



UNIVERSADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

LUCIANA MOURA REINALDO

ESTIMAÇÃO CLÁSSICA E BAYESIANA PARA DADOS EM PAINEL

BRASÍLIA

2017

LUCIANA MOURA REINALDO

ESTIMAÇÃO CLÁSSICA E BAYESIANA PARA DADOS EM PAINEL

Dissertação apresentada ao Programa de Pós-graduação em Estatística do Instituto de Ciências Exatas da Universidade de Brasília, como parte dos requisitos necessários para a obtenção do título de Mestre em Estatística.

Orientador: Prof. Dr. Bernardo Borba de Andrade

BRASÍLIA

2017

LUCIANA MOURA REINALDO

ESTIMAÇÃO CLÁSSICA E BAYESIANA PARA DADOS EM PAINEL

Dissertação apresentada ao Programa de Pós-graduação em Estatística do Instituto de Ciências Exatas da Universidade de Brasília, como parte dos requisitos necessários para a obtenção do título de Mestre em Estatística.

Aprovada em: 30/06/ 2017.

BANCA EXAMINADORA

Prof. Dr. Bernardo Borba de Andrade (Orientador)
Universidade de Brasília (UnB)

Prof. Dr. Eduardo Yoshio Nakano
Universidade de Brasília (UnB)

Prof. Dr. Vicente Lima Crisóstomo
Universidade Federal do Ceará (UFC)

Agradecimentos

Aos meus pais, Tim e Bete, por todo amor e carinho, por me ensinarem o valor do estudo, por serem sempre o meu oxigênio, minha motivação, por considerarem a quantidade quatro pequena (para número de filhos) e desejarem uma quinta filha.

À minha reca de irmãos, Ninha, Beleza, Nívia, Marquim e Emídio, pelo apoio, incentivo, conselhos, cuidados e momentos....Amo vocês imensamente!!!

Aos meus sobrinhos (por ordem alfabética para não ter problema...) Alice, Amor, Carol, Guilherme, Gustavo, Matheus e Pietra por despertarem o meu desejo de sempre voltar para casa.

Ao Enjado pela paciência, cumplicidade, por soltar minha mão, mas sempre estar ao meu lado para acudir o tombo ou celebrar a conquista.

Ao Tobonildo por me ensinar silenciosamente com um olhar o significado do amor.

Ao meu amigo Nilo Sérgio pelas conversas leves...momentos agradáveis. Lamento tanto que você tenha ido embora cedo demais e não ter visto o desfecho do negócio...espero que onde você esteja consiga ver que deu certo...sinto sua falta todos os dias.

Agradeço ao meu orientador, o professor Bernardo Borba de Andrade, por ter me aceitado como sua orientanda e ter confiado no meu trabalho. Aos professores participantes da banca examinadora Eduardo Yoshio Nakano e Vicente Lima Crisóstomo pelo tempo, pelas valiosas colaborações e sugestões.

Agradeço ao Prof. Dr. Raul Yukihiro Matsushita, por ter participado da banca examinadora na qualificação deste trabalho. Aos professores Antônio Eduardo, Cira, Gilardon, Juliana e Maria Eduarda pelo conhecimento transmitido, fundamental para a construção desse trabalho.

A todos os colegas e amigos do mestrado que compartilharam momentos de ansiedade, preocupações e também de alegria neste período. Em especial, Alex e Marcílio, quero vocês na minha vida para sempre...obrigada por tudo rapazes, vocês agregam valor!!!

À minha família adotiva de Brasília... João, Patrícia, Yan, Yarla, Neto, Cleo, Seu Tomás, Antônia, Sílvia, Juliana, Giselda, Dani, Dona Fátima e Seu Edmilton por terem feito os dias ficarem mais leves e divertidos, por me sentir tão acolhida mesmo tão distante de casa.

À minha amiga Nazaré por todo o fôlego, palavra, energia, ajuda, torcida e conselhos. Aos meus queridos amigos Cintia, Emílson, Rapha e Adelano, que mesmo distante nunca foram ausentes, por acreditarem na minha capacidade...obrigada pela força sempre!!!

Finalmente, aos meus colegas do Departamento de Administração da Universidade Federal do Ceará pelo imenso apoio.

Resumo

Estudos das mais diversas áreas de conhecimento utilizam várias metodologias de análises de dados quantitativos para verificar tendências e evoluções no comportamento de unidades de observação. Nesse sentido, a utilização de modelos que envolvam dados provenientes de várias unidades experimentais ao longo do tempo vem crescendo gradativamente na pesquisa científica. A metodologia de dados em painel permite a análise longitudinal de diversas unidades de observação em um único painel, possibilitando a identificação de padrões e a própria evolução das unidades de observação. Esse trabalho tem por objetivo sistematizar o conhecimento das estratégias de inferência relacionadas aos dados em painel, com o intuito de proporcionar uma linguagem clara e acessível àqueles que, embora não sendo econométristas, necessitam se apropriar dos métodos de análise dos dados em painel para aplicá-los na sua prática de pesquisa. Para facilitar a compreensão dos métodos, foram apresentados alguns exemplos implementados em um *software* gratuito, R, um ambiente de cálculos estatísticos, utilizando conjuntos de dados contidos nesse *software* e uma base de dados reais aplicando tanto a abordagem de inferência clássica quanto a abordagem de inferência bayesiana.

Palavras-chave: Dados em painel. Inferência Bayesiana. Inferência Clássica.

Abstract

Studies of the most diverse areas of knowledge use several methodologies of quantitative data analysis to verify trends and evolutions in the behavior of observation units. In this sense, the use of models involving data from several experimental units over time has been growing gradually in scientific research. The panel data methodology allows the longitudinal analysis of several units of observation in a single panel, allowing the identification of patterns and the evolution of observation units themselves. This work aims to systematize the knowledge of inference strategies related to panel data, with the aim of providing a clear and accessible language to those who, although not being econometricians, need to appropriate the methods of panel data analysis to apply them in their research practice. To facilitate the comprehension of the method, we have presented some examples implemented in a free software, R, a environment for statistical computing, from datasets contained in this software and a real database using as much the classical approach as the bayesian inference approach.

Keywords: Panel data. Bayesian Inference. Classical Inference.

Lista de ilustrações

Figura 1 – MQO <i>versus</i> MQVD	45
Figura 2 – Evento A	57
Figura 3 – Diagrama de árvore para o Exemplo 8	59
Figura 4 – Resumo do procedimento bayesiano	62
Figura 5 – Prioris conjugadas Gama(a,b) e suas posterioris	70
Figura 6 – Exemplo de modelo bayesiano hierárquico	71
Figura 7 – Testes de comparação entre os modelos	88
Figura 8 – Heterogeneidade entre firmas e entre os anos	91
Figura 9 – Investimento ao longo dos anos por firma	92
Figura 10 – Densidades estimadas para as posterioris dos parâmetros	103
Figura 11 – Trajetória das posterioris dos parâmetros	103
Figura 12 – Variáveis regressoras por firma	104
Figura 13 – Comparação interceptos individuais diferentes modelos	107
Figura 14 – Gráfico da trajetória das posterioris dos parâmetros	107
Figura 15 – Densidades a posteriori dos parâmetros	108
Figura 16 – Funções de autocorrelação dos parâmetros	108
Figura 17 – Trajetórias, densidades e acfs dos hiperparâmetros	109

Lista de tabelas

Tabela 1 – Exemplos da estrutura de dados em painel	27
Tabela 2 – Algumas distribuições a priori conjugadas	68
Tabela 3 – Descrição das variáveis utilizadas no modelo de investimento	91
Tabela 4 – Resultados obtidos através do método bayesiano e do MQO	102
Tabela 5 – Estimativas dos parâmetros - Modelo não hierárquico (EF)	106
Tabela 6 – Estimativas dos parâmetros - Modelo hierárquico (EA)	106
Tabela 7 – Coeficientes estimados sob a perspectiva clássica e bayesiana	109

Lista de abreviaturas e siglas

EA	Efeitos Aleatórios
EF	Efeitos Fixos
EFF	Estimador de Efeitos Fixos
MCMC	Monte Carlo via Cadeia de Markov
HPD	<i>Highest Posterior Density</i>
MQO	Mínimos Quadrados Ordinários
MQVD	Mínimos Quadrados com Variável <i>Dummy</i>
MQG	Mínimos Quadrados Generalizados
LSDV	<i>Least Squares Dummy Variable</i>
PD	Primeiras Diferenças

Sumário

1	INTRODUÇÃO	21
2	MODELAGEM CLÁSSICA PARA DADOS EM PAINEL	25
2.1	Dados em painel	25
2.2	Modelo para dados agrupados (Pooled)	28
2.2.1	Estimador de mínimos quadrados ordinários	29
2.2.2	Estimador <i>pooled</i>	32
2.3	Modelo de efeitos fixos	36
2.3.1	Estimador de efeitos fixos	37
2.3.2	Estimador de mínimos quadrados com variável <i>dummy</i>	40
2.3.3	Estimador de primeiras diferenças	45
2.4	Modelo de efeitos aleatórios	47
2.4.1	Estimador de mínimos quadrados generalizados	51
3	INFERÊNCIA BAYESIANA	55
3.1	Teorema de Bayes	56
3.2	Princípios gerais da inferência bayesiana	60
3.2.1	Princípio da verossimilhança	60
3.2.2	Princípio da suficiência	61
3.2.3	Princípio da condicionalidade	62
3.3	Elementos da Inferência Bayesiana	62
3.4	Distribuição a priori	64
3.4.1	Priori não informativa	64
3.4.1.1	Método de Bayes-Laplace	65
3.4.1.2	Método de Jeffreys	66
3.4.2	Priori conjugada	67
3.4.3	Prioris Hierárquicas	71
3.5	Estimativa pontual e intervalar	74
4	MODELOS BAYESIANOS PARA DADOS EM PAINEL	75
4.1	Modelo para dados agrupados (Pooled)	75
4.2	Modelos de efeitos individuais	78
4.2.1	Função de verossimilhança	79
4.2.2	Modelos de efeitos fixos	79
4.2.3	Modelo de efeitos aleatórios	80

5	APLICAÇÃO NO R	83
5.1	Modelo para dados agrupados	83
5.1.1	Descrição da base de dados	83
5.1.2	Modelo para dados agrupados <i>Pooled</i>	85
5.2	Pacote plm	87
5.2.1	Testes para efeito individual e efeito temporal	88
5.3	Modelos de dados em painel com plm	89
5.3.1	Descrição dos dados e o modelo de investimento	90
5.3.2	Efeitos fixos	92
5.3.3	Efeitos aleatórios	96
5.4	Abordagem Bayesiana para Dados em Painel	99
5.4.1	Modelo para dados agrupados	99
5.4.2	Modelo de efeitos individuais	104
6	CONSIDERAÇÕES FINAIS	111
	REFERÊNCIAS	113

1 Introdução

O mundo da informação é o mundo em que vivemos. A obtenção e uso da informação, como interpretá-la e servir-se dela, é um debate necessário no processo de tomada de decisão, e, de fato, é para tomar decisões que coletamos e analisamos dados. Pesquisas científicas das mais diversas áreas de conhecimento utilizam várias metodologias de análises de dados quantitativos para verificar tendências e evoluções no comportamento de unidades de observação: da biologia à economia; da geologia à física quântica, da medicina a pesquisas base para planejamento de políticas públicas.

Nesse sentido, a utilização de modelos que envolvam dados provenientes de várias unidades experimentais ao longo do tempo vem crescendo gradativamente na pesquisa científica. A disponibilidade cada vez maior de dados em escalas temporais amplia a importância de métodos e ferramentas relacionados a modelos que os envolvam (Bond e Reenen, 2007).

Dados em painel, dados longitudinais ou combinação de séries temporais e dados de corte transversal são termos empregados em econometria e estatística para caracterizar a estrutura de dados que apresentam repetidas observações da mesma unidade de corte transversal (e.g. uma empresa, pessoa, família ou município) ao longo do tempo. Um painel pode representar informações individuais, como pacientes em ensaios clínicos ou indivíduos em uma pesquisa de acompanhamento, ou unidades agregadas, tais como grupos etários da população ou áreas geográficas. Nesse contexto as mesmas variáveis para os mesmos indivíduos são acompanhadas em diferentes momentos do tempo.

A metodologia de dados em painel permite a análise longitudinal, ou seja, ao longo do tempo, de diversas unidades de observação em um único painel, possibilitando a identificação de padrões e a própria evolução das unidades de observação. Segundo Hsiao (2014), Wooldridge (2008) e Marques et al. (2000) a análise de dados em painel apresenta vantagens muito maiores que análises de séries temporais e análises transversais *cross-section*, pois disponibilizam maior quantidade de informação, maior variabilidade de dados, menor colinearidade entre as variáveis, maior número de graus de liberdade e maior eficiência na estimação.

Uma das principais contribuições da utilização dessa metodologia de análise de dados é que permite ao pesquisador investigar efeitos que não podem ser identificados apenas com o uso de dados em corte transversal ou séries temporais, visto que a análise de dados em painel permite isolar os efeitos de ações específicas, tratamentos ou políticas gerais (Fitrianto e Musakkal, 2016; Akbar et al., 2011). Os dados em painel proporcionam a incorporação de informações sobre a variação individual na análise.

Um dos fatores que diferenciam conjuntos de dados de cortes transversais e dados

em painel é a estrutura de covariância dos mesmos. No primeiro caso, corte transversal, há uma total independência entre as observações, enquanto que no segundo, dados em painel, espera-se alguma correlação entre as observações realizadas na mesma unidade de investigação. Ao observar unidades de corte transversal ao longo do tempo, é natural que possa haver heterogeneidade para os diferentes indivíduos como também dependência nas observações. As técnicas de estimação em painel podem levar em conta explicitamente essas variáveis individuais específicas (Baltagi, 2008). O uso dos dados em painel permite o controle de alguns tipos de variáveis omitidas sem observá-las (Stock e Watson, 2006).

A metodologia de dados em painel permite o tratamento da heterogeneidade não observada dos indivíduos, ou seja, aquelas peculiaridades específicas de cada indivíduo observado que se mantém ao longo do tempo. Para que se controle por variáveis não observadas, a técnica de dados em painel assume que além dos efeitos observados das variáveis explicativas presentes no modelo, há efeitos não observados específicos relativos ao indivíduo e ao tempo. Esses efeitos não observados podem ser tratados como invariantes ao longo do tempo. Este é o tratamento feito pela metodologia de efeitos fixos (EF). De forma alternativa, estes efeitos não observados são tratados como variáveis aleatórias, como feito pela metodologia de modelos de efeitos aleatórios (EA) (Hsiao, 2014). A utilização de dados sobre a forma de um painel conduz a uma grande variedade de modelos cujas diferenças dependem das hipóteses colocadas sobre os parâmetros desses modelos e sobre o comportamento do termo de erro.

Observa-se que a literatura sobre a técnica de dados em painel concentra-se na abordagem clássica. Por outro lado, a abordagem bayesiana, ao permitir uma completa inferência probabilística de todos os parâmetros sem depender de qualquer suposição de normalidade, surge como uma metodologia alternativa, talvez até mais adequada para muitas aplicações. A literatura tem sugerido que a inferência bayesiana abrirá novos horizontes para análise de alguns fenômenos estudados em finanças (D'Espallier e Guariglia, 2015; D'Espallier, Huybrechts e Iturriaga, 2011).

Este trabalho foi desenvolvido considerando a necessidade de um conhecimento estruturado sobre a metodologia de dados em painel, a nível de conceitos e práticas, em especial em pesquisas econométricas. O estudo visa cumprir este objetivo principal e preencher uma lacuna pouco explorada pela literatura em nível de ensino, pesquisa e extensão.

Dessa forma, o objetivo desse trabalho é sistematizar o conhecimento das estratégias de inferência relacionadas aos dados em painel, com o intuito de proporcionar uma linguagem clara e acessível àqueles que, embora não sendo econométricos, necessitam se apropriar dos métodos de análise dos dados em painel para aplicá-los na sua prática de pesquisa.

A compreensão é facilitada pelos exemplos implementados no software R, utilizando conjunto de dados contidos nesse software bem como em um conjunto de dados reais. A

escolha desse programa foi motivada por ser uma plataforma de *software* livre que funciona em diversos sistemas operacionais e que apresenta pacotes e funções disponíveis para a estimação de dados em painel.

O trabalho está organizado em seis seções principais, sendo a primeira, essa introdução. A segunda seção contempla uma necessária revisão da literatura sobre a modelagem clássica de dados em painel onde são dispostos o modelo para dados agrupados, modelo de efeitos fixos e o modelos de efeitos aleatórios. A terceira seção trata do referencial teórico relacionado à inferência bayesiana, que servirá de base as discussões da quarta seção que tratará dos modelos que utilizam a inferência bayesiana na metodologia de dados em painel. A quinta seção ocupa-se das aplicações da metodologia de dados em painel utilizando dados empíricos. Pondo termo ao trabalho, seguem-se as considerações finais e referências bibliográficas utilizadas.

2 Modelagem Clássica para Dados em Painel

Este capítulo visa apresentar uma visão geral dos principais modelos de regressão utilizados para dados em painel: modelos para dados agrupados, modelo de efeitos fixos e modelo de efeitos aleatórios (Seções 2.2, 2.3, 2.4, respectivamente), bem como os métodos de estimação adequados para cada um desses modelos. Utilizou-se como referências para essas seções: Cameron e Trivedi (2005), Greene (2008), Baltagi (2008), Hsiao (2014).

2.1 Dados em painel

A análise da mudança entre as observações individuais ou nos grupos desempenha um papel importante na pesquisa social e biomédica sendo fundamental para a compreensão dos mecanismos causais da doença ou patologia social, na avaliação do impacto das intervenções políticas. A utilização dessa forma de dados proporciona ao pesquisador captar a heterogeneidade¹ entre as unidades amostrais, aumentar a eficiência das estimativas, além de captar a dinâmica do comportamento das unidades. A estrutura de dados em painel é caracterizada pela combinação de várias séries temporais provenientes de diferentes unidades amostrais, ou seja, referem-se a dados de n entidades diferentes observadas em T períodos de tempo diferentes. Uma definição formal para esse tipo de dado é apresentada a seguir.

Definição 1. *Dados em painel consistem em observações sobre as mesmas n entidades em T períodos de tempo ($T \geq 2$). Se o conjunto de dados contiver observações sobre as k variáveis independentes X_1, X_2, \dots, X_k e a variável independente Y , então denotam-se os dados por*

$$(X_{1,it}, X_{2,it}, \dots, X_{k,it}, Y_{it}), \quad i = 1, \dots, n \text{ e } t = 1, \dots, T$$

em que o subscrito i refere-se à entidade observada e o subscrito t refere-se ao momento em que é observada (Stock e Watson, 2006).

Estudos utilizando a metodologia de dados em painel foram desenvolvidos em diversas áreas do conhecimento, embora seja amplamente utilizada nos artigos científicos das ciências sociais aplicadas e, principalmente, na economia. Em medicina, Rohde et al. (2016) estimaram o impacto da insegurança econômica sobre a saúde mental de adultos australianos. No marketing, Kaswengi e Diallo (2015) investigaram a propensão à compra de marca nacional em relação a variáveis relacionadas ao marketing e características pessoais do consumidor na França em momentos de recessão econômica. Sociologia, avaliando a

¹ Efeito individual.

relação entre capital social e a autopercepção de saúde na África do Sul (Lau e Ataguba, 2015).

Sachs (2015) analisou a contribuição das inovações biológicas para produtividade agrícola da cana-de-açúcar a partir da metodologia de dados em painel em seis regiões do estado de São Paulo ao longo dos anos de 1998 a 2009. Modelos econométricos com dados em painel foram construídos para avaliar como aspectos referentes à pobreza, renda, educação e saúde impactam na dinâmica do Programa Bolsa Família nas cinco regiões brasileiras, no período de 2004 a 2010 (Pinto, Coronel e Filho, 2015).

Ghinis e Fochezatto (2013) analisaram os efeitos da construção civil sobre a redução da pobreza dos estados brasileiros utilizando modelos dinâmicos de dados em painel. Enquanto Kea, Saksenaa e Hollyb (2011) utilizaram dados em painel de 143 países em desenvolvimento durante os anos de 1995 a 2008 para compreender a trajetória de despesas com saúde nesses locais.

Conforme Hsiao (2014) as vantagens desse método em relação aos dados em corte transversal e séries temporais são:

- a) Maior capacidade para construção de hipóteses comportamentais mais realistas;
- b) Permite a observação de relações dinâmicas entre os indivíduos;
- c) Controle do impacto de variáveis omitidas;
- d) Gera previsões mais precisas para os resultados individuais;
- e) Simplifica a implementação computacional e inferência estatística.

Para Baltagi (2008) as vantagens dos dados em painel são mais graus de liberdade, melhores previsões, menos multicolinearidade, mais variação nos dados que resulta em estimadores mais eficientes. Além de permitir controlar a heterogeneidade, a dinâmica de estudo e de testes de hipóteses comportamentais mais complicados do que é possível com uma única série temporal ou seção transversal.

De acordo com Cameron e Trivedi (2005) a análise de dados em painel tem como maior vantagem o aumento da precisão da estimação dos parâmetros. Esse ganho de precisão é resultado do aumento do número de observações devido à combinação de vários períodos de tempo de cada indivíduo.

Existem vários tipos de estrutura de dados em painel. Uma delas é o painel balanceado no qual o número de períodos T é o mesmo para todos os indivíduos i , isto é, os indivíduos são observados em todos os períodos em consideração ($T_i = T \forall i$), totalizando nT observações ($n \times T$). Em contrapartida no painel desbalanceado cada unidade *cross section* apresenta diferentes números de observações temporais, ou seja, a dimensão temporal difere de indivíduo para indivíduo e o tamanho da amostra é $\sum_{i=1}^N T_i$.

Definição 2. Um painel é dito balanceado se cada unidade cross section apresenta o mesmo período de tempo, $t = 1, 2, \dots, T$. Para um painel desbalanceado, a dimensão do tempo, denotada por T_i , é específica para cada indivíduo.

A Tabela 1 apresenta exemplos da estrutura de dados em painel. A Tabela 1a corresponde a um painel balanceado, no qual cada unidade de corte transversal apresenta o mesmo número de observações, observa-se que para cada indivíduo (1, 2 e 3) há a mesma quantidade de observações (2013 - 2015). Na Tabela 1b, os indivíduos 1, 2 e 3 apresentam 3, 5 e 1 observações, respectivamente, nesse caso o painel é dito desbalanceado.

Tabela 1 – Exemplos da estrutura de dados em painel

Indivíduo	Ano	Y	X1	X2
1	2013	6.0	7.8	5.8
1	2014	4.6	0.6	7.9
1	2015	9.4	2.1	1.1
2	2013	9.1	1.4	4.1
2	2014	8.3	0.9	5.9
2	2015	0.6	9.6	7.2
3	2013	9.3	0.2	6.4
3	2014	4.8	5.3	7.3
3	2015	5.9	2.1	3.2

(a) Painel balanceado

Indivíduo	Ano	Y	X1	X2
1	2013	6.0	7.8	5.8
1	2014	4.6	0.6	7.9
1	2015	9.4	2.1	1.1
2	2013	9.1	1.4	4.1
2	2014	8.3	0.9	5.9
2	2015	0.6	9.6	7.2
2	2016	9.3	0.2	6.4
2	2014	4.8	5.3	7.3
3	2015	5.9	2.1	3.2

(b) Painel desbalanceado

Outra classificação existente na literatura para dados em painel é painel curto e painel longo. Em um painel curto, o número de indivíduos de corte transversal é maior que o número de períodos de tempo. Em um painel longo, o número de períodos de tempo é superior ao número de indivíduos (Gujarati e Porter, 2011). Como ilustração, apresentam-se a seguir exemplos de banco de dados de painéis balanceado e desbalanceado contidos nos pacotes **AER** (Kleiber e Zeileis, 2008a) e **pscl** (Jackman, 2015) do R (R Core Team, 2016).

Exemplo 1. (Painel balanceado e desbalanceado)

```
#Balanceado
> data(Fatalities,package="AER")
> table(Fatalities$year)
1982 1983 1984 1985 1986 1987 1988
48   48   48   48   48   48   48

#Desbalanceado
> data(presidentialElections,package="pscl")
```

```
> table(presidentialElections$year)
1932 1936 1940 1944 1948 1952 1956 1960 1964 1968 1972 1976 1980 1984 1988 1992
48   48   48   48   47   48   48   50   50   51   51   51   51   51   51   51
1996 2000 2004 2008 2012
51   51   51   51   51
```

Observa-se no Exemplo 1, que o primeiro conjunto de dados representa um painel balanceado, ou seja, número de observações igual para todas as unidades de análise. Diferente do que ocorre no segundo conjunto de dados, existem dados faltantes por observação ao longo do tempo.

Os dados em painel se caracterizam pela sequência temporal de duas ou mais observações em cada indivíduo, por tal estrutura supõe-se que as observações entre os indivíduos sejam independentes, já as de cada indivíduo apresentam a característica de dependência com erros correlacionados.

A suposição de erros correlacionados exige uma modelagem que contemple tal característica, uma vez que o modelo de regressão linear ignora tal correlação. Este modelo trata as observações como independentes. Desse modo, existem diversos modelos específicos para dados em painel e a estimação dos parâmetros do modelo depende de premissas a respeito do intercepto, dos coeficientes angulares e dos termos de erro.

Nas seções a seguir são discutidos os principais modelos nesse contexto: dados agrupados, efeitos fixos e efeitos aleatórios. Para fins de aplicação da metodologia será considerado apenas painéis balanceados, contudo as metodologias aplicam-se igualmente para painéis desbalanceados.

2.2 Modelo para dados agrupados (*Pooled*)

O modelo inicial considerado o mais restrito é o modelo para dados agrupados (*pooled*), que despreza as dimensões temporal e espacial dos dados e considera todos os coeficientes constantes ao longo do tempo e entre os indivíduos, assume a forma de dados empilhados. Nesse caso, o método habitual para estimação dos parâmetros é o método dos mínimos quadrados ordinários e a formulação geral do modelo com n observações em T períodos e k variáveis é dado por (Cameron e Trivedi, 2005):

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T. \quad (2.1)$$

em que o subscrito i denota as diferentes unidades de corte transversal e t o período de tempo que está sendo analisado; y_{it} é a variável dependente do indivíduo i no tempo t ; \mathbf{x}'_{it} um vetor de variáveis explicativas de ordem $1 \times p$; $\boldsymbol{\beta}$ um vetor de ordem $p \times 1$ dos coeficientes das variáveis; α o intercepto e ε_{it} o termo de erro do i -ésimo indivíduo no tempo t . Se pressupõe que os erros ε_{it} são independentes e identicamente distribuídos com

média zero e variância σ^2 . Ressalta-se que nesse modelo os coeficientes $\alpha, \beta_1, \beta_2, \dots, \beta_p$ são os mesmos para todas as unidades individuais (observe a ausência do subscrito i e t). As suposições feitas para este modelo são:

S.1 Observações independentes;

S.2 $E[\varepsilon_{it} | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{it}] = 0, \quad t = 1, 2, \dots, T;$

S.3 $\text{Var}[\varepsilon_{it} | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{it}] = \sigma^2, \quad t = 1, 2, \dots, T;$

S.4 $\text{Cov}[\varepsilon_{it}, \varepsilon_{js} | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{it}] = 0$ se $i \neq j$ ou $t \neq s$.

A suposição **S.2** impõe exogeneidade estrita nas variáveis explicativas, **S.3** refere-se aos erros homoscedáticos, em que a variância de ε_{it} não pode depender de qualquer elemento de \mathbf{x}'_{it} e deve ser constante ao longo das observações, **S.4** é a suposição de correlação serial, nos quais os erros não podem ser correlacionados ao longo das observações (Greene, 2008). Pelas suposições verifica-se que o modelo *pooled* assume os mesmos pressupostos do modelo de regressão linear clássico e caso o modelo seja especificado corretamente e os regressores não sejam correlacionados com o termo de erro, pode-se estimar consistentemente os parâmetros a partir do método de mínimos quadrados ordinários (MQO).

2.2.1 Estimador de mínimos quadrados ordinários

A palavra “regressão” foi introduzida pela primeira vez em 1885 por Sir Francis Galton num estudo que demonstrava que a estatura de uma descendência não se aproxima dos progenitores, mas para a estatura média de ambos (Rocha, 2001). Um dos procedimentos mais usados para obter estimadores é aquele que se baseia no princípio dos mínimos quadrados (Bussab e Morettin, 2013; Reed e Ye, 2011), que consiste no critério para estimar os coeficientes de modo a minimizar a soma dos quadrados dos desvios. A descoberta do método de regressão baseado nos mínimos quadrados é atribuída a Carl Friedrich Gauss, que usou o procedimento no início do século XIX (Rocha, 2001).

