

Modelos Dinâmicos de Resposta Binária para Dados em Painel

Eveliny Barroso da Silva

Modelos Dinâmicos de Resposta Binária para Dados em Painel

Eveliny Barroso da Silva

Orientadora: Prof. Dra. Maria Aparecida de Paiva Franco

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

UFSCar - São Carlos

Julho/2010

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

S586md

Silva, Eveliny Barroso da.

Modelos dinâmicos de resposta binária para dados em painel / Eveliny Barroso da Silva. -- São Carlos : UFSCar, 2008.

109 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2008.

1. Análise de regressão. 2. Dados em painel. 3. Modelo logito. 4. Modelo probito. 5. Modelo de efeitos fixos. 6. Modelos de efeitos aleatórios. I. Título.

CDD: 519.536 (20^a)

Agradecimentos

Agradeço, antes de tudo, a Deus.

Agradeço também:

a meus pais e demais familiares,

ao Thiago, pelo amor, amizade e total apoio,

a todos os amigos que sempre estiveram presentes, contribuindo com críticas e sugestões,

à professora Maria Aparecida de Paiva Franco, pela orientação segura, e pelo incentivo durante todo o curso de pós-graduação,

aos professores Cecília Cândolo e Aquiles Kalatis, membros da banca examinadora do exame de qualificação, pelas sugestões apresentadas,

ao Departamento de Estatística da Universidade Federal de São Carlos pelo ambiente acolhedor,

às professoras Estela Maris Pereira Beretta e Teresa Cristina Martins Dias e aos professores Luiz Aparecido Milan e Francisco Louzada Neto,

aos diversos professores do Departamento de Estatística da Universidade Federal do Ceará, em especial aos professores João Maurício Araújo Mota e João Welliandre Carneiro Alexandre pela amizade, ensinamentos e pelo incentivo que me levaram a fazer a pós-graduação,

aos funcionários Luiza (Tia Luiza) e Robson pelo carinho, amizade e atenção,

à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio concedido para este trabalho.

Resumo

Neste trabalho é apresentado inicialmente um levantamento da literatura referente a modelos de regressão não lineares quando a variável resposta é binária e as observações são um painel de dados. Tais modelos podem incluir efeitos de várias fontes: variáveis específicas de interesse, heterogeneidade não observável dos indivíduos e valores defasados da variável resposta. A parte original do trabalho consiste nos estudos por simulação usando programação criada para esse fim no software R, visando comparar duas propostas recentes da literatura para ajustar, por máxima verossimilhança condicional, modelos dinâmicos que incluem os três tipos de efeitos mencionados. Também é original o estudo empírico, usando a metodologia de reamostragem “bootstrap”, de características da distribuição conjunta dos estimadores dos parâmetros em análises intermediárias dos dados. A justificativa do trabalho é a atualidade do tema e a inexistência de programas de ajuste de modelos dinâmicos de resposta binária na maioria dos softwares comerciais.

Palavras-chave: Dados em painel; Modelo Logito; Modelo Probit ; Modelos de Efeitos Aleatórios; Modelos de Efeitos Fixos.

Abstract

A summary of the state of the art relative to regression models for binary response variable and panel data is presented in this work. Those models may include effects from several sources: specific variables of interest, heterogeneity between individuals and lagged values of the response variable. The original contributions of the author are simulation studies to compare two different approaches to maximum likelihood estimation of parameters of dynamic models with all three kinds of effects, and also a study of properties of such estimators in group sequential analysis, using the bootstrap methodology. Original codes were developed in R for implementation of simulation studies. The relevance of the subject and the non availability of appropriate codes in commercial software for fitting dynamic models for binary response justify the choice of the theme.

Key words: Fixed and Random effects; Logit Model; Panel Data; Probit Model.

Lista de Abreviaturas

EQM: Erro Quadrático Médio

GHK: Geweke, Hajivassilou e Keane

lnL: Log-verossimilhança

MPD: Modelo Probita Dinâmico

MPDH: Modelo Probita Dinâmico de Heckman

MPDW: Modelo Probita Dinâmico de Wooldridge

MPE: Modelo Probita Estático

MQG: Mínimos Quadrados Generalizados

MQO: Mínimos Quadrados Ordinários

MQVD: Mínimos Quadrados com Variáveis “Dummy”

MV: Máxima Verossimilhança

MVC: Máxima Verossimilhança Condicional

MVS: Máxima Verossimilhança Simulada

PMT: Variável Indicadora da Participação no Mercado de Trabalho

Sumário

1	Introdução	1
1.1	Objetivo	2
1.2	Notação	3
1.3	Especificação de Modelos de Regressão para Variável Resposta Binária	5
1.4	Organização dos Capítulos	7
2	Modelos Estáticos de Regressão: Variável Resposta Contínua e Dados em Painel	9
2.1	Modelo de Efeitos Fixos	10
2.2	Modelo de Efeitos Aleatórios	17
2.2.1	Estimação por Mínimos Quadrados Generalizados (MQG)	19
2.2.2	Estimação via MQG Factível	20
3	Modelos Estáticos de Regressão para Variável Resposta Binária	23
3.1	Modelos de Resposta Binária para observações independentes	23
3.1.1	Estimação dos parâmetros	26
3.1.2	Efeitos marginais	28
3.2	Modelos de Resposta Binária para Dados em Painel	29
3.2.1	Modelo de Efeitos Fixos	30

3.2.2	Modelo de Efeitos Aleatórios	34
4	Modelos Dinâmicos de Regressão: Variável Resposta Binária e Dados em Painel	39
4.1	Função de Verossimilhança	43
4.1.1	Modelo de Efeitos Fixos	44
4.1.2	Modelo de Efeitos Aleatórios	44
4.2	Estimação de alguns Modelos Dinâmicos de Resposta Binária	45
4.2.1	Modelo de Efeitos Fixos segundo Honoré e Kyriazidou	45
4.2.2	Modelo Probit Dinâmico de Heckman com Erros Autocorrelacionados	48
4.2.3	Modelo Probit Dinâmico para Dados em Painel com Erros Não-Correlacionados, segundo Wooldridge	58
5	Resultados	62
5.1	Programa para ajuste de um Modelo Probit Estático (MPE)	62
5.2	Ajuste de um Modelo Probit Dinâmico(MPD) segundo Wooldridge usando o programa para (MPE)	64
5.3	Estudo por simulação das propriedades dos estimadores por máxima verossimilhança do MPD de Wooldridge	69
5.4	Estimando o MPD de Heckman usando a sugestão de Wooldridge	70
5.5	Análises intermediárias em um modelo probit dinâmico: estudo via <i>bootstrap</i>	72
6	Considerações Finais	76
	Referências Bibliográficas	79
A	Matrizes Subdivididas	81

B	Programas	83
B.1	Programa para o Ajuste de um MPE	83
B.2	Programa para o Ajuste de um MPD de Wooldridge	85
B.3	Programa para gerar um painel de dados segundo o MPD de Wooldridge	87
B.4	Programa para gerar um painel de dados segundo o MPD de Heckman	89
B.5	Programa em STATA para ajustar o MPD de Wooldridge	92
B.6	Programa para reamostragem <i>bootstrap</i> de uma análise de dados em painel com análises intermediárias	92

Capítulo 1

Introdução

O efeito de k variáveis explicativas indicadas genericamente por X sobre a média da distribuição de probabilidade de uma variável resposta Y pode ser o objeto de uma investigação. Com essa finalidade, N indivíduos podem ser selecionados de uma população para prover informações sobre as variáveis envolvidas e, para controlar a possível existência de efeitos individuais não observáveis que possam afetar a resposta Y , é interessante que se considere a variação da resposta para cada um dos indivíduos selecionados, sob diversas condições. O conjunto de observações obtidas para os N indivíduos em T condições sobre todas as variáveis, constitui um painel. A população considerada no estudo pode ser um conjunto de países, escolas, firmas, pessoas etc., e as condições geralmente são diferentes momentos do tempo.

Conjuntos de dados em painel podem ser coletados por órgãos governamentais para avaliar políticas públicas, por empresas em geral para determinar covariáveis importantes na avaliação de clientes, por pesquisadores da área médica, para avaliar se um tratamento é eficiente ou não (nos também denominados estudos longitudinais) etc. Dados em painel sobre variáveis no campo da Economia têm sido coletados por órgãos de governo em vários países, podendo ser citado como exemplo, o PSID (*Michigan Panel Study of Income Dynamics*), nos Estados Unidos da América, em que cerca de 6.000 famílias e 15.000 pessoas têm sido entrevistadas periodicamente a respeito de algumas variáveis, desde 1968. No Brasil tem sido feito, por exemplo, o acompanhamento de dados sobre

educação pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) que é responsável pela construção das Sinopses Estatísticas da Educação Básica (SAEB), que consistem no registro de avaliações bianuais, feitas a partir de 1998, sobre o desempenho dos alunos de escolas públicas em provas de língua portuguesa e matemática e no levantamento simultâneo de variáveis explicativas como o gasto *per capita* anual com os alunos da rede pública e a proporção de professores com diploma de curso superior em relação ao total de matrículas. (SOBREIRA, R. e CAMPOS, B.C, 2005).

Muitos estudos com dados em painel no campo da Economia têm a finalidade de determinar o efeito de covariáveis sobre uma variável resposta Y binária que é a indicadora da ocorrência ou não de um determinado evento, cuja distribuição marginal pode ser suposta como uma distribuição de Bernoulli. A determinação de variáveis que possam influenciar um trabalhador a se filiar ou não ao sindicato de sua categoria em países onde isso é optativo, ou a busca de variáveis influentes na decisão de uma mulher estar ou não no mercado de trabalho, são exemplos de questões envolvendo uma variável resposta binária e que podem ser respondidas a partir de um levantamento de um painel de dados relativos a essas variáveis. Diversas investigações usando dados em painel e com variável resposta binária podem ser encontrados na literatura atual como os trabalhos de DARGAY, J.M.; VYTHOULKAS, P.C. (1999) e SANTOS, M.J; KASSOUF, A.L.(2007). No primeiro, ajustou-se um modelo probito para uma variável resposta indicadora da realização da compra de um carro por um indivíduo, num ano t , no período de 1994 – 2001 em 9 países da comunidade européia. No segundo trabalho, fez-se um estudo sob uma perspectiva econômica da relação existente entre o mercado de drogas e a criminalidade, utilizando dados em painel de estados brasileiros.

1.1 Objetivo

A definição de um modelo dinâmico de regressão de resposta binária para dados em painel é tema muito tratado na literatura atual, havendo diversas

propostas concorrentes. Duas propostas para o processo gerador de dados de tal modelo são apresentadas aqui. A primeira tem origem em HECKMAN (1986) e seu ajuste por máxima verossimilhança tem difícil implementação computacional. A segunda delas, devida a WOOLDRIDGE (2005) leva a uma função de verossimilhança semelhante a de um modelo não-dinâmico (estático) que é facilmente ajustável e por isso tem sido usada na prática. A questão central abordada neste trabalho é a das propriedades dos estimadores dos coeficientes de um modelo de regressão segundo HECKMAN (1986), usando o processo de estimação sugerido por WOOLDRIDGE (2005). Para isso, foi criado e testado um programa em R para ajuste de um modelo de resposta binária estático. Foram feitos estudos por simulação para avaliar a compatibilidade das estimativas com as fornecidas por softwares comerciais. Este programa foi usado a seguir para ajustar o modelo de Wooldridge a conjunto de dados gerados segundo este mesmo modelo. Foram feitos estudos por simulação para avaliar propriedades dos estimadores. A seguir, foi abordada a questão da estimação dos parâmetros de um modelo de HECKMAN (1986), a partir de dados gerados segundo esse modelo, usando entretanto um programa para ajuste do modelo de WOOLDRIDGE (2005). Outra questão de interesse é a da distribuição conjunta dos estimadores de um dado coeficiente do modelo de WOOLDRIDGE (2005), obtidos em análises intermediárias do painel de dados em várias ocasiões. Essa questão é abordada neste trabalho usando amostragem “bootstrap” de um painel analisado em três ocasiões sucessivas.

1.2 Notação

Em um painel de observações de N indivíduos em T momentos do tempo, em relação a um conjunto de covariáveis (X) e uma variável resposta de interesse (Y), os T tempos, podem ser indicados por t_1, t_2, \dots, t_T e a observação de uma variável no indivíduo i na condição t será denotada com o subscrito (it). Desse modo, Y_{it} indicará a variável resposta Y e \mathbf{x}_{it} representará o vetor de valores de todas as covariáveis do indivíduo i no tempo t . O vetor de todas as k variáveis em X sobre o mesmo indivíduo i em todos os tempos será indicado por \mathbf{x}_i .

