tempo são altura, gênero e anos de estudo.

que é equivalente a (veja o apêndice para mais detalhes):

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}} \quad (1.45)$$

onde $\tilde{\mathbf{y}} = \mathbf{M}\mathbf{y}$, $\tilde{\mathbf{X}} = \mathbf{M}\mathbf{X}$ e $\tilde{\boldsymbol{\epsilon}} = \mathbf{M}\boldsymbol{\epsilon}$. O estimador¹³ $\hat{\boldsymbol{\beta}}_{WT} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}$ é consistente para $\boldsymbol{\beta}$, independentemente da correlação entre \mathbf{c} e qualquer covariável.

Para captar a variação entre grupos, (1.43) é modificada de modo que

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\mathbf{Z}\boldsymbol{\delta} + \mathbf{P}\mathbf{c} + \mathbf{P}\boldsymbol{\epsilon}$$

que é equivalente a $y_i = x_i\boldsymbol{\beta} + z_i\boldsymbol{\delta} + c_i + \epsilon_i$. Não é possível, todavia, estimar $\boldsymbol{\delta}$ de forma consistente se $E(\mathbf{c}|\mathbf{X}, \mathbf{Z}) \neq \mathbf{0}$. A forma mais segura de verificar que isso acontece é por meio da rejeição do teste de Hausman. Caso ele seja rejeitado, a alternativa mais plausível para estimar $\boldsymbol{\theta}$ é o modelo de Hausman-Taylor (Hausman and Taylor, 1981). Sendo assim, o procedimento para estimar tal modelo é:

Calcular os resíduos do modelo intra-grupos $\hat{d}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_{WT}$; Obter o estimador $\hat{\boldsymbol{\delta}} = (\mathbf{Z}'\mathbf{P}_A\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_A\hat{d}$, onde $\mathbf{P}_A = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}$ e $\mathbf{A} = [\mathbf{X}_1|\mathbf{Z}_1]$; Calcular as variâncias dos efeitos aleatórios¹⁴:

$$\begin{aligned} \hat{\sigma}_v^2 &= \frac{\tilde{\mathbf{y}}'\mathbf{P}_A\tilde{\mathbf{y}}}{N(T-1)} \\ \hat{\sigma}_1^2 &= \frac{1}{N}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\hat{\boldsymbol{\delta}})'\mathbf{P}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\hat{\boldsymbol{\delta}}) \end{aligned}$$

Uma vez que os componentes da matriz de covariâncias $\boldsymbol{\Omega}$ foram estimados, o modelo de Hausman-Taylor é obtido por meio de mínimos quadrados generalizados em

$$\hat{\boldsymbol{\Omega}}^{-\frac{1}{2}}\mathbf{y} = \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} + \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}}\mathbf{Z}\boldsymbol{\delta} + \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}}\boldsymbol{\epsilon} \quad (1.46)$$

A condição necessária para que $\boldsymbol{\delta}$ seja identificável é $\text{Posto}(\mathbf{P}_A[\mathbf{XZ}]) \leq \text{Posto}(\mathbf{P}_A\mathbf{X}) + \text{Posto}(\mathbf{P}_A\mathbf{Z})$ (Hausman and Taylor, 1981). Para mais detalhes do modelo de Hausman-Taylor, veja Baltagi (2008), Hausman (1978) e Hausman and Taylor (1981).

A aplicabilidade desse modelo, todavia, é, de certo modo, limitada pela dificuldade em encontrar variáveis exógenas que variem ao longo do tempo e sejam, simultaneamente, independentes do efeito individual.

¹³Este estimador é conhecido como intra-grupos (*within-groups*).

¹⁴Efeito individual e erro idiossincrático.

Capítulo 2

Econometria Espacial

Em aplicações práticas, eventualmente nos deparamos com dados espacialmente referenciados, de modo que a localização influencia as características individuais de cada unidade observacional. Isso acontece porque a localização de uma determinada observação pode determinar o nível de interação com unidades adjacentes. Se a modelagem estiver sendo realizada no nível do município, por exemplo, é intuitivo pensar que municípios vizinhos tendem a realizar trocas comerciais mais intensamente por vários motivos, como economia de tempo e menor custo com transporte. Sendo assim, é interessante incorporar no processo de modelagem a possível influência que a distância entre as observações tem na variável de interesse.

No contexto dos modelos econométricos, se houver uma dependência entre as unidades observacionais advinda de sua localização espacial, não é possível utilizar técnicas clássicas de regressão (corte transversal ou painel) porque um dos pressupostos mais restritivos dessas é a independência entre as observações. Caso não se atente para este fato, todo o processo de estimação estará comprometido e seus resultados não serão satisfatórios (estimadores viesados, ineficientes e, às vezes, inconsistentes). A estimação dos modelos espaciais pode ser realizada tanto sob a perspectiva Bayesiana quanto sob a Clássica. A abordagem Clássica será utilizada predominante ao longo deste trabalho. Exemplos utilizando modelagem Bayesiana podem ser encontrados em Parent and LeSage (2008) e Lichstein et al. (2002).

Há várias formas de levar a estrutura espacial dos dados em consideração. Uma forma é por meio de uma matriz de proximidades \mathbf{W} . Proximidade, todavia, pode ser um conceito subjetivo e que, além disso, pode mudar conforme a situação. Num contexto em que a situação de interesse leva em consideração as trocas comerciais entre cidades, não é razoável utilizar apenas a distância euclideana entre eles, uma vez que tenderíamos a achar que

duas cidades do Pará, por exemplo, distantes 200 km entre si, conectadas unicamente por transporte fluvial e duas cidades de São Paulo, também distantes 200 km, mas com rodovias amplas e em excelente estado de conservação, estão ligadas de forma idêntica. Nesse tipo de situação, a distância deixa de ser uma boa medida do grau de vizinhança, e medidas como o tempo médio gasto no percurso entre as cidades seriam mais relevantes.

2.1 A matriz de proximidades \mathbf{W}

Os elementos da matriz \mathbf{W} tentam captar o grau de conectividade entre as observações de várias formas distintas e que dependem do problema em estudo. Algumas maneiras de definir \mathbf{W} (elemento típico w_{ij}) são (Silva, 2006):

1. $w_{ij} = 1$ se as fronteiras das observações i e j são adjacentes e $w_{ij} = 0$ caso contrário;
2. $w_{ij} = 1$ se o centróide da região i está a menos de k quilômetros da região j e $w_{ij} = 0$ caso contrário (o método *ad hoc* para escolha da distância máxima de corte k pode incitar muitos questionamentos);
3. $w_{ij} = 1$ se o centróide da área j é um dos k centróides mais próximos do centróide de i e $w_{ij} = 0$ caso contrário;
4. $w_{ij} = f(d_{ij})$ em f é uma função decrescente da distância d_{ij} entre i e j . Tipicamente, $f(d_{ij}) = \frac{1}{1+d_{ij}}$;

Costuma-se transformar as linhas de \mathbf{W} de modo que as linhas somem 1, isto é, obtêm-se novos $w_{ij}^* = \frac{w_{ij}}{w_{i.}}$, e $w_{i.} = \sum_j w_{ij}$. No entanto, é interessante testar estatisticamente a presença de dependência espacial no conjunto de dados para que o uso de metodologia que utiliza a matriz \mathbf{W} seja justificável formalmente. Isso pode ser feito de duas maneiras: utilizando indicadores globais ou locais.

2.2 Indicadores globais de dependência espacial

2.2.1 O índice I de Moran

O índice I de Moran (LeSage, 1999) é a medida mais utilizada para identificação de autocorrelação (dependência) espacial e é calculado da seguinte forma:

$$I = \frac{N (\mathbf{y} - \bar{\mathbf{y}})' \mathbf{W} (\mathbf{y} - \bar{\mathbf{y}})}{S (\mathbf{y} - \bar{\mathbf{y}})' (\mathbf{y} - \bar{\mathbf{y}})} \quad (2.1)$$

onde S é a soma de todos os elementos da matriz \mathbf{W} e N é a dimensão de \mathbf{W} . Note que, se a matriz for padronizada, i.e, se as linhas somarem 1, então $S = N$.

Assim como o coeficiente de correlação de Pearson, o I de Moran varia no intervalo $[-1, 1]$, onde 1 (-1) indica autocorrelação espacial perfeita positiva (negativa) e 0 indica ausência de autocorrelação espacial.

2.2.2 O índice C de Geary

Outro índice utilizado e mais indicado quando existe uma pequena quantidade de vizinhanças (Lembo, 2005) é o C de Geary, o qual varia no intervalo $[0, 2]$, sendo que valores próximos de 0 indicam autocorrelação espacial positiva e valores próximos de 2 indicam autocorrelação espacial negativa, ao passo que valores em torno de 1 evidenciam ausência de autocorrelação espacial. A estatística C é dada por:

$$C = \frac{n-1}{2} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (y_i - y_j)^2}{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N \sum_{j=1}^N w_{ij}} \quad (2.2)$$

Para mais detalhes em índices de associação espacial, veja Silva (2006). Todos esses índices visam orientar o pesquisador no sentido de utilizar, ou não, modelagem espacial, caso esta seja pertinente.

2.3 O modelo espacial autoregressivo de primeira ordem

Este modelo incorpora apenas informações da vizinhança da observação, isto é, não leva em conta outras variáveis explicativas que possam influenciar o nível de y_i ¹. Neste texto, este modelo será referido como *FAR* (*First Order Autoregressive Regression*). Em adição, a matriz \mathbf{W} é padronizada de modo que suas linhas somem 1 e \mathbf{y} é corrigido pela média para suprimir o intercepto. A especificação do modelo espacial autoregressivo de primeira ordem tem a forma:

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (2.3)$$

¹ \mathbf{y} é a variável de interesse.

Se fosse aplicado o método de mínimos quadrados ordinários ao modelo (2.3), o estimador obtido seria

$$\hat{\rho} = (\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y})^{-1}\mathbf{y}'\mathbf{W}'\mathbf{y} \quad (2.4)$$

Este estimador, todavia, é viesado, uma vez que:

$$\begin{aligned} E(\hat{\rho}) &= E[(\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y})^{-1}\mathbf{y}'\mathbf{W}'\mathbf{y}] \\ &= E[(\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y})^{-1}\mathbf{y}'\mathbf{W}'(\rho\mathbf{W}\mathbf{y} + \boldsymbol{\epsilon})] \\ &= E[(\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y})^{-1}\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y}\rho + (\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y})^{-1}\mathbf{y}'\mathbf{W}'\boldsymbol{\epsilon}] \\ &= \rho + E[(\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y})^{-1}\mathbf{y}'\mathbf{W}'\boldsymbol{\epsilon}] \end{aligned} \quad (2.5)$$

Sendo assim, não se pode garantir que $\mathbf{W}\mathbf{y}$ é fixo porque \mathbf{y} é gerado por um processo espacial e, portanto, não pode ser excluído do termo da expectância em (2.5). Uma analogia direta desse problema é a autocorrelação temporal dos resíduos. Dessa forma, deve-se buscar outros métodos de estimação para ρ . Um método natural é o da máxima verossimilhança que, apesar de depender da distribuição de probabilidade escolhida para \mathbf{y} , gera estimadores com propriedades ótimas, no mínimo, assintoticamente. Seja a verossimilhança e seu logaritmo dado por:

$$\mathcal{L}(\mathbf{y}|\rho, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} |\mathbf{I} - \rho\mathbf{W}| \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \rho\mathbf{W}\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}\mathbf{y})\right\} \quad (2.6)$$

$$l(\mathbf{y}|\rho, \sigma^2) = C - \frac{N}{2}(\mathbf{y} - \rho\mathbf{W}\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}\mathbf{y}) + \ln|\mathbf{I} - \rho\mathbf{W}| \quad (2.7)$$

uma vez que $\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \rho\mathbf{W}\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}\mathbf{y}) = \frac{1}{n} \sum_{i=1}^N \hat{\epsilon}_i^2$. Entretanto, após estimar ρ , é necessário voltar e estimar $\hat{\sigma}^2 = g(\hat{\rho}, \mathbf{y})$.

Note, porém, que maximizar \mathcal{L} implica em otimizar uma função do produto de vetores $(\mathbf{y} - \rho\mathbf{W}\mathbf{y})'(\mathbf{y} - \rho\mathbf{W}\mathbf{y})$ e do determinante $|\mathbf{I} - \rho\mathbf{W}|$ de dimensão N . Em aplicações econômicas, N é tipicamente grande, complicando a estimação de ρ do ponto de vista computacional.

É interessante ressaltar que o processo de otimização da verossimilhança deve calcular, de forma simultânea, a matriz hessiana, dada por $H[l(\theta)] = \frac{\partial^2 l}{\partial \theta' \partial \theta}$, onde $\theta = (\rho, \sigma^2)$. Isso acontece porque $\mathcal{I}(\theta) = -E[H(l(\theta))]$, em que $\mathcal{I}(\theta)$ é a matriz de informação de Fisher e suas componentes fornecem estimativas da dispersão de θ assintoticamente, uma vez que $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, \mathcal{I}(\theta)^{-1})$.

Evidentemente, para calcular $H[l(\theta)]$ são necessárias operações em uma matriz de dimensão N (no caso de corte transversal), e isso é inviável quando N é grande. Para resolver

este problema, uma das soluções possíveis é o uso de algoritmos desenvolvidos para matrizes esparsas.

O modelo *FAR*, de modo geral, é muito simples e não incorpora informações adicionais a respeito da unidade observacional em estudo e, portanto, tem utilidade restrita. É razoável considerar um modelo que mensure o efeito de características da observação i e de seus vizinhos no nível de y_i . Tal modelo é conhecido como modelo espacial regressivo-autoregressivo.

2.4 O modelo espacial regressivo-autoregressivo

Similar ao modelo autoregressivo de primeira ordem, o modelo espacial regressivo-autoregressivo (*SAR - Spatial Autoregressive Model*) apenas incorpora covariáveis ao primeiro. Sendo assim, a especificação típica do *SAR* é

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (2.8)$$

Para estimar o modelo *SAR*, alguns passos são necessários (Anselin, 1988):

1. realizar mínimos quadrados ordinários em $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}_0$;
2. realizar mínimos quadrados em $\mathbf{W}\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_L + \boldsymbol{\epsilon}_L$;
3. obter $\boldsymbol{\epsilon}_0$ e $\boldsymbol{\epsilon}_L$;
4. maximizar ρ na verossimilhança $\mathcal{L} = \left(\frac{-n}{2}\right) \ln \pi - \left(\frac{n}{2}\right) \ln \frac{1}{n} (\boldsymbol{\epsilon}_0 - \rho \boldsymbol{\epsilon}_L)' (\boldsymbol{\epsilon}_0 - \rho \boldsymbol{\epsilon}_L) + \ln |\mathbf{I} - \rho \mathbf{W}|$;
5. dado o $\hat{\rho}$, obter $\hat{\boldsymbol{\beta}} = (\boldsymbol{\beta}_0 - \hat{\rho} \hat{\boldsymbol{\beta}}_L)$, $\hat{\sigma}_\epsilon = \frac{1}{n} (\boldsymbol{\epsilon}_0 - \hat{\rho} \boldsymbol{\epsilon}_L)' (\boldsymbol{\epsilon}_0 - \hat{\rho} \boldsymbol{\epsilon}_L)$.

As seções 2.4.1 e 2.4.2 abordam o processo de estimação com maior profundidade.