Considere o modelo para a i -ésima observação com k variáveis independentes:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

ou

$$Y_i = \alpha + \sum_{j=1}^k \beta_j X_{ji} + \varepsilon_i \tag{2.2}$$

Também pode ser escrito como

$$\begin{aligned} Y_1 &= \alpha + \beta_1 X_{11} + \beta_2 X_{21} + \dots + \beta_k X_{k1} + \varepsilon_1 \\ Y_2 &= \alpha + \beta_1 X_{12} + \beta_2 X_{22} + \dots + \beta_k X_{k2} + \varepsilon_2 \\ Y_3 &= \alpha + \beta_1 X_{13} + \beta_2 X_{23} + \dots + \beta_k X_{k3} + \varepsilon_3 \\ &\vdots \\ Y_n &= \alpha + \beta_1 X_{1n} + \beta_2 X_{2n} + \dots + \beta_k X_{kn} + \varepsilon_n \end{aligned}$$

As igualdades anteriores podem ser alocadas em dois vetores colunas ($n \times 1$), descritos a seguir:

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{(n \times 1)} = \underbrace{\begin{bmatrix} \alpha + \beta_1 X_{11} + \dots + \beta_k X_{k1} + \varepsilon_1 \\ \alpha + \beta_1 X_{12} + \dots + \beta_k X_{k2} + \varepsilon_2 \\ \vdots \\ \alpha + \beta_1 X_{1n} + \dots + \beta_k X_{kn} + \varepsilon_n \end{bmatrix}}_{(n \times 1)}$$

Ainda,

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{(n \times 1)} = \underbrace{\begin{bmatrix} \alpha + \beta_1 X_{11} + \dots + \beta_k X_{k1} \\ \alpha + \beta_1 X_{12} + \dots + \beta_k X_{k2} \\ \vdots \\ \alpha + \beta_1 X_{1n} + \dots + \beta_k X_{kn} \end{bmatrix}}_{(n \times 1)} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{(n \times 1)}$$

Finalmente,

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{(n \times 1)} = \underbrace{\begin{bmatrix} 1 & X_{11} & \dots & X_{k1} \\ 1 & X_{12} & \dots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \dots & X_{kn} \end{bmatrix}}_{(n \times (k+1))} \times \underbrace{\begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}_{((k+1) \times 1)} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{(n \times 1)}$$

O modelo de regressão linear composto por um componente sistemático ($\mathbf{X}\boldsymbol{\beta}$) e um componente aleatório ($\boldsymbol{\varepsilon}$) pode ser convenientemente descrito em notação matricial como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.3)$$

em que

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}, \quad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{k1} \\ 1 & X_{12} & \dots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \dots & X_{kn} \end{bmatrix}_{n \times (k+1)}, \quad \boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}.$$

Segundo [Hoffmann e Vieira \(1998\)](#) as suposições para o modelo (2.3) são:

1. A variável dependente (Y_i) é função linear das variáveis independentes ($X_{ji}, j = 1, \dots, k$);
2. Os valores das variáveis independentes são fixos;
3. $E(\varepsilon_i) = 0$, ou seja, $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, em que $\mathbf{0}$ representa um vetor de zeros;
4. Erros são homocedásticos, isto é, $E(\varepsilon_i^2) = \sigma^2, \forall i = 1, 2, \dots, n$;
5. Erros são não-correlacionados entre si, isto é, $\text{Cov}(\varepsilon_i, \varepsilon_s) = 0$ para $i \neq s$;
6. Os erros tem distribuição normal.

Combinando as pressuposições (4) e (5) tem-se

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{I}\sigma^2 \quad (2.4)$$

O estimador de mínimos quadrados ordinários para o vetor de parâmetros $\boldsymbol{\beta}$ é dado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (2.5)$$

Pode-se representar as matrizes $\mathbf{X}'\mathbf{X}$ e $\mathbf{X}'\mathbf{y}$ por somas individuais:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{X}'_1 & \mathbf{X}'_2 & \dots & \mathbf{X}'_n \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} = \sum_{i=1}^n \mathbf{X}'_i\mathbf{X}_i, \quad (2.6)$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} \mathbf{X}'_1 & \mathbf{X}'_2 & \dots & \mathbf{X}'_n \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{bmatrix} = \sum_{i=1}^n \mathbf{X}'_i\mathbf{y}_i. \quad (2.7)$$

Substituindo (2.6) e (2.7) em (2.5), tem-se que:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^n \mathbf{X}'_i\mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}'_i\mathbf{y}_i. \quad (2.8)$$

A partir do modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ com $\boldsymbol{\varepsilon} \sim N_n(0, \sigma^2\mathbf{I})$, segue que \mathbf{y} tem distribuição normal multivariada com média $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ e variância $\text{Var}(\mathbf{y}) = \sigma^2\mathbf{I}$. O estimador de mínimos quadrados $\hat{\boldsymbol{\beta}}$ tem distribuição normal p -variada com média $\boldsymbol{\beta}$ e matriz de variância-covariância dada por

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \quad (2.9)$$

O estimador de mínimos quadrados é considerado um estimador consistente para o modelo dado pela Equação (2.1) caso $\text{Cov}[\varepsilon_{it}, \mathbf{x}_{it}] = \mathbf{0}$, ou seja, se os regressores não são correlacionados com o termo de erro.

2.2.2 Estimador *pooled*

O modelo *pooled* é aquele no qual os dados sobre diferentes unidades são em conjunto sem a suposição de diferenças individuais.

$$Y_{it} = \alpha + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + \varepsilon_{it} \quad (2.10)$$

em que Y_{it} = variável dependente; x_{jit} = j -ésima variável explicativa; ε_{it} = termo de erro/distúrbios; α = intercepto; β_1, \dots, β_k parâmetros a serem estimados.

Na forma de vetor:

$$y = X\beta + u \quad (2.11)$$

em que y é um vetor de dimensões $nT \times 1$, β um vetor de ordem $(k + 1) \times 1$, X tem dimensões $nT \times (k + 1)$ e ε tem dimensões $nT \times 1$ e:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix}, \quad y = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}.$$

Sendo \mathbf{x}_{ij} o vetor de observações da j -ésima variável independente para a unidade i ao longo do tempo e \mathbf{y}_i o vetor de observações da variável dependente da unidade i . Então, o estimador *pooled* dado por:

$$\hat{\beta}^{POOLED} = (X'X)^{-1}X'y = \begin{bmatrix} \hat{\alpha}^{POOLED} \\ \hat{\beta}_1^{POOLED} \\ \vdots \\ \hat{\beta}_k^{POOLED} \end{bmatrix} \quad (2.12)$$

As premissas do modelo *pooled* são:

$$E(\varepsilon) = 0 \quad (2.13)$$

$$\varepsilon\varepsilon' = \sigma_u^2 I \quad (2.14)$$

$$posto(X) = k + 1 < nT \quad (2.15)$$

$$E(\varepsilon|X) = 0 \quad (2.16)$$

Destaca-se:

- (2.16) X é não estocástico e não correlacionado com ε .
- (2.14) O termo de erro (ε) é homoscedástico e não autocorrelacionado.
- (2.16) Exogeneidade estrita para as variáveis independentes.
- Caso as premissas de (2.13) - (2.16) sejam satisfeitas então $\hat{\beta}^{POOLED}$ é um estimador não-viciado de variância mínima.

Ressalta-se que o estimador *pooled* é similar ao estimador de mínimos quadrados ordinários utilizado na regressão linear múltipla.

Para ilustrar o modelo de regressão para dados agrupados utiliza-se o conjunto de dados *Fatalities* composto por 336 observações sobre 34 variáveis para 48 estados americanos entre 1982 e 1988.

Exemplo 2. (*Pooled*)

Considere o conjunto de dados `Fatalities` disponível no pacote `AER` (Kleiber e Zeileis, 2008a) contendo informações sobre mortes no trânsito de 48 estados americanos ao longo de 7 anos (1982-1988), totalizando 336 observações. Deseja-se modelar a taxa de mortalidade no trânsito (variável dependente) em função das variáveis regressoras: imposto sobre cerveja, consumo de bebidas e dois fatores econômicos: taxa de desemprego e renda per capita. O modelo é dado por:

$$\begin{aligned} \text{frate}_{it} &= \beta_1 \text{beertax}_{it} + \beta_2 \text{spirits}_{it} + \beta_3 \text{unemp}_{it} + \beta_4 \text{income}_{it} + \alpha + u_{it}, \\ i &= 1, 2, \dots, 48, \\ t &= 1, 2, \dots, 7. \end{aligned} \tag{2.17}$$

Portanto,

- Unidade de corte transversal: estados americanos (48 estados, $n = 48$);
- Dimensão temporal (t): 1982 a 1988 ($T = 7$);
- Painel balanceado ($nT = 48 \times 7 = 336$) observações;
- Variáveis: taxa de mortalidade no trânsito (`frate`), imposto sobre cerveja (`beertax`), consumo de bebidas (`spirits`), taxa de desemprego (`unemp`) e renda per capita (`income`).

A estimação do modelo de regressão *Pooled* no R:

Call:

```
lm(formula = frate ~ beertax + spirits + unemp + income, data = Fatalities)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.22581	-0.35100	-0.05238	0.27829	1.94364

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.119e+00	2.970e-01	13.868	< 2e-16 ***
beertax	9.720e-02	6.155e-02	1.579	0.115256
spirits	1.623e-01	4.325e-02	3.754	0.000206 ***
unemp	-2.910e-02	1.272e-02	-2.289	0.022731 *
income	-1.584e-04	1.699e-05	-9.327	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4793 on 331 degrees of freedom
 Multiple R-squared: 0.3019, Adjusted R-squared: 0.2934
 F-statistic: 35.78 on 4 and 331 DF, p-value: < 2.2e-16

As variáveis consumo de bebidas (*spirits*), taxa de desemprego (*unemp*) e renda per capita (*income*) apresentam significância estatística a 5%. Combinam-se todas as 336 observações e pressupõem-se que os coeficientes de regressão sejam os mesmos para todos os estados americanos. Algumas desvantagens desse modelo são: *i*) não há distinção entre os estados, *ii*) a modelagem ignora a estrutura temporal presente nos dados. Assim, ao analisar os diferentes estados como se fossem dados de cortes transversais “camufla-se” a heterogeneidade que possa existir entre os estados americanos, desconsiderando diferenças importantes existentes entre os estados. Utilizar o modelo *pooled* é assumir que as observações são independentes, o que em algumas situações não parece razoável.

Nesse contexto, o modelo descrito em (2.1) assume todos os coeficientes constantes para todos os indivíduos em todos os períodos de tempo, e não permite uma possível heterogeneidade individual. Além disso, o termo de erro tem média zero e variância constante, eles não estão correlacionados ao longo do tempo e nem entre indivíduos, também não são correlacionados com as variáveis explicativas, então não há nada que o diferencie do modelo de regressão múltipla.

A aplicação do modelo de dados agrupados de forma a ignorar a natureza do painel é restritiva de várias maneiras. Uma delas é considerar a falta de correlação entre erros correspondentes ao mesmo indivíduo.

Devido a estrutura dos dados em painel, precisa-se lidar com a heterogeneidade individual. Quando esses efeitos individuais estão correlacionados com as variáveis explicativas do modelo, o estimador de mínimos quadrados ordinários torna-se inconsistente, porque pode haver fatores que determinam a variável dependente, mas que não estão sendo considerados. Nesses casos costuma-se utilizar o estimador de efeitos fixos pois permanece consistente e viável.

Do modelo de regressão para dados agrupados surgem duas extensões: o modelo de efeito fixo e o modelo de efeito aleatório. Conforme [Cameron e Trivedi \(2005\)](#) os modelos de efeitos fixos apresentam a complicação adicional de que os regressores sejam correlacionados com os efeitos do nível do indivíduo e, portanto, uma estimação consistente dos parâmetros do modelo requer uma eliminação ou controle dos efeitos fixos. No modelo de efeitos aleatórios, por outro lado, assume-se que o efeito individual é puramente aleatório e não é correlacionado com os regressores. Assim, considere o modelo

$$\mathbf{y}_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{u}_{it} \quad (2.18)$$

em que a estrutura do erro \mathbf{u}_{it} é dado por

$$\mathbf{u}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_{it} \quad (2.19)$$

Nessa formulação assume-se que o termo de erro \mathbf{u}_{it} é composto por duas componentes não observáveis: o efeito individual que é constante ao longo do tempo $\boldsymbol{\alpha}_i$; e os erros idiossincráticos $\boldsymbol{\varepsilon}_{it}$ para cada indivíduo i e período t e $\boldsymbol{\varepsilon}_{it}$ é não correlacionado com \mathbf{X}_{it} . A grande proporção de aplicações empíricas envolve uma das seguintes suposições sobre o efeito individual:

1. Modelo de efeito aleatório: $\boldsymbol{\alpha}_i$ é não correlacionado com \mathbf{X}_{it} .
2. Modelo de efeito fixo: $\boldsymbol{\alpha}_i$ é correlacionado com \mathbf{X}_{it}

A distinção relevante entre os dois modelos não é se o efeito é fixo ou não, mas se o efeito está correlacionado com as variáveis explicativas. Os modelos de efeitos fixos referem-se aos modelos que permitem uma correlação arbitrária entre o efeito individual não observado $\boldsymbol{\alpha}_i$ e as variáveis explicativas observadas \mathbf{x}_{it} , contrastando como o modelo de efeito aleatório em que o efeito permanente é independente dos regressores (Wooldridge, 2008). A descrição dos modelos e suas equações são mostradas nas seções a seguir.

2.3 Modelo de efeitos fixos

Os modelos de efeitos fixos para dados em painel permitem que os interceptos variem entre as unidades observacionais, contemplando a heterogeneidade entre indivíduos, mas que são constantes ao longo do tempo. Além de serem utilizados em situações em que não é possível dissociar o efeito individual α_i das variáveis independentes. Se $\boldsymbol{\alpha}_i$ está correlacionado com qualquer das variáveis explicativas do modelo, o estimador de mínimos quadrados ordinários resultará em estimativas viesadas e inconsistentes. De acordo com Hsiao (2014) a maneira usual de contornar esse problema é empregar o estimador de efeitos fixos, que tem sido amplamente usado nas análises feitas com dados em painel para tratar a questão de heterogeneidade não observada.

A especificação de um modelo de efeitos fixos consiste em:

$$y_{it} = \boldsymbol{\alpha}_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \quad (2.20)$$

em que y_{it} é a variável dependente, $\boldsymbol{\alpha}_i$ ($i = 1, \dots, n$) refere-se ao efeito específico individual que capta a heterogeneidade não observada entre as unidades em análise que possivelmente estão correlacionados com os regressores que controlam as características invariantes no tempo. O subscrito i sugere que os interceptos podem ser diferentes em cada unidade, \mathbf{x}_{it} é um vetor $p \times 1$ que representa o conjunto de variáveis explicativas, $\boldsymbol{\beta}$ um vetor de ordem $p \times 1$ de parâmetros a serem estimados e ε_{it} é o termo de erro. Cada $i = 1, \dots, n$ é um

indivíduo e $t = 1, \dots, T$ a observação de uma característica desse indivíduo no tempo. Portanto, cada indivíduo é um *cluster* formado por um conjunto de T observações no tempo resultando em $n \times T$ observações.

O modelo de efeitos fixos apresenta n interceptos, um para cada indivíduo, os quais absorvem os efeitos de todas as variáveis omitidas que diferem entre as unidades, mas são fixas no tempo, ou seja, o modelo supõe a existência de características que variam entre os indivíduos, mas são constantes ao longo do tempo. Contudo, os parâmetros β são únicos para todos as unidades observacionais e em todos os períodos de tempo.

Nesse contexto, uma estimação consistente dos parâmetros do modelo requer uma eliminação ou controle dos efeitos fixos, e as estratégias de estimação usuais para para essa finalidade são:

1. Estimador de efeitos fixos ou *within*;
2. Mínimos quadrados com variável *dummy*;
3. Primeiras diferenças.

Os estimadores (1) e (2), efeito fixo e mínimos quadrados com variável *dummy*, respectivamente, utilizam estratégias diferentes de estimação, mas suas estimativas são iguais.

2.3.1 Estimador de efeitos fixos

O estimador de efeitos fixos ou *within* utiliza uma transformação no modelo para eliminar o efeito do componente não observado α_i e em seguida, estima os coeficientes por MQO no modelo transformado.

Seja o modelo ² dado por

$$y_{it} = \mathbf{x}'_{it}\beta + (\alpha_i + \varepsilon_{it}), \quad i = 1, 2, \dots, n; \quad t = 1, 2, \dots, T \quad (2.21)$$

em que y_{it} é a variável dependente; \mathbf{x}'_{it} o vetor das variáveis dependentes; β o vetor de parâmetros a serem estimados; α_i o componente de corte transversal ou específico dos indivíduos; ε_{it} é o elemento de erro combinado da série temporal e corte transversal, chamado de erro idiossincrático. Ressalta-se que $\alpha_i + \varepsilon_{it}$ foram colocados entre parênteses para enfatizar que esses termos são não observados.

Os pressupostos a respeito desses termos são:

Pressuposto 1. α_i é livremente correlacionado com \mathbf{x}_{it} .

Pressuposto 2. $E(\mathbf{x}_{it}\varepsilon_{is}) = \mathbf{0}$ para $s = 1, 2, \dots, T$ (*Exogeneidade estrita*).

² Assume-se que T é constante entre os indivíduos, ou seja, que trata-se de um painel balanceado.

Observa-se que se α_i é correlacionado com alguma variável do vetor \mathbf{x}_{it} , ocorrerá o problema de endogeneidade que poderia afetar as estimativas de MQO. Assim, sob os pressupostos (1) e (2), pode-se utilizar o estimador de efeitos fixos (EEF) ou de primeiras diferenças (PD) para obter estimativas consistentes de β , permitindo assim que α_i seja livremente correlacionado com \mathbf{x}_{it} . Destaca-se que caso \mathbf{x}_{it} tenham variáveis dependentes defasadas ($y_{i,t-1}, y_{i,t-2}, \dots$), tanto EEF como PD não produzirão estimativas consistentes.

Dessa forma, o estimador efeitos fixos resolve o problema de endogeneidade que poderia contaminar as estimativas MQO subtraindo de cada *cluster* de indivíduos sua média temporal, ou seja,

$$\bar{y}_i = \bar{\mathbf{x}}_i' \beta + \alpha_i + \bar{\varepsilon}_i, \quad i = 1, 2, \dots, n, \quad (2.22)$$

em que $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$, $\bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^T \mathbf{x}_{it}$ e $\bar{\varepsilon}_i = T^{-1} \sum_{t=1}^T \varepsilon_{it}$.

Subtraindo (2.21) de (2.22):

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it}' - \bar{\mathbf{x}}_i') \beta + \underbrace{(\alpha_i - \alpha_i)}_{=0} + (\varepsilon_{it} - \bar{\varepsilon}_i), \quad (2.23)$$

Logo,

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it}' - \bar{\mathbf{x}}_i') \beta + (\varepsilon_{it} - \bar{\varepsilon}_i). \quad (2.24)$$

Usando a notação: $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{\mathbf{x}}_{it} = (\mathbf{x}_{it}' - \bar{\mathbf{x}}_i')$ e $\tilde{\varepsilon}_{it} = (\varepsilon_{it} - \bar{\varepsilon}_i)$ pode-se escrever a Equação (2.24) como

$$\tilde{y}_{it} = \tilde{\mathbf{x}}_{it} \beta + \tilde{\varepsilon}_{it}, \quad i = 1, 2, \dots, n; \quad t = 1, 2, \dots, T \quad (2.25)$$

Empilhando as observações para $t = 1, \dots, T$

$$\begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \vdots \\ \tilde{\mathbf{y}}_n \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \vdots \\ \tilde{\mathbf{x}}_n \end{bmatrix} \beta + \begin{bmatrix} \tilde{\boldsymbol{\varepsilon}}_1 \\ \vdots \\ \tilde{\boldsymbol{\varepsilon}}_n \end{bmatrix}.$$

ou

$$\underset{Tn \times 1}{\tilde{\mathbf{y}}} = \underset{Tn \times p}{\tilde{\mathbf{x}}} \times \underset{p \times 1}{\beta} + \underset{Tn \times 1}{\tilde{\boldsymbol{\varepsilon}}}$$

Assim, os efeitos fixos α_i são eliminados, juntamente com os regressores invariantes no tempo sendo que $\mathbf{x}_{it} - \bar{\mathbf{x}}_i = \mathbf{0}$ se $\mathbf{x}_{it} = \mathbf{x}_i$ para todo t . Esta transformação é conhecida por *transformação within*. O estimador de efeitos fixos ou *within* ($\hat{\beta}_{EF}$) é obtido aplicando-se o estimador de MQO *pooled* à equação (Equação 2.24) e é dado por

$$\begin{aligned} \hat{\beta}_{EF} &= \left[\sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right]^{-1} \left[\sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i) \right] \\ &= \left[\sum_{i=1}^n \sum_{t=1}^T (\tilde{\mathbf{x}}_{it})(\tilde{\mathbf{x}}_{it})' \right]^{-1} \left[\sum_{i=1}^n \sum_{t=1}^T (\tilde{\mathbf{x}}_{it})(\tilde{y}_{it}) \right] \end{aligned} \quad (2.26)$$

O EEF é consistente sob a hipótese de exogeneidade estrita. O efeito fixo individual α_i pode ser estimado por

$$\hat{\alpha}_i^{FE} = \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}_{EF}, \quad i = 1, 2, \dots, n.$$

em que $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$, $\bar{\mathbf{x}}_i = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_{it}$ e $\hat{\boldsymbol{\beta}}_{EF}$ dada por (2.26). A estimativa $\hat{\alpha}_i$ é uma estimativa não viesada para α_i .

O Exemplo 3 ilustra a estimação *within* para o conjunto de dados apresentado no Exemplo 2.

Exemplo 3. (*Estimador within*)

Neste exemplo, a taxa de mortalidade no trânsito é modelada em função de fatores como: imposto sobre a cerveja, consumo de bebidas, taxa de desemprego e renda per capita.

Oneway (individual) effect Within Model

Call:

```
plm(formula = frate ~ beertax + spirits + unemp + income, data = pfat,
     model = "within")
```

Balanced Panel: n=48, T=7, N=336

Residuals :

	Min.	1st Qu.	Median	3rd Qu.	Max.
	-0.444000	-0.079200	0.000788	0.067600	0.569000

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)	
beertax	-4.8407e-01	1.6251e-01	-2.9787	0.003145	**
spirits	8.1697e-01	7.9212e-02	10.3137	< 2.2e-16	***
unemp	-2.9050e-02	9.0274e-03	-3.2180	0.001441	**
income	1.0471e-04	2.0599e-05	5.0834	6.738e-07	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 10.785

Residual Sum of Squares: 6.9816

R-Squared: 0.35265

Adj. R-Squared: 0.2364

F-statistic: 38.6774 on 4 and 284 DF, p-value: < 2.22e-16

O resultado da estimação do modelo de efeitos fixos (*within*) indica que todos os fatores explicativos são significativos, com a taxa de desemprego (**unemp**) tendo um efeito negativo sobre a taxa de mortalidade (talvez porque aqueles que estão desempregados tem o rendimento limitado e dirigem menos), e renda per capita (**income**) um efeito positivo (como esperado).

2.3.2 Estimador de mínimos quadrados com variável *dummy*

Uma outra alternativa para estimar o modelo de efeito fixos representado pela Equação (2.20) é incluir um conjunto de variáveis *dummy* para cada indivíduo permitindo que cada um tenha seu próprio intercepto e que suas especificidades sejam capturadas por esse termo.

Seja \mathbf{y}_i e \mathbf{X}_i as T observações para a i -ésima unidade i , sendo a dimensão da matriz \mathbf{X}_i $T \times k$, $\boldsymbol{\beta}$ um vetor $k \times 1$ dos coeficientes, $\boldsymbol{\varepsilon}_i$ o vetor $T \times 1$ dos erros de cada unidade i e $\boldsymbol{\iota}$ um vetor $T \times 1$ em que seus elementos são todos iguais a 1 (um), ou seja, $\boldsymbol{\iota} = (1 \ 1 \ \dots \ 1)'$. Então,

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\iota} \alpha_i + \boldsymbol{\varepsilon}_i \quad (2.27)$$

Empilhando os n indivíduos:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\iota} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\iota} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\iota} \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix}$$

ou

$$\mathbf{y} = \begin{bmatrix} \mathbf{X} & \mathbf{d}_1 & \mathbf{d}_2 & \dots & \mathbf{d}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} + \boldsymbol{\varepsilon} \quad (2.28)$$

em que \mathbf{d}_i é uma variável *dummy* que indica o i -ésimo indivíduo. Pode-se agrupar as variáveis *dummy* de todos os indivíduos da amostra numa matriz \mathbf{D} de dimensão $nT \times n$ em que

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \dots & \mathbf{d}_n \end{bmatrix}'$$

Em seguida, o agrupamento de todas as nT linhas torna-se

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{D} \boldsymbol{\alpha} + \boldsymbol{\varepsilon} \quad (2.29)$$

em que \mathbf{y} tem dimensão $nT \times 1$, a matriz \mathbf{X} tem dimensão $nT \times k$. Este modelo dado pela Equação (2.29) é geralmente referido como modelo de mínimos quadrados de variáveis *dummy*, embora a parte “mínimos quadrados” se refira à técnica normalmente usada para estimá-lo e trata-se de um modelo de regressão clássico, não sendo necessários novos resultados para analisá-lo. Se n é pequeno o suficiente, então o modelo é estimado por mínimos quadrados ordinários com k regressores em \mathbf{X} e n colunas em \mathbf{D} como uma

regressão múltipla com $k + n$ parâmetros, e pode-se determinar o estimador de mínimos quadrados $\hat{\beta}$ como (Greene, 2008):

$$\begin{aligned}\hat{\beta} &= [\mathbf{X}'\mathbf{M}_D\mathbf{X}]^{-1} [\mathbf{X}'\mathbf{M}_D\mathbf{y}], \\ \mathbf{M}_D &= \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\end{aligned}\quad (2.30)$$

\mathbf{M}_D é uma matriz simétrica ($\mathbf{M}' = \mathbf{M}$), idempotente ($\mathbf{M}_D^2 = \mathbf{M}_D$) e devido à estrutura da matriz \mathbf{D} que tem as colunas ortogonais ($\mathbf{d}'_i\mathbf{d}_j = 0, \forall i \neq j$)

$$\begin{aligned}\mathbf{D}'\mathbf{D} &= \begin{pmatrix} \mathbf{d}'_1 \\ \mathbf{d}'_2 \\ \vdots \\ \mathbf{d}'_n \end{pmatrix} \begin{pmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \dots & \mathbf{d}_n \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{d}'_1\mathbf{d}_1 & 0 & \dots & 0 \\ 0 & \mathbf{d}'_2\mathbf{d}_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \mathbf{d}'_n\mathbf{d}_n \end{pmatrix} \\ &= \begin{pmatrix} T & 0 & \dots & 0 \\ 0 & T & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & T \end{pmatrix}\end{aligned}\quad (2.31)$$

então \mathbf{M}_D torna-se

$$\mathbf{M}_D = \begin{pmatrix} \mathbf{M}^0 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^0 & \mathbf{0} & \dots & \mathbf{0} \\ & & \dots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{M}^0 \end{pmatrix}\quad (2.32)$$

Cada matriz \mathbf{M}^0 na diagonal é dada por

$$\mathbf{M}^0 = \mathbf{I}_T - \frac{1}{T}\mathbf{u}\mathbf{u}'\quad (2.33)$$

\mathbf{M}_D define os resíduos obtidos pela regressão de \mathbf{y} nas n *dummies*.

Uma vez estimado β pode-se usar as equações normais para estimar α

$$\mathbf{D}'\hat{\mathbf{y}} = \mathbf{D}'\mathbf{X}\hat{\beta} + \mathbf{D}'\mathbf{D}\hat{\alpha}\quad (2.34)$$

Logo

$$\hat{\alpha} = [\mathbf{D}'\mathbf{D}]^{-1} \mathbf{D}'(\mathbf{y} - \mathbf{X}\hat{\beta})\quad (2.35)$$

Isso implica que para cada indivíduo i ,

$$\hat{\alpha}_i = \bar{y}_i - \bar{\mathbf{x}}'_i\hat{\beta}\quad (2.36)$$

Exemplo 4. *Mínimos quadrados com variável dummy*

Greene (1997) com o objetivo de estimar uma função de custo fornece um pequeno conjunto de dados com informações sobre custos e produção de 6 empresas diferentes, em 4 diferentes períodos de tempo (1955, 1960, 1965 e 1970). A estimação de mínimos quadrados com variável *dummy* no R para esse conjunto de dados é apresentada a seguir.

Inicialmente, considerando a firma 1 como categoria de base ou referência, inclui-se 5 variáveis *dummy* para representar as seis firmas:

$$D_{2i} = \begin{cases} 1, & \text{para a firma 2,} \\ 0, & \text{caso contrário.} \end{cases}, \quad D_{3i} = \begin{cases} 1, & \text{para a firma 3,} \\ 0, & \text{caso contrário.} \end{cases}, \quad D_{4i} = \begin{cases} 1, & \text{para a firma 4,} \\ 0, & \text{caso contrário.} \end{cases},$$

$$D_{5i} = \begin{cases} 1, & \text{para a firma 5,} \\ 0, & \text{caso contrário.} \end{cases}, \quad D_{6i} = \begin{cases} 1, & \text{para a firma 6,} \\ 0, & \text{caso contrário.} \end{cases}$$

O modelo de regressão pode ser descrito por:

$$\begin{aligned} \ln C_{it} &= \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \alpha_5 D_{5i} + \alpha_6 D_{6i} + \beta \ln x_{it} + \varepsilon_{it}, \\ i &= 1, 2, \dots, 6, \\ t &= 1, 2, 3, 4. \end{aligned} \tag{2.37}$$

em que $\ln C_{it}$ representa a variável resposta, $\ln x_{it}$ variável explicativa e u_{it} os erros, $D_{2i} = 1$ para a firma 2, 0 caso contrário; $D_{3i} = 1$ para a firma 3, 0 caso contrário e assim por diante. Especifica-se uma constante e cinco variáveis *dummy*, cada coeficiente da variável *dummy* seria igual à diferença entre o intercepto de seu indivíduo e o intercepto do indivíduo base no qual não foi especificado uma variável *dummy*. Os resultados do modelo a partir do R são:

Call:

```
lm(formula = lnc ~ lnx + factor(firma), data = greene)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.214606	-0.061549	-0.006332	0.068760	0.224034

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-2.69353	0.38279	-7.037	2.00e-06	***
lnx	0.67428	0.06113	11.030	3.61e-09	***
factor(firma)2	-0.21820	0.10520	-2.074	0.0536	.
factor(firma)3	0.25357	0.17167	1.477	0.1579	
factor(firma)4	0.55904	0.19829	2.819	0.0118	*

```

factor(firma)5  0.38269    0.19331    1.980    0.0642 .
factor(firma)6  0.79001    0.24369    3.242    0.0048 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 0.1246 on 17 degrees of freedom
Multiple R-squared:  0.9924,      Adjusted R-squared:  0.9897
F-statistic: 368.8 on 6 and 17 DF,  p-value: < 2.2e-16

```

O intercepto α_1 é o valor do intercepto da firma 1 e os outros coeficientes α representam quanto os valores de intercepto das outras firmas diferem da primeira (referência). Portanto, $\alpha_2 = -0,21820$ indica quanto o valor do intercepto da segunda firma difere de α_1 , a soma ($\alpha_1 + \alpha_2 = -2,69353 - 0,21820 = -2,91173$) dá o valor real do intercepto da firma 2, já o intercepto da firma 3 é dado por ($-2,69353 + 0,25357 = -2,43996$). Os valores de intercepto das outras firmas podem ser calculados de forma análoga. Uma maneira direta para verificar os interceptos é dada por:

Call:

```
lm(formula = lnc ~ lnx + factor(firma) - 1, data = greene)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.214606	-0.061549	-0.006332	0.068760	0.224034

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
lnx	0.67428	0.06113	11.030	3.61e-09	***
factor(firma)1	-2.69353	0.38279	-7.037	2.00e-06	***
factor(firma)2	-2.91173	0.43958	-6.624	4.30e-06	***
factor(firma)3	-2.43996	0.52869	-4.615	0.000247	***
factor(firma)4	-2.13449	0.55880	-3.820	0.001371	**
factor(firma)5	-2.31084	0.55325	-4.177	0.000632	***
factor(firma)6	-1.90351	0.60808	-3.130	0.006095	**

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 0.1246 on 17 degrees of freedom
Multiple R-squared:  0.9991,      Adjusted R-squared:  0.9987
F-statistic: 2582 on 7 and 17 DF,  p-value: < 2.2e-16

```

A diferença entre o modelo de regressão para dados agrupados (*pooled*) (MQO) e o modelo de mínimos quadrados com variável *dummy* (MQVD) pode-se observar graficamente (Figura 1).