Como exemplo de um painel de dados de resposta binária, sejam $T = 4$ e $N = 4$. A Tabela 1.1 mostra um conjunto de dados em painel referente à função indicadora (Y) de filiação em sindicato por parte de trabalhadores, identificados por id , o ano em que foi feita a observação, idade e anos de escolaridade.

TABELA 1.1: *Painel de dados da variável binária filiação a sindicato e duas covariáveis.*

id	ano	Y_{it}	$idade$	esc
1	1980	1	27	11
1	1981	0	28	12
1	1982	1	29	13
1	1983	1	30	14
2	1980	0	28	12
2	1981	1	29	13
2	1982	0	30	13
2	1983	1	31	14
3	1980	1	28	12
3	1981	0	29	13
3	1982	1	30	14
3	1983	0	31	14
4	1980	1	30	13
4	1981	0	31	13
4	1982	1	32	14
4	1983	0	33	14

A Tabela 1.1 mostra que 4 indivíduos foram observados em 4 anos, tendo a primeira observação sido coletada em 1980 e a última em 1983. A variável resposta de interesse Y vale 1 se no ato da pesquisa o indivíduo está filiado a algum sindicato e vale 0 caso contrário. A variável id é a identificação do indivíduo observado, $idade$ é a idade em anos completos e esc representa os anos de escolaridade do indivíduo entrevistado. Como exemplo, para $i = 1$ e $t = 2$, $Y_{it} = 0$, $\mathbf{x}'_{it} = (28, 12)$ e $\mathbf{x}'_i = (27, 11, 28, 12, 29, 13, 30, 14)$.

1.3 Especificação de Modelos de Regressão para Variável Resposta Binária

Seja P_{it} a probabilidade de uma variável binária Y_{it} assumir o valor 1 para o indivíduo i no tempo t , dados os valores de todas as outras variáveis observadas sobre o mesmo indivíduo, em todo $s \leq t$. A questão que se coloca ao construir um modelo de regressão neste contexto que possa ser usado para predição, é encontrar uma função que exprima P_{it} em termos dessas variáveis observáveis e de alguma variável não observável que caracterize o indivíduo, dentro da idéia de que haja heterogeneidade entre os diversos indivíduos. Um modelo linear generalizado para P_{it} que tenha coeficientes de cada variável constantes no decorrer do tempo é relativamente simples de ser construído a partir de suposições sobre a distribuição conjunta de todas as variáveis envolvidas, o que permitirá também que seja estimável pelo método de máxima verossimilhança.

Como já explicitado acima, o presente trabalho visa estudar modelos de regressão linear generalizados para uma variável resposta binária no tempo t em termos dos valores observados em todos os outros tempos ($s \leq t$) de algumas covariáveis (X) e de valores da variável resposta a partir da observação de um painel de dados sobre essas variáveis. O que se entende por um modelo de regressão é o sistema de equações teóricas para a Média condicional de Y_{it} , derivado de um conjunto de suposições acerca da distribuição conjunta das variáveis Y_{it} , para ($t = 1, 2, \dots, T$) condicionalmente aos valores (\mathbf{x}_i) observados (e todos os valores Y_{is} , para ($s \leq t$)). Dependendo das suposições acima mencionadas, pode ocorrer que a função que relaciona uma função de ligação da média condicional da variável resposta (Y_{it}) do indivíduo i no tempo t em termos das variáveis condicionantes seja dada em forma de uma função linear de um subconjunto dessas variáveis.

Quando $T = 1$, isto é, quando só é feita uma observação de cada variável em cada um dos N indivíduos, e levando em conta que tais observações de diferentes indivíduos são independentes entre si, um modelo que pode ser ajustado por máxima verossimilhança, mas que pode não exprimir o verdadeiro coeficiente de regressão de cada uma das variáveis sobre a média de Y , pelo fato de poder

haver omissão de variáveis, é o modelo linear generalizado tipo Probit, Logito, e outros. Os dados obtidos dizem então respeito a um “corte transversal” da população em um único momento do tempo.

Quando há suspeita de existência de variáveis omitidas, em particular, supondo que haja um componente de heterogeneidade aleatório na população que assume um valor constante nas observações de um mesmo indivíduo, é aconselhável que seja levantado um painel de dimensão temporal $T \geq 2$. Um corte transversal da população é então acompanhado por T momentos do tempo gerando o painel de dados. Um modelo de regressão para Y_{it} que inclui entre as regressoras uma variável representando a heterogeneidade, mas que não inclui valores defasados da variável resposta é denominado estático. Se esse modelo ainda apresenta omissão de variáveis, não explicando a resposta observada, é possível fazer a suposição de um efeito dos valores passados da própria variável resposta sobre o valor atual. Um modelo é denominado dinâmico se a Esperança condicional de Y_{it} depender expressamente das covariáveis em X , de uma variável aleatória representando a heterogeneidade e de um ou mais valores defasados (por exemplo, $Y_{i,t-m}$, para algum $m \in 1, 2, \dots, t-1$) da variável resposta. Uma razão para se considerar tais modelos dinâmicos com efeitos individuais é que em observações de tais processos ocorre muitas vezes a persistência da mesma resposta no tempo. Essa persistência pode resultar de uma real dependência do estado passado (resposta do indivíduo no tempo anterior) e isso pode ser modelado através da inclusão de uma **variável resposta defasada** ($Y_{i,t-1}$) no modelo. Por outro lado, a persistência observada pode vir também de uma heterogeneidade permanente e não-observada entre os indivíduos que têm diferentes propensões a experimentar um determinado estado. (CHAY, K.Y.; HYSLOP, D.,2000)

Ilustrando com o caso onde $Y = 1$ indica a participação de um indivíduo em um programa, o efeito da participação passada sobre a participação atual no programa seria estimada sem vício se fosse possível realizar um experimento controlado em que o pesquisador atribuisse os dois tipos de participação ao acaso a um grupo de indivíduos e observasse diferenças em participação nos períodos subsequentes. A aleatorização dos indivíduos asseguraria que a participação passada não estaria relacionada com a heterogeneidade observada e qualquer

dependência serial seria interpretada como dependência do estado anterior. Na impossibilidade dessa atribuição aleatória, usam-se dados em painel para estimar modelos econométricos, em particular os modelos dinâmicos de resposta binária com efeitos aleatórios que tentam controlar o confundimento da heterogeneidade com a verdadeira dependência serial.¹

1.4 Organização dos Capítulos

Nos capítulos dessa dissertação serão tratados os modelos de regressão de variável resposta binária para um corte transversal da população, e duas classes de modelos de efeitos aleatórios para resposta binária e dados em painel, a saber, os modelos estáticos e os dinâmicos. Em qualquer um desses casos, um modelo de regressão para uma resposta binária Y_{it} tomando valores 0 e 1 pode derivar de um modelo para uma variável resposta latente contínua Y_{it}^* , desde que se considere que Y_{it} assumo o valor 1 se e somente se Y_{it}^* for positiva e esse é o modo mais comum de criar modelos de resposta binária na literatura. Por esse motivo, é conveniente indicar como é tratado o tema de regressão com efeitos aleatórios para variáveis contínuas, pelo menos no caso estático. Esta é a razão de haver o capítulo 2 com os principais resultados acerca de modelos de efeitos aleatórios estáticos e seu ajuste, para variáveis contínuas.

No capítulo 3, inicia-se a explanação de modelos de resposta binária e de seu ajuste por máxima verossimilhança, sendo tratados os modelos para observações em um corte transversal, e a seguir, os modelos estáticos com efeitos aleatórios para dados em painel.

O capítulo 4 é dedicado exclusivamente ao tema de modelos dinâmicos de efeitos fixos e aleatórios para variável resposta binária. Primeiramente é apresentado o modelo de efeitos fixos de HONORÉ, BO E.; KYRIAZIDOU, E.(2000). São apresentados a seguir dois modelos de efeitos aleatórios, o mais tradicional, que se baseia em uma proposta de Heckman para modelar as condições iniciais, isto é, para $P(Y_{i0} = 1|\alpha_i, \mathbf{x}_0)$ cujo ajuste por máxima verossimilhança é

¹Ver CHAY, K.Y.; HYSLOP, D.,2000

computacionalmente muito trabalhoso e um modelo recentemente proposto por Wooldridge que não considera a especificação $P(Y_{i0} = 1|\alpha_i, \mathbf{x}_0)$, mas, ao contrário, faz hipóteses sobre $P(Y_{i0} = 1|\alpha_i, \mathbf{x}_0)$ o que reduz o problema computacional do ajuste do modelo dinâmico de resposta binária com efeitos aleatórios, por torná-lo semelhante ao ajuste de um modelo estático, como será explicado no capítulo 4.

No capítulo 5 são apresentados os resultados dos estudos por simulação feitos para estudar propriedades dos estimadores obtidos. No capítulo 6 são feitas as conclusões.

Capítulo 2

Modelos Estáticos de Regressão: Variável Resposta Contínua e Dados em Painel

Neste capítulo serão apresentados os modelos de regressão com variável resposta do tipo contínuo utilizando dados em painel. Serão apresentadas também as técnicas que geralmente são usadas para analisar esse tipo de conjunto de dados em painel em modelos econométricos estáticos, ou seja, sem variável resposta defasada no modelo.

Um modelo de regressão básico pode ser dado por: ¹

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_i\alpha + U_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T. \quad (2.1)$$

No modelo acima \mathbf{x}'_{it} é um vetor linha de dimensão k de variáveis explicativas não constantes (para t fixo), $\boldsymbol{\beta}$ é o vetor $k \times 1$ de coeficientes, Y_{it} é a variável resposta (contínua) de interesse e U_{it} é uma variável aleatória chamada de erro transitório.

A heterogeneidade, ou efeito individual é $\mathbf{z}'_i\alpha$ onde \mathbf{z}_i contém um termo constante e um grupo de variáveis constantes no decorrer do tempo, que podem ser:

¹A fonte principal de referência desse capítulo é dada por Greene (2003).

- **observadas**, por exemplo: raça, sexo, locação etc;
- **não observadas**, tais como: características específicas da família, ou de determinada firma, heterogeneidade individual em habilidades ou preferências etc.

Se \mathbf{z}_i é observado para todos os indivíduos, então o modelo completo pode ser tratado como um modelo linear ordinário e ajustado por mínimos quadrados. Se na equação (2.1) \mathbf{z}_i contém somente um termo constante, então o método dos mínimos quadrados ordinários (MQO) estima consistente e eficientemente α e o vetor de inclinação β e este ajuste também é conhecido como **Regressão Agrupada** (“*Pooled*”). Quando \mathbf{z}_i não é observado para todos os indivíduos, outros dois modelos podem ser utilizados e serão apresentados a seguir de forma mais detalhada: modelos de **Efeitos Fixos** e modelos de **Efeitos Aleatórios**.

2.1 Modelo de Efeitos Fixos

Se na equação (2.1) \mathbf{z}_i é não-observado, mas é correlacionado com \mathbf{x}_{it} , então o estimador de mínimos quadrados de β (dado pela regressão “*pooled*”) é viciado e inconsistente devido à omissão de variáveis. Um modelo de regressão onde $\alpha_i = \mathbf{z}'_i \alpha$ é considerado um termo específico constante, individual estável no tempo é dado por:

$$Y_{it} = \mathbf{x}'_{it} \beta + \alpha_i + U_{it}. \quad i = 1, 2, \dots, N, \quad t = 1, \dots, T. \quad (2.2)$$

Na equação (2.2) α_i é tratado como um parâmetro não conhecido a ser estimado e Y_{it} é a variável resposta de interesse.

Desta forma o modelo (2.2) de efeitos fixos para dados em painel não-dinâmico (sem variáveis respostas defasadas) pode ser escrito também como:²

$$\mathbf{y}_i = \mathbf{x}_i \beta + \mathbf{i} \alpha_i + \mathbf{u}_i. \quad i = 1, 2, \dots, N. \quad (2.3)$$

Em (2.3), \mathbf{y}_i é o vetor $T \times 1$ da variável resposta, \mathbf{x}_i é a matriz $T \times k$ de variáveis explicativas, β é um vetor $k \times 1$ de coeficientes, \mathbf{i} é o vetor $T \times 1$ de uns e \mathbf{u}_i é

²Ver Greene (2003).

um vetor $T \times 1$ de erros aleatórios que tem uma distribuição não necessariamente conhecida. A expressão (2.3) pode ser escrita como:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{i} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{i} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{i} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{bmatrix}. \quad (2.4)$$

Seja $\mathbf{D} = [\mathbf{d}_1 \cdots \mathbf{d}_N]$ uma matriz de dimensão $NT \times N$, onde \mathbf{d}_i é um vetor ($NT \times 1$) de observações de uma variável “*Dummy*” que representa a i -ésima unidade, $\mathbf{d}_1 = [\mathbf{i}' \quad \mathbf{0}' \quad \cdots \quad \mathbf{0}']'$, $\mathbf{d}_2 = [\mathbf{0}' \quad \mathbf{i}' \quad \cdots \quad \mathbf{0}']'$, \cdots , $\mathbf{d}_N = [\mathbf{0}' \quad \mathbf{0}' \quad \cdots \quad \mathbf{i}']'$ e $\mathbf{0}$ é o vetor $T \times 1$ de zeros.