2.4.1 O modelo *SAR*: aspectos probabilísticos

A função de verossimilhança do modelo *SAR* pode ser obtida por meio do método do jacobiano para transformações de variáveis aleatórias, a saber:

Modelo de Corte Transversal convencional: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

$$f_Y(\mathbf{y}) = (2\pi)^{-\frac{n}{2}} |\sigma^2 \mathbf{I}| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \quad (2.9)$$

Modelo SAR: $(\mathbf{I} - \rho \mathbf{W})\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

Se $\mathbf{S} = (\mathbf{I} - \rho\mathbf{W})$, então

$$\begin{aligned} f(\mathbf{S}\mathbf{y}) &= f_Y(\mathbf{S}\mathbf{y}) \left| \frac{\partial \mathbf{S}\mathbf{y}}{\partial \mathbf{y}} \right| \\ f(\mathbf{S}\mathbf{y}) &= f_Y(\mathbf{S}\mathbf{y}) |\mathbf{S}| \\ f(\mathbf{S}\mathbf{y}) &= (2\pi)^{-\frac{n}{2}} |\sigma^2|^{-\frac{n}{2}} |\mathbf{S}| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{S}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{S}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \end{aligned}$$

No caso de \mathbf{y} observado, a verossimilhança em termos de $\mathbf{S}\mathbf{y}$ é dada por

$$\mathcal{L}(\beta, \sigma^2, \rho | \mathbf{y}) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} |\mathbf{S}| \exp \left\{ -\frac{1}{2\sigma^2} \hat{\boldsymbol{\epsilon}}' \hat{\boldsymbol{\epsilon}} \right\} \quad (2.10)$$

onde $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \rho\mathbf{W}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$. Os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e σ^2 são facilmente obtidos por meio de $\hat{\boldsymbol{\beta}}_{MV} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \hat{\rho}_{MV}\mathbf{W})\mathbf{y}$ e $\hat{\sigma}_{MV}^2 = \frac{1}{n}(\mathbf{I} - \hat{\rho}_{MV}\mathbf{W})\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MV}'(\mathbf{I} - \hat{\rho}_{MV}\mathbf{W})\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MV}$. Note que para obter tais estimadores é necessário obter o estimador $\hat{\rho}_{MV}$. Em geral, se utiliza a verossimilhança de perfil concentrada apenas em ρ . Este método consiste em fixar $\boldsymbol{\beta}$ e σ^2 em (2.10) e, em seguida, maximizar a verossimilhança de perfil em relação a ρ . Mais especificamente, o procedimento é formado pelos seguintes passos:

1. Fixar $\hat{\boldsymbol{\beta}}_{MQO} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ e $\hat{\sigma}^2 = \frac{1}{n}((\mathbf{I} - \hat{\rho}_{MV}\mathbf{W})\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MQO})'((\mathbf{I} - \hat{\rho}_{MV}\mathbf{W})\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MQO})$;

2. Reescrever (2.10) como

$$\mathcal{L}(\rho | \mathbf{y}) = (2\pi)^{-\frac{n}{2}} \left[\frac{1}{n} ((\mathbf{I} - \hat{\rho}_{MV}\mathbf{W})\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MQO})' ((\mathbf{I} - \hat{\rho}_{MV}\mathbf{W})\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MQO}) \right]^{-\frac{n}{2}} |\mathbf{I} - \rho\mathbf{W}| \exp \left\{ -\frac{1}{2\sigma^2} n\sigma^2 \right\}$$

Se $\mathbf{M} = (\mathbf{I} - \mathbf{H})$, onde $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, então $(\mathbf{I} - \hat{\rho}_{MV}\mathbf{W})\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MQO}'(\mathbf{I} - \hat{\rho}_{MV}\mathbf{W})\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MQO} = \mathbf{y}'(\mathbf{I} - \rho\mathbf{W}')\mathbf{M}'\mathbf{M}(\mathbf{I} - \rho\mathbf{W})\mathbf{y}$;

3. Otimizar a verossimilhança concentrada em ρ

$$\begin{aligned} \mathcal{L}(\rho | \mathbf{y}) &= K |\mathbf{I} - \rho\mathbf{W}| \left(\frac{1}{n} \mathbf{y}' (\mathbf{I} - \rho\mathbf{W}') \mathbf{M} (\mathbf{I} - \rho\mathbf{W}) \mathbf{y} \right)^{-\frac{n}{2}} \\ l(\rho | \mathbf{y}) &= \log K + \log |\mathbf{I} - \rho\mathbf{W}| - \frac{n}{2} \log \left(\frac{1}{n} \mathbf{y}' (\mathbf{I} - \rho\mathbf{W}') \mathbf{M} (\mathbf{I} - \rho\mathbf{W}) \mathbf{y} \right) \end{aligned}$$

O valor de $\hat{\rho}_{MV}$, todavia, deve estar contido no intervalo $\left(\frac{1}{\omega_{min}}, \frac{1}{\omega_{max}} \right)$, onde ω é o vetor de autovalores de \mathbf{W} (Ord, 1975), para que a matriz de covariâncias inerente a $\mathbf{S}\mathbf{y}$ seja positiva definida. Sendo assim, encontrar $\hat{\rho}_{MV}$ é um problema de otimização da função (2.11).

O grande desafio na obtenção de $\hat{\rho}_{MV}$ é avaliar o log-determinante $|\mathbf{I} - \rho\mathbf{W}|$ em cada iteração. Ord (1975) propôs o procedimento descrito na seção (2.4.2).

2.4.2 O modelo SAR: aspectos computacionais

Avaliar diretamente $|\mathbf{I} - \rho\mathbf{W}|$ em cada iteração do processo de estimação pode ser prático em aplicações em que há poucas unidades observacionais (N pequeno). Em aplicações econômicas, N é tipicamente grande e, portanto, é interessante utilizar um método que seja suficientemente flexível para comportar qualquer situação.

Assim, seja $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$. Se \mathbf{W} tem autovalores $\omega = (\omega_1, \dots, \omega_N)'$, então $|\omega\mathbf{I} - \mathbf{W}| = \prod_{i=1}^N (\omega - \omega_i)$. Se $\rho = \frac{1}{\omega}$, então $|\mathbf{A}| = \prod_{i=1}^N (1 - \rho\omega_i)$. A vantagem desse procedimento é que ω deve ser obtido apenas uma vez podendo, assim, ser utilizado em todas as iterações.

A log-verossimilhança (2.11) pode ser escrita como

$$l(\rho|\mathbf{y}) = K_1 + \sum_{i=1}^N \log(1 - \rho\omega_i) - \frac{n}{2} \log(\mathbf{y}'(\mathbf{I} - \rho\mathbf{W})'\mathbf{M}(\mathbf{I} - \rho\mathbf{W})\mathbf{y}) \quad (2.11)$$

Note que

$$\begin{aligned} \mathbf{y}'(\mathbf{I} - \rho\mathbf{W})'\mathbf{M}(\mathbf{I} - \rho\mathbf{W})\mathbf{y} &= \mathbf{y}'\mathbf{M}(\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \rho\mathbf{y}'\mathbf{W}'\mathbf{M}(\mathbf{I} - \rho\mathbf{W})\mathbf{y} \\ &= \mathbf{y}'\mathbf{M}\mathbf{y} - \rho\mathbf{y}'\mathbf{M}\mathbf{W}\mathbf{y} \\ &\quad - \rho\mathbf{y}'\mathbf{W}'\mathbf{M}\mathbf{y} + \rho^2\mathbf{y}'\mathbf{W}'\mathbf{M}\mathbf{W}\mathbf{y} \\ &= \mathbf{y}'\mathbf{M}\mathbf{y} - 2\rho\mathbf{y}'\mathbf{M}\mathbf{W}\mathbf{y} + \rho^2\mathbf{y}'\mathbf{W}\mathbf{M}\mathbf{W}\mathbf{y} \end{aligned}$$

O algoritmo de Newton-Raphson proposto por Ord (1975) consiste em

1. Seja $l(\rho) = \sum_{i=1}^N \log(1 - \rho\omega_i) - \frac{n}{2} \log(\mathbf{y}'\mathbf{M}\mathbf{y} - 2\rho\mathbf{y}'\mathbf{M}\mathbf{W}\mathbf{y} + \rho^2\mathbf{y}'\mathbf{W}\mathbf{M}\mathbf{W}\mathbf{y})$;
2. $\rho_{j+1} = \rho_j - \frac{l(\rho_j)}{l'(\rho_j)}$, onde ρ_j é ρ avaliado na j -ésima iteração.

Note que

$$\begin{aligned} l'(\rho) &= \frac{\partial l(\rho)}{\partial \rho} \\ &= - \sum_{i=1}^N \frac{\omega_i}{1 - \rho\omega_i} - \frac{n}{2} \left(\frac{-2\mathbf{y}'\mathbf{M}\mathbf{W}\mathbf{y} + 2\rho\mathbf{y}'\mathbf{W}'\mathbf{M}\mathbf{W}\mathbf{y}}{\mathbf{y}'\mathbf{M}\mathbf{y} - 2\rho\mathbf{y}'\mathbf{M}\mathbf{W}\mathbf{y} + \rho^2\mathbf{y}'\mathbf{W}\mathbf{M}\mathbf{W}\mathbf{y}} \right) \end{aligned}$$

Utilizando a propriedade que $\sqrt{N}(\hat{\theta}_{MV} - \theta) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}(\hat{\theta})^{-1})$, tem-se que $Var(\hat{\rho}_{MV}) = - \left[E \frac{\partial^2 l(\hat{\rho})}{\partial \rho^2} \right]^{-1}$. Esta não será derivada neste trabalho, mas pode ser obtida de forma direta.

2.5 O modelo espacial de erros autoregressivos

Eventualmente, a estrutura de dependência espacial é evidenciada nos resíduos. A

especificação típica de tal modelo é dada por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (2.12)$$

$$\mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (2.13)$$

O parâmetro λ desempenha papel semelhante ao ρ presente no modelo *SAR*, i.e., mensura a relevância da estrutura autoregressiva. Este modelo é usualmente chamado de *SEM - Spatial Autoregressive Error Model*.

Semelhante ao problema da autocorrelação temporal na regressão clássica, a autocorrelação espacial também é detectada por meio de teste estatísticos. O teste *I* de Moran, já abordado anteriormente, aplicado aos resíduos da regressão é o teste mais amplamente utilizado. Caso esteja-se utilizando uma matriz \mathbf{W} padronizada², a estatística do teste é dada por:

$$I = (\mathbf{e}'\mathbf{e})^{-1}(\mathbf{e}'\mathbf{W}\mathbf{e}) \quad (2.14)$$

onde \mathbf{e} é o vetor de resíduos. Sua distribuição assintótica é (Anselin, 1988):

$$\hat{I} \xrightarrow{\mathcal{L}} N \left(\frac{\text{tr}(\mathbf{M}\mathbf{W})}{n-k}, \frac{\text{tr}(\mathbf{M}\mathbf{W}\mathbf{M}\mathbf{W}') + \text{tr}(\mathbf{M}\mathbf{W})^2 + (\text{tr}(\mathbf{M}\mathbf{W}))^2}{d - \left(\frac{\text{tr}(\mathbf{M}\mathbf{W})}{n-k}\right)^2} \right) \quad (2.15)$$

onde $d = (n-k)(n-k+2)$ e $\mathbf{M} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$.

Caso a opção seja pela matriz \mathbf{W} não padronizada, a distribuição assintótica de \hat{I} é ligeiramente diferente, a saber:

$$\hat{I} \xrightarrow{\mathcal{L}} N \left(\left(\frac{n}{s}\right) \frac{\text{tr}(\mathbf{M}\mathbf{W})}{n-k}, \left(\frac{n}{s}\right)^2 \frac{\text{tr}(\mathbf{M}\mathbf{W}\mathbf{M}\mathbf{W}') + \text{tr}(\mathbf{M}\mathbf{W})^2 + (\text{tr}(\mathbf{M}\mathbf{W}))^2}{d - \left(\frac{\text{tr}(\mathbf{M}\mathbf{W})}{n-k}\right)^2} \right) \quad (2.16)$$

Alguns outros testes são também utilizados. Caso haja interesse do leitor, Anselin (1988) discute os testes de autocorrelação espacial de resíduos de maneira mais específica.

Sendo detectada a presença de autocorrelação, $V(\hat{\boldsymbol{\beta}})$ não é estimado consistentemente via mínimos quadrados ordinários. Mais especificamente, a $\Omega(\lambda) = V(\hat{\boldsymbol{\epsilon}})$ é uma função de λ , a saber:

$$\Omega = [(\mathbf{I} - \lambda \mathbf{W})'(\mathbf{I} - \lambda \mathbf{W})]^{-1} \quad (2.17)$$

Apesar disso, $\hat{\boldsymbol{\beta}}$ continua não viesado, porque $E[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}] = 0$. E sua

²i.e., que as linhas somem 1.

variância é dada por

$$\begin{aligned}
V(\hat{\beta}) &= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \\
&= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[(\mathbf{I} - \lambda\mathbf{W})(\mathbf{I} - \lambda\mathbf{W})]^{-1}\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned} \tag{2.18}$$

Uma vez que a forma da matriz de covariâncias dos erros tem forma conhecida (2.17), é viável utilizar a abordagem de mínimos quadrados generalizados (*GLS - Generalized Least Squares*). Sendo assim, o estimador *GLS* de β é dado por

$$\begin{aligned}
\hat{\beta}_{GLS} &= [\mathbf{X}'(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})\mathbf{y} \\
\hat{\beta}_{GLS} &= (\mathbf{X}^*\mathbf{X}^*)^{-1}\mathbf{X}^*\mathbf{y}^*
\end{aligned} \tag{2.19}$$

onde $\mathbf{X}^* = (\mathbf{I} - \lambda\mathbf{W})\mathbf{X}$ e $\mathbf{y}^* = (\mathbf{I} - \lambda\mathbf{W})\mathbf{y}$. Evidentemente, λ deve ser substituído por um estimador consistente. Usualmente, esse estimador é chamado de mínimos quadrados generalizados factíveis (*FGLS - Feasible Generalized Least Squares*). A variância de $\hat{\beta}_{GLS}$ é dada por:

$$\begin{aligned}
Var(\hat{\beta}_{GLS}) &= \hat{\sigma}^2[\mathbf{X}'(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})\mathbf{X}]^{-1} \\
\hat{\sigma}^2 &= \frac{1}{N}(\mathbf{y} - \mathbf{X}\hat{\beta}_{GLS})'(\mathbf{y} - \mathbf{X}\hat{\beta}_{GLS})
\end{aligned} \tag{2.20}$$

É interessante ressaltar a natureza assintótica do estimador de mínimos quadrados generalizados factíveis. A distribuição de probabilidades associada a $\hat{\beta}_{GLS}$ não está bem definida para amostras pequenas e nem há testes endossando a superioridade do estimador *GLS* ao de mínimos quadrados ordinários no sentido do erro quadrático médio (Anselin, 1988). Uma desvantagem clara na estimação *GLS* é a dependência de uma especificação correta na matriz de covariâncias do erro. Em regressão espacial, essa especificação decorre, inevitavelmente, da escolha da matriz \mathbf{W} .

Há outras formas mais complexas de se modelar uma variável de interesse, como o modelo Durbin, por exemplo, que incorpora, além da variável resposta defasada (no contexto espacial), covariáveis das unidades observacionais da vizinhança.

2.5.1 O modelo *SEM*: aspectos computacionais

O modelo *SEM*, cuja especificação

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}\beta + \mathbf{u} \\
\mathbf{y} &= \rho\mathbf{W}\mathbf{u} + \epsilon, \epsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})
\end{aligned}$$

pode ser reescrito como

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} - \rho \mathbf{W}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.21)$$

Cochrane and Orcutt (1949) propuseram um procedimento iterativo para modelagem sob autocorrelação temporal que, posteriormente, Ord (1975) adaptou para autocorrelação espacial. Este procedimento consiste em:

1. Realizar MQO³ em (2.21) e obter os resíduos \tilde{u} ;
2. Estimar ρ no modelo $\mathbf{u} = \rho \mathbf{W}\mathbf{u} + \boldsymbol{\epsilon}$, via máxima verossimilhança, conforme a seção (2.4.2) e obter $\tilde{\rho}$;
3. Construir as variáveis $\tilde{\mathbf{z}} = (\mathbf{I} - \tilde{\rho}\mathbf{W})\mathbf{y}$ e $\tilde{\mathbf{X}} = (\mathbf{I} - \tilde{\rho}\mathbf{W})\mathbf{X}$;
4. Realizar MQO no modelo $\tilde{\mathbf{z}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \mathbf{u}^*$ e calcular o estimador $\tilde{\boldsymbol{\beta}}$;
5. Obter os novos resíduos $\tilde{\mathbf{u}}^*$ e voltar ao passo 2;
6. Parar o processo quando $\tilde{\rho}$ e $\tilde{\boldsymbol{\beta}}$ estiverem estáveis entre uma iteração e a próxima.