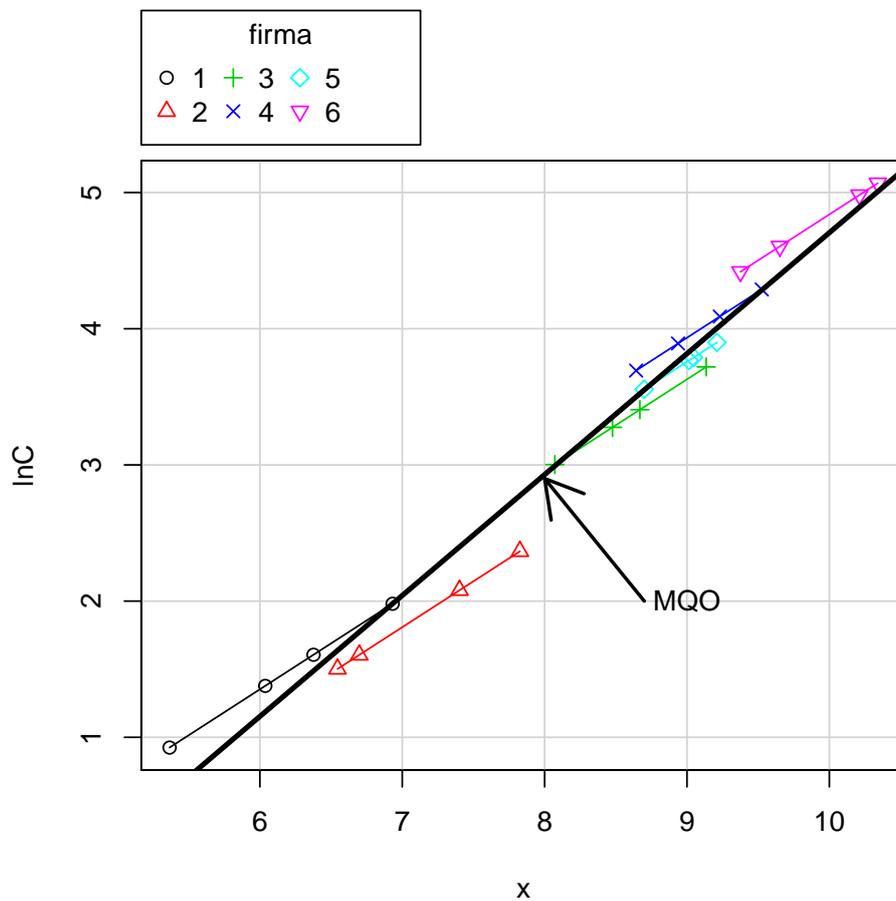
Exemplo 5. Comparação entre MQO e MQVD

Figura 1 – MQO versus MQVD

A Figura 1 apresenta o custo estimado para as 6 firmas separadamente (retas coloridas) através do métodos de mínimos quadrados com variáveis *dummy* (MQVD), bem como a estimação que considera os dados das 6 firmas agrupados (MQO) destacado no gráfico por uma seta, em que neste último, despreza-se os efeitos fixos individuais. Observa-se como a regressão com dados empilhados pode tornar tendenciosa a estimativa do coeficiente angular.

2.3.3 Estimador de primeiras diferenças

Finalmente, o último estimador para eliminar o efeito fixo individual α_i é o estimador de primeiras diferenças.

O estimador apresentado na subseção (2.3.1) foi obtido a partir da subtração do modelo original por $\bar{y}_i = \bar{x}_i' \beta + (\alpha_i + \bar{u}_i)$.

Alternativamente, uma outra transformação é subtrair o modelo original designado pela Equação (2.21) do modelo defasado em 1 período representado pela Equação (2.38),

conforme demonstrado a seguir

$$y_{i,t-1} = \mathbf{x}_{i,t-1}\boldsymbol{\beta} + \alpha_i + u_{i,t-1}, \quad (2.38)$$

Então,

$$\begin{aligned} y_{it} - y_{i,t-1} &= (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + (\alpha_i - \alpha_i) + (u_{it} - u_{i,t-1}), \\ \Delta y_{it} &= \Delta \mathbf{x}'_{it}\boldsymbol{\beta} + \Delta u_{it}, \quad t = 2, 3, \dots, T \end{aligned} \quad (2.39)$$

Aplicando MQO em (2.38), o estimador de primeiras diferenças $\hat{\boldsymbol{\beta}}_{FD}$

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{FD} &= \left[\sum_{i=1}^n \sum_{t=2}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i,t-1})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i,t-1})' \right]^{-1} \sum_{i=1}^n \sum_{t=2}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i,t-1})(y_{it} - \bar{y}_{i,t-1}) \\ &= \left[\sum_{i=1}^n \sum_{t=2}^T (\Delta \mathbf{x}_{it})(\Delta \mathbf{x}_{it})' \right]^{-1} \sum_{i=1}^n \sum_{t=2}^T (\Delta \mathbf{x}_{it})(\Delta y_{it}). \end{aligned} \quad (2.40)$$

Nota-se que existem apenas $n(T-1)$ observações nesta regressão. [Cameron e Trivedi \(2005\)](#) destacam que um erro fácil de se cometer ao implementar esse modelo é empilhar todas as nT observações e depois subtrair a primeira defasagem. Fazendo isso, apenas a observação $(1, 1)$ é descartada, enquanto que todas as T primeiras observações $(i, 1)$, $i = 1, \dots, n$, devem ser descartadas após a diferenciação.

Exemplo 6. (Primeira diferença)

Oneway (individual) effect First-Difference Model

Call:

```
plm(formula = frate ~ beertax + spirits + unemp + income, data = pfat,
     model = "fd")
```

Balanced Panel: n=48, T=7, N=336

Observations used in estimation: 288

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-0.5380	-0.1050	-0.0029	0.1020	0.5840

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
(intercept)	-0.04422662	0.01970730	-2.2442	0.02559 *
beertax	0.04956931	0.27263490	0.1818	0.85586
spirits	0.31626682	0.16759423	1.8871	0.06017 .

```

unemp      -0.00243779  0.01190617 -0.2047   0.83791
income     0.00018492  0.00004171  4.4336  1.327e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    11.213
Residual Sum of Squares: 10.123
R-Squared:              0.097231
Adj. R-Squared:        0.084471
F-statistic: 7.61999 on 4 and 283 DF, p-value: 7.6406e-06

```

2.4 Modelo de efeitos aleatórios

Na Seção (2.3) discutiu-se a estimativa de modelos de regressão linear quando os efeitos específicos individuais α_i varia entre os indivíduos i , mas são tratados como fixos e constantes ao longo do tempo. Nessa seção, assume-se novamente que todas as diferenças individuais são capturadas pelo intercepto α_i , mas tanto os efeitos individuais α_i como os erros ε_{it} serão tratados como variáveis aleatórias ao invés de fixos. Este modelo também é conhecido como modelo de componentes de variância (Gujarati e Porter, 2011).

Formalmente o modelo é dado por:

$$\begin{aligned}
 y_{it} &= \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it}, \quad i = 1, \dots, n \text{ e } t = 1, \dots, T \\
 u_{it} &= \alpha_i + \varepsilon_{it}
 \end{aligned}
 \tag{2.41}$$

o termo de erro composto u_{it} consiste de dois elementos, um componente específico individual aleatório (α_i) representando fatores não observáveis que afetam y e que não variam ao longo do tempo; e ε_{it} o erro aleatório da regressão que representa outros fatores que influenciam y , mas que variam ao longo do tempo e indivíduos; \mathbf{x}_{it} consiste no conjunto de variáveis explicativas; e $\boldsymbol{\beta}$ parâmetros a serem estimados. Como o erro composto u_{it} é formado por dois ou mais elementos de erro, o modelo de efeitos aleatórios é frequentemente chamado de *modelo de componente dos erros*.

Segundo Greene (2008) os pressupostos básicos do modelo de efeitos aleatórios são a ausência de correlação entre os efeitos individuais α_i e os regressores do modelo \mathbf{x}_{it} e ortogonalidade entre os efeitos individuais e as variáveis explicativas, isto é,

$$\begin{aligned}
 \text{Cov}[x_{it}, \alpha_i] &= 0, \quad t = 1, \dots, T \\
 \text{E}(\alpha_i | x_{i1}, x_{i1}, \dots, x_{iT}) &= 0, \forall i
 \end{aligned}
 \tag{2.42}$$

As diferenças aleatórias individuais α_i são chamados de efeitos aleatórios e tem média zero, não estão correlacionados entre os indivíduos, e tem variância constante σ_α^2 ,

de modo que

$$\begin{aligned} E(\alpha_i) &= 0, \\ \text{Cov}(\alpha_i, \alpha_j) &= 0 \quad \text{se } i \neq j, \\ \text{Var}(\alpha_i) &= \sigma_\alpha^2, \quad \forall i \end{aligned} \tag{2.43}$$

As suposições habituais do termo de erro ε_{it} são que tem média zero, variância constante igual σ_ε^2 , não estão correlacionados ao longo do tempo nem entre si, ou seja,

$$\begin{aligned} E(\varepsilon_{it}) &= 0, \\ \text{var}(\varepsilon_{it}) &= \sigma_\varepsilon^2, \\ \text{Cov}(\varepsilon_{it}, \varepsilon_{js}) &= 0 \quad \text{para } i \neq j \text{ ou } t \neq s. \end{aligned} \tag{2.44}$$

Outros pressupostos sobre o modelo de efeito aleatório são que o termo de erro não é correlacionado com as variáveis explicativas, assume-se que os efeitos individuais α_i não são correlacionados com o termo de erro de regressão ε_{it} e não correlacionado com as variáveis explicativas, ou seja:

$$\begin{aligned} \text{Cov}(\varepsilon_{it} | x_{i1}, x_{i1}, \dots, x_{iT}) &= 0, \quad \forall i, t, \\ \text{Cov}(\alpha_i, \varepsilon_{it}) &= 0 \\ \text{Cov}(\alpha_i | x_{i1}, x_{i1}, \dots, x_{iT}) &= 0, \quad \forall i. \end{aligned} \tag{2.45}$$

A partir das suposições sobre α_i e ε_{it} , podemos derivar as propriedades do termo de erro composto $u_{it} = \alpha_i + \varepsilon_{it}$, que tem média zero

$$\begin{aligned} E(u_{it}) &= E(\alpha_i + \varepsilon_{it}) \\ &= E(\alpha_i) + E(\varepsilon_{it}) \\ &= 0 + 0 \\ &= 0 \end{aligned} \tag{2.46}$$

e variância constante e homoscedástica:

$$\begin{aligned} \sigma_u^2 &= \text{Var}(u_{it}) = \text{Var}(\alpha_i + \varepsilon_{it}) \\ &= \text{Var}(\alpha_i) + \text{Var}(\varepsilon_{it}) + \underbrace{2\text{Cov}(\alpha_i, \varepsilon_{it})}_{=0} \\ &= \text{Var}(\alpha_i) + \text{Var}(\varepsilon_{it}) \\ &= \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{aligned} \tag{2.47}$$

Essas são as propriedades usuais do termo de erro, as diferenças surgem quando considera-se as correlações entre os termos de erro composto u_{it} . Existem várias correlações que podem ser consideradas.

1. A correlação entre dois indivíduos, i e j , no mesmo período de tempo, t . A covariância para este caso é dada por

$$\begin{aligned}
 \text{Cov}(u_{it}, u_{jt}) &= \text{E}(u_{it}u_{jt}) \\
 &= \text{E}[(\alpha_i + \varepsilon_{it})(\alpha_j + \varepsilon_{jt})] \\
 &= \text{E}(\alpha_i\alpha_j) + \text{E}(\alpha_i\varepsilon_{jt}) + \text{E}(\varepsilon_{it}\alpha_j) + \text{E}(\varepsilon_{it}\varepsilon_{jt}) \\
 &= 0 + 0 + 0 + 0 \\
 &= 0
 \end{aligned} \tag{2.48}$$

2. A correlação entre erros do mesmo indivíduo (i) em diferentes momentos, t e s . A covariância para este caso é dada por

$$\begin{aligned}
 \text{Cov}(u_{it}, u_{is}) &= \text{E}(u_{it}u_{is}) \\
 &= \text{E}[(\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{is})] \\
 &= \text{E}(\alpha_i^2) + \text{E}(\alpha_i\varepsilon_{is}) + \text{E}(\varepsilon_{it}\alpha_i) + \text{E}(\varepsilon_{it}\varepsilon_{is}) \\
 &= \sigma_\alpha^2 + 0 + 0 + 0 \\
 &= \sigma_\alpha^2
 \end{aligned} \tag{2.49}$$

3. A correlação entre erros para diferentes indivíduos em diferentes períodos de tempo. A covariância para este caso é dada por

$$\begin{aligned}
 \text{Cov}(u_{it}, u_{js}) &= \text{E}(u_{it}u_{js}) \\
 &= \text{E}[(\alpha_i + \varepsilon_{it})(\alpha_j + \varepsilon_{js})] \\
 &= \text{E}(\alpha_i\alpha_j) + \text{E}(\alpha_i\varepsilon_{js}) + \text{E}(\varepsilon_{it}\alpha_j) + \text{E}(\varepsilon_{it}\varepsilon_{js}) \\
 &= 0 + 0 + 0 + 0 \\
 &= 0
 \end{aligned} \tag{2.50}$$

Com efeito, a partir das suposições anteriores, tem-se:

$$\begin{aligned}
 \text{Cov}(u_{it}, u_{js}) &= \sigma_\alpha^2 + \sigma_\varepsilon^2, \quad \text{se } i = j, t = s \\
 &= \sigma_\alpha^2 \quad \text{se } i = j, t \neq s \\
 &= 0 \quad \text{se } i \neq j.
 \end{aligned} \tag{2.51}$$

Conforme (2.47) observa-se que o termo de erro u_{it} é homocedástico. Entretanto, por (2.49) demonstra-se que u_{it} e u_{is} ($t \neq s$) são correlacionadas, isto é, os termos de erros de uma dada unidade de corte transversal estão correlacionadas em dois pontos diferentes de tempo. Segundo Matyas (2008) a presença do efeito individual no erro induz para cada termo individual alguma correlação serial entre os períodos de tempo, os autores ressaltam que esta correlação serial não depende do intervalo de tempo entre as duas observações, contrariando o padrão usual de correlação serial em modelo de séries temporais.

O coeficiente de correlação denotado por $\text{Corr}(u_{it}, u_{is})$ entre os erros é dada por:

$$\text{Corr}(u_{it}, u_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2} \quad (2.52)$$

para $t \neq s$, em que $\sigma_\alpha^2 = \text{Var}(\alpha_i)$ e $\sigma_\varepsilon^2 = \text{Var}(\varepsilon_{it})$.

Por (2.52) verifica-se que para qualquer unidade de corte transversal dada, o valor da correlação entre dois termos de erro, em períodos diferentes, se mantém inalterado. Além disso, a estrutura de correlação para todos os indivíduos permanece a mesma para todas as unidades *cross section*, ou seja é, é idêntica para todos os indivíduos. De acordo com Gujarati e Porter (2011) caso não seja considerada essa estrutura de correlação e o modelo representado em (2.41) for estimado por MQO, os estimadores resultantes serão ineficientes. Para este autor o método mais adequado, neste caso, é o métodos dos *mínimos quadrados generalizados* (MQG), que será apresentado em (2.4.1).

Empilhando todas as observações relacionadas ao indivíduo i , pode-se escrever:

$$\underset{(T \times 1)}{\mathbf{y}_i} = \underset{(T \times (k+1))}{\mathbf{X}_i} \times \underset{((k+1) \times 1)}{\boldsymbol{\beta}} + \underset{(T \times 1)}{\mathbf{u}_i}, \quad i = 1, 2, \dots, n, \quad (2.53)$$

em que $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$ representa o vetor de observações da variável dependente para o i -ésimo indivíduo; \mathbf{X}_i a matriz de observações das variáveis independentes e $\mathbf{u}'_i = (u_{i1}, \dots, u_{iT})$ o vetor de erros para cada indivíduo. A presença de α_i gera correlações de u_{it} ao longo do tempo para um determinado indivíduo, embora u_{it} permaneça não correlacionado através dos indivíduos. Dadas as premissas definidas para esse modelo, o vetor de erros segue as seguintes propriedades:

$$\begin{aligned} \text{E}(u_i | x_{i1}, x_{i2}, \dots, x_{iT}) &= 0, \quad \forall i, \\ \text{Var}(u_i | x_{i1}, x_{i2}, \dots, x_{iT}) &= \boldsymbol{\Sigma} \quad \forall i, \end{aligned}$$

com a matriz de variâncias e covariâncias denotada por $\boldsymbol{\Sigma}$

$$\begin{aligned} \underset{(T \times T)}{\boldsymbol{\Sigma}} &= \begin{bmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \dots & \sigma_\alpha^2 & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{bmatrix} \\ &= \sigma_\varepsilon^2 \mathbf{I}_T + \sigma_\alpha^2 (\boldsymbol{\iota}_T \boldsymbol{\iota}'_T) \end{aligned}$$

em que \mathbf{I}_T é uma matriz identidade de ordem T e $\boldsymbol{\iota}_T = (1, \dots, 1)'$ é um vetor $T \times 1$ cujos elementos são todos iguais a um. Observa-se que a matriz $\boldsymbol{\Sigma}$ é simétrica cuja diagonal é composta pelas variâncias e os elementos fora da diagonal pelas covariâncias.

Empilhando o conjunto de vetores das observações individuais,

$$\mathbf{y} = (y_{11}, y_{12}, \dots, y_{1T}, \dots, y_{n1}, y_{n2}, \dots, y_{nT})'$$

pode-se escrever o modelo como:

$$\underset{(nT \times 1)}{\mathbf{y}} = \underset{(nT \times (k+1))}{\mathbf{X}} \times \underset{((k+1) \times 1)}{\boldsymbol{\beta}} + \underset{(nT \times 1)}{\mathbf{u}} \quad (2.54)$$

Como as observações i e j são independentes, a matriz de covariância $\boldsymbol{\Omega}$ para as nT observações é dada por:

$$\begin{aligned} \mathbf{E}(\mathbf{u}\mathbf{u}') = \underset{(nT \times nT)}{\boldsymbol{\Omega}} &= \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & & \dots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{\Sigma} \end{bmatrix} \\ &= \mathbf{I}_n \otimes \boldsymbol{\Sigma} \\ &= \mathbf{I}_n \otimes [\sigma_\varepsilon^2 \mathbf{I}_T + \sigma_\alpha^2 (\boldsymbol{\iota}_T \boldsymbol{\iota}_T')] \\ &= \mathbf{I}_n \otimes [\sigma_\varepsilon^2 (\mathbf{Q}_T + \mathbf{B}_T) + \sigma_\alpha^2 (T \times \mathbf{B}_T)] \end{aligned}$$

seja $\mathbf{Q}_T = \mathbf{I}_T - \mathbf{B}_T$ e $\mathbf{B}_T = (1/T)\boldsymbol{\iota}_T \boldsymbol{\iota}_T'$. Portanto,

$$\begin{aligned} \boldsymbol{\Omega} &= \mathbf{I}_n \otimes [\sigma_\varepsilon^2 (\mathbf{Q}_T + \mathbf{B}_T) + \sigma_\alpha^2 (T \times \mathbf{B}_T)] \\ &= \sigma_\varepsilon^2 \mathbf{I}_{nT} + T\sigma_\alpha^2 \mathbf{B} \end{aligned} \quad (2.55)$$

ou equivalente

$$= \sigma_\varepsilon^2 \mathbf{Q}_T + (T\sigma_\alpha^2 + \sigma_\varepsilon^2) \mathbf{B}_T.$$

Aqui \otimes representa o produto de Kronecker ou produto direto, \mathbf{I}_n matriz identidade de dimensão n , $\sigma_\varepsilon^2 = \text{Var}(\varepsilon_{it})$, $\sigma_\alpha^2 = \text{Var}(\alpha_i)$, \mathbf{I}_T matriz identidade de ordem T , $\boldsymbol{\iota}_T = (1, \dots, 1)'$ é um vetor $T \times 1$ cujos elementos são todos iguais a um.

Verifica-se que a matriz de variâncias-covariâncias é idêntica para todos os indivíduos. A presença de correlação entre os erros do mesmo indivíduo em períodos de tempos diferentes faz com que o método de mínimos quadrados não seja indicado para estimar os coeficientes do modelo de efeito aleatório, e neste caso o estimador de mínimos quadrados generalizados (MQG) apresenta-se como o mais apropriado por permitir obter estimadores não viesados e consistentes (Greene, 2008).

Wooldridge (2008) defende que o principal determinante para decidir entre o modelo de efeitos fixos e aleatórios é verificar se existe correlação entre o efeito não observado α_i e as variáveis explicativas. Para isso, esse autor recomenda o uso do teste de Hausman.

2.4.1 Estimador de mínimos quadrados generalizados

A matriz das variâncias-covariâncias encontradas para este modelo implica a necessidade de se utilizar o método de estimação de Mínimos Quadrados Generalizados (MQG),

uma vez que a aplicação do Método de Mínimos Quadrados conduziria a estimadores viesados.

Considera-se a forma geral do modelo:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}, \quad \text{com } E(\mathbf{U}\mathbf{U}') = \boldsymbol{\Omega}.$$

O estimador de mínimos quadrados generalizados produz estimativas de parâmetros eficientes de $\boldsymbol{\beta}$, σ_α^2 e σ_ε^2 baseado na matriz conhecida de variância-covariância $\boldsymbol{\Omega}$. O estimador MQG eficiente para $\boldsymbol{\beta}$ é dado por

$$\hat{\boldsymbol{\beta}}_{MQG} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Y} \quad (2.56)$$

e

$$\text{Var}(\hat{\boldsymbol{\beta}}_{MQG}) = \sigma_\varepsilon^2 (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \quad (2.57)$$

A seguir apresenta-se um exemplo de aplicação do método de efeitos aleatórios.

Exemplo 7. Efeitos aleatórios

Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)

Call:

```
plm(formula = frate ~ beertax + spirits + unemp + income, data = pfat,
     model = "random")
```

Balanced Panel: n=48, T=7, N=336

Effects:

```
                var std.dev share
idiosyncratic 0.02458 0.15679 0.132
individual    0.16236 0.40294 0.868
theta: 0.8545
```

Residuals :

```
    Min. 1st Qu.  Median 3rd Qu.    Max.
-0.4820 -0.1070 -0.0190  0.0763  0.8340
```

Coefficients :

```
                Estimate Std. Error t-value Pr(>|t|)
(Intercept) 2.0725e+00 3.7933e-01 5.4636 9.186e-08 ***
beertax      5.2858e-02 1.1907e-01 0.4439 0.6574
```

```
spirits      2.8937e-01  6.3780e-02  4.5369  7.994e-06 ***
unemp        -4.9694e-02  9.8597e-03 -5.0401  7.672e-07 ***
income       -1.4523e-05  1.9452e-05 -0.7466   0.4558
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    12.862
Residual Sum of Squares: 11.232
R-Squared:               0.12676
Adj. R-Squared:          0.11621
F-statistic: 12.0122 on 4 and 331 DF, p-value: 3.9765e-09
```

Os modelos tratados nesse Capítulo serão adaptados para o contexto bayesiano no Capítulo 4. Portanto, faz-se necessário uma breve revisão dos principais conceitos envolvidos na inferência bayesiana, apresentados a seguir no Capítulo 3.

3 Inferência Bayesiana

Neste Capítulo serão descritos os conceitos básicos sobre a teoria de inferência bayesiana. Na seção inicial, será descrito o teorema fundamental para o processo bayesiano. A seção seguinte trata dos princípios gerais da inferência bayesiana. A Seção (3.3) apresenta os principais elementos que regem essa abordagem, entre eles a distribuição a priori descrita na Seção (3.4). Finalmente, na Seção (3.5) são mostradas as estimativas bayesianas pontual e intervalar.

O problema fundamental da estatística é a inferência. Dados são coletados e a partir deles deseja-se fazer declarações (inferências) sobre uma ou mais características desconhecidas do mecanismo (ou processo) que deu origem aos dados observados (O'Hagan, 1994). A inferência estatística é um conjunto de técnicas que objetiva estudar a população através de evidências fornecidas por uma amostra. É a amostra que contém os elementos que podem ser observados e, a partir daí, quantidades de interesse podem ser medidas (Magalhães e Lima, 2013). Portanto, essa técnica lida com problemas de tirar conclusões sobre quantidades não observadas a partir de dados numéricos e para isso utiliza duas abordagens: a clássica (ou frequentista) e a bayesiana.

Segundo Paulino, Turkman e Murteira (2003) um aspecto importante da inferência estatística clássica consiste em reconhecer a variabilidade que se verifica de amostra para amostra, ou seja, para estabelecer inferências, os dados observados formam apenas um dos muitos conjuntos que poderiam ter sido obtidos nas mesmas circunstâncias. Com isso, o processo gerador dos dados é possivelmente controlado por um conjunto de parâmetros que pode ser representado por uma distribuição de probabilidades. Os dados observados são uma realização de uma variável aleatória X ou de um conjunto de variáveis aleatórias $X = (X_1, X_2, \dots, X_n)$ com uma função de distribuição F_0 que representa a variabilidade ou incerteza da observação X . Essa função de distribuição F_0 não é perfeitamente conhecida. Contudo, em geral, existe algum conhecimento inicial sobre a natureza do processo gerador dos dados que leva a proposição de uma família de distribuições \mathcal{F} a que pertence F_0 , denominada de modelo estatístico. Formalmente, define-se o modelo estatístico para X como (Paulino, Turkman e Murteira, 2003):

$$\mathcal{F} = \{f(x|\theta) : \theta \in \Theta\}, \quad x \in \mathcal{X}$$

em que θ representa o parâmetro, \mathcal{X} corresponde ao espaço amostral associado ao experimento e Θ é chamado espaço paramétrico. O objetivo consiste em encontrar o valor do parâmetro de interesse θ_0 e com isso determinar a função distribuição F_0 .

Logo, a inferência clássica baseia-se no princípio da repetibilidade, o qual uma vez determinado o modelo estatístico, conclusões acerca do parâmetro de interesse θ_0 são

feitas a partir da amostra observada, considerando-se as possíveis variações dos valores observados quando da coleta de diferentes amostras.

A inferência bayesiana parte da noção subjetiva ao utilizar a probabilidade para quantificar o grau de incerteza acerca de quantidades de interesse não observadas. Informalmente, pode-se definir a probabilidade subjetiva como a crença que o observador do experimento tem na ocorrência do evento de interesse. Assim, a inferência bayesiana combina toda a informação subjetiva disponível referente a um problema, com a informação proveniente dos dados observados, através de declarações probabilísticas via teorema de Bayes.

3.1 Teorema de Bayes

O Teorema de Bayes é definido como (Magalhães e Lima, 2013; Casella, 2001):

Definição 3. *Supondo que os eventos A_1, A_2, \dots, A_n estão em $(\Omega, \mathcal{F}, \mathcal{P})$, formam uma partição de Ω e todos têm probabilidade positiva. Seja B um evento qualquer com $P(B) > 0$. Então, para todo $j = 1, 2, \dots, n$, tem-se que*

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \quad (3.1)$$

Segundo Magalhães e Lima (2013) uma interpretação dessa fórmula é supor que A_i ($i = 1, \dots, n$) represente uma possível causa do resultado de um experimento aleatório com $P(A_i)$, $i = 1, \dots, n$, sendo as probabilidades à priori, realizado o experimento e obtido um resultado B , o teorema de Bayes indica como recalculas as probabilidades das causas, representadas por $P(A_i|B)$, $i = 1, \dots, n$, que são denominadas probabilidades à posteriori e podem ser usadas para avaliar o quanto cada causa A_i é responsável pela ocorrência do evento B .

O'Hagan (1994) considera que o Teorema de Bayes pode ser entendido como a fórmula de atualização da probabilidade a priori para a posterior a partir da multiplicação pela razão $P(B|A)/P(A)$. Portanto, este teorema descreve como a probabilidade muda conforme se obtém novas informações.

Supondo os eventos A e B . Pode-se expressar A como

$$A = AB \cup AB^c$$

pois, para que um resultado esteja em A , ele deve estar em A e B ou em A mas não em B (ver Figura 2). A área sombreada corresponde a $AB = (A \cap B)$, a área tracejada $AB^c = A \cap B^c$ denota o complemento ou negação de B dado por $P(B^c) = 1 - P(B)$.

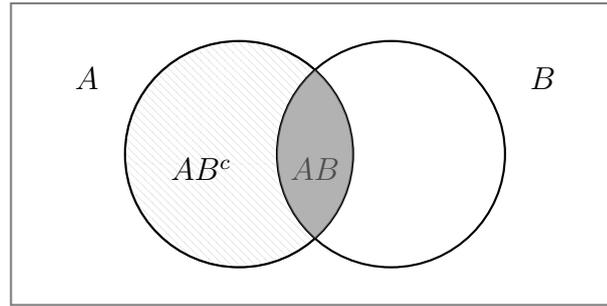


Figura 2 – Evento A

Sabe-se que os eventos AB e AB^c são mutuamente exclusivos¹, tem-se que a ocorrência do evento A poderá aumentar a probabilidade do B caso $P(A|B) > P(A)$. Utilizando a lei de probabilidade total tem que:

$$\begin{aligned} P(A) &= P(AB) \cup P(AB^c) \\ &= P(A \cap B) + P(A \cap B^c) \\ &= P(A|B)P(B) + P(A|B^c)P(B^c) \end{aligned} \quad (3.2)$$

Segundo Ross (2014) a Equação (3.2) diz que a probabilidade do evento A é uma média ponderada da probabilidade condicional de A dado que B ocorreu e da probabilidade condicional de A dado que B não ocorreu, com cada probabilidade condicional recebendo um maior peso quanto mais provável for a ocorrência do evento ao qual está relacionada.