As NT equações em (2.4) podem então ser escritas como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \mathbf{u}. \quad (2.5)$$

Em (2.5), \mathbf{y} e \mathbf{u} são vetores $NT \times 1$, \mathbf{X} a matriz $NT \times k$ e \mathbf{D} a matriz $NT \times N$, dadas em (2.4), $\mathbf{y} = [\mathbf{y}'_1, \mathbf{y}'_2, \cdots, \mathbf{y}'_N]'$, $\mathbf{X} = [\mathbf{x}'_1, \mathbf{x}'_2, \cdots, \mathbf{x}'_N]'$ e $\mathbf{u} = [\mathbf{u}'_1, \mathbf{u}'_2, \cdots, \mathbf{u}'_N]'$.

O modelo de efeitos fixos (2.5) pode ser ajustado pelo método de Mínimos Quadrados Ordinários (MQO). A implementação deste ajuste pode ser feita de dois modos:

- 1) diretamente, usando um programa que ajuste por MQO, entrando com todas as variáveis regressoras de (2.5) (isto é, todas as variáveis X e todas as variáveis “*dummy*”). Este método de implementação do ajuste é conhecido na literatura como ajuste do modelo de efeitos fixos por Mínimos Quadrados com Variáveis “*Dummy*” abreviado por MQVD.
- 2) em duas etapas, usando o teorema de “Frisch-Waugh”³ em que na 1ª etapa é feito o ajuste de MQO da variável Y e das variáveis explicativas X , em termos das variáveis “*dummy*” em D e na segunda etapa, ajusta-se por MQO uma regressão dos resíduos de Y em relação aos resíduos das variáveis explicativas X , obtidos na primeira etapa. Este modo de implementação do ajuste é conhecido na literatura como ajuste do modelo de efeitos fixos com transformação “dentro” (ou “*within*”).

³Ver Greene, (2003),pág.27.

A seguir será provada uma versão do teorema de Frisch-Waugh.

Teorema 2.1 *O modelo (2.5)*

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \mathbf{u}.$$

pode ser escrito da seguinte forma:

$$\mathbf{y} = \tilde{\mathbf{X}}\boldsymbol{\beta}^* + \mathbf{u}. \quad (2.6)$$

onde:

$$\mathbf{y} = \begin{bmatrix} \mathbf{X} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} + \mathbf{u}. \quad (2.7)$$

$\tilde{\mathbf{X}}$ é um vetor de k variáveis explicativas e \mathbf{D} é a matriz ($NT \times N$) de dummies definida acima. O estimador de MQO de $\boldsymbol{\beta}$ será dado por:

$$\mathbf{b} = \left[\mathbf{X}' \mathbf{M}_D \mathbf{X} \right]^{-1} \left[\mathbf{X}' \mathbf{M}_D \mathbf{y} \right], \quad (2.8)$$

onde $\mathbf{M}_D = \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ é a matriz fabricante de resíduos em uma regressão de qualquer variável V , em termos das variáveis “dummy”.

Uma prova foi desenvolvida:

Prova de (2.8): Para obter o estimador \mathbf{b} temos por (2.5), (2.6) e (2.7) que:

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} & \mathbf{D} \end{bmatrix}, \quad \tilde{\mathbf{X}}' = \begin{bmatrix} \mathbf{X}' \\ \mathbf{D}' \end{bmatrix}, \quad \boldsymbol{\beta}^* = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix},$$

O estimador de MQ do vetor de coeficientes de $\boldsymbol{\beta}^*$ é dado por $\hat{\boldsymbol{\beta}}^* = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}' \mathbf{y})$, com:

$$\hat{\boldsymbol{\beta}}^* = \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix},$$

Calculando

$$(\tilde{\mathbf{X}}' \tilde{\mathbf{X}}) = \begin{bmatrix} \mathbf{X}' \\ \mathbf{D}' \end{bmatrix} \begin{bmatrix} \mathbf{X} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} \\ \mathbf{D}'\mathbf{X} & \mathbf{D}'\mathbf{D} \end{bmatrix},$$

A inversa da matriz $(\tilde{\mathbf{X}}' \tilde{\mathbf{X}})$ será feita pelo método da inversa da matriz em blocos.

Considere a seguinte matriz:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

onde cada A_{ij} é uma matriz, logo A é uma matriz em blocos. A inversa da matriz em blocos é dada por:

$$A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix},$$

onde:

$$A^{11} = [A_{11} - A_{12}A_{22}^{-1}A_{21}]^{-1}$$

$$A^{21} = -A_{22}^{-1}A_{21}A^{11}$$

$$A^{12} = -A^{11}A_{12}A_{22}^{-1}$$

$$A^{22} = A_{22}^{-1} - A^{21}A_{12}A_{22}^{-1}$$

Estes resultados só podem ser utilizados se A e todas as matrizes da diagonal principal forem quadradas e A_{11} e $(A_{22} - A_{21}A_{11}^{-1}A_{12})$ forem não-singulares.

Supondo que todas as exigências foram obedecidas, seja $A^{-1} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1}$.

$$\hat{\boldsymbol{\beta}}^* = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}' \mathbf{y}) = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}' \\ \mathbf{D}' \end{bmatrix} \mathbf{y},$$

$$\hat{\boldsymbol{\beta}}^* = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}' \mathbf{y} \\ \mathbf{D}' \mathbf{y} \end{bmatrix},$$

$$\hat{\boldsymbol{\beta}}^* = \begin{bmatrix} A^{11} \mathbf{X}' \mathbf{y} + A^{12} \mathbf{D}' \mathbf{y} \\ A^{21} \mathbf{X}' \mathbf{y} + A^{22} \mathbf{D}' \mathbf{y} \end{bmatrix},$$

Logo $\mathbf{b} = A^{11} \mathbf{X}' \mathbf{y} + A^{12} \mathbf{D}' \mathbf{y}$ e $\mathbf{a} = A^{21} \mathbf{X}' \mathbf{y} + A^{22} \mathbf{D}' \mathbf{y}$

Usando as relações entre os blocos de A e de A^{-1} descritos acima, temos:

$$A^{11} \mathbf{X}' \mathbf{y} = [A_{11} - A_{12}A_{22}^{-1}A_{21}]^{-1} \mathbf{X}' \mathbf{y}$$

$$\begin{aligned} &= (\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'[I - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (\text{I}) \end{aligned}$$

$$\begin{aligned} &A^{12}\mathbf{D}'\mathbf{y} = -A^{11}A_{12}A_{22}^{-1} \\ &= -(\mathbf{X}'[I - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y} \quad (\text{II}) \end{aligned}$$

$$\mathbf{b} = (\text{I}) + (\text{II}) :$$

$$\mathbf{b} = (\mathbf{X}'[I - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{X})^{-1}(\mathbf{X}'[I - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{y})$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}(\mathbf{X}'\mathbf{M}_D\mathbf{y})$$

Assim, utilizando os dados transformados e os resultados de matriz em blocos, encontramos a expressão do estimador de mínimos quadrados para β dado em (2.8).

■

Comentários:

A matriz \mathbf{M}_D é da seguinte forma (com colunas ortogonais):

$$\mathbf{M}_D = \begin{bmatrix} \mathbf{M}^0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^0 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}^0 \end{bmatrix},$$

Na matriz acima cada elemento da diagonal é uma matriz $T \times T$ do tipo: $\mathbf{M}^0 = \mathbf{I}_T - \frac{1}{T}\mathbf{ii}'$. Ao multiplicar \mathbf{M}^0 por um vetor $T \times 1$ qualquer \mathbf{z}_i obtém-se $\mathbf{z}_i - \bar{z}\mathbf{i}$, onde \bar{z} é a média da variável em T observações para a unidade i , $\bar{z} = \sum_{t=1}^T \mathbf{z}_i / T$. Logo, a regressão de mínimos quadrados de $\mathbf{M}_D\mathbf{y}$ em termos de $\mathbf{M}_D\mathbf{X}$ é equivalente a uma regressão de $Y_{it} - \bar{Y}_i$ em termos de $\mathbf{x}_{it} - \bar{\mathbf{x}}_i$, onde \bar{Y}_i é um escalar e $\bar{\mathbf{x}}_i$ é um vetor ($K \times 1$) de médias de T observações do indivíduo i para todas as k covariáveis.

Por (2.5) temos que $\mathbf{y} = \mathbf{X}\beta + \mathbf{D}\alpha + \mathbf{u}$, $E(\mathbf{u}) = 0$ e $\text{Var}(\mathbf{u}) = \sigma^2$. Multiplicando ambos os lados por \mathbf{D}' , temos o estimador de mínimos quadrados para α dado pela equação normal de uma regressão particionada.

Temos que $\mathbb{E}(\mathbf{u}) = \mathbf{0}$, pelo método dos momentos, $\mathbb{E}(b) = \beta$ e que $\mathbb{E}(a) = \alpha$, desta forma:

$$\begin{aligned} \mathbf{D}'\mathbf{D}\mathbf{a} + \mathbf{D}'\mathbf{X}\mathbf{b} &= \mathbf{D}'\mathbf{y}, \\ \mathbf{a} &= [\mathbf{D}'\mathbf{D}]^{-1}\mathbf{D}'(\mathbf{y} - \mathbf{X}\mathbf{b}), \end{aligned}$$

Assim para cada indivíduo i ($i = 1, \dots, N$), tem-se:

$$a_i = \bar{Y}_i - \mathbf{b}'\bar{\mathbf{x}}_i. \quad (2.9)$$

Então o valor predito de Y_{it} para o modelo completo, com todas as regressoras \mathbf{X} e \mathbf{D} é dado por:

$$\hat{Y}_{it} = \mathbf{x}'_{it}\mathbf{b} + \bar{Y}_i - \mathbf{b}'\bar{\mathbf{x}}_i,$$

e o correspondente resíduo será:

$$\hat{u}_{it} = (Y_{it} - \bar{Y}_i) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\mathbf{b}.$$

É importante observar que este mesmo resíduo é obtido no ajuste do modelo de regressão sem variáveis *Dummy* e com todas as variáveis Y_{it} e \mathbf{x}_{it} transformadas em suas diferenças em torno das médias \bar{Y}_i e $\bar{\mathbf{x}}_i$, sendo $\bar{\mathbf{x}}_i$ o vetor de médias das observações de \mathbf{x} para o indivíduo i .