2.6 O modelo espacial Durbin

O modelo espacial Durbin (*SDM - Spatial Durbin Model*) é assim chamado em alusão à abordagem de Durbin à regressão com resíduos temporalmente autocorrelacionados. Este modelo incorpora informações da vizinhança da observação i tanto na variável dependente quanto no preditor linear. A especificação típica de um modelo *SDM* é:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} - \rho \mathbf{W}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.22)$$

$$\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (2.23)$$

Na prática, geralmente utiliza-se o modelo equivalente $\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{W}\mathbf{X}\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$, com a restrição $\boldsymbol{\beta}_2 = -\rho\boldsymbol{\beta}_1$. Nesse contexto, os estimadores $\hat{\boldsymbol{\beta}}_1$ e $\hat{\boldsymbol{\beta}}_2$ são dados por

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}^* \mathbf{X}^*)^{-1} \mathbf{X}^* \mathbf{y} \quad (2.24)$$

$$\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}^* \mathbf{X}^*)^{-1} \mathbf{X}^* \mathbf{W}' \mathbf{y} \quad (2.25)$$

onde $\mathbf{X}^* = [\mathbf{X}' \mathbf{W}\mathbf{X}]$. O estimador de ρ pode ser obtido maximizando a verossimilhança

$$\ln(\mathcal{L}) = C + \ln|\mathbf{I} - \rho \mathbf{W}| - \frac{n}{2} \ln(\boldsymbol{\epsilon}'_1 \boldsymbol{\epsilon}_1 - 2\rho \boldsymbol{\epsilon}'_2 \boldsymbol{\epsilon}_1 + \rho^2 \boldsymbol{\epsilon}'_2 \boldsymbol{\epsilon}_2) \quad (2.26)$$

$$\boldsymbol{\epsilon}_1 = \mathbf{y} - \mathbf{X}^* \boldsymbol{\beta}_1 \quad (2.27)$$

$$\boldsymbol{\epsilon}_2 = \mathbf{W}\mathbf{y} - \mathbf{X}^* \boldsymbol{\beta}_2 \quad (2.28)$$

³Mínimos Quadrados Ordinários.

Após obter o estimador $\hat{\rho}_{MV}$, o estimador $\hat{\beta}$ é dado por $\hat{\beta} = \hat{\beta}_1 - \hat{\rho}_{MV}\beta_2$ e $\hat{\sigma}^2 = \frac{1}{N}(\mathbf{y} - \hat{\rho}_{MV}\mathbf{W}\mathbf{y} - \mathbf{X}^*\hat{\beta})'(\mathbf{y} - \hat{\rho}_{MV}\mathbf{W}\mathbf{y} - \mathbf{X}^*\hat{\beta})$.

De modo geral, o modelo de Durbin é um excelente ponto de partida, uma vez que os modelos mais simples são apenas casos particulares dele. Outra vantagem desse modelo é que as interações entre vizinhos são levadas em consideração tanto no nível de variáveis endógenas (\mathbf{y} defasada) quanto no nível de covariáveis. Uma desvantagem inevitável desse modelo é a interpretação menos direta e intuitiva dos efeitos marginais β_k ⁴.

2.7 O modelo *MESS*

O modelo *MESS* (*Matrix Exponential Spatial Specification*) é uma adaptação do modelo *SAR* que visa otimizar o processo de estimação via máxima verossimilhança⁵ e, assim, viabilizar a modelagem espacial em conjuntos de dados extensos (LeSage and Pace, 2007).

Este modelo consiste em utilizar uma matriz $\mathbf{S}(\alpha) = e^{\alpha\mathbf{W}}$, de modo similar a $\mathbf{S} = (\mathbf{I} - \rho\mathbf{W})$ - do modelo *SAR* convencional. Algumas propriedades bastante interessantes de $\mathbf{S}(\alpha)$ são:

1. $\mathbf{S}(\alpha)$ é sempre positiva definida;
2. $\mathbf{S}(\alpha)^{-1} = (e^{\alpha\mathbf{W}})^{-1} = e^{-\alpha\mathbf{W}}$;
3. $|\mathbf{S}(\alpha)| = e^{\text{traço}(\alpha\mathbf{W})}$.

Nesse contexto, é possível mostrar que

$$e^{\alpha\mathbf{W}} = \sum_{t=0}^{\infty} \frac{\alpha^t}{t!} \mathbf{W}^t \quad (2.29)$$

$\forall \alpha < \infty$ e $\forall \mathbf{A}_{n \times n}$ (Horn and Johnson, 1988). Além disso, por causa da equivalência entre \mathbf{S} e $\mathbf{S}(\alpha)$, é possível estabelecer uma relação entre α e ρ . Seja, portanto, $|||\mathbf{S}|||$ a norma de \mathbf{S} , dada pela maior soma das linhas de \mathbf{S} . Então, $|||\mathbf{S}||| = |||\mathbf{S}(\alpha)||| \Rightarrow 1 - \rho = e^\alpha$. Na prática, se expande a série (2.29) até um certo valor q ($t = 1, 2, \dots, q$), de modo que $1 - \rho \cong e^\alpha$. As vantagens do modelo *MESS* serão abordadas com maior profundidade na seção seguinte.

⁴Para mais informações, consulte Elhorst (2010).

⁵Do ponto de vista computacional.

2.7.1 Estimando parâmetros do modelo *MESS* via máxima verossimilhança

A verossimilhança do modelo de regressão convencional $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, onde $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ é dada por

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{y}' \mathbf{M} \mathbf{y} \right\} \quad (2.30)$$

onde $\mathbf{M} = \mathbf{I} - \mathbf{H}$ e $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Seja $\mathbf{S} = (\mathbf{I} - \rho\mathbf{W})$. Então, o modelo *SAR* convencional pode ser escrito como $\mathbf{S}\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Consequentemente, pode-se reescrever a verossimilhança do modelo de regressão convencional em termos de $\mathbf{S}\mathbf{y}$ utilizando o método do jacobiano para transformações de variáveis, a saber:

$$f(g(\mathbf{y})) = f_{\mathbf{Y}}(g(\mathbf{y})) \left| \frac{\partial g(\mathbf{y})}{\partial \mathbf{y}} \right| \quad (2.31)$$

onde $g(\mathbf{y}) = \mathbf{S}\mathbf{y}$ e $\left| \frac{\partial g(\mathbf{y})}{\partial \mathbf{y}} \right| = |\mathbf{S}|$. A verossimilhança em termos de $\mathbf{S}\mathbf{y}$ é dada por

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2 | \mathbf{S}\mathbf{y}) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{y}' \mathbf{S}' \mathbf{M} \mathbf{S} \mathbf{y} \right\} |\mathbf{S}| \quad (2.32)$$

Se $\mathbf{S} = e^{\alpha\mathbf{W}}$, então

$$\mathcal{L}(\boldsymbol{\beta}, \alpha, \sigma^2 | \mathbf{S}\mathbf{y}) = (2\pi)^{-\frac{n}{2}} \exp\{\text{traço}(\alpha\mathbf{W})\} (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{y}' \mathbf{S}' \mathbf{M} \mathbf{S} \mathbf{y} \right\} \quad (2.33)$$

Como o $\text{traço}(\alpha\mathbf{W}) = e^0 = 1$, otimizar a verossimilhança acima é mais simples que otimizar a do *SAR* tradicional. Para obter $\hat{\alpha}$, é conveniente utilizar a verossimilhança de perfil, que consiste em concentrar $\mathcal{L}(\boldsymbol{\beta}, \alpha, \sigma^2 | \mathbf{S}\mathbf{y})$ em α , substituindo $\boldsymbol{\beta}$ e σ^2 por $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ e $\hat{\sigma}^2 = \frac{1}{n}\mathbf{y}'\mathbf{S}'\mathbf{M}\mathbf{S}\mathbf{y}$ obtidos via máxima verossimilhança em (2.32).

Logo, a verossimilhança concentrada em termos de α é dada por

$$\begin{aligned} \mathcal{L}(\alpha | \mathbf{S}\mathbf{y}) &= (2\pi)^{-\frac{n}{2}} \left(\frac{1}{n} \mathbf{y}' \mathbf{S}' \mathbf{M} \mathbf{S} \mathbf{y} \right)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sigma^2 \right\} \\ &= K (\mathbf{y}' \mathbf{S}' \mathbf{M} \mathbf{S} \mathbf{y})^{-\frac{n}{2}} \end{aligned} \quad (2.34)$$

onde K é uma constante que não depende de α .

Encontrar o estimador de máxima verossimilhança $\hat{\alpha}_{MV}$ é equivalente a encontrar o valor de α que minimiza a soma de quadrados dos resíduos $\mathbf{y}'\mathbf{S}'\mathbf{M}\mathbf{S}\mathbf{y}$. O procedimento para encontrar α é razoavelmente simples, se comparado com o modelo *SAR*, e compreende os seguintes passos:

1. Seja \mathbf{Y} a matriz $N \times q$, tal que $\mathbf{Y} = [\mathbf{y} \ \mathbf{W}\mathbf{y} \ \mathbf{W}^2\mathbf{y} \ \dots \ \mathbf{W}^{q-1}\mathbf{y}]$;

2. Seja, também, $\mathbf{Q} = \begin{bmatrix} \frac{1}{0!} & 0 & \dots & 0 \\ 0 & \frac{1}{1!} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \frac{1}{(q-1)!} \end{bmatrix}$ e $g(\alpha) = [1, \alpha, \alpha^2, \dots, \alpha^{q-1}]'$;

3. Então, pode-se aproximar $\mathbf{S}(\alpha)\mathbf{y} \cong \mathbf{Y}\mathbf{Q}\mathbf{g}(\alpha)$;

4. Como abordado anteriormente, maximizar a verossimilhança (2.34) consiste em minimizar a soma dos quadrados dos resíduos que, por sua vez, são função de α , de modo que

$$\begin{aligned} \hat{\boldsymbol{\epsilon}}(\alpha)' \hat{\boldsymbol{\epsilon}}(\alpha) &= \mathbf{g}(\alpha)' \mathbf{Q} \mathbf{Y}' \mathbf{M}' \mathbf{M} \mathbf{Y} \mathbf{Q} \mathbf{g}(\alpha) \\ &= \mathbf{g}(\alpha)' (\mathbf{Q} \mathbf{Y}' \mathbf{M} \mathbf{Y} \mathbf{Q}) \mathbf{g}(\alpha) \\ &= \mathbf{g}(\alpha)' \mathbf{Z} \mathbf{g}(\alpha) \end{aligned}$$

e $\mathbf{Z} = \mathbf{Q} \mathbf{Y}' \mathbf{M} \mathbf{Y} \mathbf{Q}$.

Note que a forma quadrática $P(\alpha) = \mathbf{g}(\alpha)' \mathbf{Z} \mathbf{g}(\alpha)$ é um polinômio de ordem $2q - 2$. Encontrar o valor que minimiza⁶ a soma dos quadrados dos erros consiste em encontrar o valor de α tal que $\frac{\partial P(\alpha)}{\partial \alpha} = 0$. Assim, tem-se que $\frac{\partial P(\alpha)}{\partial \alpha} = 2\mathbf{g}(\alpha)' \mathbf{Z} \left(\frac{\partial \mathbf{g}(\alpha)}{\partial \alpha} \right) = 0$ é um polinômio de ordem $2q - 3$ e, portanto, tem $2q - 3$ raízes. Encontrar raízes de polinômios é um problema bastante explorado na literatura matemática e tem soluções amplamente conhecidas.

A grande vantagem de estimar α em vez de ρ é que o primeiro exige otimizar um polinômio de ordem $2q - 3$, onde q é o nível de precisão que se deseja atingir, independentemente da dimensão N do conjunto de dados. Para mais detalhes veja LeSage and Pace (2007).

⁶ \mathbf{Z} deve ser positiva definida para garantir que o ponto α^* tal que $\frac{\partial P(\alpha^*)}{\partial \alpha^*} = 0$ é um ponto de mínimo.

Capítulo 3

Painel Espacial

As vantagens dos modelos em painel, já enumeradas anteriormente, são bastante convincentes (grande disponibilidade de graus de liberdade, maior homogeneidade intra-indivíduo e menor colinearidade). O modelos espaciais para dados em painel aliam a essas virtudes a possibilidade de modelar estruturas de dependência espacial, bem como relações entre indivíduos que não seriam captadas num painel convencional, uma vez que o processo de geração dos dados subjacente a um painel - que não leva a estrutura espacial em consideração - é a de independência das observações de corte transversal.

O foco deste trabalho será em modelos que incorporam a *dependência* espacial ao longo do tempo. Há, ainda, modelos que incorporam a *heterogeneidade* espacial. Basicamente, a heterogeneidade espacial consiste em supor que as variâncias e os coeficientes β são diferentes ao longo do espaço e, portanto, ao longo das observações de corte transversal. Os modelos mais amplamente utilizados para captar heterogeneidade espacial são *GWR* - *Geographically Weighted Regression* e *SUR* - *Seemingly Unrelated Regressions*. Para mais detalhes, veja Fotheringham et al. (2006), Elhorst (2010) e Elhorst (2003).

Do ponto de vista operacional, a estrutura do painel espacial é um pouco diferente do painel de observações (*cross-section*) independentes, uma vez que o primeiro é ordenado por período de tempo e, posteriormente, por observação (Elhorst, 2003). Visualmente, isso consiste em empilhar os conjuntos de dados de corte transversal, a saber:

$$\begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{1T} \\ y_{2T} \\ \vdots \\ y_{NT} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_T \end{bmatrix} \beta + \mathbf{D}\mathbf{c} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{N1} \\ \vdots \\ \epsilon_{1T} \\ \epsilon_{2T} \\ \vdots \\ \epsilon_{NT} \end{bmatrix}$$

onde \mathbf{c} é o efeito individual (do tempo, da observação ou de ambos), $\mathbf{D} = \mathbf{j}_T \otimes \mathbf{I}_N$ e ϵ é o erro idiossincrático.

3.1 Painel espacial de efeitos fixos

Conforme dito anteriormente, utilizar variáveis indicadoras (*dummies*) para captar o efeito individual não é a melhor escolha em aplicações econômicas porque conforme $N \rightarrow \infty$, o número de parâmetros também cresce sem limitação¹. Além disso, se o número de observações for razoavelmente grande, é complicado interpretar todos os efeitos individuais de forma satisfatória.

Nesta seção serão apresentadas as adaptações dos modelos de painel de efeitos fixos para modelos espaciais autoregressivos e de erros autoregressivos.

3.1.1 O modelo espacial autoregressivo de efeitos fixos

Seja \mathbf{I}_T a matriz identidade de ordem T , \mathbf{W}_N a matriz de proximidades², \mathbf{j}_T um vetor de 1's de ordem T e \mathbf{c} o vetor $[c_1, c_2, \dots, c_N]$ de efeitos individuais. Então, o modelo espacial autoregressivo de 1 efeito fixo (*FESAR - Fixed Effects Spatial Autoregressive Model*) pode ser escrito na forma:

$$\begin{aligned} \mathbf{y} &= \rho(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y} + (\mathbf{j}_T \otimes \mathbf{c}) + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \mathbf{S}\mathbf{y} &= (\mathbf{j}_T \otimes \mathbf{c}) + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \end{aligned}$$

¹Este fenômeno é conhecido como problema incidental dos parâmetros (*incidental parameter problem*).

²Supõe-se que a matriz de proximidades seja constante ao longo do tempo.

onde $\mathbf{S} = (\mathbf{I}_{NT} - \rho(\mathbf{I}_T \otimes \mathbf{W}))$.

Usualmente, \mathbf{X} contém um vetor de 1's para que seja estimada média geral (intercepto) e o efeito fixo c é restrito de modo que $\mathbf{j}'_N c = 0$. Assim, a interpretação do efeito fixo c_i é a diferença na média da observação i em relação à média geral.