Substituindo $P(B) = 1 - P(B^c)$ em (3.2),

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|B^c)P(B^c) \\ P(A) &= P(A|B)[1 - P(B^c)] + P(A|B^c)P(B^c) \\ P(A) &= P(A|B) - P(A|B)P(B^c) + P(A|B^c)P(B^c) \\ P(A|B) - P(A) &= [P(A|B) - P(A|B^c)]P(B^c) \end{aligned} \quad (3.3)$$

Assume-se que $P(B^c) > 0$ (caso contrário B é um evento certo, e sua probabilidade não seria de interesse), $P(A|B) > P(A)$ se e somente se $P(A|B) > P(A|B^c)$.

O teorema de Bayes pode ser generalizado supondo que B_1, B_2, \dots, B_n sejam eventos mutuamente exclusivos. Considera-se também que A tenha ocorrido e que se esteja interessado em determinar qual dos B_r eventos ocorreu. Então

$$\begin{aligned} P(B_r|A) &= \frac{P(B_r \cap A)}{P(A)} \\ &= \frac{P(B_r)P(A|B_r)}{P(A)} \\ &= \frac{P(B_r)P(A|B_r)}{\sum_r P(B_r)P(A|B_r)}. \end{aligned} \quad (3.4)$$

¹ Se $A \cap B = \emptyset$, então se diz que A e B são *mutuamente exclusivos*.

Assim, a Equação (3.4) é uma generalização de (3.1). Pode-se pensar no evento B como um conjunto de hipóteses, das quais uma e apenas uma é verdadeira. Se a hipótese r for verdadeira é dizer que o evento B ocorre. Observando o evento A altera as probabilidades anteriores $P(B_r)$ para a probabilidade posterior $P(B_r|A)$. Ressalta-se que as probabilidades posteriores somam um, uma vez que uma e apenas uma hipótese é verdadeira. O denominador $P(A)$ em (3.4) é uma média ponderada das probabilidades $P(A|B_r)$, em que os pesos $P(B_r)$ somam um. A ocorrência do evento A aumenta a probabilidade de B_r se $P(A|B_r)$ for maior que todas as médias $P(A|B_r)$. (O’Hagan, 1994)

Exemplo 8. (*Teorema de Bayes*)

Os atletas de elite são frequentemente testados quanto à presença de substâncias que melhoram o desempenho. Suponha que um desses testes tenha uma taxa de falso negativo de 0,05 e uma taxa de falso positivo de 0,10. Trabalhos anteriores sugerem que cerca de 3% do grupo de sujeitos usa um determinado medicamento proibido. Seja U o evento em que “o sujeito usa a substância proibida”; U^c denota o evento contrário. Supondo sorteado aleatoriamente um atleta para submissão ao teste, e que retorna um teste positivo, em que + representa este evento. Qual a probabilidade do atleta ter usado a substância? (Jackman, 2009)

O teste ideal seria aquele que fornecesse resultados sempre corretos em todos os atletas em que fosse aplicado, ou seja, positivo para a presença de substâncias que melhoram o desempenho e negativo para a ausência. Ocorre que existe a possibilidade de erro, portanto as situações possíveis são:

1. o teste é positivo e o atleta usou a substância: verdadeiro-positivo;
2. o teste é positivo, mas o atleta não usou a substância: falso-positivo;
3. o teste é negativo, mas o atleta usou a substância: falso-negativo;
4. o teste é negativo e o paciente não usou a substância: verdadeiro-negativo.

Defina U como sendo o evento em que o atleta usa a substância proibida. Então, U^c é o evento em que o atleta não usa a substância proibida. Seja “-” o evento em que o teste é negativo, “+” o evento em que o teste é positivo. A probabilidade requerida pode ser denotada pelo probabilidade condicional $P(U|+)$. Da definição do problema, as seguintes informações encontram-se disponíveis:

$$\begin{aligned} P(U) &= 0,03 & P(U^c) &= 0,97 \\ P(+|U) &= 0,95 & P(+|U^c) &= 0,10 \\ P(-|U) &= 0,05 & P(-|U^c) &= 0,90 \end{aligned}$$

Observa-se que a $P(U|+)$ não é uma probabilidade condicional conhecida. Entretanto, pode-se determinar usando a regra da multiplicação da probabilidade:

$$\begin{aligned} P(+ \cap U) &= P(U)P(+|U) \\ &= (0,03)(0,95) \\ &= 0,0285 \end{aligned}$$

e

$$\begin{aligned} P(+ \cap U^c) &= P(U^c)P(+|U^c) \\ &= (0,97)(0,10) \\ &= 0,097 \end{aligned}$$

Essas duas probabilidades estão apresentadas no diagrama de árvore para o problema na Figura (3).

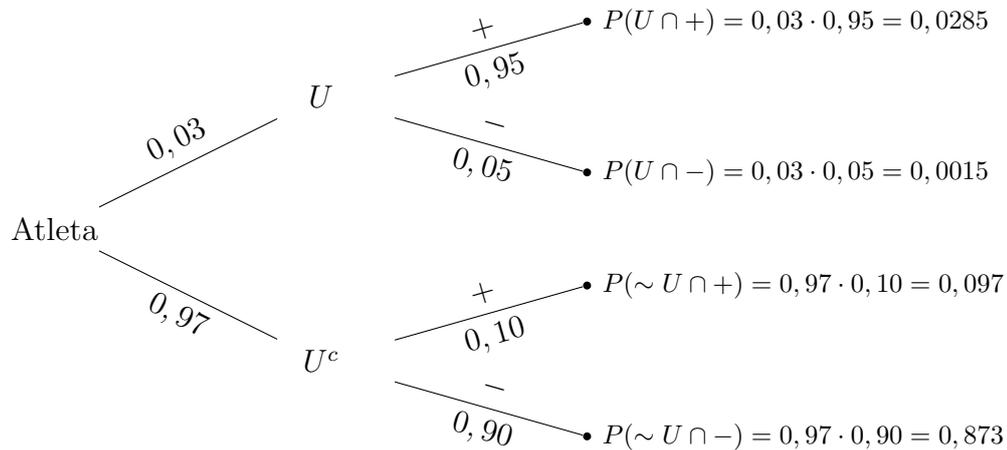


Figura 3 – Diagrama de árvore para o Exemplo 8

O evento “+” é a união de dois eventos mutuamente exclusivos, $P(+ \cap U)$ e $P(+ \cap U^c)$. Então, aplicando a regra de adição de probabilidade, tem-se:

$$\begin{aligned} P(+) &= P(+ \cap U) \cup P(+ \cap U^c) \\ &= 0,0285 + 0,097 \\ &= 0,1255 \end{aligned}$$

Do Teorema de Bayes,

$$\begin{aligned} P(U|+) &= \frac{P(+|U)P(U)}{P(+|U)P(U) + P(+|U^c)P(U^c)} \\ &= \frac{0,0285}{0,0285 + 0,097} \\ &\approx 0,23 \end{aligned}$$

Utilizando o Teorema de Bayes as probabilidades foram revisadas, assim a probabilidade do atleta ter usando a substância passou de $P(U) = 0,03$ para $P(U|+) = 0,23$.

3.2 Princípios gerais da inferência bayesiana

Ao realizar inferências sobre um parâmetro desconhecido θ um pesquisador utiliza as informações contidas em uma amostra. Assim, faz-se necessário o entendimento dos princípios que estabelecem a forma como os dados da amostra devem afetar as inferências, ou seja, princípios que dizem respeito aos aspectos dos dados e do modelo que devem ser considerados relevantes. Na inferência bayesiana os três princípios básicos são: verossimilhança, suficiência e condicionalidade.

3.2.1 Princípio da verossimilhança

A função de verossimilhança tem papel fundamental, quer na inferência clássica, quer na inferência bayesiana, como veículo portador da informação dada pela amostra. O princípio da verossimilhança sustenta que toda a informação dada pela amostra ou pela experiência está contida na função de verossimilhança (Paulino, Turkman e Murteira, 2003).

Segundo Casella (2001) o princípio da verossimilhança estabelece que se \mathbf{x} e \mathbf{y} são dois pontos amostrais tais que $L(\theta|\mathbf{x})$ é proporcional a $L(\theta|\mathbf{y})$, isto é, existe uma constante $C(\mathbf{x}, \mathbf{y})$ de modo que

$$L(\theta|\mathbf{x}) = C(\mathbf{x}, \mathbf{y})L(\theta|\mathbf{y}) \quad \text{para todo } \theta, \quad (3.5)$$

então, as conclusões obtidas a partir de \mathbf{x} e \mathbf{y} deveriam ser idênticas.

Observa-se que a constante $C(\mathbf{x}, \mathbf{y})$ em (3.5) pode assumir outros valores para diferentes pares (\mathbf{x}, \mathbf{y}) , mas $C(\mathbf{x}, \mathbf{y})$ não depende de θ . No caso de $C(\mathbf{x}, \mathbf{y}) = 1$, o princípio da verossimilhança define que se dois pontos amostrais resultam na mesma função de verossimilhança, então eles contêm as mesmas informações sobre θ .

Exemplo 9. *Exemplo retirado de Paulino, Turkman e Murteira (2003). Considera-se uma sucessão de lançamentos de uma moeda, independentes e condicionados por θ que designa a probabilidade de observar “cara”; supondo que em um dado momento se chega ao seguinte resultado ou amostra,*

$$x = \{R, C, R, R, C, C, R, C, C, C\},$$

em que R designa “cara” e C designa “coroa”. Entre outras possibilidades, este resultado poderia ter sido gerado a partir dos seguintes processos experimentais:

1. Lançar a moeda 10 vezes e contabilizar o número de caras ($X \sim \text{Binomial}(10, \theta)$).
2. Lançar a moeda até obter um total de 4 caras, contando o número de lançamentos ($Y \sim \text{Binomial-Negativa}(4, \theta)$).

Portanto, para o caso 1 a função de verossimilhança é dada por:

$$\begin{aligned} p(x|\theta) &= \frac{10!}{x!(10-x)!} \theta^x (1-\theta)^{10-x} \\ &\propto \theta^x (1-\theta)^{10-x} \end{aligned}$$

Em que o símbolo \propto significa “proporcional”. Para os dados deste Exercício, tem-se que:

$$p(4|\theta) \propto \theta^4 (1-\theta)^6.$$

No caso 2, a a função de verossimilhança é dada por:

$$\begin{aligned} p(y|\theta) &= \frac{(y-1)!}{(y-4)!3!} (1-\theta)^{y-4} \theta^4 \\ &\propto (1-\theta)^{y-4} \theta^4 \end{aligned}$$

Substituindo os dados do Exercício 9, tem-se que:

$$p(10|\theta) \propto \theta^4 (1-\theta)^6.$$

Adotando o princípio de verossimilhança, conclui-se que:

$$p(x|\theta) \propto p(y|\theta)$$

Portanto, sob a mesma priori para θ , a posteriori obtida a partir de x seria igual a posteriori obtida para y .

3.2.2 Princípio da suficiência

Uma estatística suficiente para um parâmetro θ é uma estatística que, de certa maneira, capta todas as informações sobre θ contidas na amostra (Casella, 2001).

Definição 4. *Suponha \mathbf{X} com função (de densidade) de probabilidade $p(\mathbf{x}|\theta)$. Então, $\mathbf{T} = \mathbf{T}(\mathbf{X})$ é suficiente para o parâmetro θ se:*

$$p(x|t, \theta) = p(x|\theta)$$

A definição diz que dado \mathbf{T} , \mathbf{X} quaisquer outras informações adicionais na amostra, além do valor da estatística suficiente, não apresentam mais nenhum detalhe sobre θ (Gamerman e Migon, 1993). Essas considerações levam à técnica de redução de dados conhecida como Princípio da Suficiência.

Princípio da Suficiência

Se $T(\mathbf{X})$ é uma estatística suficiente para θ , então qualquer inferência sobre θ deverá depender da amostra \mathbf{X} somente pelo valor $T(\mathbf{X})$. Isto é, se \mathbf{x} e \mathbf{y} são dois pontos amostrais, de modo que $T(\mathbf{x}) = T(\mathbf{y})$ então a inferência sobre θ deverá ser a mesma, se $\mathbf{X} = \mathbf{x}$ ou $\mathbf{X} = \mathbf{y}$ for observado (Casella, 2001).

3.2.3 Princípio da condicionalidade

Supondo que se dispõe de m experimentos possíveis de serem realizados, denotados por $E_j, j = 1, 2, \dots, m$ com o objetivo de levantar informações sobre um parâmetro θ . Supondo que um experimento foi sorteado ao acaso, entre os m , o princípio da condicionalidade estabelece que os outros experimentos que não foram sorteados são irrelevantes para se estimar θ , ou seja, apenas o experimento realizado é relevante (Paulino, Turkman e Murteira, 2003).

3.3 Elementos da Inferência Bayesiana

No modelo clássico o parâmetro θ com domínio num conjunto Θ ($\theta \in \Theta$) é um escalar ou vetor desconhecido, mas fixo. No modelo bayesiano, o parâmetro θ , $\theta \in \Theta$, é tomado como um escalar ou vetor aleatório (não observável), logo é incerto e toda a incerteza deve ser quantificada em termos de probabilidade (Koop, 2003). Nesta seção apresenta-se conceitos básicos para o estudo da inferência bayesiana.

Para inferir conclusões a respeito de um vetor de quantidades desconhecidas θ a partir de um vetor de observações $\mathbf{x} = (x_1, \dots, x_n)$, deve-se relacionar estes dois vetores de algum modo. Na abordagem bayesiana, informações prévias sobre o vetor θ são representadas usando uma distribuição de probabilidade, chamada de distribuição a priori (ou priori), representada por $h(\theta)$, que estabelece quais valores de θ são mais prováveis, segundo informações disponíveis antes de conhecer as observações.

Assim como na abordagem frequentista, toda a informação proveniente dos dados observados é carregada pela função de verossimilhança. A informação contida em $h(\theta)$ é, então, atualizada através da informação dos dados contida em $f(x|\theta)$, via teorema de Bayes, levando a distribuição posteriori de θ , representada por $h(\theta|x)$ (Ver Figura 4).

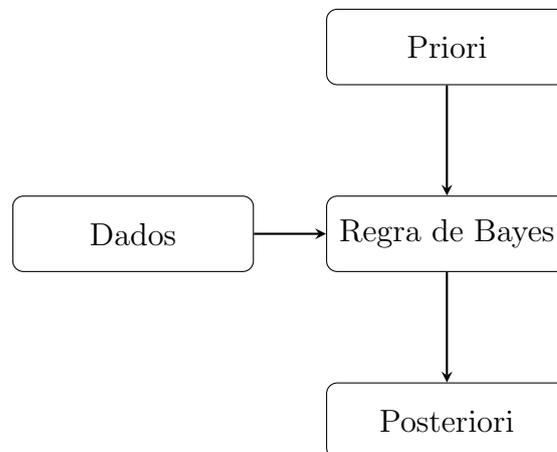


Figura 4 – Resumo do procedimento bayesiano

Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas (i.i.d) condicionalmente a θ , tem-se então que

$$f(x|\theta) = \prod_{i=1}^n f(x_i|\theta),$$

em que $f(x_i|\theta)$ é a distribuição da variável aleatória X_i condicional a θ .

Supondo-se que foi observada a amostra aleatória x . A distribuição de probabilidade conjunta para X e θ , representada por $f(x, \theta)$ é

$$f(x, \theta) = f(x|\theta)h(\theta)$$

$f(x|\theta)$ é também denominada de informação amostral.

A informação amostral é função de duas componentes, x e θ . Fixando θ , $f(\cdot|\theta)$ é uma distribuição de probabilidade. No entanto, após observar $X = x$, $f(x|\theta)$ é apenas uma função de θ e, neste caso, passa a ser denominada por função de verossimilhança de θ em relação ao conjunto de dados observados x , $L(\theta|x) = f(x|\theta)$. Assim, a função de verossimilhança desempenha um papel importante na determinação da distribuição a posteriori, pois é interpretada como um meio através do qual o conjunto de dados transforma o conhecimento a priori sobre θ . A distribuição à posteriori contém toda informação necessária ao desenvolvimento de toda a inferência bayesiana (Paulino, Turkman e Murteira, 2003). Pela regra de Bayes, se θ é contínuo, têm-se que a densidade da posteriori, $h(\theta|x)$, é dada por

$$\begin{aligned} h(\theta|x) &= \frac{f(x, \theta)}{p(x)} \\ &= \frac{f(x|\theta)h(\theta)}{\int_{\Theta} f(x, \theta)d\theta} \\ &= \frac{f(x|\theta)h(\theta)}{\int_{\Theta} f(x|\theta)h(\theta)d\theta} \end{aligned} \quad (3.6)$$

em que $p(x)$ representa a distribuição marginal de X .

No caso de θ ser discreto, tem-se:

$$\begin{aligned} h(\theta|x) &= \frac{f(x, \theta)}{p(x)} \\ &= \frac{f(x|\theta)h(\theta)}{\sum_{\theta \in \Theta} f(x|\theta)h(\theta)} \end{aligned} \quad (3.7)$$

Observa-se em (3.6) e (3.7) que o denominador, $p(x)$, não é função do parâmetro θ , portanto pode ser omitido e considerado como constante normalizadora de (3.6) e (3.7). Com isso a igualdade é substituída por uma proporcionalidade e pode-se reescrever (3.6) e (3.7) como (Koop, 2003):

$$\underbrace{h(\theta|x)}_{\text{posteriori}} \propto \underbrace{L(\theta|x)}_{\text{verossimilhança}} \times \underbrace{h(\theta)}_{\text{priori}}. \quad (3.8)$$

em que o símbolo \propto significa “proporcional”.

Ressalta-se que a priori não é uma distribuição para θ , mas sim uma distribuição que representa a incerteza do pesquisador diante do valor desconhecido θ . Um tipo de distribuição a priori é a priori conjugada que consiste na priori cuja família da posteriori é a mesma da priori. Outra maneira consiste em especificar uma priori não informativa, que ocorre em situações que se conhece pouco ou quando não há informação disponível sobre os possíveis valores do parâmetro, ou quando se espera que a informação dos dados seja dominante (Ehlers, 2011). A seguir, discute-se formas de determinação da distribuição a priori.

3.4 Distribuição a priori

Na inferência bayesiana a utilização da informação anterior ou externa requer a especificação de uma distribuição a priori. Sua determinação é, em geral, subjetiva, nada impedindo no entanto que dados de experimentos passados sejam utilizados, o único compromisso é que esta distribuição represente o conhecimento sobre θ , a quantidade desconhecida, antes de se realizar um experimento (Gamerman e Migon, 1993).

Neste contexto, Koop (2003) sugere que o especialista reúna diversas opiniões de profissionais do setor em estudo para, assim, tirar conclusões mais convincentes. Nesta seção discutem-se diferentes maneiras de se especificar a distribuição a priori.

Segundo Antoniak (1974) na escolha por uma família de distribuições a priori são desejáveis as seguintes propriedades:

1. A família de distribuições a priori deve ser capaz de expressar qualquer informação ou conhecimento sobre o vetor de parâmetros;
2. A família de distribuições a priori deve ser parametrizada de forma a produzir uma interpretação clara das crenças a priori.

Alguns tipos de distribuições a priori são mostradas a seguir.

3.4.1 Priori não informativa

A inferência bayesiana difere da frequentista pela incorporação na análise da informação prévia que se dispõe sobre as quantidades desconhecidas do problema sendo estudado. Mas em algumas situações o especialista pode ter a convicção que a informação disponível para avaliar a distribuição a priori não existe. Nestes casos que se deseja representar a ausência de informação utiliza-se uma classe de prioris denominadas não-informativas.

De acordo com Paulino, Turkman e Murteira (2003) este tipo de distribuição a priori pode:

- deduzir as crenças a posteriori para quem parte de um conhecimento escasso e, nessa medida, se acha incapaz de determinar subjetivamente uma distribuição razoável;
- permitir a comparação com os resultados da inferência clássica que utiliza a informação amostral;
- averiguar a influência nas inferências da distribuição a priori subjetiva que descreve a informação realmente existente, quando confrontada com as que resultam do uso da distribuição a priori de referência.

A seguir, apresentam-se métodos para obtenção de distribuições a priori não-informativas.

3.4.1.1 Método de Bayes-Laplace

O primeiro método para gerar prioris não informativas foi proposto por Bayes e Laplace a partir do Princípio da Razão Insuficiente. Este método estabelece que na ausência de razão suficiente para privilegiar uma possibilidade em detrimento de outras, devido a escassez informativa a priori, deve-se adotar a distribuição uniforme em que todos os possíveis valores de θ são igualmente prováveis, não favorecendo nenhum valor particular de θ . Assim no caso de Θ finito, por exemplo, $\Theta = \{\theta_1, \dots, \theta_k\}$, a distribuição não-informativa gerada para esse argumento é a distribuição uniforme discreta (Paulino, Turkman e Murteira, 2003).

$$h(\theta) = \frac{1}{k}, \theta \in \Theta \quad (3.9)$$

Neste caso, nenhum valor particular de θ é favorecido. Entretanto, Gamerman e Migon (1993) alerta sobre algumas dificuldades intrínsecas a esta escolha. Se θ é contínuo, então:

- (i) $h(\theta)$ é uma distribuição imprópria, isto é, a integral sobre todos os possíveis valores de θ não converge

$$\int h(\theta) d\theta \rightarrow \infty$$

- (ii) se $\phi = \phi(\theta)$ é uma transformação 1 a 1 não linear de um parâmetro θ , e se θ tem distribuição uniforme, então as distribuições de θ e ϕ não são, em geral, probabilisticamente compatíveis. De fato, considerando-se $h(\theta)$ uma distribuição a priori para θ , pelo teorema de transformações de variáveis, a densidade de ϕ é dada por:

$$h(\phi) = h[\theta(\phi)] \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right|.$$

Quando a distribuição a priori $h(\theta)$ é uniforme, percebe-se que a reparametrização de θ , $h(\phi)$, não é necessariamente uniforme.

3.4.1.2 Método de Jeffreys

Uma alternativa à não-invariância da priori de Bayes-Laplace é o método de Jeffreys, proposto por [Jeffreys \(1946\)](#) e é obtido a partir da medida de Informação de Fisher sobre θ .

A distribuição a priori de Jeffreys para o caso uniparamétrico é definida por:

$$h(\theta) \propto I_F(\theta)^{\frac{1}{2}} \quad (3.10)$$

Segundo [Bolfarine e Sandoval \(2010\)](#) a Informação de Fisher sobre θ para uma variável aleatória X com função (densidade) de probabilidade $f(x|\theta)$ é dada por:

$$I_F(\theta) = \text{E} \left[\left(\frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 \right] = -\text{E} \left[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right] \quad (3.11)$$

No caso de um vetor paramétrico $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ ([Paulino, Turkman e Murteira, 2003](#)):

$$h(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{\frac{1}{2}} \quad (3.12)$$

em que $|\cdot|$ é o determinante e $I(\boldsymbol{\theta})$ é a matrix de Informação de Fisher para o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$, que é representada por:

$$\begin{aligned} I(\boldsymbol{\theta}) &= \text{E} \left[\left(\frac{\partial \log f(x|\boldsymbol{\theta})}{\partial \theta_i} \right) \left(\frac{\partial \log f(x|\boldsymbol{\theta})}{\partial \theta_j} \right) \right] \\ &= -\text{E} \left[\frac{\partial^2 \log f(x|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] \end{aligned} \quad (3.13)$$

Jeffreys justificou seu método pelo fato de que ele satisfaz a exigência de reparametrização invariante, ou seja:

$$\begin{aligned} I_F(\psi) &= I(\boldsymbol{\theta}(\psi)) \left(\frac{d\boldsymbol{\theta}}{d\psi} \right)^2 \\ h(\boldsymbol{\theta}) &\propto I(\boldsymbol{\theta}(\psi))^{1/2} \left| \frac{d\boldsymbol{\theta}}{d\psi} \right| = h(\boldsymbol{\theta}(\psi)) \left| \frac{d\boldsymbol{\theta}}{d\psi} \right| \end{aligned}$$

Esta última equação é a fórmula de transformação em que $\psi = f(\boldsymbol{\theta})$.

Exemplo 10. *Supondo que a variável aleatória X tem distribuição de Bernoulli para a qual o parâmetro θ é desconhecido ($0 < \theta < 1$). Determine a distribuição a priori de Jeffreys para θ .*

Se $X|\theta \sim \text{Bernoulli}(\theta)$ então,

$$f(x|\theta) = \theta^x (1 - \theta)^{1-x} \quad \text{para } x = 1 \text{ ou } x = 0.$$

Então,

$$l(x|\theta) = \log f(x|\theta) = x \log \theta + (1 - x) \log(1 - \theta)$$

e

$$l'(x|\theta) = \frac{x}{\theta} - \frac{1-x}{1-\theta} \quad \text{e} \quad l''(x|\theta) = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

Seja $E(X) = \theta$, a Informação de Fisher é

$$\begin{aligned} I(x|\theta) &= -E[l''(x|\theta)] \\ &= -E\left[-\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}\right] \\ &= \frac{E(X)}{\theta^2} + \frac{1-E(X)}{(1-\theta)^2} \\ &= \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} \\ &= \frac{1}{\theta} + \frac{1}{1-\theta} \end{aligned}$$

Logo, a distribuição a priori de Jeffreys é

$$\begin{aligned} h(\theta) &\propto I(\theta)^{1/2} \\ &\propto \theta^{-1/2}(1-\theta)^{-1/2}, \theta \in (0, 1) \end{aligned}$$

[Bernardo \(1989\)](#) destaca alguns aspectos da distribuição a priori de Jeffreys:

- (i) A principal motivação intuitiva da distribuição a priori de Jeffreys é ser invariante, a qual é condição necessária mas não suficiente para determinar uma referência de distribuição a priori.
- (ii) A existência da distribuição a priori de Jeffreys requer condições de regularidade forte.

3.4.2 Priori conjugada

Em algumas situações o desenvolvimento matemático e computacional da distribuição a posteriori utilizando determinadas prioris pode ser difícil ou resultar em distribuições que apresentam forma desconhecida. Nesses casos, pode-se fazer uso de prioris conjugadas, em que a distribuição à posteriori pertence a mesma família de distribuições da priori e portanto a atualização do conhecimento sobre o parâmetro θ envolve apenas uma mudança nos hiperparâmetros ([O'Hagan, 1994](#)). Estes parâmetros indexadores da distribuição a priori são chamados de hiperparâmetros para distingui-los do parâmetro de interesse θ .

Definição 5. *Supondo que a distribuição a priori $p(\theta)$ pertença a uma classe paramétrica de distribuições \mathcal{F} . Então a distribuição a priori é chamada de conjugada com respectiva distribuição de verossimilhança $p(\mathbf{y}|\theta)$ se a distribuição a posteriori $p(\theta|\mathbf{y})$ também pertencer a \mathcal{F} ([Jackman, 2009](#)).*

Conforme [Gelman et al. \(2014\)](#) a família conjugada é matematicamente conveniente porque a distribuição a posteriori segue uma forma paramétrica conhecida, para isso é

necessário que a distribuição a priori e a distribuição de verossimilhança tenham o mesmo núcleo (kernel). Por exemplo, a função densidade univariada da distribuição normal

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right]$$

e o núcleo da distribuição (para σ conhecido) é

$$\exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right]$$

em que $\frac{1}{\sqrt{2\pi}\sigma}$ é a constante normalizadora. A Tabela (2) apresenta algumas distribuições conjugadas.

Tabela 2 – Algumas distribuições a priori conjugadas

Verossimilhança	Parâmetro	Priori	Posteriori
$X \sim \text{Binomial}(n, p)$	$0 \leq p \leq 1$	Beta(a, b) $a > 0, b > 0$	Beta(a^*, b^*) $a^* = a + x$ $b^* = b + n - x$
$X = X_1, \dots, X_n$ $X_i \stackrel{iid}{\sim} \text{Poisson}(\lambda)$	$\lambda > 0$	Gama(a, b) $a > 0, b > 0$	Gama(a^*, b^*) $a^* = a + n\bar{x}$ $b^* = b + n$
$X = X_1, \dots, X_n$ $X_i \stackrel{iid}{\sim} \text{Exponencial}(\lambda)$	$\lambda > 0$	Gama(a, b) $a > 0, b > 0$	Gama(a^*, b^*) $a^* = a + n$ $b^* = b + n\bar{x}$
$X = X_1, \dots, X_n$ $X_i \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma^2)$ σ^2 conhecido	$\infty < \mu < \infty$	Normal(a, b^2) $-\infty < a < \infty$ $b > 0$	Normal(a^*, b^{*2}) $a^* = \frac{nb^2\bar{x} + \sigma^2 a}{nb^2 + \sigma^2}$ $b^{*2} = \frac{\sigma^2 b^2}{nb^2 + \sigma^2}$

Exemplo 11. (Distribuição Poisson) Se θ tem distribuição Gama com parâmetro de forma, $a > 0$, e parâmetro de escala, $b > 0$, denotada por $\theta \sim \text{Gama}(a, b)$, sua função densidade de probabilidade é dada por:

$$h(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}, \quad (3.14)$$

e se $\mathbf{x}|\theta$ é uma amostra independente e identicamente distribuída da distribuição Poisson, cuja verossimilhança é

$$f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n \left(\frac{e^{-\theta} \theta^{x_i}}{x_i!} \right), \theta > 0.$$

Mostre que a distribuição a posteriori $h(\theta|\mathbf{x}) \sim \text{Gama}(a^*, b^*)$, em que $a^* = S + a$, $b^* = n + b$ e $S = \sum_{i=1}^n x_i$.