Teorema 2.2 *Sob as hipóteses do Teorema 2.1 tem-se que:*

$$\text{Var}[\mathbf{b}] = \sigma^2 \left(\mathbf{X}'\mathbf{M}_D\mathbf{X} \right)^{-1}.$$

O estimador da variância σ^2 é:

$$s^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \mathbf{x}'_{it}\mathbf{b} - a_i)^2}{NT - N - k}. \quad (2.10)$$

Corolário 2.1 *Greene (2003) apresenta uma estimativa da variância de \mathbf{b} :*

$$\text{Var}[\mathbf{b}] = s^2[\mathbf{X}'\mathbf{M}_D\mathbf{X}]^{-1}. \quad (2.11)$$

Prova: Pelo resultado dado em (2.8), temos que:

$$\begin{aligned}
 \mathbf{b} &= [\mathbf{X}'\mathbf{M}_D\mathbf{X}]^{-1} [\mathbf{X}'\mathbf{M}_D\mathbf{y}] \\
 \mathbf{b} &= [\mathbf{X}'\mathbf{M}_D\mathbf{X}]^{-1} [\mathbf{X}'\mathbf{M}_D(\mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \mathbf{u})] \\
 \mathbf{b} &= \boldsymbol{\beta} + [\mathbf{X}'\mathbf{M}_D\mathbf{X}]^{-1} [\mathbf{X}'\mathbf{M}_D(\mathbf{D}\boldsymbol{\alpha} + \mathbf{u})] \\
 (\mathbf{b} - \boldsymbol{\beta}) &= [\mathbf{X}'\mathbf{M}_D\mathbf{X}]^{-1} [\mathbf{X}'\mathbf{M}_D(\mathbf{D}\boldsymbol{\alpha} + \mathbf{u})] \\
 (\mathbf{b} - \boldsymbol{\beta})' &= [\mathbf{X}'\mathbf{M}_D(\mathbf{D}\boldsymbol{\alpha} + \mathbf{u})]' \left[[\mathbf{X}'\mathbf{M}_D\mathbf{X}]^{-1} \right]'
 \end{aligned}$$

$$\text{Var}[\mathbf{b}|X] = \mathbb{E}[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'|X]$$

$$\begin{aligned}
 \text{Var}[\mathbf{b}|X] &= \mathbb{E} \left[\left([\mathbf{X}'\mathbf{M}_D\mathbf{X}]^{-1} (\mathbf{X}'\mathbf{M}_D\mathbf{D}\boldsymbol{\alpha} + \mathbf{X}'\mathbf{M}_D\mathbf{u}) \times \right. \right. \\
 &\quad \left. \left. (\boldsymbol{\alpha}'\mathbf{D}'\mathbf{M}'_D\mathbf{X} + \mathbf{u}'\mathbf{M}'_D\mathbf{X}) [\mathbf{X}'\mathbf{M}_D\mathbf{X}]^{-1} \right) |X \right] \\
 &= (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1} \times \mathbb{E} \left[(\mathbf{X}'\mathbf{M}_D\mathbf{D}\boldsymbol{\alpha}\boldsymbol{\alpha}'\mathbf{D}'\mathbf{M}'_D\mathbf{X} + \mathbf{X}'\mathbf{M}_D\mathbf{D}\boldsymbol{\alpha}\mathbf{u}'\mathbf{M}'_D\mathbf{X} + \right. \\
 &\quad \left. \mathbf{X}'\mathbf{M}_D\mathbf{u}\boldsymbol{\alpha}'\mathbf{D}'\mathbf{M}'_D\mathbf{X} + \mathbf{X}'\mathbf{M}_D\mathbf{u}\mathbf{u}'\mathbf{M}'_D\mathbf{X}) |X \right] \times (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1} \\
 &= (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1} \times \left\{ \mathbb{E} \left[(\mathbf{X}'\mathbf{M}_D\mathbf{D}\boldsymbol{\alpha}\boldsymbol{\alpha}'\mathbf{D}'\mathbf{M}'_D\mathbf{X}) |X \right] + \mathbb{E} \left[(\mathbf{X}'\mathbf{M}_D\mathbf{D}\boldsymbol{\alpha}\mathbf{u}'\mathbf{M}'_D\mathbf{X}) |X \right] + \right. \\
 &\quad \left. \mathbb{E} \left[(\mathbf{X}'\mathbf{M}_D\mathbf{u}\boldsymbol{\alpha}'\mathbf{D}'\mathbf{M}'_D\mathbf{X}) |X \right] + \mathbb{E} \left[(\mathbf{X}'\mathbf{M}_D\mathbf{u}\mathbf{u}'\mathbf{M}'_D\mathbf{X}) |X \right] \right\} \times (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}
 \end{aligned}$$

Como $\mathbf{M}_D = \mathbf{0}$ pois $\mathbf{M}_D = \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$, a primeira parcela da expressão acima se anula. Lembrando que $\mathbb{E}[\mathbf{u}|X] = \mathbf{0}$ e $\mathbb{E}[\mathbf{u}\mathbf{u}'|X] = \boldsymbol{\sigma}^2\mathbf{I}_{NT}$, a variância de \mathbf{b} dado X será: ⁴

$$\text{Var}[\mathbf{b}|X] = \boldsymbol{\sigma}^2 (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}.$$

$$s^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \mathbf{x}'_{it}\mathbf{b} - a_i)^2}{NT - N - k}. \quad (2.12)$$

■

⁴Para maiores detalhes ver Greene (2003), pág.68

De (2.9) segue que a variância assintótica do estimador de um efeito individual, a_i é dado por:

$$\text{Var.Ass}[a_i] = \frac{\sigma^2}{T} + \bar{\mathbf{x}}_i' \{ \text{Var.Ass}[\mathbf{b}] \} \bar{\mathbf{x}}_i.$$

Nesta seção foram apresentados detalhes da estimação por MQO do modelo de regressão linear de efeitos fixos. A seguir será considerada a estimação por Mínimos Quadrados do modelo de efeitos aleatórios.

2.2 Modelo de Efeitos Aleatórios

Supondo que a heterogeneidade não-observável (ou efeito específico, citado anteriormente) seja não-correlacionada com as variáveis explicativas do modelo, o modelo a ser utilizado é o modelo de efeitos aleatórios.

A equação (2.1) pode ser escrita da seguinte forma:

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + (\alpha^* + \alpha_i) + U_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T. \quad (2.13)$$

onde $\mathbf{z}'_i\boldsymbol{\alpha} = (\alpha^* + \alpha_i)$, sendo α^* a média de $\mathbf{z}'_i\boldsymbol{\alpha}$ e α_i um elemento aleatório específico do grupo, que pode ter uma distribuição de probabilidade não necessariamente conhecida, \mathbf{x}'_{it} é um vetor linha de variáveis explicativas de dimensão k , não constantes, $\boldsymbol{\beta}$ é o vetor de coeficientes de dimensão $k \times 1$, Y_{it} é a variável resposta (contínua) de interesse e U_{it} é uma variável aleatória chamada de erro transitório.

Sendo \mathbf{X} a matriz $NT \times k$ de observações das k variáveis de N indivíduos em T tempos, para o modelo (2.13) assume-se que:

$$\mathbb{E}[U_{it}|\mathbf{X}] = \mathbb{E}[\alpha_i|\mathbf{X}] = 0,$$

$$\mathbb{E}[U_{it}^2|\mathbf{X}] = \sigma_u^2,$$

$$\mathbb{E}[\alpha_i^2|\mathbf{X}] = \sigma_\alpha^2,$$

$$\mathbb{E}[U_{it}\alpha_j|\mathbf{X}] = 0 \quad \forall \quad i, \quad t \quad \text{e} \quad j, i = 1, \dots, N \quad \text{e} \quad t = 1, \dots, T. \quad (2.14)$$

$$\mathbb{E}[U_{it}U_{js}|\mathbf{X}] = 0 \quad \text{se} \quad t \neq s \quad \text{ou} \quad i \neq j,$$

$$\mathbb{E}[\alpha_i, \alpha_j | \mathbf{X}] = 0 \quad \text{se } i \neq j.$$

Definindo:

$$\delta_{it} = \alpha_i + U_{it},$$

o modelo (2.13) pode ser escrito como ($\alpha^* = 0$):

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \delta_{it}, \quad i = 1, \dots, N \quad \text{e} \quad t = 1, \dots, T. \quad (2.15)$$

Devido a forma de δ_{it} , (2.13) tem sido chamado de “modelo de componentes de erro”. Para este modelo:

$$\mathbb{E}[\delta_{it}^2 | \mathbf{X}] = \sigma_u^2 + \sigma_\alpha^2,$$

$$\mathbb{E}[\delta_{it}\delta_{is} | \mathbf{X}] = \sigma_\alpha^2 \quad t \neq s,$$

$$\mathbb{E}[\delta_{it}\delta_{js} | \mathbf{X}] = 0 \quad \forall \quad t \quad \text{e} \quad s \quad \text{para} \quad i \neq j.$$

Para T observações para a unidade i , seja \mathbf{i}_T um vetor coluna de uns de dimensão $T \times 1$, $\boldsymbol{\delta}_i = [\delta_{i1}, \delta_{i2}, \dots, \delta_{iT}]'$ e $\boldsymbol{\Sigma}_i = \mathbb{E}[\boldsymbol{\delta}_i\boldsymbol{\delta}'_i | \mathbf{X}]$. Então, sendo \mathbf{I}_T a matriz identidade $T \times T$,

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_u^2 + \sigma_\alpha^2 \end{bmatrix}_{T \times T} = \sigma_u^2 \mathbf{I}_T + \sigma_\alpha^2 \mathbf{i}_T \mathbf{i}'_T, \quad (2.16)$$

■

Com as suposições feitas a respeito das variáveis δ_{it} para $i = 1, \dots, N$ e $t = 1, \dots, T$, conclui-se que sua matriz de covariância conjunta $\boldsymbol{\Omega}$ é uma matriz bloco diagonal que pode ser dada pelo produto de Kronecker de uma matriz identidade $N \times N$, indicada por \mathbf{I}_N por $\boldsymbol{\Sigma}_i$, isto é:

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma} \end{bmatrix}_{NT \times NT} = \mathbf{I}_N \otimes \boldsymbol{\Sigma}. \quad (2.17)$$

No modelo de efeitos fixos o método de estimação utilizado foi o de Mínimos Quadrados Ordinários. No modelo de efeitos aleatórios o método de estimação a ser utilizado é o de Mínimos Quadrados Generalizados (MQG) e na sub-seção a seguir são apresentados detalhes desse ajuste.

2.2.1 Estimação por Mínimos Quadrados Generalizados (MQG)

No modelo (2.15), supondo $\mathbf{\Omega}$ conhecido, o estimador de mínimos quadrados generalizados de $\boldsymbol{\beta}$ é dado por:

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

Na expressão acima \mathbf{y} é o vetor ($NT \times 1$) de observações da variável resposta, \mathbf{X} é a matriz ($NT \times K$) de observações das k variáveis explicativas em T tempos para N indivíduos e $\hat{\boldsymbol{\beta}}$ é o estimador de $\boldsymbol{\beta}$.

Uma maneira alternativa de calcular $\hat{\boldsymbol{\beta}}$ é usar o método de mínimos quadrados ordinários para estimar $\boldsymbol{\beta}$ no modelo transformado: $P\mathbf{y} = P\mathbf{X}\boldsymbol{\beta} + P\mathbf{u}$ que é homocedástico se $P = \mathbf{\Omega}^{-1/2}$, com $\mathbf{\Omega}^{-1/2} = [\mathbf{I}_N \otimes \boldsymbol{\Sigma}]^{-1/2}$ e

$$\boldsymbol{\Sigma}^{-1/2} = \frac{1}{\sigma_u} \left[\mathbf{I} - \frac{\theta}{T} \mathbf{i}_T \mathbf{i}_T' \right],$$

onde

$$\theta = 1 - \frac{\sigma_u}{\sigma_u^2 + T\sigma_\alpha^2}.$$

Sendo \mathbf{y}_i o vetor $T \times 1$ das observações da variável resposta Y no indivíduo i , observa-se que \mathbf{y}_i é transformado por P em $P\mathbf{y}_i$ dado por:

$$P\mathbf{y}_i = \boldsymbol{\Sigma}^{-1/2}\mathbf{y}_i = \frac{1}{\sigma_u} \begin{bmatrix} Y_{i1} - \theta\bar{Y}_i \\ Y_{i2} - \theta\bar{Y}_i \\ \vdots \\ Y_{iT} - \theta\bar{Y}_i \end{bmatrix}. \quad (2.18)$$

Uma forma análoga vale para $P\mathbf{x}_i$.

Para o conjunto de dados como um todo, a estimação por mínimos quadrados generalizados pode então ser calculada utilizando a regressão dos desvios parciais de Y_{it} ($Y_{it} - \theta\bar{Y}_i$) em relação aos desvios parciais ($\mathbf{x}_{it} - \theta\bar{\mathbf{x}}_i$) de \mathbf{x}_{it} . Este procedimento pode ser comparado ao utilizado na seção 2.1 onde foi mostrada a estimação para o modelo de mínimos quadrados com variáveis Dummy (MQVD), que usa $\theta = 1$.

Quando os componentes de variância (σ_u^2 e σ_α^2) são conhecidos, o método de estimação a ser utilizado é o MQG. Geralmente esses componentes são não-conhecidos, e desta forma deve-se primeiro estimá-los e em seguida, estimando Σ , adotar um procedimento factível de estimação por mínimos quadrados generalizados para os coeficientes. Isto será apresentado na sub-seção a seguir.