Do ponto de vista prático, ao utilizar a *Within Transformation*, a presença do intercepto não faz diferença, uma vez que todos os efeitos constantes ao longo do tempo são suprimidos. Desse modo, o modelo equivalente pode ser representado como

$$\mathbf{M}_{NT}\mathbf{y} = \mathbf{M}_{NT}\mathbf{X}\boldsymbol{\beta} + \mathbf{M}_{NT}\boldsymbol{\epsilon} \quad (3.1)$$

$$\mathbf{M}_{NT} = \mathbf{I}_{NT} - \left(\frac{1}{T} \mathbf{j}_T \mathbf{j}'_T \otimes \mathbf{I}_N \right) \quad (3.2)$$

A função de verossimilhança do modelo (3.1) sob a *Within Transformation* é dada por

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2, \rho) = (2\pi)^{-\frac{NT}{2}} (\sigma^2)^{-\frac{NT}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{S}\mathbf{M}\mathbf{y} - \mathbf{M}\mathbf{X}\boldsymbol{\beta})' (\mathbf{S}\mathbf{M}\mathbf{y} - \mathbf{M}\mathbf{X}\boldsymbol{\beta}) \right\} |\mathbf{S}|$$

De modo análogo à seção 2.4.1, pode-se fixar um valor para $\boldsymbol{\beta}$ e σ^2 e concentrar a verossimilhança apenas em ρ . Sendo assim, sejam $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{M}\mathbf{X})^{-1} \mathbf{X}'\mathbf{M}\mathbf{y}$ e $\tilde{\sigma}^2 = \frac{1}{NT} \mathbf{y}'\mathbf{S}'\mathbf{P}\mathbf{S}\mathbf{y}$, onde \mathbf{P} é a matriz de projeção ortogonal $\mathbf{P} = \mathbf{I}_{NT} - \mathbf{M}\mathbf{X}(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1} \mathbf{X}'\mathbf{M}$. Desse modo, a verossimilhança de perfil, concentrada em ρ é dada por

$$\begin{aligned} \mathcal{L}(\rho, \sigma^2, \boldsymbol{\beta}) &= (2\pi)^{-\frac{NT}{2}} (\sigma^2)^{-\frac{NT}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y}'\mathbf{S}'\mathbf{P}\mathbf{S}\mathbf{y})' (\mathbf{y}'\mathbf{S}'\mathbf{P}\mathbf{S}\mathbf{y}) \right\} |\mathbf{S}| \\ \mathcal{L}(\rho) &= (2\pi)^{-\frac{NT}{2}} \left(\frac{1}{T} \mathbf{y}'\mathbf{S}'\mathbf{P}\mathbf{S}\mathbf{y} \right)^{-\frac{NT}{2}} \exp \left\{ -\frac{1}{2} NT \right\} |\mathbf{S}| \end{aligned}$$

Note que $\mathbf{I}_{NT} - \rho(\mathbf{I}_T \otimes \mathbf{W}) = \mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})$, e, portanto, $|\mathbf{S}| = |\mathbf{I}_T \otimes (\mathbf{I}_N - \rho\mathbf{W})| = |\mathbf{I}_N - \rho\mathbf{W}|^T$. Então, pode-se escrever (3.3) como

$$\begin{aligned} \mathcal{L}(\rho) &= k (\mathbf{y}'\mathbf{S}'\mathbf{P}\mathbf{S}\mathbf{y})^{-\frac{NT}{2}} |\mathbf{I}_N - \rho\mathbf{W}|^T \\ l(\rho) &= \log k - \frac{NT}{2} \log (\mathbf{y}'\mathbf{S}'\mathbf{P}\mathbf{S}\mathbf{y}) + T \log |\mathbf{I}_N - \rho\mathbf{W}| \end{aligned}$$

No mesmo contexto da seção 2.4.1, pode-se reescrever (3.3) como

$$l(\rho) = \log k - \frac{NT}{2} \log (\mathbf{y}'\mathbf{S}'\mathbf{P}\mathbf{S}\mathbf{y}) + T \sum_{i=1}^N \log (1 - \rho\omega_i) \quad (3.3)$$

onde ω_i é o i -ésimo autovalor da matriz \mathbf{W} .

Note que

$$\mathbf{y}'\mathbf{S}'\mathbf{P}\mathbf{S}\mathbf{y} = \mathbf{y}'\mathbf{P}\mathbf{y} - \rho\mathbf{y}'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{y} - \rho\mathbf{y}'(\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}\mathbf{y} + \rho^2\mathbf{y}'(\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{y}$$

porque

$$\begin{aligned}
\mathbf{y}'\mathbf{S}'\mathbf{P}\mathbf{S}\mathbf{y} &= \mathbf{y}'(\mathbf{I}_{NT} \otimes (\mathbf{I}_N - \rho\mathbf{W}))'\mathbf{P}(\mathbf{I}_{NT} \otimes (\mathbf{I}_N - \rho\mathbf{W}))\mathbf{y} \\
&= \mathbf{y}'(\mathbf{I}_{NT} \otimes (\mathbf{I}_N - \rho\mathbf{W}'))\mathbf{P}(\mathbf{I}_{NT} \otimes (\mathbf{I}_N - \rho\mathbf{W}))\mathbf{y} \\
&= \mathbf{y}'(\mathbf{I}_{NT} - \rho\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}(\mathbf{I}_{NT} - \rho\mathbf{I}_T \otimes \mathbf{W})\mathbf{y} \\
&= (\mathbf{y}' - \rho\mathbf{y}'\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}(\mathbf{y} - \rho\mathbf{I}_T \otimes \mathbf{W}\mathbf{y}) \\
&= \mathbf{y}'\mathbf{P}(\mathbf{y} - \rho(\mathbf{I}_T \otimes \mathbf{W})\mathbf{y}) - \rho\mathbf{y}'(\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}(\mathbf{y} - \rho(\mathbf{I}_T \otimes \mathbf{W})\mathbf{y}) \\
&= \mathbf{y}'\mathbf{P}\mathbf{y} - \rho\mathbf{y}'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{y} - \rho\mathbf{y}'(\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}\mathbf{y} + \rho^2\mathbf{y}'(\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{y}
\end{aligned}$$

Então, a verossimilhança (3.3) se resume a

$$\begin{aligned}
l(\rho) &= \log k + T \sum_{i=1}^N \log(1 - \rho\omega_i) \\
&\quad - \frac{NT}{2} \log [\mathbf{y}'\mathbf{P}\mathbf{y} - \rho\mathbf{y}'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{y} - \rho\mathbf{y}'(\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}\mathbf{y} + \rho^2\mathbf{y}'(\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{y}]
\end{aligned}$$

De modo análogo à seção 2.4.2, é razoável estimar ρ por meio de um processo iterativo Newton-Raphson tal como

$$\rho_{j+1} = \rho_j - \frac{l(\rho)}{l'(\rho)}$$

onde

$$\begin{aligned}
l'(\rho) &= \frac{\partial l(\rho)}{\partial \rho} = -\mathbf{y}'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{y} - \mathbf{y}'(\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}\mathbf{y} \\
&\quad + 2\rho\mathbf{y}'(\mathbf{I}_T \otimes \mathbf{W}')\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{y} - T \sum_{i=1}^N \frac{\omega_i}{1 - \rho\omega_i}
\end{aligned}$$

Uma vez estimado ρ , o estimador $\hat{\beta} = (\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}(\mathbf{I}_{NT} - \rho\mathbf{I}_T \otimes \mathbf{W})\mathbf{y}$ é obtida de forma análoga ao Capítulo 2. Caso o modelo de interesse seja o de 2 efeitos fixos, uma pequena alteração na matriz \mathbf{M} se faz necessária, a saber

$$\mathbf{M} = \mathbf{I}_N \otimes \mathbf{I}_T - \frac{1}{T}\mathbf{j}_T\mathbf{j}'_T \otimes \mathbf{I}_N - \frac{1}{N}\mathbf{I}_N \otimes \mathbf{j}_N\mathbf{j}'_N + \frac{1}{NT}\mathbf{j}_{NT}\mathbf{j}'_{NT}$$

O procedimento de estimação é idêntico ao de 1 efeito fixo, a menos da matriz \mathbf{M} que deve ser adaptada para 2 efeitos fixos.

Note que foram apresentados métodos de estimação que desconsideram os efeitos individuais. Caso haja o interesse explícito em estimar esses efeitos e interpretá-los, basta incorporar as variáveis indicadoras da unidade observacional e/ou do período no tempo³ na matriz de variáveis exógenas. Na aplicação abordada neste texto, estimar efeitos individuais é inviável, em virtude do tamanho da amostra ser bastante grande.

³Geralmente, o interesse é exclusivamente nos parâmetros β .

3.1.2 O modelo espacial autoregressivo de efeitos aleatórios

O modelo espacial autoregressivo de efeitos aleatórios (*RESAR - Random effects Spatial Autoregressive Model*) incorpora a dependência espacial de maneira muito natural, uma vez que a dependência entre observações vizinhas pode ser levada em consideração por meio de um *SAR* nos termos de erro. Mais especificamente, seja o conjunto de dados referente ao período de tempo t . Então, a componente aleatória v_t (de dimensão $N \times 1$) pode ser decomposta em $v_t = c + \epsilon_t$, onde c é o efeito individual aleatório ($c \sim N(0, \sigma_c^2 \mathbf{I}_N)$). Posteriormente,

$$\epsilon_t = \lambda \mathbf{W} \epsilon_t + u_t \quad (3.4)$$

$$\epsilon_t = (\mathbf{I}_N - \lambda \mathbf{W}_N)^{-1} u_t \quad (3.5)$$

$$\mathbf{B}_N = \mathbf{I}_N - \lambda \mathbf{W}_N \quad (3.6)$$

$$\epsilon_t = \mathbf{B}_N^{-1} u_t \quad (3.7)$$

onde $u_t \sim N(0, \sigma_u^2 \mathbf{I}_N)$.

Na forma $NT \times 1$, tem-se que

$$v = (j_T \otimes \mathbf{I}_N) c + (\mathbf{I}_T \otimes \mathbf{B}_N^{-1}) u \quad (3.8)$$

$$Var(v) = E(vv') = \sigma_c^2 (j_T j_T' \otimes \mathbf{I}_N) + \sigma_u^2 [\mathbf{I}_T \otimes (\mathbf{B}_N' \mathbf{B}_N)^{-1}] \quad (3.9)$$

Note que, assim como no painel com observações independentes, o efeito c induz a correlação temporal (na mesma unidade observacional), ao passo que o efeito u induz a correlação espacial (na dimensão do corte transversal).

Uma grande vantagem do *RESAR* é que este não sofre do problema incidental dos parâmetros, uma vez que c_i é a realização da variável aleatória C . Sendo assim, quando $N \rightarrow \infty$, toda a teoria assintótica é válida.

É interessante observar que modelos de efeitos aleatórios com autocorrelação espacial podem ser considerados casos específicos de não esfericidade da matriz de covariâncias nos termos aleatórios, i.e., elementos fora da diagonal da matriz de covariâncias dos erros são diferentes de zero.

3.2 O modelo *MESS* em painel

De modo análogo ao modelo *MESS* para dados de corte transversal, é possível adaptar a matriz $S(\alpha) = e^{\alpha \mathbf{W}}$ para atender às características do conjunto de dados em painel, agora

denominado *MESPS* (*Matrix Exponential Spatial Panel Specification*). Sendo assim, seja $\mathbf{S} = \mathbf{I}_T \otimes S(\alpha)$ a matriz exponencial de proximidades adaptada ao painel.

Então, a verossimilhança do modelo

$$\mathbf{S}\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{c} + \boldsymbol{\epsilon}$$

é dada por

$$\mathcal{L}(\sigma^2, \boldsymbol{\beta}, \alpha | \mathbf{y}) = (2\pi)^{-\frac{NT}{2}} (\sigma^2)^{-\frac{NT}{2}} |\mathbf{S}| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{S}\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{c})' (\mathbf{S}\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{c}) \right\} \quad (3.10)$$

Em virtude do grande número de observações, é inviável estimar um efeito individual para cada uma delas, além de sofrer com o problema incidental dos parâmetros. Assim, a *Within Transformation* é a melhor opção para obter estimadores consistentes para $\boldsymbol{\beta}$. As matrizes que transformam \mathbf{y} e \mathbf{X} em desvios de suas respectivas médias são

- $\mathbf{M} = \mathbf{I}_{NT} - \frac{1}{T} \mathbf{j}_T \mathbf{j}'_T \otimes \mathbf{I}_N$, no caso de 1 efeito fixo;
- $\mathbf{M} = \mathbf{I}_{NT} - \frac{1}{T} \mathbf{j}_T \mathbf{j}'_T \otimes \mathbf{I}_N - \frac{1}{N} \mathbf{I}_T \otimes \mathbf{j}_N \mathbf{j}'_N + \frac{1}{NT} \mathbf{j}_{NT} \mathbf{j}'_{NT}$, no caso de 2 efeitos fixos.

A verossimilhança (3.10) pode, portanto, ser reescrita como

$$\mathcal{L}(\sigma^2, \boldsymbol{\beta}, \alpha | \mathbf{y}) = (2\pi)^{-\frac{NT}{2}} (\sigma^2)^{-\frac{NT}{2}} |\mathbf{S}| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{S}\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})' (\mathbf{S}\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) \right\}$$

em que $\tilde{\mathbf{y}} = \mathbf{M}\mathbf{y}$ e $\tilde{\mathbf{X}} = \mathbf{M}\mathbf{X}$.

Fixando $\tilde{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'\mathbf{S}\tilde{\mathbf{y}}$, tem-se

$$\mathbf{S}\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta} = (\mathbf{I}_{NT} - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}') \mathbf{S}\tilde{\mathbf{y}}$$

Se $\mathbf{P} = \mathbf{I}_{NT} - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'$, então

$$\begin{aligned} (\mathbf{S}\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})' (\mathbf{S}\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) &= \tilde{\mathbf{y}}' \mathbf{S}' \mathbf{P}' \mathbf{P} \mathbf{S} \tilde{\mathbf{y}} \\ &= \tilde{\mathbf{y}}' \mathbf{S}' \mathbf{P} \mathbf{S} \tilde{\mathbf{y}} \end{aligned}$$

uma vez que \mathbf{P} é idempotente.