Conforme os dados do Exemplo, a função de verossimilhança será

$$\begin{aligned} f(x_1, \dots, x_n | \theta) &= \prod_{i=1}^n \left(\frac{e^{-\theta} \theta^{x_i}}{x_i!} \right) \\ &= \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \\ &\propto e^{-n\theta} \theta^{\sum_{i=1}^n x_i} \\ &\propto e^{-n\theta} \theta^S \end{aligned}$$

Usando o Teorema de Bayes (3.6) e (3.8) para obtenção da distribuição à posteriori:

$$\begin{aligned} h(\theta | \mathbf{x}) &\propto f(\mathbf{x} | \theta) h(\theta) \\ &\propto \frac{\theta^S}{\prod_{i=1}^n x_i!} e^{-n\theta} \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \\ &\propto \theta^S e^{-n\theta} \theta^{a-1} e^{-b\theta} \\ &\propto \theta^{S+a-1} e^{-\theta(n+b)} \end{aligned} \tag{3.15}$$

Observa-se em (3.15) que foram suprimidos as funções que não dependiam do parâmetro θ . Portanto, a expressão corresponde ao núcleo de uma densidade Gama ($a + \sum x_i, n + b$), ou seja, a distribuição a posteriori de θ é proporcional a uma distribuição Gama com parâmetros $\alpha^* = S + a = \sum_{i=1}^n x_i + a$ e $\beta^* = n + b$

$$h(\theta | \mathbf{x}) \sim \text{Gama}(\alpha^*, \beta^*)$$

E conclui-se que a família Gama é uma priori conjugada para o parâmetro θ da distribuição Poisson.

Exemplo 12. *Seja X_1, \dots, X_{10} uma amostra da distribuição Poisson com média (θ).*

$$X_i \stackrel{iid}{\sim} \text{Poisson}(\theta)$$

Determine a distribuição a posteriori considerando uma distribuição priori Gama com parâmetros $a = 2$ e $b = 0.2$ e os dados observados são $x = (12, 14, 15, 12, 16, 14, 27, 10, 14, 16)$.

A distribuição a priori é:

$$h(\theta) \sim \text{Gama}(2, 0.2)$$

A distribuição a priori proporcional é dada por:

$$\begin{aligned} h(\theta) &\propto \theta^{a-1} e^{-b\theta} \\ &\propto \theta^{2-1} e^{-0.2\theta} \end{aligned}$$

A partir dos dados $n = 10$, $\sum_{i=1}^{10} x_i = 150$ e a função de verossimilhança é proporcional a

$$\begin{aligned} f(\mathbf{x}|\theta) &\propto e^{-n\theta} \theta^{\sum_{i=1}^n x_i} \\ &\propto e^{-10\theta} \theta^{\sum_{i=1}^{10} x_i} \\ &\propto e^{-10\theta} \theta^{400} \end{aligned}$$

Com isso a distribuição a posteriori é proporcional a:

$$\begin{aligned} h(\theta|\mathbf{x}) &\propto \theta^{2+150-1} e^{-(10+0.2)\theta} \\ &\propto \theta^{152-1} e^{-10.2\theta} \end{aligned}$$

Isso corresponde a uma distribuição Gama (152, 10.2).

A Figura (5) mostra diferentes distribuições a priori e suas respectivas distribuições a posteriori para os dados do Exemplo 12.

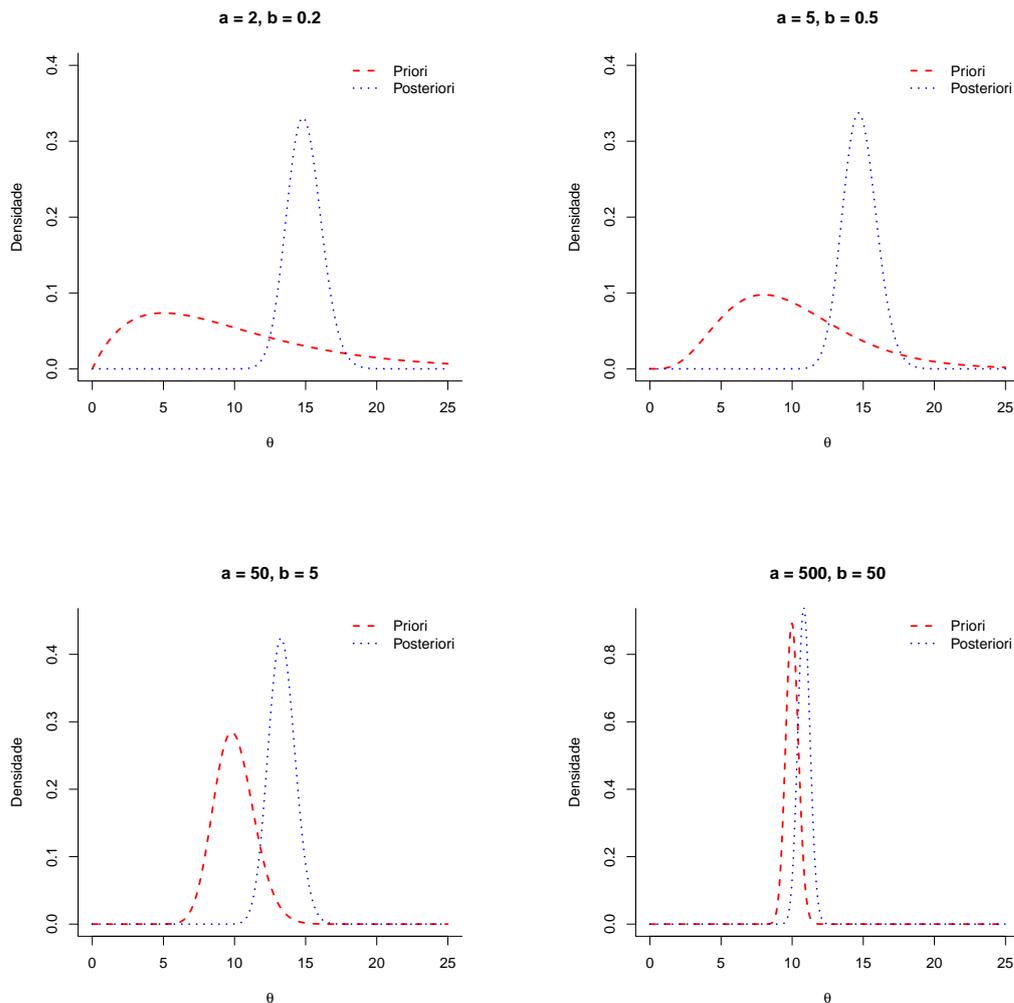


Figura 5 – Prioris conjugadas Gama(a,b) e suas posteriores

3.4.3 Prioris Hierárquicas

Modelos hierárquicos constituem um promissor caminho para a expansão dos modelos bayesianos (Banerjee, Carlin e Gelfand, 2014; Royle, 2008; Clark, 2006). Eles representam uma estrutura de modelagem com capacidade de explorar diversas fontes de informação, modelar problemas com estruturas complexas de dependência, acomodar influências que são desconhecidas, traçar inferência com grande número de variáveis latentes e parâmetros que descrevem relacionamentos complexos (Clark, 2005). A ideia de utilizar a estrutura hierárquica com múltiplos estágios para a distribuição a priori foi formalizada por Lindley e Smith (1972).

O método consiste em dividir a especificação da distribuição a priori em estágios. Na opinião de Gamerman e Migon (1993) esta divisão em estágios é um artifício probabilístico que auxilia a identificação e a especificação coerente da priori. A Figura (6) ilustra um exemplo de modelo hierárquico, em que a distribuição a priori de θ depende de outro parâmetro ϕ , que também é desconhecido e pode ter uma probabilidade a priori associada.

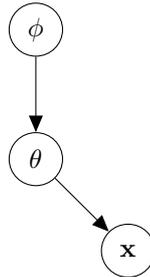


Figura 6 – Exemplo de modelo bayesiano hierárquico

Para Paulino, Turkman e Murteira (2003) a metodologia hierárquica bayesiana modela a incerteza nos parâmetros que auxiliam a especificação da priori, denominados hiperparâmetros, através de novas distribuições induzindo assim uma decomposição da distribuição a priori em dois ou mais níveis. Para este autor a decomposição a priori é geralmente justificada pela incapacidade de se quantificar exatamente a priori e pelo interesse em incorporar a incerteza decorrente sobre os hiperparâmetros.

Na abordagem hierárquica, além de especificar o modelo de distribuição $f(\mathbf{y}|\boldsymbol{\theta})$ para os dados observados $\mathbf{y} = (y_1, \dots, y_n)$ dado um vetor de parâmetros desconhecidos $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$, supondo que $\boldsymbol{\theta}$ é uma amostra aleatória de uma distribuição a priori $h(\boldsymbol{\theta}|\boldsymbol{\lambda})$, em que $\boldsymbol{\lambda}$ é um vetor de hiperparâmetros, denominação que se dá aos parâmetros pertencentes à distribuição a priori especificada para $\boldsymbol{\theta}$. Se $\boldsymbol{\lambda}$ é conhecido, a inferência sobre $\boldsymbol{\theta}$ baseia-se na sua distribuição a posteriori ² (Banerjee, Carlin e Gelfand, 2014),

$$h(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\lambda}) = \frac{h(\mathbf{y}, \boldsymbol{\theta}|\boldsymbol{\lambda})}{h(\mathbf{y}|\boldsymbol{\lambda})} = \frac{h(\mathbf{y}, \boldsymbol{\theta}|\boldsymbol{\lambda})}{\int h(\mathbf{y}, \boldsymbol{\theta}|\boldsymbol{\lambda})d\boldsymbol{\theta}} = \frac{f(\mathbf{y}|\boldsymbol{\theta})h(\boldsymbol{\theta}|\boldsymbol{\lambda})}{\int f(\mathbf{y}|\boldsymbol{\theta})h(\boldsymbol{\theta}|\boldsymbol{\lambda})d\boldsymbol{\theta}}. \quad (3.16)$$

² No caso contínuo

Observa-se a contribuição para a distribuição a posteriori tanto dos dados observados (na forma da distribuição de verossimilhança f) e do conhecimento ou opinião externa (na forma da priori). Caso $\boldsymbol{\lambda}$ seja desconhecido, será necessário especificar um segundo estágio (ou hiperpriori) representado pela distribuição $h(\boldsymbol{\lambda})$, e (3.16) será substituída por:

$$h(\boldsymbol{\theta}|\mathbf{y}) = \frac{h(\mathbf{y}, \boldsymbol{\theta})}{h(\mathbf{y})} = \frac{\int f(\mathbf{y}|\boldsymbol{\theta})h(\boldsymbol{\theta}|\boldsymbol{\lambda})h(\boldsymbol{\lambda})d\boldsymbol{\lambda}}{\int \int f(\mathbf{y}|\boldsymbol{\theta})h(\boldsymbol{\theta}|\boldsymbol{\lambda})h(\boldsymbol{\lambda})d\boldsymbol{\theta}d\boldsymbol{\lambda}}. \quad (3.17)$$

Observa-se em (3.17) uma estrutura hierárquica implícita, ou seja, três níveis de hierarquia com interesse primário no nível $\boldsymbol{\theta}$. Pode-se resumir a estrutura básica e uma extensão do modelo hierárquico:

Estrutura básica:

$$\begin{aligned} y &\sim h(y|\theta) \\ \theta &\sim h(\theta|\phi) \\ \phi &\sim h(\phi) \end{aligned}$$

Extensão com um nível a mais na hierarquia:

$$\begin{aligned} y &\sim h(y|\theta) \\ \theta &\sim h(\theta|\phi) \\ \phi &\sim h(\phi|\psi) \\ \psi &\sim h(\psi) \end{aligned}$$

Assim a distribuição a posteriori é proporcional a

$$h(\theta, \phi, \psi|y) \propto h(y|\theta)h(\theta|\phi)h(\phi|\psi)h(\psi)$$

A seguir apresentam-se alguns exemplos básicos da utilização do modelo hierárquico.

Exemplo 13. (*Modelo beta/binomial hierárquico*)

Seja $\mathbf{x} = (x_1, \dots, x_n)$, em que

$$x_i|\theta_i \sim \text{Bin}(n_i, \theta_i)$$

e x_i são independentes de θ_i . E ainda que

$$\theta_i \stackrel{\text{iid}}{\sim} \text{Beta}(\alpha, \beta).$$

Agrupam-se todas as probabilidades de sucesso em um vetor $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ e atribuindo-se uma distribuição a priori para os hiperparâmetros α e β , $P(\alpha, \beta)$. Desse

modo, o modelo resumido é dado por:

$$\begin{aligned}x_i|\theta &\sim \text{Bin}(n_i, \theta) \\ \theta|\alpha, \beta &\sim \text{Beta}(\alpha, \beta) \\ \alpha|a, b &\sim \Gamma(a, b) \\ \beta|a, b &\sim \Gamma(a, b), \quad \text{com } a \text{ e } b \text{ conhecidos.}\end{aligned}$$

A distribuição a posteriori de θ torna-se proporcional a:

$$\begin{aligned}h(\boldsymbol{\theta}, \alpha, \beta|\mathbf{x}) &\propto h(\mathbf{x}|\boldsymbol{\theta}, \alpha, \beta)h(\boldsymbol{\theta}, \alpha, \beta) \\ &\propto h(\alpha, \beta) \prod_{i=1}^n \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{\alpha-1} (1 - \theta_i)^{\beta-1} \prod_{i=1}^n \theta_i^{x_i} (1 - \theta_i)^{n_i - x_i}\end{aligned}$$

A distribuição condicional de θ_i é

$$h(\theta_i|\mathbf{x}, \alpha, \beta, \boldsymbol{\theta}_{-i}) \sim \theta^{x_i - \alpha - 1} (1 - \theta_i)^{n_i - x_i + \beta - 1}$$

Portanto,

$$\theta_i|\mathbf{x}, \alpha, \beta, \boldsymbol{\theta}_{-i} \sim \text{Beta}(x_i + \alpha, n_i - x_i + \beta)$$

Exemplo 14. (*Normal hierárquica; Gamerman e Migon, 1993*) Supondo que Y_1, \dots, Y_n são tais que $Y_i \sim N(\theta_i, \sigma^2)$, com σ^2 conhecido. A especificação da priori de $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ pode ser baseada nas seguintes hipóteses:

- os θ_i 's são independentes, isto é, $p(\boldsymbol{\theta}) = \prod_i p(\theta_i)$; ou
- os θ_i 's constituem uma amostra de uma população $p(\boldsymbol{\theta}|\boldsymbol{\lambda})$ em que $\boldsymbol{\lambda} = (\mu, \tau^2)$ contém os parâmetros que descrevem a população.

Daí, para a última hipótese

$$h(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \prod_{i=1}^n p(\theta_i|\boldsymbol{\lambda}).$$

Essa especificação constitui o primeiro estágio. Para complementar a priori é necessário especificar o segundo estágio: a distribuição de $\boldsymbol{\lambda}$, $p(\boldsymbol{\lambda})$. Ressalta-se que $h(\boldsymbol{\lambda})$ independe do primeiro estágio. Tendo isso, pode-se obter a distribuição a priori (marginal) de $\boldsymbol{\theta}$.

$$h(\boldsymbol{\theta}) = \int h(\boldsymbol{\theta}, \boldsymbol{\lambda}) d\boldsymbol{\lambda} = \int h(\boldsymbol{\theta}|\boldsymbol{\lambda}) h(\boldsymbol{\lambda}) d\boldsymbol{\lambda} = \int \prod_{i=1}^n h(\theta_i|\boldsymbol{\lambda}) h(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$$

em que $\boldsymbol{\lambda}$ são os hiperparâmetros. Assim, os θ_i 's são supostos permutáveis, que podem seguir a seguinte estrutura ou hierarquia sob θ_i :

$$\begin{aligned}\theta_i|\boldsymbol{\lambda} &\sim h(\theta_i|\boldsymbol{\lambda}) \quad \text{modelo hierárquico para } \theta_i \\ \boldsymbol{\lambda} &\sim h(\boldsymbol{\lambda}) \quad \text{priori dos hiperparâmetros } \boldsymbol{\lambda}\end{aligned}$$

3.5 Estimativa pontual e intervalar

Nesta seção discutem-se dois procedimentos de estimação de parâmetros: estimação pontual e intervalar.

A distribuição a posteriori apresenta tudo o que pode ser obtido em termos de informação sobre o parâmetro de interesse (θ). No entanto, às vezes é necessário resumir a informação disponível através de uns poucos números para comunicação externa. O caso mais simples é o de estimação pontual em que procura-se determinar um único valor de θ que resuma a distribuição como um todo (Gamerman e Migon, 1993).

Na abordagem bayesiana adota-se a função de perda para a escolha do estimador de θ . Segundo Ehlers (2011) para cada possível valor de θ e cada possível estimativa $a \in \Theta$, associa-se uma perda $\mathcal{L}(\alpha, \theta)$, de modo que quanto maior a distância entre a e θ maior o valor de perda. Segundo este autor, a função perda determina a perda sofrida ao se tornar a decisão α dado o real estado $\theta \in \Theta$. Neste caso, a perda esperada a posteriori é expressa como um número real e é definida por:

$$E[\mathcal{L}(\alpha, \theta|\mathbf{x})] = \int \mathcal{L}(\alpha, \theta)h(\theta|\mathbf{x})d\theta \quad (3.18)$$

em que α , denominado estimador de Bayes, é escolhido de tal forma que a perda esperada a posteriori seja minimizada.

O estimador de Bayes depende da função de perda que é adotada. Assim, caso a função perda seja definida por $\mathcal{L}(\alpha, \theta) = (\alpha - \theta)^2$ então o estimador de Bayes é a média a posteriori dada por $\hat{\theta} = E(\theta|x)$ ((Ehlers,2011). Mas caso seja adotada a função perda absoluta ($\mathcal{L}(\alpha, \theta) = |\alpha - \theta|$) tem-se que o estimador é a mediana a posteriori. Finalmente, se a função perda denominada 0-1 é adotada, o estimador de Bayes será a moda a posteriori. Esse estimador de θ também é chamado de estimador de máxima verossimilhança generalizado (Ehlers,2011; O'Hagan,1994; Gamerman e Migon, 1993).

A outra maneira é associar aos estimadores pontuais uma medida que forneça a incerteza associada a esse estimador, que é o escopo do intervalo de credibilidade.

Segundo Gamerman e Migon (1993) C é um intervalo de confiança bayesiano ou intervalo de credibilidade de $100(1 - \gamma)\%$ para θ se $P(\theta \in C \geq 1 - \gamma)$. Nessa caso, $(1 - \gamma)$ é chamado de nível de confiança ou credibilidade. Este intervalo é obtido de uma região de Θ que contenha uma parte substancial da massa probabilística a posteriori (Paulino, Turkman e Murteira, 2003).

No caso, o interesse está naquele intervalo que apresenta o menor comprimento possível. De acordo com Ehlers (2011), pode-se mostrar que os intervalos de comprimento mínimos são obtidos tomando-se os valores de θ com maior densidade a posteriori. Esses intervalos são denominados intervalos *Highest Posterior Density* (HPD).

No próximo capítulo discute-se modelos e métodos apropriados da inferência bayesiana quando o pesquisador utiliza dados em painel.

4 Modelos Bayesianos para Dados em Painel

O objetivo deste capítulo é apresentar como os métodos bayesianos são usados para modelar e analisar dados em painel a partir das abordagens desenvolvidas por [Koop, 2003](#); [Morawetz, 2006](#); [Congdon, 2010](#).

Nas seções seguintes discutem-se os modelos e métodos da inferência bayesiana no contexto dos dados em painel. Para isso combinam-se aspectos dos vários modelos já apresentados. As seções são organizadas de acordo com a estrutura colocada nos coeficientes de regressão, assim, inicia-se com o modelo que assume que os coeficientes são os mesmos para todos os indivíduos (modelo de dados agrupados) e em seguida, os modelos que permitem que os coeficientes variem entre os indivíduos (modelo de efeito fixo e aleatório).

Os modelos que serão apresentados seguirão a seguinte notação. Supondo que y_{it} e ε_{it} denotam a t -ésima observação ($t = 1, \dots, T$) das variáveis dependentes e dos erros, respectivamente, para o i -ésimo indivíduo ($i = 1, \dots, n$). E que y_i e ε_i representam vetores das T observações das variáveis dependentes e dos erros, respectivamente, para o i -ésimo indivíduo.

Em algumas aplicações deste capítulo é importante distinguir entre o intercepto e o coeficiente de inclinação. Para isso, define-se X_i uma matriz $T \times k$ contendo as T observações para cada uma das k variáveis explicativas (incluindo o intercepto) para o i -ésimo indivíduo. Denota-se a matriz \tilde{X}_i de dimensão $T \times (k - 1)$ igual a matriz X_i sem o intercepto, logo $X_i = [\iota_T \tilde{X}_i]$, em que ι_T é um vetor de 1's. Empilhando as observações para todos os n indivíduos, obtem-se um vetor Tn :

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \text{e} \quad \varepsilon = \begin{bmatrix} \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

De modo análogo, empilhando todas as observações das variáveis explicativas produzem uma matriz $Tn \times k$:

$$X = \begin{bmatrix} X_1 \\ \dots \\ X_n \end{bmatrix}$$

4.1 Modelo para dados agrupados (Pooled)

O modelo para dados agrupados baseia-se em um painel balanceado com T observações para os n indivíduos. Assim, as observações no painel podem ser representadas na forma (y_{it}, x_{it}) , $i = 1, \dots, n$; $t = 1, \dots, T$, em que o índice i denota a unidade individual e o índice t o período do tempo.

A expressão vetorial do modelo para dados agrupados para o i -ésimo indivíduo pode ser expresso como:

$$y_i = X_i\beta + \varepsilon_i, \quad i = 1, \dots, n. \quad (4.1)$$

A variável dependente y_i é um vetor de comprimento T correspondente aos valores de y para a unidade i , X_i é uma matriz $T \times k$ das variáveis explicativas, β é um vetor $k \times 1$ dos coeficientes comuns para todos os indivíduos, incluindo o intercepto e o termo de erro ε_i é um vetor de dimensão $T \times 1$ que segue uma distribuição normal multivariada.

A especificação da função de verossimilhança depende dos pressupostos sobre os erros:

1. ε_i segue uma distribuição normal multivariada com média 0_T e matriz de covariância $h^{-1}I_T$, em que 0_T é um vetor $T \times 1$ que apresenta todos os elementos igual a zero, I_T é uma matriz identidade $T \times T$, h é a precisão e é dada por $h = \sigma^{-2}$. A notação é dada por $\varepsilon \sim N(0_T, h^{-1}I_T)$.
2. ε_i e ε_j são independentes uns dos outros para $i \neq j$. Neste capítulo assume-se que $i, j = 1, \dots, n$.
3. Todos os elementos de X_i são fixos, isto é, não são variáveis aleatórias, ou, se forem variáveis aleatórias, são independentes de todos os elementos de ε_i com função densidade de probabilidade, $p(X_i|\lambda)$, em que λ é um vetor de parâmetros que não são incluídos em β e h .

A matriz de variância e covariâncias de um vetor é uma matriz simétrica que contém as variâncias de todos os elementos do vetor na diagonal e as covariâncias entre diferentes elementos completam as demais posições da matriz, ou seja,

$$\begin{aligned} \text{var}(\varepsilon) = \Sigma_{n \times n} &\equiv \begin{bmatrix} \text{var}(\varepsilon_1) & \text{cov}(\varepsilon_1, \varepsilon_2) & \dots & \text{cov}(\varepsilon_1, \varepsilon_n) \\ \text{cov}(\varepsilon_1, \varepsilon_2) & \text{var}(\varepsilon_2) & \dots & \cdot \\ \cdot & \text{cov}(\varepsilon_2, \varepsilon_3) & \dots & \cdot \\ \cdot & \cdot & \dots & \text{cov}(\varepsilon_{n-1}, \varepsilon_n) \\ \text{cov}(\varepsilon_1, \varepsilon_n) & \cdot & \dots & \text{var}(\varepsilon_n) \end{bmatrix} \\ &= \begin{bmatrix} h^{-1} & 0 & \dots & 0 \\ 0 & h^{-1} & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & 0 \\ 0 & \cdot & \dots & h^{-1} \end{bmatrix} \end{aligned} \quad (4.2)$$

A afirmação que $\text{var}(\varepsilon) = h^{-1}I_T$ é uma notação compactada para $\text{var}(\varepsilon_i) = h^{-1}$ e $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ para $i, j = 1, \dots, T$ e $i \neq j$. Neste caso, pode-se escrever como:

$$\varepsilon \sim N\left(0_T, \frac{1}{h}I_T\right)$$

O pressuposto que os erros são independentes em todos os indivíduos e períodos de tempo implica que o modelo de dados agrupados se reduz ao modelo de regressão linear múltipla. Portanto, a função de verossimilhança para o modelo de dados agrupados a partir da definição da função densidade da distribuição normal multivariada é dada por:

$$\begin{aligned} p(y_i|\beta, h) &= \prod_{i=1}^n \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[-\frac{h}{2} (y_i - X_i\beta)' (y_i - X_i\beta) \right] \right\} \\ &= \frac{h^{\frac{nT}{2}}}{(2\pi)^{\frac{nT}{2}}} \left\{ \exp \left[-\frac{h}{2} (y - X\beta)' (y - X\beta) \right] \right\} \\ &\propto h^{\frac{nT}{2}} \left\{ \exp \left[-\frac{h}{2} (y - X\beta)' (y - X\beta) \right] \right\} \end{aligned} \quad (4.3)$$

A função de verossimilhança pode ser escrita também da seguinte forma:

$$p(y|\beta, h) = \frac{1}{(2\pi)^{\frac{nT}{2}}} \left\{ h^{\frac{k}{2}} \exp \left[-\frac{h}{2} (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right] \right\} \left\{ h^{\frac{\nu}{2}} \exp \left[-\frac{h\nu}{2s^2} \right] \right\} \quad (4.4)$$

Em que $\hat{\beta}$, s^2 e ν são os estimadores de mínimos quadrados para β , erro padrão e graus de liberdade, respectivamente, e são dados por:

$$\hat{\beta} = (X'X)^{-1} X'y, \quad (4.5)$$

$$s^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{\nu}, \quad (4.6)$$

$$\nu = nT - k. \quad (4.7)$$

A especificação da distribuição a priori é uma questão relevante na implementação bayesiana, a escolha de uma priori inadequada pode levar a flutuações nas estimativas dos parâmetros (Kass e Raftery, 1995). Dessa forma, em alguns casos a escolha de prioris conjugadas facilitam a interpretação e podem facilitar a implementação computacional. A forma da verossimilhança dada pela Equação (4.5) sugere o núcleo de uma densidade normal para β , e a segunda parte o núcleo de uma densidade gama para h . Com isso, para as aplicações do próximo capítulo utilizou-se a priori conjugada normal-gama, ou seja,

$$\begin{aligned} \beta|h &\sim \text{Normal}(\underline{\beta}, \underline{V}) \\ &\text{e} \\ h &\sim \text{Gama}(\underline{s}^{-2}, \underline{\nu}) \end{aligned} \quad (4.8)$$

Então, a priori conjugada natural para β e h é denotado por:

$$\beta, h \sim \text{Normal-Gama}(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}) \quad (4.9)$$

Assim, o pesquisador para expressar suas informações prévias escolhe os valores referentes aos hiperparâmetros: $\underline{\beta}$, \underline{V} , \underline{s}^{-2} , $\underline{\nu}$, definidos a seguir.

Neste trabalho utiliza-se a barra embaixo dos parâmetros, por exemplo, $\underline{\beta}$, para indicar os parâmetros da distribuição a priori, e a barra acima dos parâmetros, por exemplo $\overline{\beta}$ denotam os parâmetros da distribuição posteriori.

A distribuição a posteriori resume as informações a respeito dos parâmetros β e h contidas tanto nos dados como na priori e é proporcional ao produto da distribuição a priori (4.9) e da verossimilhança (4.5). Logo, a distribuição a posteriori conjugada normal-gama é dada por

$$\begin{aligned}\beta, h|y &\sim \text{Normal-Gama}(\overline{\beta}, \overline{V}, \overline{s}^{-2}, \overline{\nu}) \\ \overline{V} &= (\underline{V}^{-1} + X_i'X_i)^{-1} \\ \overline{\beta} &= \overline{V} (\underline{V}^{-1}\underline{\beta} + X_i'X_i\hat{\beta}) \\ \overline{\nu} &= \underline{\nu} + nT\end{aligned}\tag{4.10}$$

No Capítulo (5) apresentam-se a implementações do modelo para dados agrupados e suas distribuições (4.9), (4.5) e (4.10) obtidas nessa Seção.

4.2 Modelos de efeitos individuais

Os modelos de efeitos individuais são caracterizados por permitirem que cada unidade individual tenha seu próprio intercepto e são representados por:

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, n; t = 1, \dots, T.\tag{4.11}$$

em que y_{it} é a variável dependente, x_{it} é uma matriz conhecida k -dimensional que acomoda variáveis explicativas, α_i são os interceptos específicos para cada indivíduo, esses interceptos diferentes entre os indivíduos são uma maneira de modelar a heterogeneidade. Assume-se que $\varepsilon_{it} \sim N(0, h^{-1})$ e $\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0$ a menos que $i = t$ e $j = s$. Nota-se que a distribuição de ε_{it} foi parametrizada em termos da precisão e não da variância.

Na literatura econométrica clássica se o efeito específico α_i for considerado uma variável aleatória é chamado de “efeito aleatório” e, caso seja não aleatório mas desconhecido é chamado de “efeito fixo”.

Sob a perspectiva bayesiana não há distinção entre os efeitos específicos individuais, eles são considerados quantidades aleatórias a serem estimadas. Nesse contexto, não há distinção entre os modelos de efeitos fixos e aleatório, e eles são caracterizados conforme a distribuição a priori atribuídas aos efeitos específicos individuais. Assim, para o modelo de efeitos aleatórios utiliza-se priori com uma estrutura hierárquica, nos modelos de efeitos fixo considera-se prioris não hierárquicas.

4.2.1 Função de verossimilhança

A função de verossimilhança para o indivíduo i é baseada na seguinte equação de regressão

$$y_i = \alpha_i \iota_T + \tilde{X}_i \tilde{\beta} + \varepsilon_i \quad (4.12)$$

em que nesta notação y_i é um vetor $T \times 1$ das variáveis dependentes para o i -ésimo indivíduo, α_i denota o intercepto para o i -ésimo indivíduo, ι_T uma matriz de 1's, \tilde{X}_i uma matriz $T \times k$ das variáveis independentes, $\tilde{\beta}$ representa o vetor $k \times 1$ dos coeficientes de inclinação que são iguais para todos os indivíduos. O termo ε_i vetor $T \times 1$ e \tilde{X} tem distribuição normal, não correlacionado e independente de \tilde{X}_i, α_i e β . A Equação (4.12), sob os pressupostos definidos na seção anterior, implica na seguinte função de verossimilhança

$$p(y|\alpha, \beta, h) = \prod_{i=1}^n \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[-\frac{h}{2} (y_i - \alpha_i - \tilde{X}_i \tilde{\beta})' (y_i - \alpha_i - \tilde{X}_i \tilde{\beta}) \right] \right\} \quad (4.13)$$

em que $\alpha = (\alpha_1, \dots, \alpha_n)'$.