2.2.2 Estimação via MQG Factível

Estimação dos componentes de variância:

Do modelo de efeitos aleatórios dado em 2.13: ⁵

$$Y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha^* + \alpha_i + U_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (2.19)$$

segue para o i -ésimo indivíduo:

$$\bar{Y}_i = \bar{\mathbf{x}}'_i\boldsymbol{\beta} + \alpha^* + \alpha_i + \bar{U}_i. \quad (2.20)$$

Como $\boldsymbol{\beta}$ é fixo para todos os indivíduos em todos os tempos e $\bar{U}_i = \sum_{t=1}^T U_{it}/T$,

$$Y_{it} - \bar{Y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \boldsymbol{\beta} + (U_{it} - \bar{U}_i), \quad (2.21)$$

obtém-se uma equação que não envolve α_i .

$$\mathbb{E}\left[\sum_{i=1}^T (U_{it} - \bar{U}_i)^2\right] = (T-1)\sigma_u^2.$$

Estimação de σ_u^2 :

Se $\boldsymbol{\beta}$ fosse observado, $(U_{it} - \bar{U}_i)$ também o seria e um estimador não-viciado de σ_u^2 baseado em T observações do grupo i seria dado por:

$$\hat{\sigma}_u^2(i) = \frac{\sum_{t=1}^T (U_{it} - \bar{U}_i)^2}{T-1}. \quad (2.22)$$

Um estimador consistente de (2.22) pode ser obtido fazendo um ajuste por MQO de um modelo de regressão com variáveis dummy (MQVD), calculando os resíduos dessa regressão auxiliar, calculando para cada indivíduo i um quadrado médio

⁵Ver Greene (2003)

($s_u^2(i)$) dos T resíduos referentes as observações de y_{it} e, finalmente, tomando a média desses quadrados médios, em relação aos N diferentes indivíduos do painel:

$$s_u^2(i) = \frac{\sum_{t=1}^T (\hat{U}_{it} - \bar{U}_i)^2}{T - K - 1}. \quad (2.23)$$

Na expressão (2.23), $\bar{U}_i = \sum_{t=1}^T \hat{U}_{it}/T$.

Há N estimadores semelhantes e fazendo a média obtém-se:

$$\begin{aligned} \bar{s}_u^2 &= \frac{1}{N} \sum_{i=1}^N s_u^2(i) = \frac{1}{N} \sum_{i=1}^N \left[\frac{\sum_{t=1}^T (\hat{U}_{it} - \bar{U}_i)^2}{T - K - 1} \right]. \\ \bar{s}_u^2 &= \frac{\sum_{i=1}^N \sum_{t=1}^T (\hat{U}_{it} - \bar{U}_i)^2}{NT - NK - N}. \end{aligned} \quad (2.24)$$

Um estimador não-viciado é dado por: ⁶

$$\hat{\sigma}_u^2 = s_{MQVD}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (\hat{U}_{it} - \bar{U}_i)^2}{NT - K - N}. \quad (2.25)$$

Como a equação (2.23) é igual a equação (2.12), conclui-se que o estimador do componente de variância σ_u^2 do modelo (2.13) é dado pelo quadrado médio residual de um ajuste por mínimos quadrados ordinários de um modelo de regressão com efeitos fixos (MQVD).

Estimação de σ_α^2 .

Usando o quadrado médio de resíduos ($s_{Agrupado}^2$) de um ajuste por mínimos quadrados ordinários de um modelo denominado Regressão Agrupada (“Pooled”), que é um modelo obtido de (2.1) tirando os efeitos individuais (\mathbf{z}_i') e adicionando um intercepto, obtém-se um estimador consistente da variância de Y_{it} nesse modelo, que é igual a variância de Y_{it} no modelo (2.13), isto é, $\sigma_u^2 + \sigma_\alpha^2$.

Quando isto ocorre outra opção para obter estimadores consistentes (embora viciados) dos componentes de variância, consiste em retirar as correções feitas para os graus de liberdade em (2.24) e (2.25) (que foram feitas para obter estimadores não-viesados da variância). Desta forma, os dois estimadores de variância serão não-negativos, já que a soma de quadrados de resíduos no modelo MQVD é menor do que a da regressão simples com somente um termo constante. Estas estimativas de variância estão, por exemplo, no software LIMDEP.

⁶Ver Judge (1984), p. 484

Outra sugestão para estimar os componentes de variância é usar o quadrado médio de resíduos de um ajuste do modelo por Mínimos Quadrados Ordinários da expressão (2.20) como um estimador consistente da variância de \bar{y}_i que é indicada por $\hat{\sigma}_{Entre}^2 = \hat{\sigma}_u^2/T + \hat{\sigma}_\alpha^2$. Dessa expressão segue que $\hat{\sigma}_\alpha^2 = \hat{\sigma}_{Entre}^2 - \hat{\sigma}_u^2/T$. Pode-se encontrar estimadores de σ_u^2 e σ_α^2 através das expressões a seguir:

$$\hat{\sigma}_u^2 = \frac{T}{T-1}(\hat{\sigma}_{Agrupado}^2 - \hat{\sigma}_{Entre}^2)$$

$$\hat{\sigma}_\alpha^2 = \frac{T}{T-1}\hat{\sigma}_{Entre}^2 - \frac{1}{T-1}\hat{\sigma}_{Agrupado}^2 = \varpi\hat{\sigma}_{Entre}^2 + (1-\varpi)\hat{\sigma}_{Agrupado}^2$$

onde $\varpi > 1$. Estes estimadores podem produzir uma estimativa negativa de $\hat{\sigma}_u^2$, o que sugere a necessidade de considerar outros estimadores para os componentes de variância.

Fuller e Battese (1974) propuseram outros estimadores para o componente de variância σ_α^2 baseado no método “ajuste de constantes” devido a Hensdenson. Estas estimativas estão implementadas no software SAS.

Estimação dos parâmetros de regressão por MQGF

A partir das estimativas dos componentes de variância, obtém-se uma estimativa $(\tilde{\theta}, \tilde{\Sigma})$ de (θ, Σ) e portanto também se obtém uma estimativa \tilde{P} de $P = \Omega^{-1/2}$. A estimativa por MQG factível de β é obtida por MQO aplicados aos dados transformados através de \tilde{P} como em 2.18.

Neste capítulo foram apresentados os modelos de regressão para uma variável resposta contínua usando dados em painel. No próximo capítulo serão apresentados os modelos de regressão para variável resposta do tipo Bernoulli para observações independentes e para dados em painel.

Capítulo 3

Modelos Estáticos de Regressão para Variável Resposta Binária

Neste capítulo será feito um levantamento de modelos de regressão propostos para o caso em que a variável resposta Y é uma variável resposta binária. Um exemplo de variável resposta binária em modelos de economia é a variável indicadora da participação do indivíduo no mercado de trabalho (PMT): em uma pesquisa sobre desemprego, uma amostra de N indivíduos é entrevistada em uma determinada data e cada indivíduo deve responder se, naquela data está empregado ou não. No primeiro caso, o valor observado de Y é 1, e é 0 caso contrário. As variáveis explicativas indicadas por X que podem influenciar ou afetar a variável resposta seriam por exemplo: sexo, idade, estado civil, número de filhos etc.

3.1 Modelos de Resposta Binária para observações independentes

Os N indivíduos selecionados no exemplo de pesquisa acima descrito constituem um corte transversal da população de interesse e os dados obtidos podem ser apresentados como na Tabela 3.1 abaixo, onde $Y = PMT$.

TABELA 3.1: *Dados de um corte transversal da variável binária PMT e covariáveis.*

Indivíduo	PMT	Sexo	Idade	Nº de filhos
1	1	M	19	1
2	0	F	32	3
3	1	M	25	2
⋮	⋮	⋮	⋮	⋮
N	1	F	52	4

No exemplo da Tabela 3.1 a probabilidade de um indivíduo participar do mercado de trabalho, $P(PMT = 1)$, pode não ser constante mas sim depender das outras variáveis observáveis. Essa dependência pode ser expressa através de um modelo de regressão, que é uma expressão do tipo: $P[PMT = 1] = \text{Função}(\text{constante, idade, renda familiar, educação, número de filhos})$

Esse modelo assume que a probabilidade da variável resposta “Participação no Mercado do Trabalho” (PMT) valer um é uma função das variáveis explicativas: idade, renda familiar, educação, número de filhos e de uma constante.

De um modo geral, um modelo de regressão para uma variável resposta binária Y pode ser dado por:

$$P(Y = 1|\mathbf{x}) = F(\mathbf{x}, \boldsymbol{\beta}) \quad (3.1)$$

$$P(Y = 0|\mathbf{x}) = 1 - F(\mathbf{x}, \boldsymbol{\beta}).$$

Usualmente $F(\mathbf{x}, \boldsymbol{\beta}) = F(\mathbf{x}'\boldsymbol{\beta})$.

Uma razão para se utilizar a formulação acima está ligada à idéia de uma variável Y^* não observável, latente, do tipo contínuo, que determina a variável observável Y que toma os valores 1 (quando $Y^* > 0$) ou 0 (quando $Y^* < 0$). A **variável latente** pode ser por exemplo o cálculo que o indivíduo faz para decidir se irá cometer crimes: se seus benefícios forem maiores que seus custos, o indivíduo tem grande chance de cometer o crime. Outro exemplo seria a tomada de decisão de um indivíduo para fazer uma grande compra: a teoria do comportamento do consumidor estabelece que o consumidor considera o benefício de fazer ou não a

compra ou utilizar o dinheiro para outra coisa. A teoria econômica diz que não é possível observar o benefício de se efetuar uma compra calculado por uma pessoa, só sendo possível observar se a compra é efetuada ou não, assim como também não é possível observar o benefício que uma pessoa julga obter se cometer um crime, só sendo possível observar se o crime foi realizado ou não.

Supondo que exista um modelo de regressão linear para a variável latente Y^* em termos de covariáveis ficará definido um modelo linear generalizado para Y . De fato, se Y^* for dado pela expressão (3.2), onde U tem distribuição logística com média zero e variância conhecida, por exemplo $\pi^2/3$, ou normal com média zero e variância um, \mathbf{x}' é o vetor $1 \times k$ de variáveis explicativas e $\boldsymbol{\beta}$ é o vetor $k \times 1$ de coeficientes, então (3.3) definirá um modelo de regressão para Y .

$$Y^* = \mathbf{x}'\boldsymbol{\beta} + U. \quad (3.2)$$

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y^* > 0) = \mathbb{P}(\mathbf{x}'\boldsymbol{\beta} + U > 0)$$

$$\mathbb{P}(Y = 1) = \mathbb{P}(U > -(\mathbf{x}'\boldsymbol{\beta}))$$

Se a função densidade de U for simétrica, sendo $F(\cdot)$ a função de distribuição de U , segue que:

$$\mathbb{P}(Y = 1) = \mathbb{P}(U < (\mathbf{x}'\boldsymbol{\beta})) = F(\mathbf{x}'\boldsymbol{\beta}) \quad (3.3)$$

$$\mathbb{P}(Y = 0) = 1 - \mathbb{P}(Y = 1) = 1 - F(\mathbf{x}'\boldsymbol{\beta})$$

Este modelo também pode ser visto como um modelo de regressão para a média da variável resposta Y :

$$\mathbb{E}[Y|\mathbf{x}] = 0[1 - F(\mathbf{x}'\boldsymbol{\beta})] + 1[F(\mathbf{x}'\boldsymbol{\beta})] = F(\mathbf{x}'\boldsymbol{\beta}). \quad (3.4)$$

Se $F(\cdot)$ (em 3.3) for a função distribuição da Normal padrão, o modelo de regressão é o modelo **Probito**:

$$\mathbb{P}(Y = 1|\mathbf{x}) = \int_{-\infty}^{\mathbf{x}'\boldsymbol{\beta}} \phi(t)dt = \Phi(\mathbf{x}'\boldsymbol{\beta}). \quad (3.5)$$

Se $F(\cdot)$ (em 3.3) for a função distribuição da logística, o modelo de regressão é o modelo **Logito**:

$$\mathbb{P}(Y = 1|\mathbf{x}) = \frac{e^{\mathbf{x}'\boldsymbol{\beta}}}{1 + e^{\mathbf{x}'\boldsymbol{\beta}}} = \Lambda(\mathbf{x}'\boldsymbol{\beta}). \quad (3.6)$$

3.1.1 Estimação dos parâmetros

A estimação dos modelos de resposta binária é usualmente baseada no método da máxima verossimilhança. Cada observação é tratada como uma retirada simples de uma distribuição de Bernoulli. N observações independentes deste modelo com probabilidades de sucesso $F(\mathbf{x}'_i\boldsymbol{\beta})$, $i = 1, \dots, N$, têm função densidade conjunta dada por:

$$P(Y_1 = y_1, \dots, Y_N = y_N | \mathbf{x}) = \prod_{y_i=0} [1 - F(\mathbf{x}'_i\boldsymbol{\beta})] \prod_{y_i=1} [F(\mathbf{x}'_i\boldsymbol{\beta})]. \quad (3.7)$$

A função de verossimilhança para uma amostra de N observações pode ser dada por:

$$L(\boldsymbol{\beta} | \text{dados}) = \prod_{i=1}^N [1 - F(\mathbf{x}'_i\boldsymbol{\beta})]^{1-y_i} [F(\mathbf{x}'_i\boldsymbol{\beta})]^{y_i}. \quad (3.8)$$

Aplicando logaritmo obtém-se:

$$\ln L = \sum_{i=1}^N \{ (1 - y_i) \ln [1 - F(\mathbf{x}'_i\boldsymbol{\beta})] + (y_i) \ln [F(\mathbf{x}'_i\boldsymbol{\beta})] \}. \quad (3.9)$$

A derivada da log-verossimilhança com relação a $\boldsymbol{\beta}$ é:

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \left[\frac{y_i f_i}{F_i} + (1 - y_i) \frac{-f_i}{(1 - F_i)} \right] \mathbf{x}_i, \quad (3.10)$$

onde $f_i = \frac{\partial F_i}{\partial (\mathbf{x}'_i\boldsymbol{\beta})}$, $\mathbf{x}'_i\boldsymbol{\beta}$ é o argumento de f_i .