Assim, para obter a verossimilhança concentrada apenas em α , é razoável fixar o parâmetro $\tilde{\sigma}^2 = \frac{1}{NT} \tilde{\mathbf{y}}' \mathbf{S}' \mathbf{P} \mathbf{S} \tilde{\mathbf{y}}$. Portanto, a verossimilhança concentrada em α é dada por

$$\begin{aligned} \mathcal{L}(\alpha) &= (2\pi)^{-\frac{NT}{2}} \left[\frac{1}{NT} \tilde{\mathbf{y}}' \mathbf{S}' \mathbf{P} \mathbf{S} \tilde{\mathbf{y}} \right]^{-\frac{NT}{2}} |\mathbf{S}|^T \exp \left\{ -\frac{NT}{2} \right\} \\ &= k |S(\alpha)|^T \left[\frac{1}{NT} \tilde{\mathbf{y}}' \mathbf{S}' \mathbf{P} \mathbf{S} \tilde{\mathbf{y}} \right]^{-\frac{NT}{2}} \\ l(\alpha) &= \log k + T \log |S(\alpha)| - \frac{NT}{2} \log \left[\frac{1}{NT} \tilde{\mathbf{y}}' \mathbf{S}' \mathbf{P} \mathbf{S} \tilde{\mathbf{y}} \right] \end{aligned}$$

mas $|S(\alpha)| = e^{\text{traço}(\alpha\mathbf{W})} = e^0 = 1$. Assim, a forma final da verossimilhança concentrada em α é

$$l(\alpha) = \log(k - \frac{NT}{2} \log(\tilde{\mathbf{y}}' \mathbf{S}' \mathbf{P} \mathbf{S} \tilde{\mathbf{y}})) \quad (3.11)$$

Note que $\mathbf{S} = \mathbf{I}_T \otimes S(\alpha) = \mathbf{I}_{NT} + \alpha \mathbf{I}_T \otimes \mathbf{W} + \frac{\alpha^2}{2} \mathbf{I}_T \otimes \mathbf{W}^2 + \dots + \frac{\alpha^j}{j!} \mathbf{I}_T \otimes \mathbf{W}^j + \dots$. Então, ao parametrizar

- $\tilde{\mathbf{Y}} = (\tilde{\mathbf{y}}, \mathbf{I}_T \otimes \mathbf{W}\tilde{\mathbf{y}}, \mathbf{I}_T \otimes \mathbf{W}^2\tilde{\mathbf{y}}, \dots, \mathbf{I}_T \otimes \mathbf{W}^j\tilde{\mathbf{y}})$;

- $\mathbf{Q} = \begin{bmatrix} \frac{1}{0!} & 0 & \dots & 0 \\ 0 & \frac{1}{1!} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \frac{1}{(j-1)!} \end{bmatrix}$ e

- $\mathbf{g}(\alpha) = [1, \alpha, \alpha^2, \dots, \alpha^{j-1}]'$,

pode-se reescrever a soma de quadrados em (3.11) como

$$l(\alpha) = \log k - \frac{NT}{2} \log \mathbf{g}(\alpha)' \mathbf{Q}' \tilde{\mathbf{Y}}' \mathbf{P} \tilde{\mathbf{Y}} \mathbf{Q} \mathbf{g}(\alpha)$$

Se $\mathbf{Z} = \mathbf{Q}' \tilde{\mathbf{Y}}' \mathbf{P} \tilde{\mathbf{Y}} \mathbf{Q}$, então⁴

$$l(\alpha) = \log k - \frac{NT}{2} \log \mathbf{g}(\alpha)' \mathbf{Z} \mathbf{g}(\alpha)$$

Encontrar $\hat{\alpha}$ consiste em minimizar a forma quadrática $\mathbf{g}(\alpha)' \mathbf{Z} \mathbf{g}(\alpha)$ com relação a α . Uma característica importante dessa forma quadrática é que as antidiagonais da \mathbf{Z} compõem os coeficientes do polinômio associado a $\mathbf{g}(\alpha)$. Seja, para efeito meramente didático, $\mathbf{g}(\alpha) =$

$$(1, \alpha, \alpha^2)' \text{ e } \mathbf{Z} = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 5 & 1 \\ 2 & 1 & 6 \end{bmatrix}. \text{ Então,}$$

$$\begin{aligned} \mathbf{g}(\alpha)' \mathbf{Z} \mathbf{g}(\alpha) &= (1, \alpha, \alpha^2) \begin{bmatrix} 2 & 1 & 3 \\ 4 & 5 & 1 \\ 2 & 1 & 6 \end{bmatrix} \begin{pmatrix} 1 \\ \alpha \\ \alpha^2 \end{pmatrix} \\ &= (2 + 4\alpha + 2\alpha^2, 1 + 5\alpha + \alpha^2, 3 + \alpha + 6\alpha^2) \begin{pmatrix} 1 \\ \alpha \\ \alpha^2 \end{pmatrix} \\ &= 2 + 4\alpha + 2\alpha^2 + \alpha + 5\alpha^2 + \alpha^3 + 3\alpha^2 + \alpha^3 + 6\alpha^4 \\ &= 2 + 5\alpha + 10\alpha^2 + 2\alpha^3 + 6\alpha^4 \end{aligned}$$

⁴Note que \mathbf{Q} é simétrica.

Note que as antidiagonais (2, 4 1, 3 5 2, 1 1, 6), se somadas uma a uma, definem os coeficientes do polinômio associado a $\mathbf{g}(\alpha)' \mathbf{Z} \mathbf{g}(\alpha)$. A maneira mais intuitiva de identificar se dois elementos z_{ij} e z_{kw} ($i \neq k$ e $j \neq w$) pertencem à mesma antidiagonal é verificar se $i + j = k + w$. Para mais detalhes, veja Horn and Johnson (1988).

No caso do modelo *MESS*, em que $\mathbf{g}(\alpha)' \mathbf{Z} \mathbf{g}(\alpha)$ define um polinômio de ordem $2q - 2$, a antidiagonal tal que $i + j = m$ define o coeficiente do termo α^{m-2} . Mais especificamente,

$$\mathbf{g}(\alpha)' \mathbf{Z} \mathbf{g}(\alpha) = \sum_{m=2}^{2q} c_m \alpha^{m-2}$$

onde c_m é a soma dos termos na antidiagonal tal que $i + j = m$.

A condição de otimalidade de primeira ordem é $\frac{\partial}{\partial \alpha} \mathbf{g}(\alpha)' \mathbf{Z} \mathbf{g}(\alpha) = 0$. Isso implica que $\sum_{m=3}^{2q} c_m (m-2) \alpha^{m-3} = 0$. Sendo assim, pode-se obter $\hat{\alpha}$ por meio de um procedimento Newton-Raphson, de modo que

$$\alpha_{k+1} = \alpha_k - \frac{\sum_{m=3}^{2q} c_m (m-2) \alpha_k^{m-3}}{\sum_{m=4}^{2q} c_m (m-2) (m-3) \alpha_k^{m-4}}$$

onde α_k é o valor de α avaliado na k -ésima iteração. O estimador de máxima verossimilhança $\hat{\alpha}_{MV}$ é tal que $|\alpha_{k+1} - \alpha_k| \leq \epsilon$, e ϵ é pequeno o suficiente.

Derivar a distribuição de probabilidades de $\hat{\alpha}$ é bastante complicado, tornando a obtenção da variância exata de $\hat{\alpha}$ desnecessariamente complexa, uma vez que, no contexto de dados econômicos, as amostras costumam ser grandes. Sendo assim, é possível estimar $Var(\hat{\alpha})$ assintoticamente por meio das propriedades do estimador de máxima verossimilhança. Em grandes amostras, $Var(\hat{\alpha}) \approx \left[-\frac{\partial^2 l(\hat{\alpha})}{\partial \alpha^2} \right]^{-1}$.

Mais especificamente,

$$\begin{aligned} l(\alpha) &= k - \frac{NT}{2} \log(g(\alpha)' \mathbf{Z} g(\alpha)) \\ &= k - \frac{NT}{2} \log \left(\sum_{m=2}^{2q} c_m \alpha^{m-2} \right) \\ \frac{\partial l(\alpha)}{\partial \alpha} &= -\frac{NT}{2} \left\{ \frac{\sum_{m=3}^{2q} (m-2) c_m \alpha^{m-3}}{\sum_{m=2}^{2q} c_m \alpha^{m-2}} \right\} \end{aligned}$$

se $g_1(\alpha) = \sum_{m=2}^{2q} c_m \alpha^{m-2}$ e $g_2(\alpha) = \sum_{m=3}^{2q} (m-2) c_m \alpha^{m-3}$, então

$$\begin{aligned} \frac{\partial l(\alpha)}{\partial \alpha} &= -\frac{NT}{2} \{g_2(\alpha)[g_1(\alpha)]^{-1}\} \\ \frac{\partial^2 l(\alpha)}{\partial \alpha^2} &= -\frac{NT}{2} \{g_2'(\alpha)[g_1(\alpha)]^{-1} - g_2(\alpha)[g_1(\alpha)]^{-2} g_1'(\alpha)\} \\ &= -\frac{NT}{2} \left\{ \frac{g_2'(\alpha)[g_1(\alpha)]^{-1} - g_2(\alpha)g_1'(\alpha)}{[g_1(\alpha)]^2} \right\} \end{aligned}$$

note que $g'_1(\alpha) = g_2(\alpha)$ e $g'_2(\alpha) = \sum_{m=4}^{2q} (m-2)(m-3)c_m\alpha^{m-4}$.

Avaliando no ponto $\hat{\alpha}$, tem-se que

$$Var(\hat{\alpha}) \approx \left[-\frac{\partial^2 l(\hat{\alpha})}{\partial \alpha^2} \right]^{-1} \approx \frac{2}{NT} \left\{ \frac{[g_1(\alpha)]^2}{g'_2(\alpha)[g_1(\alpha)]^{-1} - g_2(\alpha)g'_1(\alpha)} \right\}$$

Uma vez obtido o valor estimado de α , obter $\hat{\beta}$ é análogo estimador de máxima verossimilhança no modelo *SAR*. Em detalhes,

$$\hat{\beta} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \mathbf{S} \tilde{\mathbf{y}} \quad (3.12)$$

A matriz de covariâncias de $\hat{\theta} = (\hat{\beta}, \hat{\sigma}^2)'$ pode ser obtida utilizando o inverso da informação de Fisher. Mais especificamente, tem-se que

$$Var(\hat{\theta}) = \left[E \left(-\frac{\partial^2 l}{\partial \theta \partial \theta'} \right) \right]^{-1} \quad (3.13)$$

Note que é relativamente simples obter as covariâncias entre os estimadores por meio de (3.13)⁵. Adotando $\gamma = \sigma^2$ para tornar a notação menos carregada, tem-se que:

$$\begin{aligned} E \left(-\frac{\partial^2 l}{\partial \beta_0^2} \right) &= \frac{NT}{\gamma} \\ E \left(-\frac{\partial^2 l}{\partial \beta_0 \partial \beta_i} \right) &= \frac{\mathbf{j}' \tilde{\mathbf{X}}_i}{\gamma} \\ E \left(-\frac{\partial^2 l}{\partial \beta_i \partial \beta_i} \right) &= \frac{\tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i}{\gamma} \\ E \left(-\frac{\partial^2 l}{\partial \beta_i \partial \gamma} \right) &= 0 \\ E \left(-\frac{\partial^2 l}{\partial \gamma^2} \right) &= \frac{NT}{2\gamma^2} \end{aligned}$$

Representando os resultados acima de modo matricial, pode-se observar:

$$Var(\hat{\theta}) = \gamma \begin{bmatrix} NT & & & & \\ \mathbf{j}' \tilde{\mathbf{X}}_1 & \mathbf{X}'_1 \tilde{\mathbf{X}}_1 & & & \\ \vdots & & \ddots & & \\ \mathbf{j}' \tilde{\mathbf{X}}_k & \dots & & \tilde{\mathbf{X}}'_k \tilde{\mathbf{X}}_k & 0 \\ 0 & \dots & & 0 & \frac{NT}{2\gamma} \end{bmatrix}^{-1}$$

Nesse caso, substitui-se γ por um estimador consistente $\hat{\gamma} = \frac{1}{NT} (\mathbf{S} \tilde{\mathbf{y}} - \tilde{\mathbf{X}} \hat{\beta})' (\mathbf{S} \tilde{\mathbf{y}} - \tilde{\mathbf{X}} \hat{\beta})$.

⁵Para mais detalhes, veja o Apêndice A.2

Capítulo 4

Painel Espacial - Aplicações

Com o objetivo de comparar os resultados dos principais modelos¹ descritos neste texto e, especialmente, avaliar o desempenho do modelo *MESPS*, utilizou-se alguns conjuntos de dados com características especiais que permitem observar o funcionamento de cada modelo em situações distintas. Os modelos analisados foram Painel convencional (incluindo o modelo combinado), *SAR* e *MESPS* - todos com 1 e 2 efeitos fixos. Esses foram aplicados em quatro conjuntos de dados, a saber:

Municípios de Rondônia: baixa dependência espacial;

Municípios de São Paulo: baixa dependência espacial e base relativamente grande;

Consumo de cigarros nos Estados Unidos: alta dependência espacial;

Municípios do Brasil: base extensa com baixa dependência espacial.

As variáveis que compõem cada base de dados serão descritas nas seções seguintes.

4.1 Estudo de caso 1: Municípios de Rondônia

As bases de dados cuja unidade observacional é o município são compostas por 4 períodos de tempo equiespaçados (1999 - 2002). As variáveis disponíveis (no município i aferido no ano t) são:

pib - Produto interno bruto;

¹Neste momento, algumas questões acerca da especificação de cada modelo podem ser levantadas, como a forma funcional das variáveis, presença de formas quadráticas. Estes assuntos serão abordados com certo grau de superficialidade por, evidentemente, não ser esse o foco deste trabalho.

agro - Valor da produção agrícola;

ind - Valor da produção industrial.

A matriz de proximidades utilizada contém entrada 0, para municípios que não são vizinhos e 1 para aqueles que o são. Além disso, a matriz é ponderada de modo que as linhas somem 1.

Sendo assim, foram realizadas as estimações dos modelos Combinado, Painel convencional de 1 e 2 efeitos fixos, MESPS 1 e 2 efeitos fixos. A especificação é tal que *agro* e *ind* são variáveis explicativas para *pib*. Os resultados² estão apresentados na Tabela 4.1.

Tabela 4.1: Resultados dos Modelos - Municípios de Rondônia

Modelo combinado (<i>Pooled</i>)			
Estimadores			
<i>Intercepto</i>	8.266,22 (4.596,36)		
$\hat{\beta}_{agro}$	0,8119 (0,220)		
$\hat{\beta}_{ind}$	3,2617 (0,050)		
Modelos com 1 efeito fixo			
Estimadores	Painel convencional	SAR	MESPS
<i>Intercepto</i>	20.448(4.059,9)	26.789,78 (3.735,835)	18.435,68 (3.481,249)
$\hat{\beta}_{agro}$	1,5769 (0,196)	1,6824 (0,181)	1,6852 (0,168)
$\hat{\beta}_{ind}$	2,3579 (0,086)	2,3604 (0,074)	2,3612 (0,074)
$\hat{\rho}$		-0,0581 (0,047) ^{NS}	-0,0602
$\hat{\alpha}$			0,0585 (0,048) ^{NS}
Modelos com 2 efeitos fixos			
Estimadores	Painel convencional	SAR	MESPS
<i>Intercepto</i>	17.202 (4.365,2)	15.616,57 (3.782,25)	17.405,16 (3.718,89)
$\hat{\beta}_{agro}$	1,6786 (0,203)	1,6620 (0,1821)	1,6626 (0,1734)
$\hat{\beta}_{ind}$	2,4086 (0,089)	2,4115 (0,0773)	2,4114 (0,0766)
$\hat{\rho}$		0,0126 (0,0586) ^{NS}	0,0120
$\hat{\alpha}$			-0,0121 (0,0596) ^{NS}

²O valor dentro dos parênteses representa o erro-padrão do respectivo estimador e a sigla *NS* significa *não significativo*.

Note que os resultados do *MESPS* são bastante similares aos do *SAR*, diferindo, em média, na terceira casa decimal. Apesar de essa discrepância não ser desprezível em alguns casos, a vantagem advinda da menor intensidade computacional do *MESPS* faz com que ele seja uma alternativa bastante razoável em conjuntos de dados extensos. A ressalva a ser feita nesse caso é que a estrutura espacial pode ser desprezada, uma vez que tanto $\hat{\alpha}$ quanto $\hat{\rho}$ geraram valores não significativos. Esse fato é ilustrado pela Figura 4.1, em que não se pode, na média, observar concentrações de dependência espacial organizadas geograficamente.

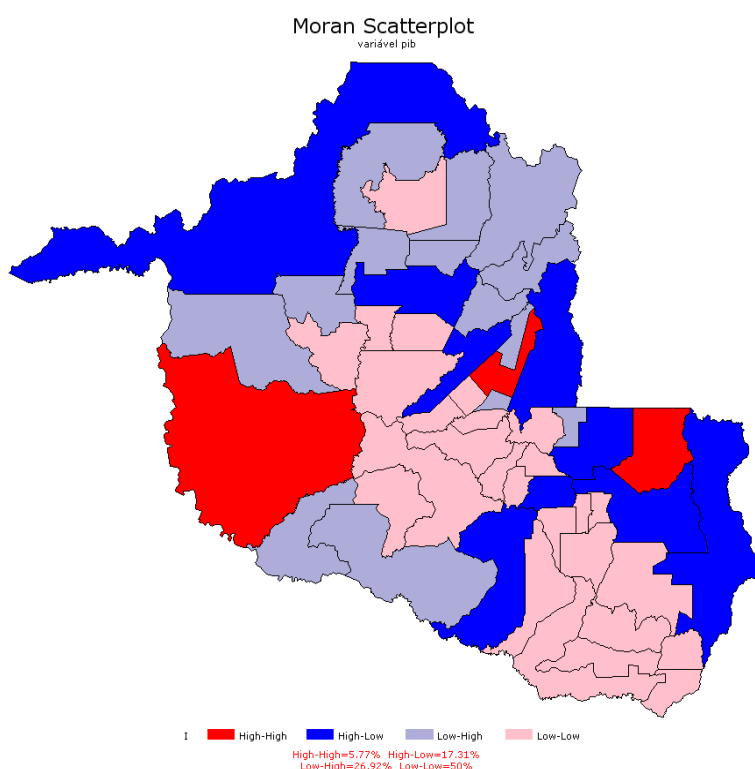


Figura 4.1: Dependência espacial dos municípios do Estado de Rondônia

Os tipos de dependência espacial indicados na Figura (4.1) devem ser interpretados da seguinte forma:

High-High: Região com altos valores da variável resposta y e vizinhos com altos valores de y ;

High-Low: Região com altos valores de y e vizinhos com baixos valores de y ;

Low-High: Região com baixos valores de y e vizinhos com altos valores de y ;

Low-Low: Região com baixos valores de y e vizinhos com baixos valores de y .