4.2.2 Modelos de efeitos fixos

Para o modelo de efeitos fixos a Equação (4.12) pode ser reescrita como:

$$y = X^* \beta^* + \varepsilon \quad (4.14)$$

em que X^* é uma matriz $Tn \times (n + k - 1)$ dada por

$$X^* = \begin{bmatrix} \iota_T & 0_T & \cdot & \cdot & 0_T & \tilde{X}_1 \\ 0_T & \iota_T & \cdot & \cdot & \cdot & \tilde{X}_2 \\ \cdot & 0_T & \cdot & \cdot & \cdot & \cdot \\ 0_T & \cdot & \cdot & \cdot & \iota_T & \tilde{X}_n \end{bmatrix}$$

e

$$\beta^* = \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \alpha_n \\ \tilde{\beta} \end{bmatrix}$$

Para determinação da distribuição a posteriori do modelo de efeitos fixos qualquer umas das prioris explícitas na Seção (3.4) podem ser aplicadas a parâmetro β^* e sua precisão h). Conforme Koop (2003) a utilização de uma priori não hierárquica leva um modelo que é análogo ao modelo de efeitos fixos. Constata-se isso ao verificar que a matriz X^* , que inclui as variáveis explicativas associadas a uma matriz de variáveis *dummy* para cada indivíduo.

Na aplicação do modelo de efeito fixo são assumidas as seguintes distribuições a priori não hierárquicas normal-gama para o parâmetro β^* e sua precisão

$$\beta^* \sim \text{Normal}(\underline{\beta}^*, \underline{V}) \quad (4.15)$$

e

$$h \sim \text{Gama}(\underline{s}^{-2}, \underline{\nu}) \quad (4.16)$$

Considerando as distribuições a priori (4.15) e (4.16) independentes entre si, obtém-se a distribuição a posteriori conjunta normal-gama para estimação dos parâmetros (Koop, 2003; D'Espallier, Huybrechts e Iturriaga, 2011)

$$\beta^* | y, h \sim \text{Normal}(\bar{\beta}^*, \bar{V}) \quad (4.17)$$

e

$$h | y, \beta^* \sim \text{Gama}(\bar{s}^{-2}, \bar{\nu}) \quad (4.18)$$

em que

$$\begin{aligned} \bar{V} &= (\underline{V}^{-1} + hX^{*'}X^*)^{-1} \\ \bar{\beta}^* &= \bar{V}(\bar{V}^{-1}\underline{\beta}^* + hX^{*'}y) \\ \bar{\nu} &= Tn + \bar{\nu} \\ \bar{s}^2 &= \frac{\sum_{i=1}^n (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta})' (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta}) + \nu s^2}{\bar{\nu}} \end{aligned}$$

A distribuição a posteriori (4.18) pode ser estimada empiricamente por meio de métodos MCMC (*Markov Chain Monte Carlo*) utilizando o software R (R Core Team, 2016).

4.2.3 Modelo de efeitos aleatórios

O modelo de efeitos aleatórios contém $n + k$ parâmetros, isto é, n interceptos α , $k - 1$ coeficientes de inclinação em $\tilde{\beta}$ mais o parâmetro de precisão, h , isso sugere que uma priori hierárquica possa ser apropriada (Banerjee, Carlin e Gelfand, 2014; Jackman, 2009; Koop, 2003).

Uma priori hierárquica conveniente para o modelo de efeito aleatório,

$$\alpha_i \sim \text{Normal}(\mu_\alpha, V_\alpha), \quad i = 1, \dots, n. \quad (4.19)$$

com α_i e α_j são independentes para qualquer $i \neq j$. A estrutura hierárquica surge caso os parâmetros μ_α e V_α são tratados como desconhecidos que requerem suas próprias prioris. Supondo que as prioris são independentes com

$$\mu_\alpha \sim \text{Normal}(\underline{\mu}_\alpha, \underline{\sigma}_\alpha^2) \quad (4.20)$$

e

$$V_\alpha^{-1} \sim \text{Gama}(\underline{V}_\alpha^{-1}, \underline{\nu}_\alpha) \quad (4.21)$$

Para os demais parâmetros, assumu-se prioris não hierárquica com distribuição normal-gama. Portanto,

$$\tilde{\beta} \sim \text{Normal}(\underline{\beta}, \underline{V}_\beta) \quad (4.22)$$

e

$$h \sim \text{Gama}(\underline{s}^{-2}, \underline{\nu}) \quad (4.23)$$

Comparando com a abordagem clássica, tal estrutura hierárquica leva a um modelo análogo ao de efeito fixos (Koop, 2003).

Combinando a verossimilhança (4.12) com as prioris (4.19) e (4.23), pelo Teorema de Bayes, obtém-se a posteriori conjunta normal-gama de todos os parâmetros simultaneamente (Koop, 2003):

$$\tilde{\beta}|y, h, \alpha, \mu_\alpha, V_\alpha \sim \text{Normal}(\bar{\beta}, \bar{V}_\beta) \quad (4.24)$$

e

$$h|y, \tilde{\beta}, \alpha, \mu_\alpha, V_\alpha \sim \text{Gama}(\bar{s}^{-2}, \bar{\nu}) \quad (4.25)$$

em que

$$\begin{aligned} \bar{V}_\beta &= \left(\underline{V}_\beta^{-1} + h \sum_{i=1}^n \tilde{X}_i' \tilde{X}_i \right)^{-1} \\ \bar{\beta} &= \bar{V} \left(\underline{V}_\beta^{-1} \underline{\beta} + h \sum_{i=1}^n \tilde{X}_i' [y_i - \alpha \iota_T] \right) \\ \bar{\nu} &= Tn + \underline{\nu} \\ \bar{s}^2 &= \frac{\sum_{i=1}^n (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta})' (y_i - \alpha_i \iota_T - \tilde{X}_i \tilde{\beta}) + \underline{\nu} \underline{s}^2}{\underline{\nu}} \end{aligned}$$

A distribuição posteriori condicional para cada α_i é independente de α_j para $i \neq j$ e é dada por

$$\alpha_i|y, \tilde{\beta}, h, \mu_\alpha, V_\alpha \sim \text{Normal}(\bar{\alpha}_i, \bar{V}_i) \quad (4.26)$$

na qual

$$\bar{V}_i = \frac{V_\alpha h^{-1}}{TV_\alpha + h^{-1}}$$

e

$$\bar{\alpha} = \frac{V_\alpha (y_i - \tilde{X}_i \tilde{\beta})' \iota_T + h^{-1} \mu_\alpha}{(TV_\alpha + h^{-1})}$$

Finalmente, a distribuição posteriori condicional para os parâmetros hierárquicos, μ_α e V_α são

$$\mu_\alpha|y, \tilde{\beta}, h, \alpha, V_\alpha \sim \text{Normal}(\bar{\mu}_\alpha, \bar{\sigma}_\alpha^2) \quad (4.27)$$

e

$$V_\alpha^{-1} | y, \tilde{\beta}, h, \alpha, V_\alpha \sim \text{Gama}(\bar{V}_\alpha^{-1}, \bar{\nu}_\alpha) \quad (4.28)$$

na qual

$$\begin{aligned} \bar{\sigma}^2 &= \frac{V_\alpha \sigma_\alpha^2}{V_\alpha + n \sigma_\alpha^2} \\ \bar{\mu}_\alpha &= \frac{V_\alpha \mu_\alpha + \sigma_\alpha^2 \sum_{i=1}^n \alpha_i}{V_\alpha + n \sigma_\alpha^2} \\ \bar{\nu}_\alpha &= \nu_\alpha + n \\ \bar{V}_\alpha &= \frac{\sum_{i=1}^n (\alpha_i - \mu_\alpha)^2 + V_\alpha \nu_\alpha}{\bar{\nu}_\alpha} \end{aligned}$$

Ressalta-se que a inferência sobre cada parâmetro é feita com as distribuições marginais. Para tanto, a partir das distribuições a posteriori condicionais conjuntas (4.24) a (4.28), pode-se obter as distribuições a posteriori condicionais completas para cada parâmetro, para implementação do algoritmo do Amostrador de Gibbs, o qual permite obter aproximações das distribuições marginais.

5 Aplicação no R

O objetivo deste capítulo é reproduzir as estimações dos modelos de dados em painel no contexto clássico e bayesiano no ambiente computacional estatístico R ([R Core Team, 2016](#)) ou somente R como é conhecido entre os usuários. A escolha desse programa foi motivada por ser uma plataforma de *software* livre que funciona em diversos sistemas operacionais e que apresenta pacotes e funções disponíveis para a estimação de dados em painel.

Os métodos descritos nos Capítulos 2 e 4 são aplicados em dois conjunto de dados: FCInvBR usado por [Crisostomo, Iturriaga e González \(2014\)](#) e [Grunfeld](#) ([Kleiber e Zeileis, 2008b](#)) que é usado em diversos livros de econometria ([Gujarati e Porter, 2011](#); [Baltagi, 2008](#); [Greene, 2008](#); [Zellner, 1996](#)).

Este capítulo está subdividido em quatro seções. Na primeira seção descreve-se o conjunto de dados a ser implementado no modelos de dados agrupados e como desenvolver a estimação com o auxílio da função `lm()`. A seção seguinte apresenta um breve descrição da biblioteca `plm` desenvolvida por [Croissant e Millo \(2008\)](#) para utilização das técnicas básicas de dados em painel no contexto clássico. O emprego dos comandos desse pacote em um conjunto de dados reais pode ser verificada na Seção (5.3). Finalmente a Seção (5.4) traz a aplicação bayesiana para os modelos de dados em painel.

5.1 Modelo para dados agrupados

Nesta primeira aplicação examina-se um modelo que relaciona o investimento bruto real de uma empresa com a valor de mercado da empresa e o estoque de capital. Este conjunto de dados contém informações de 11 grandes empresas de americanas no período de 20 anos (de 1934 a 1954).

5.1.1 Descrição da base de dados

O conjunto de dados chamado `Grunfeld`, contém dados anuais para 11 companhias americanas, entre 1935 e 1954. O problema consiste em encontrar os determinantes do investimento por empresa, $invest_{it}$, entre os regressores, como o valor de empresa, $value_{it}$, e o estoque de capital ($capital_{it}$). Foi utilizado pela primeira vez por [Grunfeld, 1958](#), totalizando 220 observações. Trata-se de um conjunto de dados de painel balanceado de dados anuais coletados de 11 empresas americanas, entre 1935 e 1954. Este conjunto de dados está disponível no pacote `AER`. Pode-se carregar este conjunto de dados com o seguinte comando:

```
> data("Grunfeld", package = "AER")
> attach(Grunfeld)
```

Pode-se verificar a dimensão deste banco de dados com o comando:

```
> dim(Grunfeld)
```

```
[1] 220  5
```

Os nomes das variáveis do conjunto de dados podem ser obtidos a partir do comando `names`:

```
> names(Grunfeld)
```

```
[1] "invest" "value" "capital" "firm" "year"
```

O conjunto de dados inclui as seguintes variáveis ¹:

invest o investimento bruto

value o valor de mercado da empresa, definido como o preço da ação

capital estoque de capital

firm 11 General Motors, US Steel, General Electric, Chrysler, Atlantic Refining, IBM, Union Oil, Westinghouse, Goodyear, Diamond Match, American Steel

year período de tempo (1 = 1935, ..., 20 = 1954)

Um resumo das principais medidas descritivas de todas as variáveis do banco de dados é obtido a partir do comando `summary()`.

```
> summary(Grunfeld)
```

invest	value	capital	firm
Min. : 0.93	Min. : 30.28	Min. : 0.8	General Motors : 20
1st Qu.: 27.38	1st Qu.: 160.32	1st Qu.: 67.1	US Steel : 20
Median : 52.37	Median : 404.65	Median : 180.1	General Electric : 20
Mean : 133.31	Mean : 988.58	Mean : 257.1	Chrysler : 20
3rd Qu.: 99.78	3rd Qu.: 1605.92	3rd Qu.: 344.5	Atlantic Refining: 20
Max. : 1486.70	Max. : 6241.70	Max. : 2226.3	IBM : 20
			(Other) : 100
year			
Min. : 1935			
1st Qu.: 1940			
Median : 1944			

¹ Esta informação também está disponível na documentação deste conjunto de dados, que pode ser obtida através do comando `help("Grunfeld", package = "AER")`.

```

Mean    :1944
3rd Qu.:1949
Max.    :1954

```

A função `head()` apresenta as primeiras linhas do banco de dados:

```
> head(Grunfeld)
```

```

invest value capital      firm year
1  317.6 3078.5     2.8 General Motors 1935
2  391.8 4661.7    52.6 General Motors 1936
3  410.6 5387.1   156.9 General Motors 1937
4  257.7 2792.2   209.2 General Motors 1938
5  330.8 4313.2   203.4 General Motors 1939
6  461.2 4643.9   207.2 General Motors 1940

```

Enquanto que para visualização das últimas linhas do banco:

```
> tail(Grunfeld)
```

```

invest value capital      firm year
215  6.433 39.961  73.827 American Steel 1949
216  4.770 36.494  75.847 American Steel 1950
217  6.532 46.082  77.367 American Steel 1951
218  7.329 57.616  78.631 American Steel 1952
219  9.020 57.441  80.215 American Steel 1953
220  6.281 47.165  83.788 American Steel 1954

```

5.1.2 Modelo para dados agrupados *Pooled*

O estimador POOLED OLS ignora a estrutura de dados em painel, trata todas as observações como sendo não correlacionadas para um dado indivíduo, com erros homoscedásticos entre os indivíduos. Assim, todas as 220 observações são empilhadas, desprezando a natureza de corte transversal e de séries temporais dos dados.

Considere o modelo a seguir:

$$\begin{aligned}
 invest_{it} &= \alpha + \beta_1 value_{it} + \beta_2 capital_{it} + \varepsilon_{it} \\
 i &= 1, 2, \dots, 11 \\
 t &= 1, 2, \dots, 20
 \end{aligned} \tag{5.1}$$

em que i é a i -ésima empresa e t é o período de tempo para as variáveis definidas anteriormente.

Neste caso, combinam-se todas as 220 observações, mas pressupondo que os coeficientes de regressão sejam os mesmos para todas as empresas. Ou seja, não há distinção entre as empresas americanas, uma empresa é tão boa quanto a outra. Supõe-se ainda que o termo de erro seja $\varepsilon_{it} \sim iid(0, \sigma_\varepsilon^2)$, isto é, que ele seja distribuído idêntica e independentemente com média zero e variância constante.

O comando para estimar modelos lineares no R é `lm`. O primeiro argumento do comando `lm` especifica o modelo que deve ser estimado. Este deve ser um objeto de fórmula que consiste no nome da variável dependente, seguido por um til (`~`) e o(s) nome(s) da(s) variável(is) explicativa(s). O argumento `data` especifica o conjunto de dados. Neste caso, a variável dependente é `invest` e as variáveis explicativas, `value` e `capital`.

```
> ols <- lm( invest ~ value + capital, data = Grunfeld)
> ols
```

Call:

```
lm(formula = invest ~ value + capital, data = Grunfeld)
```

Coefficients:

(Intercept)	value	capital
-38.4101	0.1145	0.2275

Com o comando `summary` pode-se obter mais informações sobre a regressão linear.

```
> summary(ols)
```

Call:

```
lm(formula = invest ~ value + capital, data = Grunfeld)
```

Residuals:

Min	1Q	Median	3Q	Max
-290.33	-25.76	11.06	29.74	377.94

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-38.410054	8.413371	-4.565	8.35e-06	***
value	0.114534	0.005519	20.753	< 2e-16	***
capital	0.227514	0.024228	9.390	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 90.28 on 217 degrees of freedom

Multiple R-squared: 0.8179, Adjusted R-squared: 0.8162
F-statistic: 487.3 on 2 and 217 DF, p-value: < 2.2e-16

Verifica-se nos resultados da estimação POOLED OLS que os coeficientes são altamente significativos e também que o valor de R^2 é muito alto. Destaca-se que esse modelo não faz distinção entre as diversas empresas nem diz se a resposta do investimento às variáveis explicativas ao longo do tempo é a mesma para todas as empresas. Assim, ao agrupar diferentes empresas em períodos diferentes, camufla-se a heterogeneidade que possa existir entre as empresas.

5.2 Pacote *plm*

Nesta seção aborda-se a estimação clássica dos modelos de dados em painel utilizando alguns comandos básicos disponibilizados pelo R e pelo pacote *plm* (Croissant e Millo, 2008). Este pacote fornece uma série de funções e estruturas de dados que são especialmente projetadas para dados em painel, esta biblioteca é carregada usando:

```
> library(plm)
```

Neste pacote os dados são armazenados em um objeto da classe `pdata.frame`, que é um `data.frame` com atributos adicionais que descrevem a estrutura dos dados. Portanto, faz-se necessário a transformação do conjunto de dados para um formato adequado para usar as funções do pacote *plm*, isto ocorre a partir da função `pdata.frame`.

Um `pdata.frame` pode ser criado a partir de `data.frame` usando a função `pdata.frame`. A função `pdata.frame` tem 2 argumentos principais:

- o nome do `data.frame`,
- `index`: um vetor (de tamanho um ou dois) indicando os índices individual e temporal. Assim, é preciso especificar um atributo para a dimensão individual e outro para a dimensão temporal.

Ilustra-se o uso da função `pdata.frame` com os dados Grunfeld. Neste exemplo, são definidos a dimensão individual (variável *firm*) e a temporal (variável *year*):

```
> data(Grunfeld)
> Grun <- pdata.frame(Grunfeld, index = c("firm", "year"))
```

Na estrutura básica da função *plm* deve-se indicar a fórmula do modelo, os dados e o tipo de modelo a ser estimado, ou seja

```
plm(formula, data, model = c("within", "random", "ht", "between", "pooling",  
"fd"), em que:
```

- `formula`: representa a descrição simbólica do modelo a ser estimado,
- `data`: o objeto `pdata.frame` que contém os dados,
- `model`: o tipo de modelo a estimar. Vários modelos podem ser estimados com a função `plm`, por exemplo:
 - modelo para dados agrupados: `pooling`,
 - modelo de efeito fixo: `within`,
 - modelo primeira-diferenças: `fd`,
 - modelo between: `between`,
 - modelo de efeito aleatório: `random`.

O comportamento padrão da função `plm` é estimar os modelos utilizando o efeito individual, adicionando o argumento `effect` pode-se também apresentar:

- o efeito temporal: `effect=time`,
- o efeito individual e temporal: `effect=twoways`.

5.2.1 Testes para efeito individual e efeito temporal

A biblioteca `plm` tem implementado alguns testes para comparação entre os modelos. A Figura (7) apresenta os testes e quais os modelos que eles comparam.

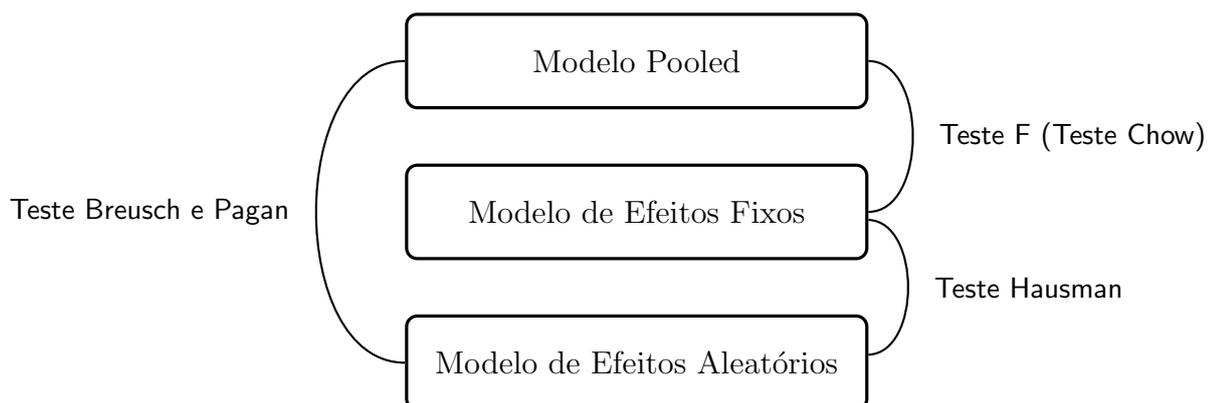


Figura 7 – Testes de comparação entre os modelos

O Teste F ou teste F de Chow para efeito individual e/ou temporal é baseado na comparação entre o modelo de efeito fixo (*within*) e o modelo para dados agrupados (*pooled*). A hipótese nula deste teste é de que há igualdade de interceptos e inclinações para todos os indivíduos, que corresponde a característica do modelo de dados agrupados. A função dedicada para este teste é `pFtest()`.

Breusch e Pagan (1980) desenvolveram um teste baseado no multiplicador de Lagrange para confrontar as estimativas entre o modelo de dados agrupados e modelos de efeitos aleatórios. O teste Breusch e Pagan consiste em verificar se $\sigma_\alpha^2 = 0$, as hipóteses definidas para esse teste são:

$$H_0 : \sigma_\alpha^2 = 0$$
$$H_1 : \sigma_\alpha^2 \neq 0$$

Caso a hipótese nula seja aceita, o modelo para dados agrupados é preferível ao modelo de efeitos aleatórios. O comando para execução desse teste é `plmtest` adicionado pelo argumento `type = bp`.

A escolha do modelo mais adequado se o de efeito fixo ou efeito aleatório pode ser feita através de um teste de especificação, o teste de Hausman (Hausman, 1978). De acordo com Baltagi (2008) o teste de Hausman testa a hipótese:

$$H_0 : \alpha_i \text{ não são correlacionados com } X_{it}$$
$$H_1 : \alpha_i \text{ são correlacionados com } X_{it}$$

Caso se rejeite a hipótese nula, o modelo de efeito fixo é o mais adequado. O teste de Hausman é dado pela função `phptest()`, em que os argumentos consistem nos modelos de efeitos fixo e efeito aleatório.

Na próxima seção detalha-se a utilização do pacote `plm` com o auxílio de um conjunto de dados e a utilização dos testes de comparação entre os modelos, por exemplo, o `pFtest`.

5.3 Modelos de dados em painel com `plm`

Na seção anterior descreve-se a mecânica e as funções básicas do pacote `plm`. Nesta seção, ilustra-se como estimar os modelos de dados em painel – modelo de efeitos fixos e modelo de efeitos aleatórios – com o auxílio do pacote `plm` utilizando um conjunto de dados reais.

A ilustração desta seção considera os dados de uma subamostra do conjunto de dados `FCInvBR` discutido no artigo de Crisostomo, Iturriaga e González (2014). O conjunto original contém informações anuais de 199 empresas brasileiras não financeiras no período de 12 anos, entre 1995 e 2006. A subamostra é um painel balanceado e representa os dados de 8 empresas no mesmo período de tempo.

Com isso, estimam-se dois modelos de regressão para dados em painel em que a variável dependente é o investimento e há quatro regressores: o fluxo de caixa, o nível de produção, dívida e o Q de Tobin. A Tabela (3) apresenta a descrição das variáveis. Ressalta-se que essas regressões não são estimadas por Crisostomo, Iturriaga e González

(2014), uma vez que o objetivo não é replicar os resultados encontrados por estes autores, mas exemplificar os recursos do pacote `plm` utilizando um conjunto de dados reais.

Na Seção (5.3.1) apresentam-se o banco de dados que fornece dados financeiros de uma amostra de empresas brasileiras para o período de 1995- 2006, a qual servirá de base de dados para a ilustração do uso do pacote `plm` e também o modelo de investimento que será estimado na próxima subseção pela função `plm` deste pacote. A regressão com efeitos fixos é uma extensão da regressão linear múltipla que explora dados em painel para o controle de variáveis que diferem entre entidades, mas não constantes ao longo do tempo (Stock e Watson, 2006). A regressão com efeitos fixos será apresentada na Seção (5.3.2). Na Seção (5.3.3) é tratado o modelo de efeitos aleatórios, em que os efeitos individuais são tratados como variáveis aleatórias em vez de constantes fixas.

5.3.1 Descrição dos dados e o modelo de investimento

No Capítulo 2 foi dito que dados em painel referem-se a dados de n entidades diferentes observadas em T períodos de tempo diferentes. Os conjuntos de dados examinados nesta seção são dados em painel.

Estes dados anuais foram extraídos do programa Economática, que fornece dados do balanço patrimonial das empresas com ações negociadas na Bolsa de Valores de São Paulo (BM&FBOVESPA). Foram coletadas diversas informações dos demonstrativos financeiros, por exemplo: ativo total, ativo imobilizado, depreciações, entre outras, estas informações compõem o cálculo das variáveis dos modelos.

Na descrição de dados em painel para acompanhar tanto a entidade quando o período de tempo utilizam-se dois subscritos: o primeiro, i , refere-se à entidade, e o segundo, t , refere-se ao período de tempo da observação. Portanto, Y_{it} representa a variável Y observada para a i -ésima das n entidades no t -ésimo dos T períodos.

Os dados em painel desta seção referem-se a $n = 8$ entidades (empresas), nos quais a cada entidade é observada em $T = 12$ períodos de tempo (cada um dos anos, 1995, ..., 2006), totalizando $8 \times 12 = 96$ observações.

Com o auxílio do R confirma-se que a base de dados sobre os investimentos corporativos inclui observações para todas as 8 firmas para todos os 12 anos, de modo que é um painel balanceado. Se, entretanto, faltassem dados, por exemplo, caso não tivessem dados sobre os investimentos para algumas empresas em 2003, a base de dados seria um painel desbalanceado. Os métodos apresentados neste trabalho são descritos para um painel balanceado; contudo, todos esses métodos podem ser utilizados em um painel desbalanceado.

O modelo de investimento que será utilizado nesta e na próxima seção é uma

versão adaptada do modelo proposto por [Crisostomo, Iturriaga e González \(2014\)](#)²

$$Inv_{it} = \beta_1 FC_{it} + \beta_2 R_{it} + \beta_3 D_{it} + \beta_4 Q_{it} + \alpha_i + \varepsilon_{it}, \quad (i = 1, \dots, 8; t = 1, \dots, 12). \quad (5.2)$$

Em que Inv_{it} é o investimento da firma i no ano t , FC_{it} representa a variável fluxo de caixa da firma i no ano t , R_{it} é o faturamento da empresa i no ano t , D_{it} as dívidas da firma i no ano t , Q_{it} Q de Tobin da firma i no ano t , α_i ($i = 1, \dots, 8$) é o termo relacionado com efeitos fixos da empresa, ou seja, captura os fatores não observáveis específicos da empresa e constantes ao longo do tempo, tratados como interceptos desconhecidos a serem estimados, um para cada firma; $\varepsilon_{i,t}$ se refere a erros aleatórios. A Tabela (3) expõe a descrição das variáveis $Inv_{it}, FC_{it}, R_{it}, D_{it}, Q_{it}$ usados no modelo de investimento.

Tabela 3 – Descrição das variáveis utilizadas no modelo de investimento

Variável	Descrição
Investimento (Inv)	Diferença entre o estoque de capital atual e o estoque no período anterior
Fluxo de caixa (FC)	Soma do lucro líquido e as depreciações
Nível de produção (R)	Aproximado pelo faturamento da empresa
Dívida (D)	Dívida bancária da empresa
Q de Tobin (Q)	Capacidade da empresa gerar valor com investimento

Na Figura (8) ilustra-se a heterogeneidade entre as firmas e entre os anos. Observa-se que o ano de 2003 apresenta um maior intervalo de variação da variável investimento, o mesmo ocorrendo com a empresa 243. A análise desta empresa leva a conclusão de que trata-se de um valor atípico (*outlier*) que apresenta para observações discrepantes das demais empresas da amostra.

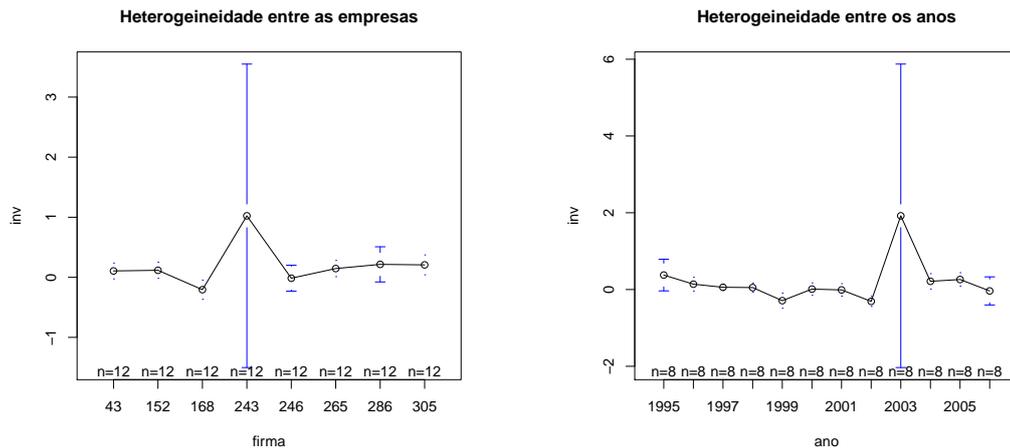


Figura 8 – Heterogeneidade entre firmas e entre os anos

² O modelo aqui postulado é adequado ao propósito desta seção que é apresentar um modelo de dados em painel a ser estimado utilizando os recursos do R.

O comportamento da variável dependente (Inv_{it}) através do anos para cada uma das empresas pode ser observado na Figura (9) a seguir.