Modelo Logito

No caso do modelo Logito (3.6), obtém-se:

$$\begin{aligned} \frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^N \left[\frac{y_i \frac{e^{\mathbf{x}'_i\boldsymbol{\beta}}}{(1+e^{\mathbf{x}'_i\boldsymbol{\beta}})^2}}{\frac{e^{\mathbf{x}'_i\boldsymbol{\beta}}}{(1+e^{\mathbf{x}'_i\boldsymbol{\beta}})}} + \frac{(1-y_i) \frac{-e^{\mathbf{x}'_i\boldsymbol{\beta}}}{(1+e^{\mathbf{x}'_i\boldsymbol{\beta}})^2}}{\frac{1}{(1+e^{\mathbf{x}'_i\boldsymbol{\beta}})}} \right] \mathbf{x}_i \\ &= \sum_{i=1}^N \left[\frac{y_i}{(1+e^{\mathbf{x}'_i\boldsymbol{\beta}})} + (1-y_i) \frac{-e^{\mathbf{x}'_i\boldsymbol{\beta}}}{(1+e^{\mathbf{x}'_i\boldsymbol{\beta}})} \right] \mathbf{x}_i \\ &= \sum_{i=1}^N \left[\frac{y_i - e^{\mathbf{x}'_i\boldsymbol{\beta}} + y_i e^{\mathbf{x}'_i\boldsymbol{\beta}}}{(1+e^{\mathbf{x}'_i\boldsymbol{\beta}})} \right] \mathbf{x}_i \\ &= \sum_{i=1}^N \left[\frac{y_i(1+e^{\mathbf{x}'_i\boldsymbol{\beta}})}{(1+e^{\mathbf{x}'_i\boldsymbol{\beta}})} - \frac{e^{\mathbf{x}'_i\boldsymbol{\beta}}}{(1+e^{\mathbf{x}'_i\boldsymbol{\beta}})} \right] \mathbf{x}_i \end{aligned}$$

Finalmente,

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N (y_i - \Lambda_i) \mathbf{x}_i. \quad (3.11)$$

A Equação de log-verossimilhança é não linear e exigirá uma solução iterativa.

Modelo Probit

Para o modelo **probit**, pela equação (3.7), a log-verossimilhança é:

$$\ln L = \sum_{y_i=0} \ln[1 - \Phi(\mathbf{x}'_i \boldsymbol{\beta})] + \sum_{y_i=1} \ln[\Phi(\mathbf{x}'_i \boldsymbol{\beta})], \quad (3.12)$$

Derivando (3.12) em relação aos elementos do vetor $\boldsymbol{\beta}$ obtém-se:

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{y_i=0} \frac{-\phi_i}{1 - \Phi_i} \mathbf{x}_i + \sum_{y_i=1} \frac{\phi_i}{\Phi_i} \mathbf{x}_i = \sum_{y_i=0} \lambda_i^0 \mathbf{x}_i + \sum_{y_i=1} \lambda_i^1 \mathbf{x}_i, \quad (3.13)$$

Se $F(\cdot)$ tem densidade simétrica em torno do zero, $1 - F(\mathbf{x}' \boldsymbol{\beta}) = F(-\mathbf{x}' \boldsymbol{\beta})$. Seja $q_i = 2y_i - 1$. Então $\ln L = \sum_i \ln F(q_i \mathbf{x}'_i \boldsymbol{\beta})$. Dessa forma a expressão (3.13) pode ser escrita como:

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \left[\frac{q_i \phi(q_i \mathbf{x}'_i \boldsymbol{\beta})}{\Phi(q_i \mathbf{x}'_i \boldsymbol{\beta})} \right] \mathbf{x}_i = \sum_{i=1}^N \lambda_i \mathbf{x}_i, \quad (3.14)$$

Cálculo das derivadas de segunda ordem:

Modelo Logito

$$H = \frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = - \sum_i \Lambda_i (1 - \Lambda_i) \mathbf{x}_i \mathbf{x}'_i. \quad (3.15)$$

Modelo Probit

Para o modelo **Probit** tem-se uma simplificação (Greene 2003) utilizando a variável $\lambda_i = \lambda(y_i, \mathbf{x}'_i \boldsymbol{\beta})$ que é definida em (3.14). A matriz Hessiana pode ser obtida utilizando o seguinte resultado: $d\phi(z)/dz = -z\phi(z)$. Logo:

$$H = \frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^N -\lambda_i (\lambda_i + \mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}_i \mathbf{x}'_i. \quad (3.16)$$

H é uma matriz negativa definida tanto no modelo Probit quanto no Logito, garantindo a concavidade da função de verossimilhança e a existência de um ponto de máximo global. A matriz de variância e covariância dos EMV (Estimadores de Máxima Verossimilhança) pode ser estimada pelo inverso da estimativa da matriz hessiana.

Na próxima seção serão apresentados os efeitos marginais.

3.1.2 Efeitos marginais

Definição 3.1 *Efeito marginal de uma covariável do tipo contínuo X_j sobre a variável resposta em um modelo de regressão é a derivada da função de regressão (ou do modelo) em relação a X_j .*

É importante verificar que os parâmetros de um modelo de regressão não linear não são os efeitos marginais. No caso do modelo de resposta binária, o efeito marginal de \mathbf{x}_j é:

$$\frac{\partial \mathbb{E}[Y|\mathbf{x}]}{\partial \mathbf{x}_j} = \frac{\partial F[\mathbf{x}'\boldsymbol{\beta}]}{\partial \mathbf{x}_j} = \left\{ \frac{\partial F(\mathbf{x}'\boldsymbol{\beta})}{\partial (\mathbf{x}'\boldsymbol{\beta})} \right\} \boldsymbol{\beta}_j = f(\mathbf{x}'\boldsymbol{\beta})\boldsymbol{\beta}_j, \quad (3.17)$$

onde f é a função densidade correspondente à função distribuição acumulada F . O estimador do efeito marginal de \mathbf{x}_j será dado por: $g_j(\mathbf{x}, \hat{\boldsymbol{\beta}}) = f(\mathbf{x}'\hat{\boldsymbol{\beta}})\hat{\boldsymbol{\beta}}_j$.

Se $F(\cdot)$ for a função acumulada da **logística**:

$$\frac{\partial \mathbb{E}[Y|\mathbf{x}]}{\partial \mathbf{x}_j} = \Lambda(\mathbf{x}'\boldsymbol{\beta})[1 + \Lambda(\mathbf{x}'\boldsymbol{\beta})]\boldsymbol{\beta}_j. \quad (3.18)$$

Se $F(\cdot)$ for a função distribuição acumulada da **normal**,

$$\frac{\partial \mathbb{E}[Y|\mathbf{x}]}{\partial \mathbf{x}_j} = \frac{\partial F[\mathbf{x}'\boldsymbol{\beta}]}{\partial \mathbf{x}_j} = \phi(\mathbf{x}'\boldsymbol{\beta})\boldsymbol{\beta}_j, \quad (3.19)$$

onde ϕ é a densidade da distribuição normal padrão.

Se for conhecida a matriz de **covariância** do vetor de estimadores $\hat{\boldsymbol{\beta}}$, a variância do estimador de um efeito marginal de \mathbf{x}_j pode ser obtida pelo método delta. A função $g_j(\mathbf{x}, \boldsymbol{\beta}) = f(\mathbf{x}'\boldsymbol{\beta})\boldsymbol{\beta}_j$ tem derivadas parciais dadas por:

$$\frac{\partial g_j(\mathbf{x}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_l} = \frac{\partial f(\mathbf{x}'\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_l} \boldsymbol{\beta}_j + f(\mathbf{x}'\boldsymbol{\beta}) \frac{\partial \boldsymbol{\beta}_j}{\partial \boldsymbol{\beta}_l}, \quad l = 1, \dots, k. \quad (3.20)$$

Seja $\frac{\partial g_j(\mathbf{x}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ o vetor $1 \times k$ de derivadas parciais. Usando o método delta a variância do efeito marginal da covariável \mathbf{x}_j é aproximadamente dada por:

$$\left(\frac{\partial g_j(\mathbf{x}, \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \right) \mathbf{V} \left(\frac{\partial g_j(\mathbf{x}, \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \right)^t. \quad (3.21)$$

sendo \mathbf{V} a matriz $k \times k$ de covariância dos estimadores dos coeficientes $\boldsymbol{\beta}$.

A expressão do efeito marginal de uma covariável depende dos valores observados das covariáveis \mathbf{x}_j e assim, após a estimação do vetor $\boldsymbol{\beta}$, é possível

encontrar um estimador do *efeito marginal* para cada elemento da amostra. Uma média desses valores ou um *efeito marginal na média* das covariáveis, pode ser usado para mostrar a mudança em $F(\mathbf{x}'\boldsymbol{\beta})$ devido a uma mudança de uma unidade em \mathbf{x}_j .

Na próxima seção serão apresentados os métodos de estimação para os modelos de resposta binária para dados em painel. Diferentemente do que foi apresentado na seção 3.1 as observações dos dados em painel são variáveis aleatórias dependentes.

3.2 Modelos de Resposta Binária para Dados em Painel

Um modelo de regressão para uma variável binária Y_{it} observada nos tempos $t = 1, \dots, T$ em um conjunto de dados em painel pode ser dada em termos de uma variável latente Y_{it}^* :

$$Y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + V_{it}, \quad i = 1, 2, \dots, N \quad e \quad t = 1, 2, \dots, T, \quad (3.22)$$

$$Y_{it} = 1, \text{ se } Y_{it}^* > 0 \quad Y_{it} = 0, \text{ caso contrário,}$$

onde \mathbf{x}_{it}' é um vetor linha de variáveis explicativas não-constante de dimensão k , $\boldsymbol{\beta}$ é o vetor de coeficientes de dimensão $k \times 1$, Y_{it}^* é a variável latente não-observada, Y_{it} é a variável resposta (discreta) de interesse e V_{it} é uma variável aleatória. Na abordagem mais utilizada nos modelos de efeitos fixos ou aleatórios considera-se $V_{it} = U_{it} + \alpha_i$, de forma que o modelo (3.22) passa a ser:

$$Y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + U_{it} + \alpha_i, \quad i = 1, \dots, N \quad t = 1, \dots, T, \quad (3.23)$$

onde α_i é a heterogeneidade individual não-observada. Distinguem-se os modelos de efeitos fixos dos modelos de efeitos aleatórios pelo relacionamento entre α_i e \mathbf{x}_{it} . Se α_i é não relacionado com \mathbf{x}_{it} , a distribuição condicional $f(\alpha_i|\mathbf{x}_{it})$ não depende de \mathbf{x}_{it} , o que produz um modelo de efeitos aleatórios. Se a distribuição é não restrita, e α_i e \mathbf{x}_{it} podem ser correlacionados, então tem-se o modelo de efeitos

fixos. Na subseção a seguir são apresentados mais detalhes sobre os modelos de regressão de efeitos fixos para variável resposta binária para dados em painel.

3.2.1 Modelo de Efeitos Fixos

Se a heterogeneidade não-observável for correlacionada com \mathbf{x}_{it} , a estratégia a ser adotada é basear as análises nas estimativas obtidas pelo modelo de Efeitos Fixos. Para isso pressupõe-se que $\alpha_i = \mathbf{z}'_i \boldsymbol{\alpha}$ seja um termo constante específico individual estável no tempo. Assim, α_i é um parâmetro desconhecido a ser estimado. Pode-se reescrever o modelo (3.22) como:

$$Y_{it}^* = \mathbf{x}'_{it} \boldsymbol{\beta} + \alpha_i + U_{it}, \quad i = 1, \dots, N \quad t = 1, \dots, T, \quad (3.24)$$

$$\begin{cases} Y_{it} = 1 & \text{se } Y_{it}^* > 0 \\ Y_{it} = 0 & \text{caso contrário.} \end{cases}$$

Este modelo tem k parâmetros em $\boldsymbol{\beta}$ a serem estimados e N termos constantes individuais (α_i).