Perceba, também, que em situações em que a estrutura espacial dos dados não é relevante, isto é, conforme $|\hat{\alpha}|$ vai se aproximando de zero quando se inclui os efeitos do tempo, os estimadores $\hat{\beta}$ são bastantes semelhantes no painel convencional e no espacial. Este resultado faz todo o sentido uma vez que a estimativa de α é muito próxima de zero. Nesse contexto, é interessante utilizar o painel convencional por ser mais parcimonioso e sua aplicabilidade ser amplamente conhecida.

4.2 Estudo de caso 2: Municípios de São Paulo

Com as mesmas variáveis da base de Rondônia, São Paulo apresenta uma característica bastante própria que é a baixa dependência espacial com áreas isoladas de alta dependência espacial (ilhas de dependência). Os modelos foram bastante estáveis, do ponto de vista da similaridade entre *SAR* e *MESPS*, e o modelo em painel, supostamente, é a melhor opção sob a ótica da parcimônia.

O desempenho do *MESPS* nesta situação é excelente, se comparado ao *SAR*. Apesar de a estrutura espacial parecer irrelevante, pode-se observar que há áreas isoladas em que há alta dependência espacial. Mais especificamente, altas dependências estão aglomeradas em torno da Rodovia Anhanguera. Por causa disso, não é recomendável afirmar que não há dependência espacial, mas que a matriz \mathbf{W} não capta esta dependência da forma adequada. Um sugestão seria utilizar uma função do tempo de percurso rodoviário, uma vez que rodovias em bom estado de conservação facilitam o trânsito de mercadorias e serviços. A Figura 4.2 ilustra bem este fato.

O modelo *MESPS*, entretanto, apresenta alguns problemas, já apontados por LeSage and Pace (2007), no que diz respeito a confiabilidade dos resultados quando o valor de $|\rho|$ se aproxima de 1. Um exemplo claro deste problema é ilustrado por meio dos dados de consumo de cigarros nos Estados Unidos, apresentados por (Baltagi, 2008). No mesmo artigo, LeSage and Pace (2007) sugere aumentar a ordem de q para obter resultados mais consistentes, uma vez que haverão menos termos compondo o resíduo da série. Evidentemente, aumentar o valor de q tem impacto significativo no esforço computacional e, conseqüentemente, no tempo de processamento para estimar os parâmetros. A Figura (4.3) mostra o impacto que o valor de q tem no tempo de processamento.

A convergência para $\hat{\alpha}_{MV}$, todavia, é alcançada mesmo para valores pequenos de q . Veja na Figura 4.4 que aumentar a ordem do polinômio de α não implica em maior estabilidade da estimativa quando a dependência espacial é não significativa. O exemplo seguinte mostra

Tabela 4.2: Resultados dos Modelos - Municípios de São Paulo

Modelo combinado (<i>Pooled</i>)			
Estimadores			
<i>Intercepto</i>	-68.741 (14.507)		
$\hat{\beta}_{agro}$	0,7067 (0,236)		
$\hat{\beta}_{ind}$	2,7965 (0,00656)		
Modelos com 1 efeito fixo			
Estimadores	Painel convencional	SAR	MESPS
<i>Intercepto</i>	130.885 (6436,4)	127.446,6 (5.590,7)	131.062,9 (5.571,07)
$\hat{\beta}_{agro}$	0,8203 (0,1353)	0,8177 (0,1174)	0,8177 (0,1171)
$\hat{\beta}_{ind}$	1,9161 (0,0157)	1,9157 (0,0136)	1,9157 (0,0135)
$\hat{\rho}$		0,0021 (0,006) ^{NS}	0,00211
$\hat{\alpha}$			-0,0021 (0,0054) ^{NS}
Modelos com 2 efeitos fixos			
Estimadores	Painel convencional	SAR	MESPS
<i>Intercepto</i>	133.025,7 (7006,9)	130.021,5 (6.060,6)	132.978,6 (6.060,2)
$\hat{\beta}_{agro}$	0,7571 (0,1547)	0,760 (0,1339)	0,7599 (0,1338)
$\hat{\beta}_{ind}$	1,9159 (0,0157)	1,9157 (0,0136)	1,9157 (0,0136)
$\hat{\rho}$		0,0017 (0,0061) ^{NS}	-0,001677
$\hat{\alpha}$			-0,001679 (0,0055) ^{NS}

o quanto a magnitude de α é determinante na estabilidade da estimação.

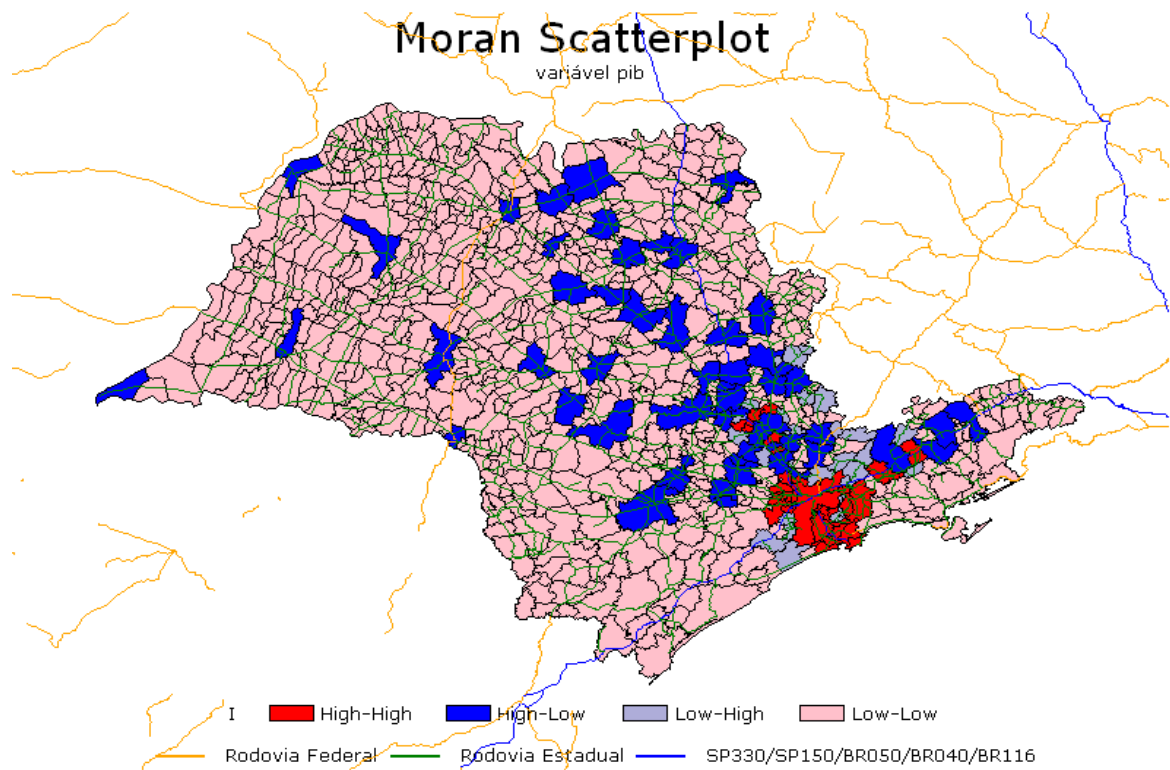


Figura 4.2: Dependência espacial entre municípios de São Paulo

4.3 Estudo de caso 3: Consumo de cigarros nos Estados Unidos

Neste contexto, o consumo de cigarros no tempo t é explicado pelo consumo de cigarros no tempo $t - 1$ ($\log c$), pelo preço do maço ($\log p$) e pela renda do estado ($\log pn$). Veja a tabela 4.3 em que essas diferenças se mostram bastante evidentes.

As diferenças entre o modelo convencional e espacial são mais evidentes, no caso de dados de consumo de cigarros, porque sua estrutura espacial é - atipicamente - relevante. Perceba que as estimativas tanto do modelo *SAR* quanto *MESPS* foram bastante próximas. Mesmo as estimativas da dispersão dos estimadores foram quase coincidentes. Isso dá certa tranquilidade para interpretar os resultados obtidos por meio do *MESPS* do mesmo ponto de vista do *SAR*. O único parâmetro que não deveria ser interpretado diretamente é $\hat{\alpha}$, uma vez que este não se relaciona com ρ por meio de uma identidade, mas por uma função

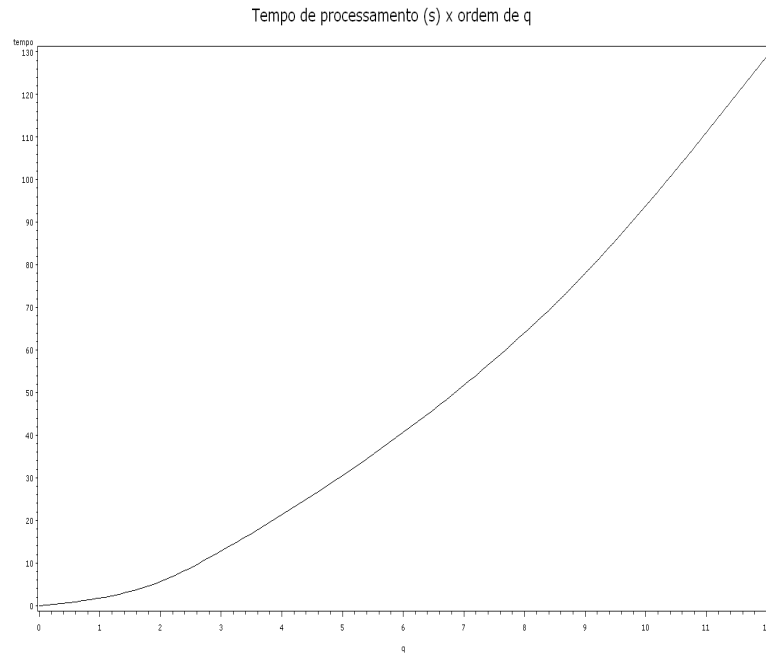


Figura 4.3: Tempo de processamento (em segundos) em função da ordem do polinômio de α

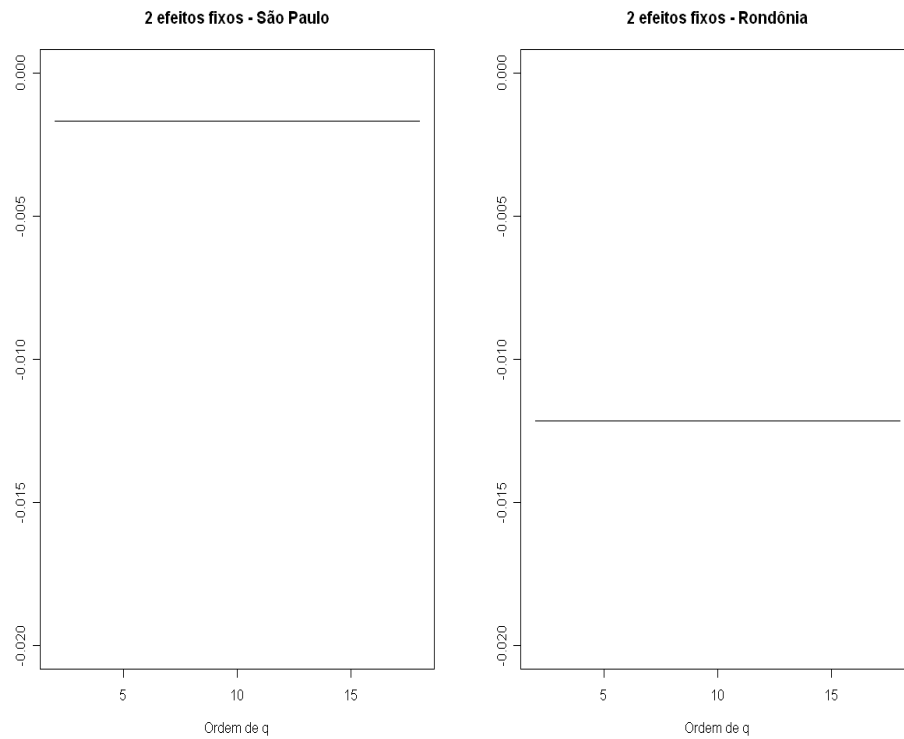


Figura 4.4: Estimativas de α em função da ordem de q

aproximada³. Como o foco do estudo é, em geral, avaliar o impacto de cada covariável na variável resposta, o *MESPS* é uma excelente alternativa. Caso o objetivo do estudo seja estimar ρ com precisão e interpretá-lo e, além disso, se houver indicação de que seu valor seja próximo de 1 (em módulo) então talvez o *SAR* seja a escolha mais inteligente. Obviamente, com a escolha do *SAR* perde-se em agilidade computacional, uma vez que ele é mais intensivo. Cabe ao pesquisador avaliar as possibilidades com parcimônia e prioridades bem definidas.

No contexto em que $|\rho|$ (e, conseqüentemente, α) difere razoavelmente⁴ de zero, é interessante verificar o impacto que a escolha de q tem na precisão das estimativas do ponto de vista da estabilidade⁵ de α . Veja, por meio da Figura 4.5 que utilizar um valor razoável para q estabiliza o valor de $\hat{\alpha}$ não sendo necessário aumentá-lo excessivamente. Por outro lado, um valor muito pequeno para q coloca a estimação numa área de instabilidade. Nesses casos, a sugestão é estimar com valores de q pequenos e ir aumentando gradativamente: teste com $q = 2$, $q = 3$, depois $q = 4$ e assim sucessivamente. Caso os valores de $\hat{\alpha}$ não varie de forma brusca, não é necessário recorrer a valores como $q = 12$. Como o tempo aumenta aproximadamente de forma exponencial em função de q , em bases de dados extensas, essa diferença pode ser de dias ou semanas de processamento.

Note, também, que o modelo com 1 efeito fixo é mais instável em relação a q . Uma explicação é o valor extremamente relevante de $\hat{\alpha}$, que demora a ser suavizado pelo denominador na série de $\mathbf{S}(\alpha)$ ⁶. Um ponto interessante a ser ressaltado é que modelos com valores atipicamente grandes para as estimativas do parâmetro de dependência espacial podem indicar má especificação. Esse exemplo evidencia o quanto o parâmetro de dependência espacial capta o efeito do tempo no caso de 1 efeito fixo (efeito individual). Quando controla-se para o efeito do tempo, a magnitude de α cai significativamente e as estimativas ficam mais estáveis conforme se incrementa o valor de q .

³LeSage and Pace (2007) exemplifica esse problema com maiores detalhes

⁴No caso dos municípios, $\hat{\alpha}$ com $q = 2$ diferia de $q = 18$ apenas na sexta casa decimal.

⁵Os tempos de processamento do *MESPS* para este conjunto de dados foram da ordem de 0,12 segundos (para $q = 18$). A amostra utilizada possui 46 unidades censitárias.

⁶Lembre-se de que o k -ésimo termo da série de $\mathbf{S}(\alpha)$ é $\frac{\alpha^k}{k!} \mathbf{W}^k$. O termo $k!$ vai dissolvendo o valor de α^k .