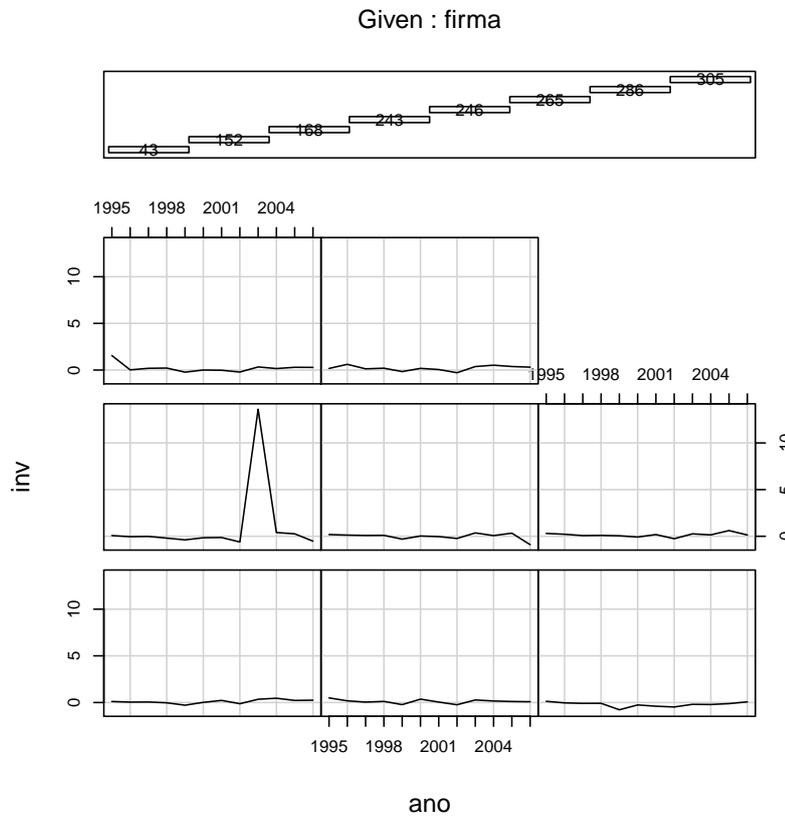


Figura 9 – Investimento ao longo dos anos por firma

No modelo descrito na Equação (5.2) busca-se identificar o papel das principais variáveis que possam impactar sobre as decisões de investimento da empresas, tais como o fluxo de caixa, receita e financiamentos. Este modelo será estimado a seguir utilizando as funções do pacote `plm`.

5.3.2 Efeitos fixos

Considere o modelo de investimento dado pela Equação (5.2) com a variável dependente (Investimento) e os regressores observados (Fluxo de caixa, Nível de Produção, Dívida e Q de Tobin) representados na forma geral de um modelo de regressão de dados em painel ao considerar:

$$Inv_{it} = Y_{it}, \quad FC_{it} = X_{1it}, \quad R_{it} = X_{2it}, \quad D_{it} = X_{3it}, \quad Q_{it} = X_{4it} \quad (5.3)$$

O modelo torna-se então:

$$Y_{it} = \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + \alpha_i + \varepsilon_{it} \quad (5.4)$$

em que $i = 1, \dots, 8$; $t = 1, \dots, 12$ (painel balanceado); X_{kit} representam as covariáveis para $k = 1, \dots, 4$; $\varepsilon_{it} \sim \text{Normal}(0, \sigma^2)$.

A Equação (5.4) é o modelo de regressão com efeitos fixos, em que $\alpha_1, \dots, \alpha_8$ são tratados como interceptos desconhecidos a serem estimados um para cada firma.

A interpretação de α_i como um intercepto específico para cada firma na Equação (5.4) vem de se considerar a reta de regressão da população para a i -ésima firma. Ressalta-se que os coeficientes de declividade da reta de regressão da população, $\beta_1, \beta_2, \beta_3, \beta_4$ são os mesmos para todas as firmas, mas o intercepto da regressão da população varia de uma firma para a outra, mas é constante ao longo do tempo.

A estrutura de dados em painel para esta aplicação significa que uma observação é definida pelo valor de duas variáveis: a firma e o ano. Em geral, se refere à unidade de corte transversal como a “entidade” (neste caso, a firma) e a variável tempo como a variável “tempo” ou “período” (neste caso, ano). Esta informação deve ser inserida no pacote *plm*.

Para utilizar a biblioteca *plm* referente a dados em painel, é necessário transformar o conjunto de dados. Para isso, em cada comando *plm*, `index=c("firma", "ano")` se define a primeira variável (firma) como a entidade e o segundo (ano) como variável tempo, ou seja,

```
> painel <- plm.data(amostra, c("firma", "ano"))
```

A próxima etapa é estimar a regressão de efeitos fixos. O código *plm* para a estimação de efeitos fixos é dada por:

```
> ef<-plm(inv ~ fcl + recl + debl + qtl, data = painel, model="within")
> summary(ef)
```

```
Oneway (individual) effect Within Model
```

```
Call:
```

```
plm(formula = inv ~ fcl + recl + debl + qtl, data = painel, model = "within")
```

```
Balanced Panel: n=8, T=12, N=96
```

```
Residuals :
```

```
    Min. 1st Qu.  Median 3rd Qu.    Max.
-0.8990 -0.2240  0.0141  0.1830  1.6700
```

```
Coefficients :
```

```
      Estimate Std. Error t-value Pr(>|t|)
fcl  -0.01957300  0.00255876  -7.6494 3.038e-11 ***
recl  0.01995603  0.00227239   8.7820 1.630e-13 ***
debl  0.01034751  0.00088812  11.6510 < 2.2e-16 ***
qtl  -0.30515635  0.03763089  -8.1092 3.666e-12 ***
```

```
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 180.37
 Residual Sum of Squares: 15.321
 R-Squared: 0.91506
 Adj. R-Squared: 0.90394
 F-statistic: 226.235 on 4 and 84 DF, p-value: < 2.22e-16

O comando `summary` mostra que tem-se 8 indivíduos ao longo de 12 anos, que dá um total de 96 observações. Trata-se de um painel balanceado. Com este modelo eliminam-se os termos que são constantes ao longo do tempo, incluindo o termo constante que pertence ao resíduo. O resultado da estimação indica que todas as covariáveis são significativas, as variáveis fluxo de caixa e Q de Tobin com um efeito negativo sobre o investimento.

Pode-se verificar os efeitos fixos (constantes para cada firma) com o seguinte comando:

```
> fixef(ef)
      43      152      168      243      246      265
0.41092027 0.18568348 -0.13112928 -2.56277820 0.05470329 0.32282391
      286      305
0.09518862 0.39720102
```

Um resumo dos efeitos individuais e dos os erros é mostrado a seguir

```
> summary(fixef(ef))
      Estimate Std. Error t-value Pr(>|t|)
43  0.410920  0.129242  3.1795 0.002067 **
152 0.185683  0.124700  1.4890 0.140221
168 -0.131129 0.167624 -0.7823 0.436247
243 -2.562778 0.201781 -12.7008 < 2.2e-16 ***
246 0.054703  0.123850  0.4417 0.659850
265 0.322824  0.125789  2.5664 0.012050 *
286 0.095189  0.131422  0.7243 0.470897
305 0.397201  0.126051  3.1511 0.002254 **
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Os interceptos específicos para cada firma no modelo de regressão com efeitos fixos também podem ser expressos pela utilização de variáveis *dummy* para representar as firmas individuais com o auxílio da função `lm`.

Call:

```
lm(formula = inv ~ fcl + recl + debl + qtl + as.factor(firma) -
    1, data = amostra)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.89884	-0.22440	0.01413	0.18312	1.66653

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
fcl	-0.0195730	0.0025588	-7.649	3.04e-11	***
recl	0.0199560	0.0022724	8.782	1.63e-13	***
debl	0.0103475	0.0008881	11.651	< 2e-16	***
qtl	-0.3051564	0.0376309	-8.109	3.67e-12	***
as.factor(firma)43	0.4109203	0.1292422	3.179	0.00207	**
as.factor(firma)152	0.1856835	0.1246998	1.489	0.14022	
as.factor(firma)168	-0.1311293	0.1676235	-0.782	0.43625	
as.factor(firma)243	-2.5627782	0.2017806	-12.701	< 2e-16	***
as.factor(firma)246	0.0547033	0.1238504	0.442	0.65985	
as.factor(firma)265	0.3228239	0.1257890	2.566	0.01205	*
as.factor(firma)286	0.0951886	0.1314224	0.724	0.47090	
as.factor(firma)305	0.3972010	0.1260507	3.151	0.00225	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4271 on 84 degrees of freedom

Multiple R-squared: 0.9215, Adjusted R-squared: 0.9102

F-statistic: 82.13 on 12 and 84 DF, p-value: < 2.2e-16

Na sequência, por meio do teste `pFtest` compara-se o modelo *Pooled* com o *within*.

```
> pooled<-plm(inv ~ fcl + recl + debl + qtl, data = painel, model="pooling")
> ef<-plm(inv ~ fcl + recl + debl + qtl, data = painel, model="within")
> pFtest(ef,pooled)
```

F test for individual effects

data: inv ~ fcl + recl + debl + qtl

F = 27.105, df1 = 7, df2 = 84, p-value < 2.2e-16

alternative hypothesis: significant effects

O teste resultou em um p-value $< 2.2e-16$, que aponta a rejeição da hipótese nula, portanto o modelo *within* é o mais apropriado.

5.3.3 Efeitos aleatórios

No modelo de efeitos aleatórios, os efeitos individuais (α_i) são considerados variáveis aleatórias em vez de como constantes fixas. Assume-se que os α_i são independentes dos erros ε_{it} e também são mutuamente independentes. Portanto, deve-se assumir que

$$\begin{aligned}\alpha_i &\overset{iid}{\sim} \text{com média } 0 \text{ e variância } \sigma_\alpha^2 \\ \varepsilon_{it} &\overset{iid}{\sim} \text{com média } 0 \text{ e variância } \sigma_\varepsilon^2\end{aligned}$$

e que α_i e ε_{it} são independentes. (iid significa independentes e identicamente distribuído.)

O modelo de efeitos aleatórios é o mesmo da Equação (5.4), exceto pelo fato de que α_i são variáveis aleatórias, ou seja,

$$Y_{it} = \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + \alpha_i + \varepsilon_{it} \quad (5.5)$$

Portanto, em vez de considerar α_i como fixo presupõe-se que ele seja uma variável aleatória. O estimador de efeitos aleatórios é calculado pelo função `plm` inserindo no argumento `model` a opção `random`, conforme dado a seguir

```
> ea<-plm(inv ~ fcl + recl + debl + qtl, data = painel, model="random")
> summary(ea)
```

```
Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)
```

Call:

```
plm(formula = inv ~ fcl + recl + debl + qtl, data = painel, model = "random")
```

```
Balanced Panel: n=8, T=12, N=96
```

Effects:

	var	std.dev	share
idiosyncratic	0.18239	0.42707	1.05
individual	-0.00862	NA	-0.05
theta:	-0.5199		

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-2.860	-0.250	0.065	0.332	4.120

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)	
(Intercept)	0.0993876	0.0774011	1.2841	0.202381	
fcl	-0.0322593	0.0044614	-7.2307	1.452e-10	***
recl	0.0220431	0.0041067	5.3675	6.052e-07	***
debl	0.0050897	0.0015050	3.3819	0.001063	**
qtl	-0.2739334	0.0544420	-5.0317	2.435e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 205.55

Residual Sum of Squares: 62.807

R-Squared: 0.69444

Adj. R-Squared: 0.68101

F-statistic: 51.7044 on 4 and 91 DF, p-value: < 2.22e-16

O modelo de efeito aleatório apresenta o termo de erro composto. Logo, o resultado da estimação dos modelo de efeito aleatório fornece informações sobre a variância dos componentes dos erros, um referente ao componente de corte transversal ou específico dos indivíduos representado por *individual*, o outro termo idiossincrático, que varia com o corte transversal e também com o tempo (*idiosyncratic*).

Com o uso do teste de Hausman pode-se decidir entre o modelo de efeito fixo ou aleatório. A hipótese nula é que os efeitos individuais não estão correlacionados com os regressores. Se a hipótese nula for rejeitada, a conclusão é que o modelo de efeitos aleatórios não é adequado, porque provavelmente os efeitos individuais aleatórios estão correlacionados com um ou mais regressores. Nesse caso, o modelo de efeitos fixos é preferível aos de efeitos aleatórios.

O pacote *plm* oferece o comando *phtest* para o teste fr Hausman automático. O comando utiliza as estimativas dos modelos de efeitos fixo e efeito aleatório obtidas anteriormente, que foram armazenadas nos objetos *ef* e *ea*, respectivamente, e realiza o teste de Hausmann.

```
> phtest(ef,ea)
```

```
Hausman Test
```

```
data: inv ~ fcl + recl + debl + qtl
```

```
chisq = 817.89, df = 4, p-value < 2.2e-16
```

```
alternative hypothesis: one model is inconsistent
```

No caso da aplicação desta seção, a utilização do teste de Hausman auxilia na rejeição da hipótese nula de que o modelo de efeitos aleatórios oferece estimativas dos parâmetros mais consistentes, conforme comando e resultado apresentado a seguir. Portanto, como resultado deve-se preferir o modelo de efeito fixo. A seguir apresentam-se os coeficientes de dados agrupados (*pooled*), efeitos aleatórios e efeitos fixos de cada variável explicativa.

Modelo de regressão e erros

```

=====
                        Dependent variable:
                        -----
                                inv
                                Pooled      EA      EF
                                (1)        (2)      (3)
                        -----
fcl                    -0.0301***  -0.0323***  -0.0196***
                        (0.0041)   (0.0045)   (0.0026)

recl                   0.0223***  0.0220***  0.0200***
                        (0.0038)   (0.0041)   (0.0023)

debl                   0.0062***  0.0051***  0.0103***
                        (0.0014)   (0.0015)   (0.0009)

qtl                    -0.2609***  -0.2739***  -0.3052***
                        (0.0551)   (0.0544)   (0.0376)

Constant              0.0192      0.0994
                        (0.0936)   (0.0774)

                        -----
Observations          96          96          96
R2                   0.7390      0.6944      0.9151
Adjusted R2          0.7275      0.6810      0.9039
F Statistic          64.4063***  51.7044***  226.2348***
=====
Note:                  *p<0.1; **p<0.05; ***p<0.01

```

Os resultados sem os erros podem ser apresentados a partir do comando

```
> results <- round(data.frame("Pooled"=pooled$coefficients[2:5],
+ "Efeitos fixos"=ef$coeff[1:4],
+ "Efeitos aleatórios"=ea$coeff[2:5]),4)
> results
```

	Pooled	Efeitos.fixos	Efeitos.aleatórios
fc1	-0.0301	-0.0196	-0.0323
rec1	0.0223	0.0200	0.0220
debl	0.0062	0.0103	0.0051
qt1	-0.2609	-0.3052	-0.2739

Observa-se que os coeficientes estimados variam de modelo para modelo. Além disso, que o vetor de regressores apresenta significância estatística em todos os modelos. Verifica-se a existência de maior R^2 e também os menores erros para o modelo de efeito aleatório, o que valida o teste de Hausman que indica a preferência desse modelo em relação aos demais.

Realizadas as estimações pela inferência clássica, na próxima seção realizam-se as estratégias de inferência bayesiana com os mesmos conjuntos de dados aqui utilizados.

5.4 Abordagem Bayesiana para Dados em Painel

Esta seção tem por objetivo ilustrar a aplicação da estimação dos modelos de dados agrupados, efeito fixo e aleatório sob o enfoque bayesiano para uma subamostra do conjunto de dados discutido no artigo de [Crisostomo, Iturriaga e González \(2014\)](#). Na subseção (5.4.1) desenvolvem-se as estratégias de estimação para o modelo de dados agrupados, enquanto que na subseção (5.4.2) as estratégias para os modelos de efeito fixo e aleatório sob a perspectiva bayesiana,

Ressalta-se que não há pretensão em discutir as vantagens ou desvantagens da utilização de métodos bayesianos em relação aos métodos clássicos. Também não é o propósito definir qual o melhor modelo a ser utilizado entre os modelos de dados em painel. Isso devido ao fato que, na prática, a escolha de um ou outro modelo depende da situação em que se está trabalhando e das variáveis que estão sendo utilizadas no modelo.

5.4.1 Modelo para dados agrupados

Nesta subseção, tem-se o interesse em desenvolver a estratégia de estimação bayesiana para um modelo de investimento apresentado na subseção (5.3.1). Para estimação dos parâmetros deste modelo foram utilizados dados de uma subamostra do conjunto de dados discutido no artigo de [Crisostomo, Iturriaga e González \(2014\)](#), em que o interesse foi verificar se o investimento corporativo está associados com as variáveis fluxo de caixa,

nível de produção, dívida e a capacidade da empresa gerar valor com investimento (Q de Tobin).

Conforme a Seção (2.2) o modelo de dados agrupados se reduz ao modelo de regressão linear múltipla. Dessa forma, o modelo de investimento para verificar a associação entre a variável dependente investimento (Y_{it}), em relação às variáveis independentes fluxo de caixa, nível de produção, dívida e Q de Tobin, será dado pelo modelo de regressão linear múltipla:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i, \quad i = 1, \dots, 96. \quad (5.6)$$

em que n é o número de indivíduos, y_i é a observação da variável dependente para o i -ésimo indivíduo, $X_i = (x_{1i}, x_{2i}, \dots, x_{ki})'$ é um vetor de observações das variáveis independentes para o i -ésimo indivíduo, $\beta = (\beta_0, \beta_1, \dots, \beta_k)'$ é um vetor de coeficientes de regressão (parâmetros) e ε_i é um componente de erro aleatório. Assume-se que os erros são independentes e seguem uma distribuição normal com média zero e variância desconhecida σ^2 .

Nesta aplicação em que $y_i = Inv_i$, $x_{1i} = FC_i$, $X_{2i} = R_i$, $X_{3i} = D_i$, $X_{4i} = Q_i$, $i = 1, \dots, 96$ e ε_i representa o erro aleatório do i -ésimo indivíduo, em que esses erros são independentes e seguem distribuição normal com média zero e variância desconhecida σ^2 .

Para a análise bayesiana dos dados da subamostra, e considerando-se o modelo definido em (5.6), e as seguintes distribuições a priori já declaradas na Equação (4.8) para $\beta = (\beta_0, \beta_1, \dots, \beta_4)$ e $h = \frac{1}{\sigma^2}$,

$$\begin{aligned} \beta_k &\sim \text{Normal}(a_k; b_k^2), \quad a \text{ e } b \text{ conhecidos}, k = 0, 1, \dots, 4 \\ h &\sim \text{Gama}(c; d), \quad c \text{ e } d \text{ conhecidos} \end{aligned} \quad (5.7)$$

Assim, a distribuição a priori para β e σ^2 são dadas por:

$$\begin{aligned} \beta_k &\sim N(a_k; b_k^2), \quad a \text{ e } b \text{ conhecidos}, k = 0, 1, \dots, 4 \\ \sigma^2 &\sim \text{GI}(c; d), \quad c \text{ e } d \text{ conhecidos} \end{aligned} \quad (5.8)$$

em que $N(a_k; b_k^2)$ denota uma distribuição normal com média a e variância b^2 e $\text{GI}(c; d)$ denota uma distribuição gama inversa com média $\frac{d}{c-1}$ e variância $\frac{d^2}{(c-1)^2(c-2)}$.

Além disso, foi assumido independência a priori para os parâmetros. Assim, a distribuição a priori conjugada é dada por,

$$h(\beta, \sigma^2) \propto \prod_{k=0}^4 \exp\left[-\frac{1}{2b_k^2}(\beta_k - a_k)^2\right] \times (\sigma^2)^{-(c+1)} \exp\left(-\frac{d}{\sigma^2}\right) \quad (5.9)$$

Considerando o modelo (5.6), a função de verossimilhança para os dados observados y_i segundo os parâmetros $\beta = (\beta_0, \dots, \beta_4)$ e σ^2 as covariáveis X_{ik} , $i = 1, \dots, 96$, $k =$

$0, \dots, 4$, é dada por,

$$\begin{aligned} p(y_i | \beta, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} \exp \left[-\frac{1}{2} \frac{(y_i - \mu_i)^2}{\sigma^2} \right] \\ &= \frac{1}{(\sqrt{2\pi\sigma})^n} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2 \right] \end{aligned} \quad (5.10)$$

em que

$$\mu_i = \beta_0 + \sum_{k=1}^4 \beta_k X_{ik}, \quad i = 1, \dots, n$$

Conforme (3.8) a distribuição a posteriori conjunta para os parâmetros é obtida combinando-se a distribuição a priori com a função de verossimilhança a partir da regra de Bayes.

As distribuições a posteriori condicionais são apresentadas a seguir:

Para β , em que $k = 0, \dots, 4$,

$$h(\beta | y) \propto \exp \left[-\frac{1}{2b_k^2} (\beta_k - a_k)^2 - \frac{1}{2\sigma^2} \left(y_i - \beta_0 - \sum_{k=1}^4 \beta_k X_{ik} \right) \right] \quad (5.11)$$

Para σ^2 ,

$$h(\sigma | y) \propto (\sigma^2)^{-(c+1)} \frac{1}{(\sqrt{2\pi\sigma})^n} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{k=1}^4 \beta_k X_{ik} \right) - \frac{d}{\sigma^2} \right] \quad (5.12)$$

Para a análise bayesiana dos dados, foi considerado os seguintes hiperparâmetros para as distribuições a priori dadas em (5.8), $a_k = 0$, $b_k^2 = 10^2$, $k = 0, 1, \dots, 4$, $c = d = 0.001$. Essa escolha foi motivada para se ter distribuições a priori não informativas.

As estimativas dos parâmetros do modelo (5.6) foram obtidas através da função `MCMCregress()` da biblioteca `MCMCpack` (Martin, Quinn e Park, 2011) do *software* R a partir da rotina a seguir.

```
> pooledbayes<-MCMCregress(inv ~ fcl + recl + debl + qtl,
+ data=amostra,b0=0,B0=100,c0=0.001,d0=0.001)
> summary(pooledbayes)
```