A log-verossimilhança para o modelo de efeitos fixos é:

$$\ln L = \sum_{i=1}^N \sum_{t=1}^T \ln \mathbb{P}(Y_{it} = y_{it} | \alpha_i + \mathbf{x}'_{it} \boldsymbol{\beta}), \quad (3.25)$$

Se $F(\cdot)$ indicar a função distribuição acumulada de U_{it} , e $F(a) = 1 - F(-a)$, então, como em (3.3), segue que:

$$\mathbb{P}(Y_{it} = y_{it} | \mathbf{x}_{it}) = F[q_{it}(\alpha_i + \mathbf{x}'_{it} \boldsymbol{\beta})],$$

sendo $q_{it} = 2y_{it} - 1$.

No caso do modelo Probit,

$$F(q_{it}(\alpha_i + \mathbf{x}'_{it} \boldsymbol{\beta})) = \Phi[q_{it}(\alpha_i + \mathbf{x}'_{it} \boldsymbol{\beta})]$$

e para o modelo Logito:

$$F(q_{it}(\alpha_i + \mathbf{x}'_{it} \boldsymbol{\beta})) = \Lambda[q_{it}(\alpha_i + \mathbf{x}'_{it} \boldsymbol{\beta})]$$

Seja $z_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta}$.

$$\text{Então } F[q_{it}(\alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta})] = F(q_{it}z_{it}),$$

e

$$\ln L = \sum_{i=1}^N \sum_{t=1}^T \ln F(q_{it}z_{it}),$$

Cálculo da derivada da log-verossimilhança com relação a α_i :

Para $1 \leq i \leq N$, temos:

$$\frac{\partial \ln L}{\partial \alpha_i} = \sum_{j=1}^N \sum_{t=1}^T \frac{1}{F(q_{jt}z_{jt})} \frac{\partial F(q_{jt}z_{jt})}{\partial \alpha_i} \frac{\partial q_{jt}z_{jt}}{\partial \alpha_i},$$

mas

$$\frac{\partial q_{jt}z_{jt}}{\partial \alpha_i} = q_{jt} \frac{\partial z_{jt}}{\partial \alpha_i} = q_{jt} \frac{\partial \alpha_j}{\partial \alpha_i},$$

onde $\frac{\partial \alpha_j}{\partial \alpha_i} = 1$ se $i = j$; 0 se $i \neq j$.

Logo a equação de verossimilhança com relação a α_i é:

$$\frac{\partial \ln L}{\partial \alpha_i} = \sum_{t=1}^T \frac{f(q_{it}z_{it})}{F(q_{it}z_{it})} q_{it} = \sum_{t=1}^T g_{it} = 0. \quad (3.26)$$

Cálculo da derivada da log-verossimilhança com relação a $\boldsymbol{\beta}$:

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{j=1}^N \sum_{t=1}^T \frac{1}{F(q_{jt}z_{jt})} f(q_{jt}z_{jt}) \frac{\partial q_{jt}z_{jt}}{\partial \boldsymbol{\beta}},$$

mas

$$\frac{\partial q_{jt}z_{jt}}{\partial \boldsymbol{\beta}} = q_{jt} \frac{\partial z_{jt}}{\partial \boldsymbol{\beta}} = q_{jt} \mathbf{x}_{it},$$

Logo a equação de verossimilhança relativa a β_i é:

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \sum_{t=1}^T \frac{q_{it} f(q_{it}z_{it})}{F(q_{it}z_{it})} \mathbf{x}_{it} = \sum_{t=1}^T g_{it} \mathbf{x}_{it} = \mathbf{0}, \quad (3.27)$$

onde $f(\cdot)$ é a densidade correspondente a $F(\cdot)$. Para o modelo Probit, g_{it} é $q_{it}\phi(q_{it}z_{it})$ e $q_{it}[1 - \Lambda(q_{it}z_{it})]$ para o modelo Logito.

Cálculo da matriz Hessiana da função de log-verossimilhança:

A segunda derivada da log-verossimilhança é uma matriz \mathbf{H} que pode ser subdividida em quatro blocos, $H_{11}, H_{22}, H_{12} = H_{21}$, dados por:

$$H_{11} = \frac{\partial^2 \ln L}{\partial \alpha_i^2} = \sum_{t=1}^T \left[\frac{f'(q_{it}z_{it})}{F(q_{it}z_{it})} - \left(\frac{f(q_{it}z_{it})}{F(q_{it}z_{it})} \right)^2 \right] = \sum_{t=1}^T h_{it} = h_{ii} < 0,$$

$$H_{12} = \frac{\partial^2 \ln L}{\partial \beta \partial \alpha_i} = \sum_{t=1}^T h_{it} \mathbf{x}_{it},$$

$$H_{22} = \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = \sum_{i=1}^N \sum_{t=1}^T h_{it} \mathbf{x}_{it} \mathbf{x}_{it}' = \mathbf{H}_{\beta\beta'},$$

As derivadas das densidades com respeito aos seus respectivos argumentos são: $-(q_{it}z_{it})\phi(q_{it}z_{it})$ para o modelo Probit e $[1 - 2\Lambda(q_{it}z_{it})]$ para o modelo Logito.

Utilizando a fórmula para a inversa de uma matriz subdividida em blocos (Ver Apêndice A), encontra-se uma submatriz $k \times k$ da inversa da Hessiana que corresponde à **matriz de covariância assintótica** dos estimadores de máxima verossimilhança de β , que é dada por:

$$\mathbf{H}^{\beta\beta'} = \left\{ \sum_{i=1}^N \left[\sum_{t=1}^T h_{it} \mathbf{x}_{it} \mathbf{x}_{it}' - \frac{1}{h_{ii}} \left(\sum_{t=1}^T h_{it} \mathbf{x}_{it} \right) \left(\sum_{t=1}^T h_{it} \mathbf{x}_{it}' \right) \right] \right\}^{-1},$$

$$\mathbf{H}^{\beta\beta'} = \left\{ \sum_{i=1}^N \left[\sum_{t=1}^T h_{it} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right] \right\}^{-1}, \quad \text{onde } \bar{\mathbf{x}}_i = \sum_{t=1}^T \frac{h_{it} \mathbf{x}_{it}}{h_{ii}}.$$

Estimação dos parâmetros por máxima verossimilhança

A solução do sistema de equações das verossimilhanças (3.26) e (3.27) descrita pode ser obtido pelo método de Newton Raphson. O estimador de β é função dos estimadores de α . Como a dimensão de α é N , quando $N \rightarrow \infty$ surge o problema da inconsistência dos estimadores de β , que é denominado “problema dos parâmetros incidentais”.

Na estimação do modelo de regressão linear de efeitos fixos com erros normais é possível estimar os parâmetros β de forma independente da estimação

dos parâmetros incidentais α_i , utilizando os desvios $(Y_{it} - \bar{Y}_i)$ e $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$, conforme foi visto no capítulo 2. As médias (\bar{Y}_i) são estatísticas suficientes para α , e portanto as diferenças $(Y_{it} - \bar{Y}_i)$ têm distribuições livres dos parâmetros α , só dependendo do vetor β .

Entretanto, no caso de modelos de resposta binária, nem sempre existe uma estatística suficiente para α . Uma exceção ocorre no modelo de efeitos fixos em que os erros têm distribuição logística. Neste caso existe uma estatística suficiente para α , conforme é mostrado a seguir.

O modelo **logito** de efeitos fixos é dado por:

$$P(Y_{it} = 1 | \mathbf{x}_{it}) = \frac{e^{\alpha_i + \mathbf{x}'_{it}\beta}}{1 + e^{\alpha_i + \mathbf{x}'_{it}\beta}}.$$

Greene (2003) cita Chamberlain(1980) que mostrou que a função de verossimilhança condicional L^c , na soma $(\sum_{t=1}^T Y_{it})_i$, dada por:

$$L^c = \prod_{i=1}^N P \left(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT} \left| \sum_{t=1}^T Y_{it} \right. \right), \quad i = 1, \dots, N. \quad (3.28)$$

é livre dos parâmetros incidentais α_i . De fato, a verossimilhança conjunta para o grupo de T observações no indivíduo i condicionada em $(\sum_{t=1}^T Y_{it})_i$ (no número de vezes em que Y_{it} toma o valor 1) é:

$$P \left(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT} \left| \sum_{t=1}^T Y_{it} \right. \right) = \frac{\exp(\sum_{t=1}^T y_{it} \mathbf{x}'_{it} \beta)}{\sum_{\sum_t d_{it} = S_i} \exp(\sum_{t=1}^T d_{it} \mathbf{x}'_{it} \beta)}. \quad (3.29)$$

Na expressão (3.29) a função que aparece no denominador é somada no conjunto de todas as $\binom{T}{S_i}$ diferentes seqüências $(d_{it})_t$ de tamanho T constituída de zeros e uns tendo a mesma soma que a seqüência y_{it} que tem soma $S_i = \sum_{t=1}^T y_{it}$.¹

Para ilustrar a explicação anterior seja $T = 3$ e $(y_{it})_t = (1, 0, 1)$. Neste caso as diversas seqüências $(d_{it})_t$ são dadas por: $(1, 0, 1)$; $(1, 1, 0)$; $(0, 1, 1)$ e $S_i = 2$.

Para $T = 2$ e $(y_{it})_t = (1, 0)$, as seqüências $(d_{it})_t$ são dadas por: $(1, 0)$; $(0, 1)$ e $S_i = 2$. Para $T = 2$ e $(y_{it})_t = (0, 0)$, só existe uma seqüência $(d_{it})_t$ igual a $(0, 0)$.

Para demonstrar a equação (3.29) para o caso $T = 2$, consideremos as 4 realizações possíveis do par (Y_{i1}, Y_{i2}) .

¹Greene (2003) recomenda ver o cálculo em Arellano(2000, p.47) ou Baltagi (1995, p.180)

1. $Y_{i1} = 0$ e $Y_{i2} = 0 \Rightarrow \mathbb{P}(Y_{i1} = 0, Y_{i2} = 0 | S_i = 0) = 1,$

2. $Y_{i1} = 1$ e $Y_{i2} = 1 \Rightarrow \mathbb{P}(Y_{i1} = 1, Y_{i2} = 1 | S_i = 2) = 1,$

O i -ésimo termo em L^c para ambos os casos (1. e 2.) é sempre 1.

3. Se $Y_{i1} = 0$ e $Y_{i2} = 1$, então:

$$\begin{aligned} \mathbb{P}(Y_{i1} = 0, Y_{i2} = 1 | S_i = 1) &= \frac{\mathbb{P}(Y_{i1} = 0, Y_{i2} = 1 \text{ e } S_i = 1)}{\mathbb{P}(S_i = 1)}, \\ &= \frac{\mathbb{P}(Y_{i1} = 0, Y_{i2} = 1)}{\mathbb{P}(Y_{i1} = 0, Y_{i2} = 1) + \mathbb{P}(Y_{i1} = 1, Y_{i2} = 0)} \end{aligned}$$

Portanto para o caso 3. a probabilidade condicional é:

$$\begin{aligned} \mathbb{P}(Y_{i1} = 0, Y_{i2} = 1 | S_i = 1) &= \frac{\frac{1}{1+e^{\alpha_i + \mathbf{x}'_{i1}\boldsymbol{\beta}}} \frac{e^{\alpha_i + \mathbf{x}'_{i2}\boldsymbol{\beta}}}{1+e^{\alpha_i + \mathbf{x}'_{i2}\boldsymbol{\beta}}}}{\frac{1}{1+e^{\alpha_i + \mathbf{x}'_{i1}\boldsymbol{\beta}}} \frac{e^{\alpha_i + \mathbf{x}'_{i2}\boldsymbol{\beta}}}{1+e^{\alpha_i + \mathbf{x}'_{i2}\boldsymbol{\beta}}} + \frac{e^{\alpha_i + \mathbf{x}'_{i1}\boldsymbol{\beta}}}{1+e^{\alpha_i + \mathbf{x}'_{i1}\boldsymbol{\beta}}} \frac{1}{1+e^{\alpha_i + \mathbf{x}'_{i2}\boldsymbol{\beta}}}}, \\ \mathbb{P}(Y_{i1} = 0, Y_{i2} = 1 | S_i = 1) &= \frac{e^{\mathbf{x}'_{i2}\boldsymbol{\beta}}}{e^{\mathbf{x}'_{i1}\boldsymbol{\beta}} + e^{\mathbf{x}'_{i2}\boldsymbol{\beta}}}. \end{aligned}$$

Logo, a probabilidade do evento $(Y_{i1} = 0, Y_{i2} = 1)$ condicional no evento $(Y_{i1} + Y_{i2} = 1)$, não contém o termo α_i .