Tabela 4.3: Resultados dos Modelos - Consumo de cigarros nos Estados Unidos

Modelo combinado (<i>Pooled</i>)			
Estimadores			
<i>Intercepto</i>	2,4767 (0,187)		
$\hat{\beta}_{logc}$	0,4186 (0,039)		
$\hat{\beta}_{logp}$	0,6663 (0,097)		
$\hat{\beta}_{logpn}$	-0,1872 (0,095)		
Modelos com 1 efeito fixo			
Estimadores	Painel convencional	SAR	MESPS
<i>Intercepto</i>	1,0752 (0,3972)	-0,3578 (0,188) ^{NS}	2,7162 (0,216)
$\hat{\beta}_{logc}$	0,6936 (0,083)	0,2571 (0,039)	0,3632 (0,045)
$\hat{\beta}_{logp}$	0,4307 (0,092)	0,1362 (0,041)	0,1953 (0,050)
$\hat{\beta}_{logpn}$	-0,5699 (0,092)	-0,0583 (0,041) ^{NS}	-0,1397 (0,050)
$\hat{\rho}$		0,8051 (0,027)	0,6488
$\hat{\alpha}$			-1,0465 (0,061)
Modelos com 2 efeitos fixos			
Estimadores	Painel convencional	SAR	MESPS
<i>Intercepto</i>	3,1541 (0,274)	0,88 (0,206)	3,30 (0,218)
$\hat{\beta}_{logc}$	0,2535 (0,057)	0,2317 (0,043)	0,2449 (0,046)
$\hat{\beta}_{logp}$	0,1459 (0,061)	0,1328 (0,046)	0,1375 (0,049)
$\hat{\beta}_{logpn}$	-0,0616 (0,062)	0,0038 (0,047) ^{NS}	-0,0048 (0,049) ^{NS}
$\hat{\rho}$		0,556 (0,0529)	0,44097
$\hat{\alpha}$			-0,5816 (0,0736)

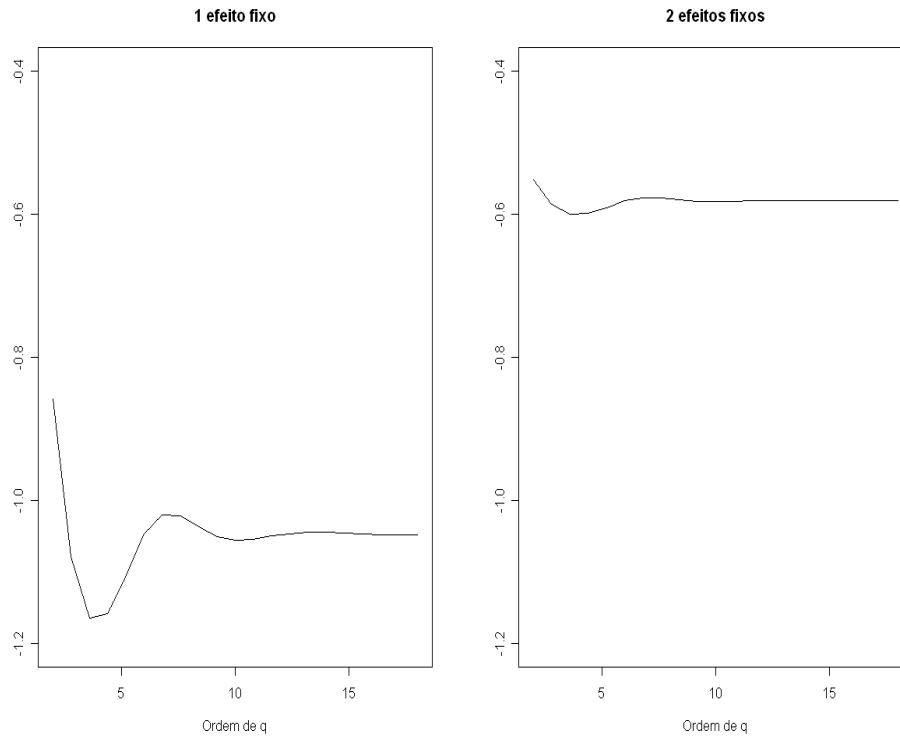


Figura 4.5: Estimativas de α em função da ordem do polinômio de α

4.4 Estudo de caso 4: Municípios do Brasil

O conjunto de dados deste estudo de caso compreende 5.560 municípios brasileiros mensurados em 4 anos (1999-2002) e, assim como os estudos de caso 1 e 2, é composto pelas variáveis *pib*, *agro* e *ind*. Apesar de a estimativa de α ser estatisticamente significativa, a magnitude dela é muito pequena para fazer com que os resultados do *MESPS* sejam diferentes do painel convencional a ponto de representar vantagem do ponto de vista da interpretação dos resultados. Para efeito de comparação, o procedimento do painel convencional de 2 efeitos fixos demorou 1,39 segundos para estimar o modelo. Em situações como esta, é mais razoável utilizar o procedimento de painel convencional por ser mais parcimonioso. Mesmo assim, é interessante ressaltar que o modelo *MESPS* gerou os resultados após 26 dias de processamento (em um computador equipado com processador Core 2 Duo[©] e 4 gb de memória RAM), ao passo que os resultados do *SAR* não puderam ser obtidos por limitações computacionais. Os resultados dos modelos estão na Tabela 4.4.

Novamente se observa que por causa da irrelevância da estrutura espacial⁷ os resultados

⁷Utilizado a matriz \mathbf{W} binária.

do modelo em painel convencional e *MESPS* foram coincidentes até a segunda casa decimal. Os modelos com 1 efeito fixo não foram estimados por falta de tempo hábil. Todavia, não há indicações de que haveriam resultados conflitantes com os que foram obtidos nas seções anteriores.

Caso o leitor esteja interessado na abordagem Bayesiana para situações como as explicadas neste capítulo, veja Parent and LeSage (2008).

Um ponto que levanta questionamentos é a escolha da matriz de proximidades na relevância da estrutura espacial. Seria interessante testar o modelo *MESPS* utilizando outro tipo de matriz \mathbf{W} (incorporando a distância rodoviária entre os municípios, por exemplo) e, então, avaliar seu desempenho. Este, entretanto, será um tema para estudos futuros.

Tabela 4.4: Resultados dos Modelos - Municípios do Brasil

Modelo combinado (<i>Pooled</i>)			
Estimadores			
<i>Intercepto</i>	-23432	(4379,2)	
$\hat{\beta}_{agro}$	1,2265	(0,136)	
$\hat{\beta}_{ind}$	2,7322	(0,005)	
Modelos com 2 efeitos fixos			
Estimadores	Painel convencional	SAR	MESPS
<i>Intercepto</i>	61.802	(2392,7)	- 59.193 (1376,1)
$\hat{\beta}_{agro}$	1,3572	(0,0787)	- 1,3556 (0,068)
$\hat{\beta}_{inb}$	1,6378	(0,0078)	- 1,6355 (0,0067)
$\hat{\rho}$			- 0,0112
$\hat{\alpha}$			-0,0113 (0,0035)

Capítulo 5

Conclusões e Recomendações

5.1 Conclusões

Conforme abordado anteriormente, os modelos em painel são extremamente úteis para estudar o comportamento das observações ao longo do tempo, aproveitando, assim, a baixa heterogeneidade intra-observação.

Na maioria dos estudos de caso apresentados neste trabalho, os modelos em painel espacial geraram resultados bastante semelhantes ao painel convencional porque a estrutura espacial era pouco relevante. É interessante, todavia, testar a hipótese de dependência espacial para não negligenciar um risco de viés por variável omitida. Dependendo do tamanho do conjunto de dados em questão, pode não ser viável estimar um modelo espacial tradicional (*SAR*) por causa do tempo requerido para estimá-lo.

Devido ao esforço computacional do modelo *SAR*, LeSage and Pace (2007) desenvolveram o Modelo *MESS*, o qual se mostrou uma grande alternativa para conjuntos de dados extensos em que modelos convencionais não se adaptariam. A grande desvantagem é a instabilidade da estimação em situações com grande dependência espacial. Um atenuante desta desvantagem é o fato de estruturas assim serem encontradas predominantemente em situações de cunho didático.

Esse trabalho mostrou o impacto nos parâmetros do modelo quando está presente a dependência espacial. No caso, o modelo em painel com 2 efeitos fixos tradicional se mostrou bastante similar ao modelo espacial com 1 efeito fixo, demonstrando o potencial desse último em recuperar a falta de especificação do modelo, por meio da dependência espacial. Outro avanço foi a derivação analítica do modelo *MESS* em painel, o qual foi denominado *MESPS* - *Matrix Exponential Spatial Panel Specification*. Como visto no Capítulo 5, o *MESPS*

permitiu que fosse estimado o parâmetro de dependência espacial num conjunto de dados com 5.560 municípios em 4 períodos de tempo. Caso o *MESPS* não tivesse características que simplificassem seu processo de estimação, o desenvolvimento de métodos para otimizar a estimação do *SAR* se daria em outro sentido, talvez aproveitando propriedades de partições de matrizes.

O problema ficou em estimar o intercepto do modelo *MESPS* e sua variância, uma vez que o procedimento computacional proposto no Apêndice A.4 leva em consideração apenas o procedimento *Within Transformation*.

5.2 Recomendações para Trabalhos Futuros

Os parâmetros do modelo *MESPS* se mostraram consistentes, quando comparados ao modelo *SAR*, no entanto o intercepto foi superestimado, exatamente por não incorporar o fator espacial. Em ambos os modelos, o erro padrão do intercepto não foi corretamente estimado por não incluir a variância de $\rho\mathbf{W}\mathbf{y}$, os quais são claramente dependentes. Os resultados não estão, de forma alguma, comprometidos por causa do intercepto, uma vez que o interesse da análise está no impacto marginal das variáveis explicativas. O desenvolvimento do intercepto e sua variância pode ser um assunto para trabalhos futuros.

Algumas extensões do modelo *MESPS* ainda não foram implementadas de forma clara em alguma linguagem computacional amplamente difundida no meio acadêmico, como o modelo de efeitos aleatórios e os modelos para painéis desbalanceados. A relevância desses dois modelos é inquestionável, uma vez que modelos de efeitos aleatórios são, entre outras coisas, extremamente robustos para captar heterogeneidade ao longo das observações. Já os painéis desbalanceados tem uma contribuição decisiva da modelagem espacial porque a persistência de algumas observações no conjunto de dados pode ser influenciada, quando não determinada, pela estrutura espacial. Para ilustrar esse fenômeno, imagine que o painel se refira a empresas que atuem em determinada atividade econômica. A permanência ou ausência de algumas empresas no mercado poderia ser determinada pela existência, ou não, de concorrentes no mesmo bairro ou cidade¹.

Seria interessante, também, adaptar o procedimento *SAR* para grandes conjuntos de dados. Infelizmente o *SAR* não possui as vantagens computacionais que o *MESPS* possui e, por isso, seria necessário um esforço no sentido de otimizar o cálculo de determinantes de

¹Proximidade, neste contexto, poderia ser simplesmente a atuação no mesmo mercado consumidor.

matrizes de grandes dimensões.

Por fim, para tornar o modelo *MESPS* mais abrangente, será imprescindível desenvolver sua versão para efeitos aleatórios, uma vez que esses efeitos controlam a heterogeneidade entre e intra-observações de maneira mais eficiente.

Referências Bibliográficas

- Anselin, L. (1988). *Spatial Econometrics*. Kluwer Academic Publishers.
- Anselin, L. & Smirnov, O. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. *Computational Statistics and Data Analysis*, 35:301–319.
- Arellano, M. (2003). *Panel data econometrics*. Oxford University Press.
- Balestra, P. & Nerlove, M. (1966). Pooling cross-section and times series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica*, 34:585–612.
- Baltagi, B. H. (2008). *Econometric analysis of Panel data*, (4th ed.). Wiley.
- Breusch, T. & Pagan, A. (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, 47:1287–1294.
- Casella, G. & Berger, R. (2001). *Statistical Inference*. Duxbury Press, 2a edição.
- Cochrane, D. & Orcutt, G. (1949). Applications of least squares regressions to relationships containing autocorrelated error terms. *Journal of American Statistical Association*, 44:32–61.
- Elhorst, P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, 26:244–268.
- Elhorst, P. (2010). *Handbook of Applied Spatial Analysis*. Springer.
- Fotheringham, A. S., Brunson, C., & Charlton, M. (2006). *Geographically Weighted Regression - the analysis of spatially varying relationships*. John Wiley & Sons Ltd.
- Frees, E. (2004). *Longitudinal and Panel Data*. Cambridge University Press.
- Greene, W. (2006). *Econometric Analysis*, (6th ed.). Pearson Education.

- Griliches, Z. (1977). Estimating returns to schooling: Some econometric problems. *Econometrica*, 45:1–22.
- Guilford, J. (1954). *Psychometric Methods*. McGraw Hill.
- Hartley, H. & Rao, J. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54:93–108.
- Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 46:1251–1272.
- Hausman, J. & Taylor, W. (1981). Panel data and unobservable individual effects. *Econometrica*, 49:1377–1398.
- Hedeker, D. & Gibbons, R. (2006). *Longitudinal Data Analysis*. Wiley.
- Horn, R. & Johnson, C. (1988). *Matrix Analysis*. Cambridge University Press.
- Hsiao, C. (2003). *Analysis of Panel Data*, (2nd ed.). Cambridge University Press.
- Johnston, J. & Dinardo, J. (1996). *Econometric Methods*. McGraw Hill/Irwin.
- Lembo, A. J. (2005). Spatial autocorrelation. Technical report.
- LeSage, J. (1999). *The theory and practice of spatial econometrics*. Department of Economics - University of Toledo.
- LeSage, J. & Pace, K. (2007). A matrix exponential spatial specification. *Journal of Econometrics*, 140:190–214.
- Lichstein, J., Simons, T. R., Shiner, S. A., & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, 72(3):445–463.
- McCulloch, C. & Searle, S. (2001). *Generalized, Linear and Mixed Models*. Wiley.
- Moral-Benito, E. (2009). Determinants of economic growth - a bayesian panel data approach. *The World Bank Development Research Group - Policy Research Working Paper*, 4830.
- Nerlove, M. (1971). A note on error components models. *Econometrica*, 39:359–382.
- Ord, J. (1975). Estimation methods for models of spatial interaction. *Journal of American Statistical Association*, 70:120–126.

- Parent, O. & LeSage, J. (2008). Using the variance structure of the conditional autoregressive spatial specification to model knowledge spillovers. *Journal of Applied Econometrics*, 23(2):235–256.
- Patterson, H. & R., T. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58:545–554.
- Silva, A. R. (2006). Avaliação de modelos de regressão espacial para análise de cenários do transporte rodoviário de carga. Master's thesis, Universidade de Brasília, Departamento de Engenharia Civil e Ambiental.
- Silva, A. R. & Alves, P. F. (2007). Computational algorithm for estimation of panel data models with two fixed effects. *Revista de Matemática e Estatística*, v.25(n.2):p.19–32.
- Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer.
- Wooldridge, J. (2001). *Econometric Analysis of Cross Section and Panel Data*, (1st ed.). The MIT Press.

Apêndice A

A.1 Within Transformation

Seja \mathbf{y} um vetor e \mathbf{X} uma matriz de variáveis explicativas que pode ser particionada em dois conjuntos \mathbf{X}_1 e \mathbf{X}_2 . Seja o modelo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (\text{A.1})$$

em que \mathbf{X} pode ser particionado em $\mathbf{X} = (\mathbf{X}_1 | \mathbf{X}_2)$, assim como $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$. Então, de acordo com o Teorema de *Frisch-Waugh*, é possível estimar os efeitos parciais β_1 por meio de regressões auxiliares. Mais especificamente:

$$\mathbf{y} = \mathbf{X}_2\boldsymbol{\alpha} + \boldsymbol{\epsilon} \quad (\text{A.2})$$

$$X_{1i} = X_{2i}\gamma + u_i \quad (\text{A.3})$$

onde X_{1i} é a i -ésima variável da partição \mathbf{X}_1 . Para obter β_1 , basta calcular a regressão

$$\boldsymbol{\epsilon} = \mathbf{U}\boldsymbol{\beta}_1 + \mathbf{e} \quad (\text{A.4})$$

e \mathbf{U} é a matriz cujas colunas são os vetores de resíduos \mathbf{u}_i .