```
Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000
```

1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

Mean	SD	Naive SE	Time-series SE
------	----	----------	----------------

(Intercept)	-0.018821	0.067841	6.784e-04	6.784e-04
fcl	-0.028458	0.004117	4.117e-05	4.117e-05
recl	0.022619	0.003845	3.845e-05	3.907e-05
debl	0.005557	0.001394	1.394e-05	1.394e-05
qtl	-0.203077	0.046788	4.679e-04	4.679e-04
sigma2	0.564776	0.086335	8.634e-04	9.092e-04

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
(Intercept)	-0.152465	-0.064131	-0.018427	0.027150	0.112543
fcl	-0.036464	-0.031240	-0.028453	-0.025694	-0.020435
recl	0.015170	0.020033	0.022679	0.025159	0.030218
debl	0.002851	0.004625	0.005556	0.006489	0.008285
qtl	-0.296228	-0.233880	-0.203550	-0.172092	-0.111611
sigma2	0.419818	0.503370	0.555748	0.616662	0.754317

A Tabela (4) mostra as estimativas dos parâmetros obtidas através do método bayesiano e do método dos mínimos quadrados ordinários e seus respectivos erros padrão. Ressalta-se que as estimativas bayesianas correspondem as médias da distribuição a posteriori, obtidas via algoritmo *Gibbs-Sampling*. Enquanto que as estimativa de mínimos quadrados ordinário (EMQ) foram estimadas através do comando `lm()` do R.

Tabela 4 – Resultados obtidos através do método bayesiano e do MQO

Parâmetro	Média a posteriori	Erro-padrão	EMQ	Erro-padrão
β_0	-0.018821	0.067841	0.019151	0.093629
β_1	-0.028458	0.004117	-0.030053	0.004119
β_2	0.022619	0.003845	0.022314	0.003801
β_3	0.005557	0.001394	0.006168	0.001404
β_4	-0.203077	0.046788	-0.260920	0.055081

Um critério de avaliação dos resultados obtidos por inferência bayesiana é o diagnostico de convergência das cadeias simuladas. Uma análise de convergência pode ser feita preliminarmente verificando gráficos ou medidas descritivas dos valores simulados dos parâmetros de interesse. Entre os gráficos usuais para essa análise estão o da estimativa da distribuição a posteriori do parâmetro de interesse θ , por exemplo a densidade kernel e o gráfico de θ ao longo das iterações.

A Figura (10) ilustra as densidades estimadas de cada um dos parâmetros: $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \sigma^2$. Já a Figura (11) apresenta o gráfico das iterações para os parâmetros.

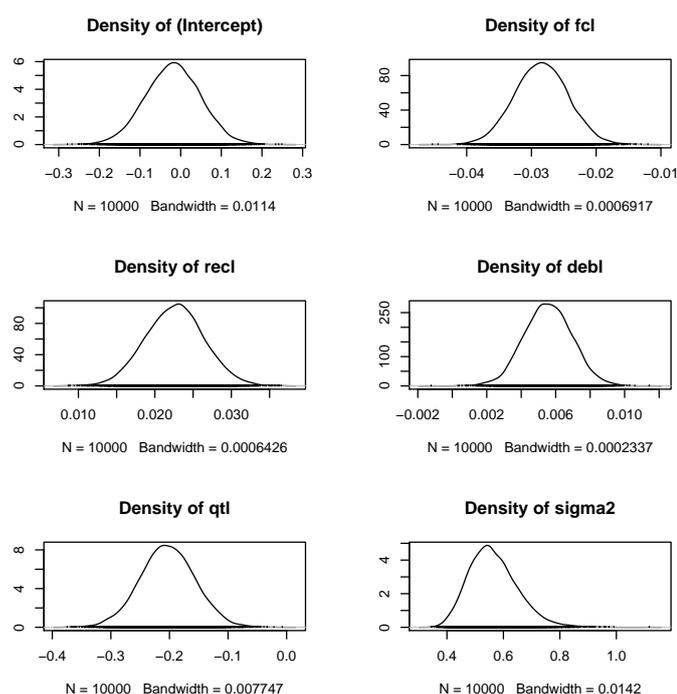


Figura 10 – Densidades estimadas para as posteriors dos parâmetros

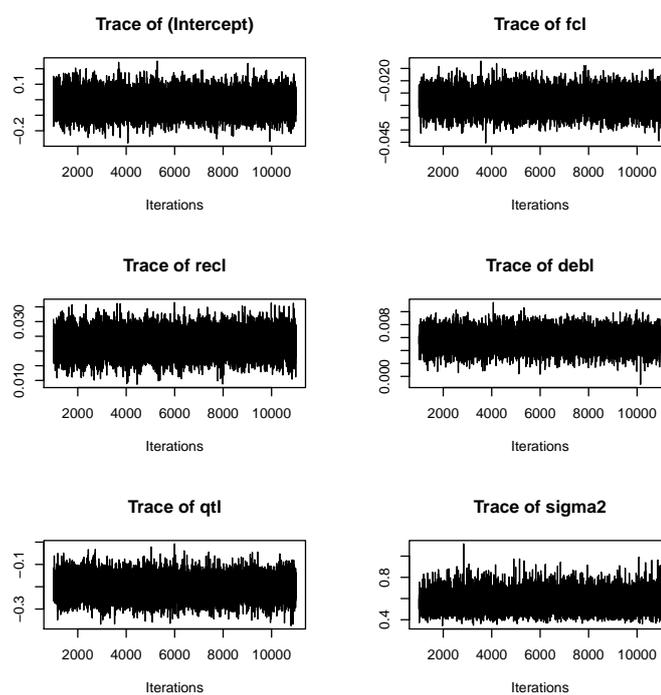


Figura 11 – Trajetória das posteriors dos parâmetros

Observa-se pelos gráficos de diagnósticos Figuras (10) e (11) um indício de convergência das cadeias simuladas. As cadeias geradas para cada parâmetro β oscilam em torno da média comum, sem apresentar tendências. Portanto, verifica-se que as densidades apresentam a forma de uma distribuição unimodal e as trajetórias dos gráficos relacionados

ao traço apresentaram a estacionariedade esperada. Na próxima subseção aplica-se a metodologia bayesiana para os modelos de efeitos individuais.

5.4.2 Modelo de efeitos individuais

Nesta subseção serão apresentadas as estratégias de inferência bayesiana para os modelos de efeito fixo e de efeito aleatório. Assim, inicia-se com a estimação do modelo não hierárquico que representa o modelo de efeito fixo, e depois estima-se o modelo hierárquico que designa o modelo de efeito aleatório.

Os dados do conjunto examinados nesta subseção é mesmo tratado na Seção anterior e referem-se a oito empresas brasileiras não financeiras, nas quais cada firma é observada em $T = 12$ períodos de tempo (1995-2006), totalizando 96 observações (painel balanceado).

Supondo que se esteja interessado em analisar um modelo para avaliar o comportamento do investimento de um grupo de empresas a partir de um conjunto de covariáveis. A variável dependente é o investimento e as variáveis explicativas são fluxo de caixa, nível de produção, dívida e Q de Tobin. Na Figura (12) pode-se verificar o comportamento das empresas em relação as variáveis regressoras.

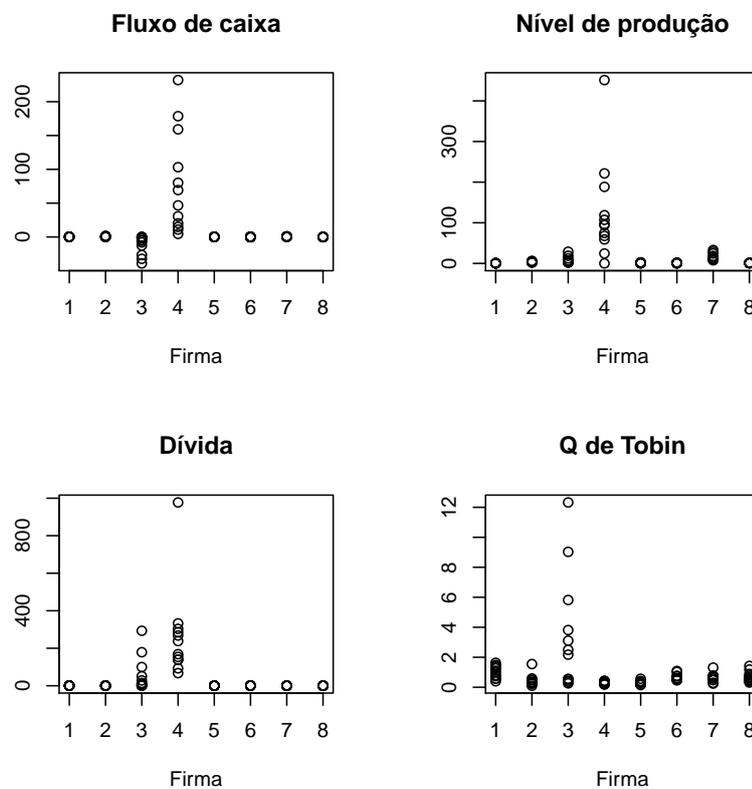


Figura 12 – Variáveis regressoras por firma

A distinção entre os modelos efeitos fixos e aleatório ocorre de acordo com a

escolha da distribuição a priori atribuídas aos efeitos específicos individuais. Assim, para o modelo de efeitos aleatórios utiliza-se priori com uma estrutura hierárquica, nos modelos de efeitos fixo considera-se prioris não hierárquicas.

Considera-se o seguinte modelo para a avaliação do investimento corporativo:

$$y_{it} = \beta_{0i} + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + \beta_4 x_{4it} + \varepsilon_{it}, \quad i = 1, \dots, 8; \quad t = 1, \dots, 12, \quad (5.13)$$

em que i é a i -ésima empresa e t é o período de tempo, y_{it} é o t -ésimo investimento para a i -ésima empresa, x_{1it} é a t -ésimo fluxo de caixa da i -ésima empresa, x_{2it} é o t -ésimo nível de produção para i -ésima empresa, x_{3it} é a t -ésima dívida para i -ésima empresa e x_{4it} é o t -ésimo Q de Tobin para i -ésima empresa.

Para o parâmetros β_k , $k = 0, 1, 2, 3, 4$ utilizam-se prioris normais independentes com média zero e com baixa precisão. Enquanto que para as precisões utilizam-se prioris gamas. Estas especificações estão detalhadas na Seção (4.2).

$$\text{Modelo: } \mathbf{y} | \boldsymbol{\beta}, \sigma^2 \sim \text{NM}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n).$$

$$\text{Especificação de prioris para } \boldsymbol{\beta} \text{ e } h = \sigma^{-2}$$

As duas estruturas de modelagens para o modelo (5.13) de dados em painel são: modelo de efeito fixo e modelo de efeito aleatório. Os modelos de efeito fixo correspondem a introdução de uma variável categórica para representar as unidades observacionais, no caso desta aplicação as empresas não financeiras, enquanto que os modelos de efeito aleatório, o intercepto é suposto ser aleatório, sendo descrito por uma distribuição de probabilidade.

Dessa forma, nesta aplicação, exploram -se os modelos:

- **Modelo de efeito fixo**

Atribuindo-se uma priori não hierárquica ao modelo (Equação 5.13) este é considerado similar ao modelo de efeitos fixos. Para assegurar a identificabilidade dos α_i e intercepto, consideram-se que:

$$\beta_{0i} = \beta_0 + \alpha_i, \quad \sum_{i=1}^8 \alpha_i = 0, \quad \varepsilon_{it} \sim \text{Normal}(0, \sigma^2)$$

As prioris para este modelo são definidas por:

$$\beta_{ki} \sim \text{Normal}(0, 0.0001), \quad k = 0, 1, 2, 3, 4; i = 1, \dots, 8.$$

$$h \sim \text{Gama}(0.01, 0.01)$$

- **Modelo de efeito aleatório**

Para o modelo de efeitos aleatórios assume-se distribuições normais com variâncias desconhecidas para os efeitos específicos individuais, e as distribuições a priori são especificadas hierarquicamente:

Distribuição a priori (1^o nível)

$$\begin{aligned}\alpha_i &\sim \text{Normal}(\mu_\alpha, V_\alpha), \quad i = 1, \dots, 8 \\ \tilde{\beta} &\sim \text{Normal}(0, 0.0001) \\ h = 1/\sigma^2 &\sim \text{Gama}(0.01, 0.01)\end{aligned}$$

Distribuição a priori (2^o nível)

$$\begin{aligned}\mu_\alpha &\sim \text{Normal}(0, 0.0001) \\ V_\alpha^{-1} &\sim \text{Gama}(0.01, 0.01)\end{aligned}$$

As estimativas pontuais para os parâmetros do modelo de efeito fixo são apresentados na Tabela (5).

Tabela 5 – Estimativas dos parâmetros - Modelo não hierárquico (EF)

Parâmetro	média a posteriori	DP
α_1	0.410920	0.129242
α_2	0.185683	0.124700
α_3	-0.131129	0.167624
α_4	-2.562778	0.201781
α_5	0.054703	0.123850
α_6	0.322824	0.125789
α_7	0.095189	0.131422
α_8	0.397201	0.126051
fluxo de caixa	-0.019573	0.002559
nível de produção	0.019956	0.002272
dívida	0.010348	0.000888
Q de Tobin	-0.305156	0.037631

A Figura (13) é uma comparação entre os interceptos individuais de cada empresa estimados pelos modelos de regressão individual, modelo *pooled* e o modelo de efeito fixo.

A Tabela (6) mostra as estimativas dos parâmetros, dos hiperparâmetros, seus respectivos erros de estimação e o erro de Monte Carlo (EMC) para o modelo de efeito aleatório (hierárquico).

Tabela 6 – Estimativas dos parâmetros - Modelo hierárquico (EA)

Parâmetro	Estimativa	Erro-padrão	EMC
β_1	-0.021861036	0.005739976	6.294e-05
β_2	0.018664230	0.004672455	5.297e-05
β_3	0.009478581	0.001902967	2.094e-05
β_4	-0.294823520	0.087238205	0.0011344
μ_α	-0.079034676	0.865919443	0.005312
V_α	0.8659194	0.527771174	0.007455

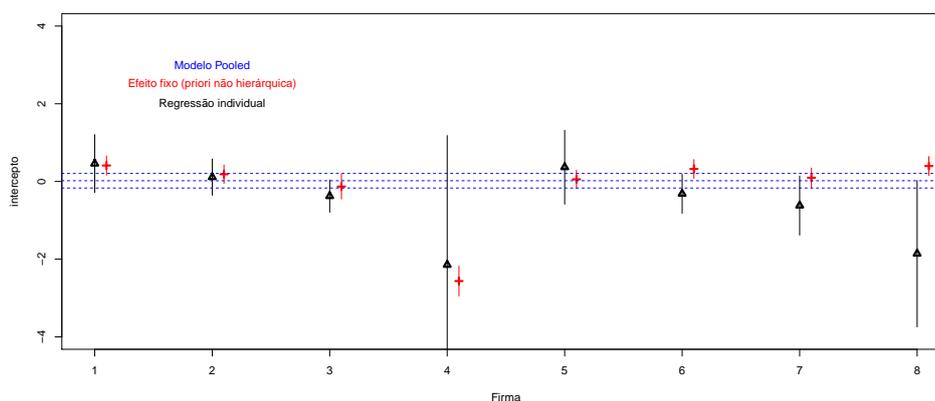


Figura 13 – Comparação interceptos individuais diferentes modelos

Nas Figuras (14), (15) e (16) estão presentes as trajetórias das cadeiras geradas, as densidades a posteriori e os gráficos das funções de autocorrelação (acf) para cada um dos parâmetros $\beta_1, \beta_2, \beta_3, \beta_4$, respectivamente. Percebe-se que as trajetórias e as autocorrelações decrescem a medida que o *lag* (defesagem) aumenta, além disso que as densidades apresentam a forma unimodal, indicando convergência do método.

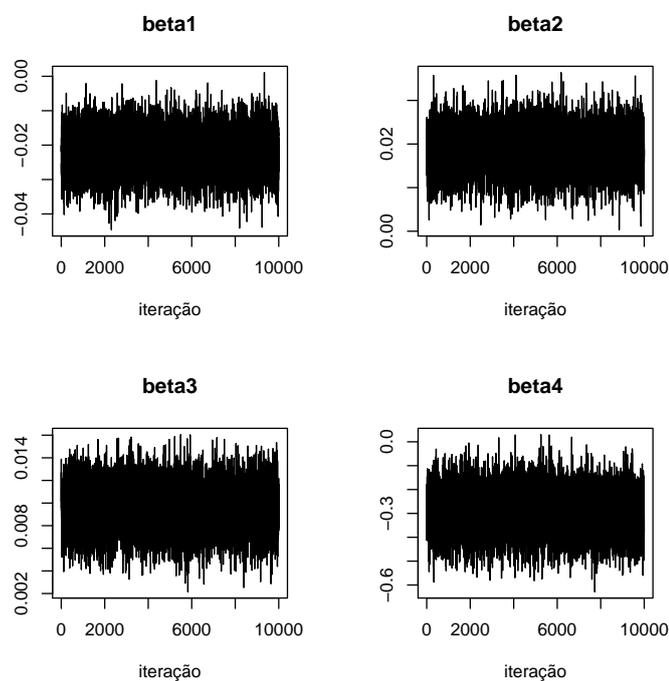


Figura 14 – Gráfico da trajetória das posteriores dos parâmetros

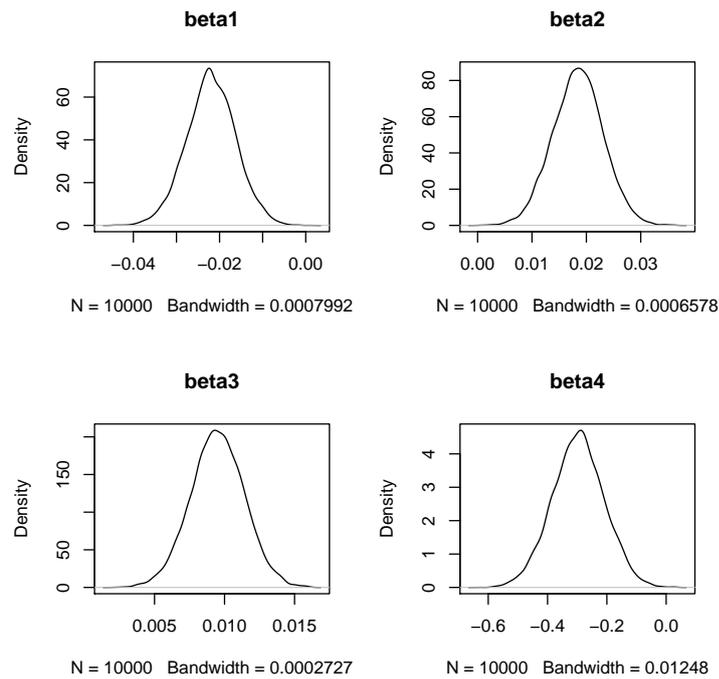


Figura 15 – Densidades a posteriori dos parâmetros

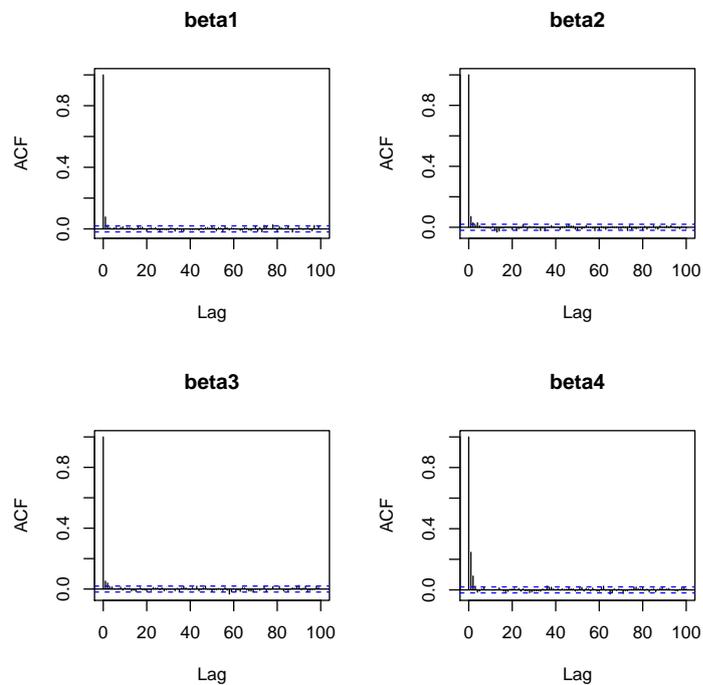


Figura 16 – Funções de autocorrelação dos parâmetros

A Figura (17) ilustra as representações gráficas das trajetórias das cadeias, as densidades a posteriori e as funções de autocorrelações (acf) das estimativas dos hiperparâmetros. Verifica-se que as trajetórias do gráfico relacionadas ao traço apresentaram

a estacionariedade esperada, as densidades uma forma unimodal e as acf um rápido decaimento, caracterizando a convergência do método.

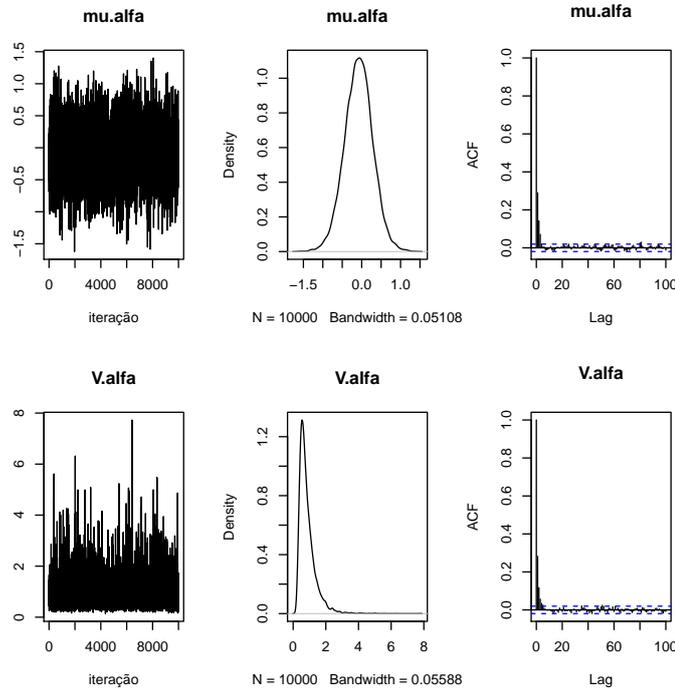


Figura 17 – Trajetórias, densidades e acfs dos hiperparâmetros

Finalmente, a Tabela (7) resume os resultados da estimação dos parâmetros para o modelo *pooled*, modelo de efeito fixo (não hierárquico) e o modelo de efeito aleatório (hierárquico) sob a perspectiva de inferência clássica e bayesiana.

Tabela 7 – Coeficientes estimados sob a perspectiva clássica e bayesiana

Método		β_1	β_2	β_3	β_3
Clássico	Pooled	-0,0301	0,0223	0,0062	-0,2609
	Efeito fixo	-0,0196	0,0199	0,0103	-0,3052
	Efeito aleatório	-0,0323	0,0220	0,0051	-0,2739
Bayesiano	Pooled	-0.0301	0.0223	0.0062	-0.2609
	Não hierárquico	-0.0196	0.0199	0.0103	-0.3052
	Hierárquico	-0.0219	0.0187	0.0095	-0.2948

Observa-se que as estimativas pontuais (médias a posteriori) dos parâmetros obtidas no contexto clássico e bayesiano estão próximas. A justificativa para isso é que ao utilizar prioris não informativas, espera-se que o resultado na abordagem bayesiana seja semelhante ao resultado na abordagem clássica. Além disso, espera-se que a verossimilhança (dados) predomine a medida que o tamanho da amostra aumente.

6 Considerações finais

Modelos de dados em painel permitem conjugar a diversidade de comportamentos individuais, com a existência de dinâmicas de ajustamento, ainda que potencialmente distintas, ou seja, permite tipificar as respostas de diferentes indivíduos a determinados acontecimentos, em diferentes momentos. Essa metodologia também possibilita avaliar a relação entre alguma variável de desempenho e diversas variáveis preditivas, permitindo que se elaborem inferências sobre as eventuais diferenças entre indivíduos e ao longo do tempo sobre a evolução daquilo que se pretende estudar.

Dadas as possibilidades, a modelagem em painel é utilizada cada vez mais na pesquisa científica em diversas áreas, contudo, é necessário, que a aplicação venha acompanhada de rigor metodológico, e a devida cautela quando da análise dos resultados, principalmente quando se buscam previsões. A qualidade da pesquisa, portanto, depende de um conhecimento estruturado por parte do pesquisador, é a partir do domínio da metodologia e suas técnicas que é possível analisar as informações de forma a identificar previsões válidas.

Este trabalho foi desenvolvido considerando a necessidade de um conhecimento estruturado sobre a metodologia de dados em painel, em especial em pesquisas econométricas. O estudo apresentou conceitos, métodos e aplicações visando possibilitar a compreensão da utilização do modelo de dados em painel, através de uma linguagem clara e acessível àqueles que, embora não sendo econométricos, necessitam se apropriar dos métodos de análise dos dados em painel para aplicá-los na sua prática de pesquisa.

As simulações de aplicações do modelo no software R complementam a exposição da modelagem, facilitando a didática do conteúdo. Ressalta-se que todas as técnicas ilustradas neste trabalho também se aplicam para painéis desbalanceados. O presente trabalho pretende contribuir em nível de ensino, pesquisa e extensão, quanto à compreensão e utilização da modelagem de dados em painel utilizando o software R em sua aplicação.

Uma das principais contribuições desse trabalho foi a exposição dos métodos bayesianos de análise de dados em painel, uma vez que observa-se que a literatura sobre esses métodos concentra-se na abordagem clássica. Destaca-se que um dos possíveis ganhos da inferência bayesiana é maior flexibilidade a medida que os modelos vão se tornando mais complexos, ainda que os modelos desse trabalho não sejam complicados o suficiente para ilustrar esse ganho.

Como sugestões de trabalhos futuros incluem-se desenvolvimento da teoria e aplicação de outros modelos utilizados no contexto de dados em painel: os modelos dinâmicos e os modelos de coeficientes aleatórios, sob as perspectivas clássicas e bayesianas. É necessário enfatizar que tais detalhamentos não foram realizados por ter dedicado esforços

maiores na compreensão e interpretação dos modelos usuais da metodologia de dados em painel, que são os modelos de dados agrupados, efeito fixo e aleatório.

Neste sentido, espera-se que os análises desenvolvidas nesse trabalho, sirvam de base para o desenvolvimento de estudos mais aprofundados dos modelos de dados em painel. Ao longo do processo de pesquisa foram percebidas diversas oportunidades de aperfeiçoamentos não possíveis de serem implementados neste projeto devido às limitações de tempo.

Referências

- AKBAR, A. et al. Determinant of economic growth in asian countries: A panel data perspective. *Pakistan Journal of Social sciences*, Citeseer, v. 31, n. 1, p. 145–157, 2011. Citado na página 21.
- ANTONIAK, C. E. Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, JSTOR, p. 1152–1174, 1974. Citado na página 64.
- BALTAGI, B. *Econometric analysis of panel data*. 4th. ed. [S.l.]: John Wiley & Sons, 2008. Citado 5 vezes nas páginas 22, 25, 26, 83 e 89.
- BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. *Hierarchical modeling and analysis for spatial data*. 2ed.. ed. [S.l.]: Crc Press, 2014. Citado 2 vezes nas páginas 71 e 80.
- BERNARDO, J. M. [the geometry of asymptotic inference]: Comment: On multivariate jeffreys' priors. *Statistical Science*, Institute of Mathematical Statistics, v. 4, n. 3, p. 227–229, 1989. ISSN 08834237. Citado na página 67.
- BOLFARINE, H.; SANDOVAL, M. C. *Introdução à inferência estatística*. 2. ed. ed. [S.l.]: SBM, 2010. Citado na página 66.
- BOND, S.; REENEN, J. V. Microeconomic models of investment and employment. *Handbook of econometrics*, Elsevier, v. 6, p. 4417–4498, 2007. Citado na página 21.
- BREUSCH, T. S.; PAGAN, A. R. The lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, [Oxford University Press, Review of Economic Studies, Ltd.], v. 47, n. 1, p. 239–253, 1980. ISSN 00346527, 1467937X. Disponível em: <<http://www.jstor.org/stable/2297111>>. Citado na página 89.
- BUSSAB, W. d. O.; MORETTIN, P. A. *Estatística básica*. [S.l.]: Saraiva, 2013. Citado na página 29.
- CAMERON, A. C.; TRIVEDI, P. K. *Microeconometrics: methods and applications*. New York: Cambridge University Press, 2005. Citado 5 vezes nas páginas 25, 26, 28, 35 e 46.
- CASELLA, R. L. B. G. *Statistical Inference*. 2°. ed. [S.l.]: Duxbury Press, 2001. ISBN 0534243126,9780534243128. Citado 3 vezes nas páginas 56, 60 e 61.
- CLARK, A. G. J. S. *Hierarchical Modelling for the Environmental Sciences: Statistical Methods and Applications (2006)*. illustrated edition. [S.l.]: Oxford University Press, USA, 2006. (Oxford Biology). ISBN 9780198569671,019856967X. Citado na página 71.
- CLARK, J. S. Why environmental scientists are becoming bayesians. *Ecology letters*, Wiley Online Library, v. 8, n. 1, p. 2–14, 2005. Citado na página 71.
- CONGDON, P. D. *Applied Bayesian hierarchical methods*. [S.l.]: CRC Press, 2010. Citado na página 75.

- CRISOSTOMO, V. L.; ITURRIAGA, F. J. L.; GONZÁLEZ, E. V. Financial constraints for investment in Brazil. *International Journal of Managerial Finance*, Emerald Group Publishing Limited, v. 10, n. 1, p. 73–92, 2014. Citado 5 vezes nas páginas 83, 89, 90, 91 e 99.
- CROISSANT, Y.; MILLO, G. Panel data econometrics in R: The plm package. *Journal of Statistical Software*, v. 27, n. 2, 2008. Disponível em: <<http://www.jstatsoft.org/v27/i02/>>. Citado 2 vezes nas páginas 83 e 87.
- D'ESPALLIER, B.; GUARIGLIA, A. Does the investment opportunities bias affect the investment–cash flow sensitivities of unlisted smes? *The European Journal of Finance*, Taylor & Francis, v. 21, n. 1, p. 1–25, 2015. Citado na página 22.
- D'ESPALLIER, B.; HUYBRECHTS, J.; ITURRIAGA, F. J. L. Analyzing firm-varying investment-cash flow sensitivities and cash-cash flow sensitivities: A bayesian approach. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad*, Taylor & Francis, v. 40, n. 151, p. 439–467, 2011. Citado 2 vezes nas páginas 22 e 80.
- EHLERS, R. S. Inferência bayesiana. *Departamento de Matemática Aplicada e Estatística, ICMC-USP*, 2011. Citado 2 vezes nas páginas 64 e 74.
- FITRIANTO, A.; MUSAKKAL, N. F. K. Panel data analysis for sabah construction industries: Choosing the best model. *Procedia Economics and Finance*, Elsevier, v. 35, p. 241–248, 2016. Citado na página 21.
- GAMERMAN, D.; MIGON, H. dos S. *Inferência estatística: uma abordagem integrada*. [S.l.]: Instituto de Matemática, Universidade Federal do Rio de Janeiro, 1993. Citado 6 vezes nas páginas 61, 64, 65, 71, 73 e 74.
- GELMAN, A. et al. *Bayesian data analysis*. [S.l.]: Chapman & Hall/CRC Boca Raton, FL, USA, 2014. v. 2. Citado na página 67.
- GHINIS, C. P.; FOCHEZATTO, A. Crescimento pró-pobre nos estados brasileiros: análise da contribuição da construção civil usando um modelo de dados em painel dinâmico, 1985-2008. *Economia Aplicada*, scielo, v. 17, p. 243 – 266, 09 2013. Citado na página 26.
- GREENE, W. H. *Econometric analysis*. 6th. ed. New Jersey: Practice Hall, 2008. Citado 6 vezes nas páginas 25, 29, 41, 47, 51 e 83.
- GRUNFELD, Y. *The Determinants of Corporate Investment*. Tese (Doutorado) — University of Chicago, 1958. Citado na página 83.
- GUJARATI, D. N.; PORTER, D. *Econometria Básica*. 5th. ed. [S.l.]: McGraw Hill Brasil, 2011. Citado 4 vezes nas páginas 27, 47, 50 e 83.
- HAUSMAN, J. A. Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 1251–1271, 1978. Citado na página 89.
- HOFFMANN, R.; VIEIRA, S. *Análise de regressão: uma introdução à econometria*. 3th. ed. [S.l.]: Editora HUCITEC, Editora da Universidade de São Paulo, 1998. (Coleção economia e planejamento. Obras didáticas). Citado na página 30.

HSIAO, C. *Analysis of panel data*. 3. ed. New York: Cambridge University Press, 2014. Citado 5 vezes nas páginas 21, 22, 25, 26 e 36.

JACKMAN, S. *Bayesian Analysis for the Social Sciences*. 1. ed. [S.l.: s.n.], 2009. (Wiley Series in Probability and Statistics). ISBN 0470011548,9780470011546. Citado 3 vezes nas páginas 58, 67 e 80.

JACKMAN, S. *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University*. Stanford, California, 2015. R package version 1.4.9. Disponível em: <<http://pscl.stanford.edu/>>. Citado na página 27.

JEFFREYS, H. An invariant form for the prior probability in estimation problems. In: THE ROYAL SOCIETY. *Proceedings of the Royal Society of London a: mathematical, physical and engineering sciences*. [S.l.], 1946. v. 186, n. 1007, p. 453–461. Citado na página 66.

KASS, R. E.; RAFTERY, A. E. Bayes factors. *Journal of the american statistical association*, Taylor & Francis Group, v. 90, n. 430, p. 773–795, 1995. Citado na página 77.

KASWENGI, J.; DIALLO, M. F. Consumer choice of store brands across store formats: A panel data analysis under crisis periods. *Journal of Retailing and Consumer Services*, v. 23, p. 70 – 76, 2015. Citado na página 25.

KEA, X.; SAKSENA, P.; HOLLYB, A. The determinants of health expenditure: a country-level panel data analysis. *Geneva: World Health Organization*, 2011. Citado na página 26.

KLEIBER, C.; ZEILEIS, A. *Applied Econometrics with R*. New York: Springer-Verlag, 2008. ISBN 978-0-387-77316-2. Disponível em: <<https://CRAN.R-project.org/package=AER>>. Citado 2 vezes nas páginas 27 e 34.

KLEIBER, C.; ZEILEIS, A. *Applied econometrics with R*. [S.l.]: Springer Science & Business Media, 2008. Citado na página 83.

KOOP, G. *Bayesian econometrics*. [S.l.]: Wiley-Interscience, 2003. Citado 7 vezes nas páginas 62, 63, 64, 75, 79, 80 e 81.

LAU, Y. K.; ATAGUBA, J. E. Investigating the relationship between self-rated health and social capital in south africa: a multilevel panel data analysis. *BMC public health*, BioMed Central, v. 15, n. 1, p. 1, 2015. Citado na página 26.

LINDLEY, D. V.; SMITH, A. F. Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 1–41, 1972. Citado na página 71.

MADEIRA, R. F. Restrições financeiras nas empresas brasileiras de capital aberto: a relevância da estrutura de capital para o investimento. *Revista do BNDES, Rio de Janeiro*, n. 39, p. 69–122, 2013. Citado na página 56.

MAGALHÃES, M. N.; LIMA, A. C. P. de. *Noções de probabilidade e estatística*. 7th. ed. [S.l.]: Editora da Universidade de São Paulo, 2013. Citado 2 vezes nas páginas 55 e 56.

- MARQUES, L. D. et al. Modelos dinâmicos com dados em painel: revisão de literatura. *Centro de estudos Macroeconômicos e Previsão, faculdade de Economia do Porto*, 2000. Citado na página 21.
- MARTIN, A. D.; QUINN, K. M.; PARK, J. H. MCMCpack: Markov chain monte carlo in R. *Journal of Statistical Software*, v. 42, n. 9, p. 22, 2011. Citado na página 101.
- MATYAS, P. S. L. *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Third edition. [S.l.]: Springer, 2008. Citado na página 49.
- MORAWETZ, U. *Bayesian modelling of panel data with individual effects applied to simulated data*. [S.l.]: Univ. für Bodenkultur, Department für Wirtschafts-u. Sozialwiss., Inst. für Nachhaltige Wirtschaftsentwicklung, 2006. Citado na página 75.
- O'HAGAN, A. *The Advanced Theory of Statistics, Vol. 2B: Bayesian Inference*. [S.l.]: Hodder Education Publishers, 1994. ISBN 0340529229,9780340529225. Citado 5 vezes nas páginas 55, 56, 58, 67 e 74.
- PAULINO, C.; TURKMAN, M.; MURTEIRA, B. *Estatística bayesiana*. [S.l.]: Fundação Calouste Gulbenkian, 2003. Citado 9 vezes nas páginas 55, 60, 62, 63, 64, 65, 66, 71 e 74.
- PINTO, N. G.; CORONEL, D.; FILHO, R. B. O programa bolsa família de 2004 a 2010: Efeitos do desenvolvimento regional no brasil e em suas regiões. *Qualitas Revista Eletrônica*, v. 16, n. 1, 2015. Citado na página 26.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: <<https://www.R-project.org/>>. Citado 3 vezes nas páginas 27, 80 e 83.
- REED, W. R.; YE, H. Which panel data estimator should i use? *Applied Economics*, Taylor & Francis, v. 43, n. 8, p. 985–1000, 2011. Citado na página 29.
- ROCHA, C. D. A. d. Algoritmo recursivo dos mínimos quadrados para regressão linear local. Universidade do Porto. Reitoria, 2001. Citado na página 29.
- ROHDE, N. et al. The effect of economic insecurity on mental health: Recent evidence from australian panel data. *Social Science & Medicine*, Elsevier, 2016. Citado na página 25.
- ROSS, S. *A First Course in Probability*. 9. ed. [S.l.]: Pearson, 2014. ISBN 9781292024929. Citado na página 57.
- ROYLE, R. M. D. J. A. *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. 1. ed. [S.l.]: Elsevier Academic Press, 2008. ISBN 0123740975,9780123740977. Citado na página 71.
- SACHS, R. C. C. *Contribuição das inovações biológicas para a produtividade da cana-de-açúcar no estado de São Paulo, 1998-2009*. Tese (Doutorado) — Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, 2015. Citado na página 26.
- STOCK, J. H.; WATSON, M. W. *Introduction to econometrics*. 2th. ed. Boston: Pearson/Addison Wesley, 2006. Citado 3 vezes nas páginas 22, 25 e 90.

WOOLDRIDGE, J. M. *Introdução à econometria: uma abordagem moderna*. 1th. ed. São Paulo: Cengage Learning, 2008. Citado 3 vezes nas páginas 21, 36 e 51.

ZELLNER, A. *An introduction to Bayesian inference in econometrics*. [S.l.]: Wiley-Interscience, 1996. (Wiley Classics Library). ISBN 9780471169376,9780471981657,0471169374,0471981656. Citado na página 83.