No contexto do modelo 1.2, pode-se definir a matriz

$$\mathbf{D} = \begin{bmatrix} \mathbf{j} & 0 & \dots & 0 \\ 0 & \mathbf{j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{j} \end{bmatrix}$$

como sendo a partição \mathbf{X}_2 representada acima. Sendo assim, utilizando estimadores de mínimos quadrados ordinários, tem-se que:

$$\begin{aligned}\epsilon &= (\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}')\mathbf{y} \\ \mathbf{U} &= (\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}')\mathbf{X} \\ \hat{\beta} &= (\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\epsilon\end{aligned}$$

Seja $\mathbf{M} = (\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}')$.¹ Então,

$$\hat{\beta} = (\mathbf{X}'\mathbf{M}'\mathbf{M}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{M}'\mathbf{M}\mathbf{y}),$$

$$\hat{\beta} = (\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{M}\mathbf{y}),$$

$$\mathbf{M} = \mathbf{I} - \frac{1}{\mathbf{T}}\mathbf{ii}'$$

$$\mathbf{D}_{NT,N} = \begin{bmatrix} \mathbf{j} & 0 & \dots & 0 \\ 0 & \mathbf{j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{j} \end{bmatrix} \quad \mathbf{D}'_{N,NT} = \begin{bmatrix} \mathbf{j}' & 0 & \dots & 0 \\ 0 & \mathbf{j}' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{j}' \end{bmatrix} \quad \mathbf{D}'\mathbf{D}_{N,N} = \begin{bmatrix} \mathbf{j}'\mathbf{j} & 0 & \dots & 0 \\ 0 & \mathbf{j}'\mathbf{j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{j}'\mathbf{j} \end{bmatrix}$$

mas $\mathbf{j}'\mathbf{j} = \sum_{j=1}^T 1 = T$.

$$\mathbf{D}'\mathbf{D}_{N,N} = \begin{bmatrix} T & 0 & \dots & 0 \\ 0 & T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T \end{bmatrix},$$

$$(\mathbf{D}'\mathbf{D})^{-1} = \frac{1}{T}\mathbf{I}_{N,N}$$

$$\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' = \frac{1}{T}\mathbf{D}\mathbf{D}'$$

$$\mathbf{D}\mathbf{D}'_{N,N} = \begin{bmatrix} \mathbf{jj}' & 0 & \dots & 0 \\ 0 & \mathbf{jj}' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{jj}' \end{bmatrix} \Rightarrow \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' = \frac{1}{T} \begin{bmatrix} \mathbf{jj}' & 0 & \dots & 0 \\ 0 & \mathbf{jj}' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{jj}' \end{bmatrix}$$

$$\frac{1}{T}\mathbf{jj}' = \begin{bmatrix} \frac{1}{T} & \frac{1}{T} & \dots & \frac{1}{T} \\ \frac{1}{T} & \frac{1}{T} & \dots & \frac{1}{T} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{T} & \frac{1}{T} & \dots & \frac{1}{T} \end{bmatrix}$$

¹Esta é a matriz que transforma o vetor resposta no vetor de resíduos. Esta matriz é simétrica e idempotente.

Seja $\mathbf{z} = (z_1, \dots, z_T)'$ o nível de uma variável z na mesma observação ao longo do tempo.

$$(\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}')\mathbf{z} = \mathbf{Mz} = \begin{pmatrix} z_1 \\ \dots \\ z_T \end{pmatrix} - \frac{1}{T} \begin{pmatrix} \sum_{j=1}^T z_j \\ \dots \\ \sum_{j=1}^T z_j \end{pmatrix}, \text{ que é equivalente a subtrair a}$$

média do indivíduo de cada observação em cada tempo.

Seja o modelo transformado

$$\mathbf{My} = \mathbf{MX}\boldsymbol{\beta} + \mathbf{MDc} + \mathbf{M}\boldsymbol{\epsilon} \quad (\text{A.5})$$

Note que este modelo é equivalente a

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\boldsymbol{\beta} + \epsilon_{it} - \bar{\epsilon}_i \quad (\text{A.6})$$

porque $\mathbf{MDc} = (\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}')\mathbf{Dc} = \mathbf{Dc} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{Dc} = \mathbf{Dc} - \mathbf{Dc} = \mathbf{0}$

A.2 Matriz de Covariâncias - Modelo *MESPS*

Para facilitar a derivação da matriz de covariâncias dos estimadores do modelo *MESPS*, será utilizada a seguinte parametrização de sua respectiva função de verossimilhança

$$l(\alpha, \gamma, \beta) = -\frac{NT}{2} \log \gamma - \frac{1}{2\gamma} \left(\mathbf{y}' \mathbf{S}' \mathbf{S} \mathbf{y} - 2 \sum_{i=0}^K \beta_i \mathbf{X}'_i \mathbf{S} \mathbf{y} + 2 \sum_{i=0}^k \sum_{j \neq i} \beta_i \beta_j \mathbf{X}'_i \mathbf{X}_j + \sum_{i=0}^K \beta_i^2 \mathbf{X}'_i \mathbf{X}_i \right) \quad (\text{A.7})$$

em que $\mathbf{X}_0 = \mathbf{j}_{NT} = (1, \dots, 1)'$, $SQR = (\mathbf{S} \mathbf{y} - \mathbf{X} \boldsymbol{\beta})' (\mathbf{S} \mathbf{y} - \mathbf{X} \boldsymbol{\beta})$ e $\gamma = \sigma^2$, para tornar a notação menos carregada. Sendo assim, tem-se que

$$\begin{aligned} \frac{\partial l}{\partial \beta_i} &= -\frac{1}{2\gamma} \left(-2 \mathbf{X}'_i \mathbf{S} \mathbf{y} + 2 \sum_{j \neq i} \beta_j \mathbf{X}'_i \mathbf{X}_j + 2 \beta_i \mathbf{X}'_i \mathbf{X}_i \right) \\ \frac{\partial^2 l}{\partial \beta_i^2} &= -\frac{1}{2\gamma} (2 \mathbf{X}'_i \mathbf{X}_i) = \frac{\mathbf{X}'_i \mathbf{X}_i}{\gamma} \\ \frac{\partial^2 l}{\partial \beta_i \partial \beta_j} &= -\frac{1}{2\gamma} (2 \mathbf{X}'_i \mathbf{X}_j) = \frac{-\mathbf{X}'_i \mathbf{X}_j}{\gamma} \\ \frac{\partial^2 l}{\partial \beta_i \partial \gamma} &= \frac{1}{2\gamma^2} (-2 \mathbf{X}'_i \mathbf{S} \mathbf{y} + 2 \sum_{j \neq i} \beta_j \mathbf{X}'_i \mathbf{X}_j + 2 \beta_i \mathbf{X}'_i \mathbf{X}_i) \\ &= -\frac{1}{\gamma^2} \mathbf{X}'_i (\mathbf{S} \mathbf{y} - \sum_{j=0}^k \beta_j \mathbf{X}_j) \\ E \left(\frac{\partial^2 l}{\partial \beta_i \partial \gamma} \right) &= -\frac{1}{\gamma^2} \mathbf{X}'_i E(\mathbf{S} \mathbf{y} - \sum_{j=0}^k \beta_j \mathbf{X}_j) = 0 \\ \frac{\partial l}{\partial \gamma} &= -\frac{NT}{2\gamma} + \frac{SQR}{2\gamma^2} \\ \frac{\partial^2 l}{\partial \gamma^2} &= \frac{NT}{2\gamma^2} - \frac{2SQR}{2\gamma^3} = \frac{NT}{2\gamma^2} - \frac{NT\gamma}{\gamma^3} \\ &= \frac{NT}{\gamma^2} \left(\frac{1}{2} - 1 \right) = -\frac{NT}{2\gamma^2} \end{aligned}$$

Evidentemente, caso seja realizada a *Within Transformation*, basta substituir \mathbf{X}_i por $\mathbf{M} \tilde{\mathbf{X}}_i = \tilde{\mathbf{X}}_i$ e \mathbf{M} deve ser escolhida de acordo com a transformação desejada (1 ou 2 efeitos fixos).

A.3 Procedimento computacional - *MESPS*

O procedimento para estimar o modelo *MESPS*, em linguagem *SAS*, em conjuntos de dados pequenos e médios é dado pelo procedimento a seguir.

```
%macro MESPS(tab=,y=,x=,year=,spatial=,spatialmatrix=,q=);
proc sort data=&tab;by &year &spatial;run;
proc iml;
use &tab;
read all var{&y} into y;
read all var{&x} into x;
read all var{&year} into T;
read all var{&spatial} into N;
use &spatialmatrix;
read all into W;
T=ncol(unique(T));
N=ncol(unique(N));
NT=N*T;
nvar=ncol(x);
nomes={Intercept &x alpha}';
M=I(NT)-J(T,T,1/T)@I(N)-I(T)@J(N,N,1/N)+j(NT,NT,1/NT);
Xtil=M*X;
ytil=M*y;
P=I(NT)-xtil*inv(xtil'*xtil)*xtil';
nq=&q;
yi=j(NT,nq,0);
Q1=1;

do i=1 to nq;
yi[,i]=(I(T)@(W**i))*(ytil);
Q1=Q1//fact(i);
end;

yMat=(ytil)||yi;
Q=inv(diag(Q1));
```



```

Z=Q*(yMat')*P*yMat*Q;

do j=1 to nrow(z);
do h=1 to nrow(z);
if j=1 & h=1 then do;c1=j+h;c2=z[j,h];end;
else do;c1=c1/(j+h);c2=c2/z[j,h];end;
end;
end;
c=c1||c2;
call sort(c,{1,2});
_nc_=(unique(c[,1]));
cm=j(1,ncol(_nc_),0);
do j=1 to ncol(_nc_);
do i=1 to nrow(c);
if c[i,1]=_nc_[j] then cm[j]=cm[j]+c[i,2];
end;
end;
v=j(ncol(cm),1,0);
do i=1 to ncol(cm);
v[i]=i;
end;
v[ncol(cm)]=0;
v1=v;
v2=v;
v1[ncol(cm)]=0;
v1[ncol(cm)-1]=0;
call sort(v,{1});
call sort(v1,{1});
alpha1=-3;
converge=1;
criteria = 0.001;
iter = 1;
itermax = 100;
do while (converge > criteria & iter < itermax);

```

```

v3=j(ncol(cm)-2,1,0);
v4=j(ncol(cm)-3,1,0);
do i=1 to ncol(cm)-2;
v3[i]=alpha1**(i);
end;
do i=1 to ncol(cm)-3;
v4[i]=alpha1**(i);
end;
v3=j(2,1,1)//v3;
v4=j(3,1,1)//v4;
alpha=alpha1-((cm*(v#v3))/(cm*(v#v1#v4)));
converge=abs(alpha-alpha1);
print iter alpha1 alpha;
alpha1=alpha;
iter = iter + 1;
end;
rho=1-exp(alpha);
print alpha rho;

galpha=1;
do i=1 to nq;
galpha=galpha//alpha**i;
end;
Sy=yamat*Q*galpha;

beta=inv(xtil'*xtil)*xtil'*sy;

/*Calculando a variancia de alpha*/
a=j(ncol(cm),1,0);
do i=1 to ncol(cm);
v[i]=i;
a[i]=alpha**(i-1);
end;
v1=j(1,1,0)/(v[1:nrow(v)-1]);

```

```

v2=j(1,1,0)/(v1[1:nrow(v1)-1]);
a1=j(1,1,1)/(a[1:nrow(a)-1]);
a2=j(1,1,1)/(a1[1:nrow(a1)-1]);

g1=cm*a;
g2=cm*(v1#a1);
g2l=cm*(v1#v2#a2);
stdalpha=sqrt(1/( (nt/2)*((g2l*g1-g2**2)/g1**2)));

/*-Hessiano para k variaveis*/
beta0=(j(nt,1,1/nt)'*y)-(j(nt,1,1/nt)'*X)*beta;
res=Sy-xtil*beta;
gamma=(1/NT)*(res'*res);
Hess=j(ncol(xtil)+1,ncol(xtil)+1,0);
do i=1 to ncol(xtil);
do j=1 to ncol(xtil);
Hess[i,j]=Xtil[,i]'*Xtil[,j];
end;
end;
Hess[ncol(xtil)+1,ncol(xtil)+1]=NT/(2*gamma);
Mat_var=gamma*inv(Hess);

stdbeta=sqrt(vecdiag(mat_var[1:nvar,1:nvar]));
bvec=beta0//beta//alpha;
stdbeta0=sqrt(abs(gamma/(NT)+(j(nt,1,1/nt)'*X)*
mat_var[1:nvar,1:nvar]*(j(nt,1,1/nt)'*X)'));
tmp=stdbeta0//stdbeta//stdalpha;
tstat = bvec/tmp;
probt=2*(1-probt(abs(tstat),NT-2));
param=bvec||tmp||tstat;
* Print;
d1={"Estimation Method" "Number of Spatial Units  "
"Number of Time Series" "Number of Observations"}';
d2={"FixTwo - MESPS"}//char(N,6)//char(T,6)//char(NT,6);

```

```

print "Dependent Variable: &y";
print "Model Description",, d1[label=''] d2[label=''];
est={'Estimate' 'Std. Error' 't Value'};
print,"Parameter Estimates",, nomes[label='Variable']
param[format=comma12.4 colname=est label='']
probt[format=pvalue6. colname='P > |t|' label=''];
quit;
%mend MESPS;

```

A.4 Procedimento computacional - *MESPS* - Grandes conjuntos de dados

As modificações necessárias para possibilitar a estimação do *MESPS* em grandes conjuntos de dados foram: primeiro substituir

```

/*****/
M=I(N,T)-J(T,T,1/T)@I(N)-I(T)@J(N,N,1/N)+j(N,T,1/NT);
Xtil=M*X;
ytil=M*y;
/*****/

```

por

```

/*****/
e=j(T,T,1/T);
k=j(N,N,1/N);
Et=(I(T)-e);
Xtil=j(N,T,nvar,0);
ytil=j(N,T,1,0);
do j=1 to T;
do i=1 to N;
En=I(N)[i,]-k[i,];
do l=1 to T;
M=Et[j,l]@En;
l1=(l-1)*N+1;

```

```

l2=1*N;
Xtil[(j-1)*N+i,]=Xtil[(j-1)*N+i,]+M*x[l1:l2,];
ytil[(j-1)*N+i]=ytil[(j-1)*N+i]+M*y[l1:l2];
end;
end;
end;
/*****/

```

Depois, substituir

```

/*****/
do i=1 to nq;
yi[,i]=(I(T)@(W**i))*(ytil);
Q1=Q1//fact(i);
end;
yMat=(ytil)||yi;
Q=inv(diag(Q1));
/*****/

por

/*****/
do ikk=1 to nq;
if ikk=1 then w2=1;
else w2=w2*w;
do k=1 to T;
IT=I(T)[k,];
do i=1 to N;
do kk=1 to T;
do j=1 to N;
if ikk=1 then do;
if kk=1 & j=1 then w31=IT[kk]@w2*w[i,j];
else w31=w31||(IT[kk]@w2*w[i,j]);
end;
else do;
if kk=1 & j=1 then w31=IT[kk]@w2[i,]*w[,j];

```

```

else w31=w31||(IT[kk]@w2[i,]*w[,j]);
end;
end;
end;
if k=1 & i=1 then yil=w31*ytil;
else yil=yil/(w31*ytil);
end;
end;
if ikk=1 then yi=yil;
else yi=yi||yil;
Q1=Q1//fact(ikk);
free w31 yil;
end;
yMat=ytil||yi;
Q=inv(diag(Q1));
/*****/

e por fim, substituir

/*****/
P=I(NT)-xtil*inv(xtil'*xtil)*xtil';
Z=Q*(yMat')*P*yMat*Q;
/*****/

por

/*****/
do k=1 to nq+1;
do i=1 to NT;
INT=j(1,NT,0);
INT[i]=1;
do j=1 to NT;
if j=1 then
P2=xtil[i,]*inv(xtil'*xtil)*xtil[j,]';
else
P2=P2||xtil[i,]*inv(xtil'*xtil)*xtil[j,]';

```

```

end;
if i=1 then
PP=yMat[,k] *(INT-P2) ;
else
PP=PP||yMat[,k] *(INT-P2) ;
end;
if k=1 then PP1=PP;
else PP1=PP1//PP;
free P2 PP;
end;
P=PP1*yMat;
Z=Q*P*Q;
/*****/

```

Com isso, foi possível estimar o modelo, mesmo após um grande período de processamento.