4. Para $Y_{i1} = 1$ e $Y_{i2} = 0$, a prova é feita de forma semelhante ao caso 3.

A função de verossimilhança (3.28), portanto, não contém os parâmetros α_i e é dada pelo produto dos fatores dados em (3.29), para $i = 1, \dots, N$. Os estimadores de $\boldsymbol{\beta}$ e de sua matriz de variância-covariância podem ser obtidos pelos métodos convencionais. O ajuste do modelo (3.28) está implementado na maioria dos pacotes computacionais econométricos.

Na subseção a seguir serão apresentados os métodos de estimação para o modelo de efeitos aleatórios para os modelos estáticos de regressão para variável resposta binária para dados em painel.

3.2.2 Modelo de Efeitos Aleatórios

Se a heterogeneidade individual não-observável existir, mas não for correlacionada com as variáveis exógenas, o modelo de Efeitos Aleatórios é preferível para fundamentar as análises. Utilizando o modelo (3.23), sendo \mathbf{x} o vetor de

todas as variáveis exógenas, \mathbf{x}_{it} , na amostra, $\forall t^2$ e considerando que α_i é não relacionado com \mathbf{x}_{it} , um modelo de efeitos aleatórios proposto por Butler e Moffitt³ especifica ainda que U_{it} e α_i são variáveis aleatórias independentes com:

$$1) \mathbb{E}(U_{it}|\mathbf{x}) = 0$$

$$2) \text{Cov}(U_{it}, U_{js}|\mathbf{x}) = \begin{cases} \text{Var}(U_{it}|\mathbf{x}) = 1 & \text{se } i = j \text{ e } t = s; \\ 0 & \text{caso contrário.} \end{cases}$$

$$3) \mathbb{E}(\alpha_i|\mathbf{x}) = 0 \quad \text{Cov}(U_{it}, \alpha_j|\mathbf{x}) = 0 \quad \forall i, t, j$$

$$4) \text{Cov}(\alpha_i, \alpha_j|\mathbf{x}) = \begin{cases} \text{Var}(\alpha_i|\mathbf{x}) = \sigma_\alpha^2 & \text{se } i = j; \\ 0 & \text{caso contrário.} \end{cases}$$

Em consequência, seguem as propriedades abaixo, onde $V_{it} = U_{it} + \alpha_i$:

$$\text{i) } \mathbb{E}[V_{it}|\mathbf{x}] = 0 \quad \text{Var}[V_{it}|\mathbf{x}] = \sigma_u^2 + \sigma_\alpha^2 = 1 + \sigma_\alpha^2,$$

$$\text{ii) } \rho = \text{Corr}[V_{it}, V_{is}|\mathbf{x}] = \frac{\sigma_\alpha^2}{1 + \sigma_\alpha^2},$$

$$\text{iii) } \sigma_\alpha^2 = \rho / (1 - \rho).$$

$$\mathbb{P}(Y_i = 1|\mathbf{x}_i) = \int_{-\infty}^{\mathbf{x}'_i \boldsymbol{\beta}} f_V(v_i) dv_i = F(\mathbf{x}'_i \boldsymbol{\beta}). \quad (3.30)$$

$$\mathbb{P}(Y_i = 0|\mathbf{x}_i) = \int_{\mathbf{x}'_i \boldsymbol{\beta}}^{\infty} f_V(v_i) dv_i = 1 - F(\mathbf{x}'_i \boldsymbol{\beta}) = F(-\mathbf{x}'_i \boldsymbol{\beta}). \quad (3.31)$$

No caso da distribuição Normal $\mathbb{P}(y_i|\mathbf{x}_i)$ pode ser escrito para $y_i = 0$ ou $y_i = 1$ como $\Phi[(2y_i - 1)\mathbf{x}'_i \boldsymbol{\beta}]$ e no modelo Logito, $\mathbb{P}(y_i|\mathbf{x}_i) = \Lambda[(2y_i - 1)\mathbf{x}'_i \boldsymbol{\beta}]$.

Estimação por Máxima Verossimilhança

Em um caso geral de dados em painel, com matriz de variância e covariância dos erros não-restrita, a contribuição do indivíduo i para a função de

²Ver Wooldridge (1999).

³Ver Greene (2003)

verossimilhança é a probabilidade conjunta de todas as observações T :

$$L_i = \mathbb{P}(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT} | \mathbf{x}) = \int_{L_{iT}}^{S_{iT}} \cdots \int_{L_{i1}}^{S_{i1}} f_V(v_{i1}, \dots, v_{iT}) dv_{i1} \dots dv_{iT} \quad (3.32)$$

sendo $(L_i, S_i) = (-\infty, \mathbf{x}'_i \boldsymbol{\beta})$ se $Y_i = 1$ e $(L_i, S_i) = (\mathbf{x}'_i \boldsymbol{\beta}, \infty)$ se $Y_i = 0$. A integração da densidade conjunta é muito complicada na maioria dos casos.

Uma simplificação do cálculo é possível, lembrando que:

$$f_V(v_{i1}, \dots, v_{iT}, \alpha_i) = f_V(v_{i1}, \dots, v_{iT} | \alpha_i) f(\alpha_i),$$

e que

$$f_V(v_{i1}, \dots, v_{iT}) = \int_{-\infty}^{+\infty} f_V(v_{i1}, \dots, v_{iT} | \alpha_i) f(\alpha_i) d\alpha_i.$$

Condicionando em α_i , os V_{it} 's são independentes, pois $V_{it} = U_{it} + \alpha_i$, sendo $(U_{it})_{t=1, \dots, T}$ uma seqüência de variáveis aleatórias independentes. Logo,

$$f_V(v_{i1}, \dots, v_{iT}) = \int_{-\infty}^{+\infty} \prod_{t=1}^T f_V(v_{it} | \alpha_i) f(\alpha_i) d\alpha_i.$$

Substituindo esse resultado em 3.32 obtém-se:

$$L_i = \int_{L_{iT}}^{S_{iT}} \cdots \int_{L_{i1}}^{S_{i1}} \int_{-\infty}^{+\infty} \prod_{t=1}^T f_V(v_{it} | \alpha_i) f(\alpha_i) d\alpha_i dv_{i1} \dots dv_{iT}.$$

Fazendo a mudança na ordem da integração:

$$L_i = \int_{-\infty}^{+\infty} \left[\int_{L_{iT}}^{S_{iT}} \cdots \int_{L_{i1}}^{S_{i1}} \prod_{t=1}^T f_V(v_{it} | \alpha_i) dv_{i1} \dots dv_{iT} \right] f(\alpha_i) d\alpha_i.$$

O termo dentro do colchete pode ser escrito como um produto das probabilidades, logo:

$$L_i = \int_{-\infty}^{+\infty} \left[\prod_{t=1}^T \left(\int_{L_{it}}^{S_{it}} f_V(v_{it} | \alpha_i) dv_{it} \right) \right] f(\alpha_i) d\alpha_i.$$

Como $\mathbb{P}(Y_{it} = y_{it} | \mathbf{x}'_{it} \boldsymbol{\beta} + \alpha_i) = \int_{L_{it}}^{S_{it}} f_V(v_{it} | \alpha_i) dv_{it}$, tem-se,

$$L_i = \int_{-\infty}^{+\infty} \left[\prod_{t=1}^T \mathbb{P}(Y_{it} = y_{it} | \mathbf{x}'_{it} \boldsymbol{\beta} + \alpha_i) \right] f(\alpha_i) d\alpha_i.$$

Greene (2003) cita Butler e Moffit (1982) que assumem que α_i é normalmente distribuído. Para o modelo Probit, $\mathbb{P}(Y_{it} = y_{it} | \mathbf{x}'_{it} \boldsymbol{\beta} + \alpha_i) = \Phi [q_{it}(\mathbf{x}'_{it} \boldsymbol{\beta} + \alpha_i)]$,

onde $\Phi(\cdot)$ é a função acumulada da distribuição normal e $q_{it} = 2y_{it} - 1$ e para o modelo Logito $\Lambda [q_{it}(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i)]$.

Se $g(\alpha_i) = \prod_{t=1}^T \mathbb{P}(Y_{it} = y_{it} | \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i)$ a função de verossimilhança (L_i) pode ser escrita como:

$$L_i = \int_{-\infty}^{+\infty} \frac{1}{\sigma_\alpha \sqrt{2\pi}} e^{-\frac{\alpha_i^2}{2\sigma_\alpha^2}} g(\alpha_i) d\alpha_i. \quad (3.33)$$

Seja $r_i = \alpha_i/\sigma_\alpha\sqrt{2}$, então $\alpha_i = r_i(\sigma_\alpha\sqrt{2}) = \theta r_i$ e $d\alpha_i = \theta dr_i$. Fazendo mudança de variável temos:

$$L_i = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-r_i^2} g(\theta r_i) dr_i.$$

Para o modelo Probit a função de verossimilhança do conjunto de dados em painel é:

$$L = \prod_{i=1}^N \left\{ \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-r_i^2} \left[\prod_{t=1}^T \Phi(q_{it}(\mathbf{x}'_{it}\boldsymbol{\beta} + \theta r_i)) \right] dr_i \right\}. \quad (3.34)$$

O cálculo dessa função é feita por Quadratura de Gauss-Hermite.⁴

Seguindo todos os passos, obtém-se a aproximação da log-verossimilhança (3.34):

$$\ln L_H = \sum_{i=1}^N \left\{ \ln \left[\frac{1}{\sqrt{\pi}} \sum_{h=1}^H \prod_{t=1}^T w_h \Phi(q_{it}(\mathbf{x}'_{it}\boldsymbol{\beta} + \theta Z_h)) \right] \right\},$$

onde H é o número de pontos para a quadratura, w_h e Z_h são os pesos e nós para a quadratura, respectivamente. A maximização desta função para o modelo Probit tem sido incorporada em muitos softwares econométricos contemporâneos e pode ser facilmente estendida para outros modelos. As 1ª e 2ª derivadas também são complexas e o seu cálculo é feito por quadratura. Uma estimativa de σ_u é obtida como sendo igual a $\theta/\sqrt{2}$ e o erro padrão por $\hat{\theta}/\sqrt{2}$.

Um problema com a abordagem de Butler e Moffit (1982) é que esta assume normalidade para a heterogeneidade α_i .⁵ Em geral, para outras distribuições existe uma grande dificuldade em encontrar uma forma fechada para a integral ou um método satisfatório para aproximar a mesma. Uma abordagem alternativa é o método da Máxima Verossimilhança Simulada que será apresentada no capítulo 4.

⁴Ver seção E.5.4 Greene (2003).

⁵Ver Greene (2003, pág.692).

Neste capítulo foram definidos os modelos estáticos de regressão de efeitos fixos e de efeitos aleatórios com variável resposta binária e o ajuste destes modelos por máxima verossimilhança foi discutido. No próximo capítulo serão considerados os modelos dinâmicos de regressão com variável resposta binária.

Capítulo 4

Modelos Dinâmicos de Regressão: Variável Resposta Binária e Dados em Painel

Um modelo de regressão para uma variável binária Y_{it} observada nos tempos $t = 1, \dots, T$ em um conjunto de dados em painel é um modelo dinâmico se o valor defasado da variável resposta ($Y_{i,t-1}$) é uma das variáveis regressoras.

A Tabela 1.1 apresentada no capítulo 1 ilustra um conjunto de dados em painel. Se for desejado incluir entre as variáveis explicativas o valor defasado da variável resposta, à Tabela 1.1 acrescenta-se uma coluna contendo os valores defasados da variável resposta para fins de uso em um programa de ajuste de modelo de regressão. A Tabela 4.1 mostra como seria a entrada de dados nesse caso.

A Tabela 4.1 mostra que 4 indivíduos foram observados em 4 anos, tendo a primeira observação sido coletada em 1980 e a última em 1983. A variável resposta de interesse Y_{it} vale 1 se no ato da pesquisa o indivíduo está filiado a algum sindicato e vale 0 caso contrário. A variável resposta defasada $Y_{i,t-1}$ é o valor da variável resposta no período $t - 1$. A variável id representa o indivíduo observado e a variável esc representa os anos de escolaridade do indivíduo entrevistado.

TABELA 4.1: *Painel de dados da variável binária filiação a sindicato, sua defasagem de ordem 1 e duas covariáveis.*

<i>id</i>	<i>ano</i>	Y_{it}	$Y_{i,t-1}$	<i>idade</i>	<i>esc</i>
1	1980	1	-	28	12
1	1981	0	1	29	13
1	1982	1	0	27	11
1	1983	1	1	31	14
2	1980	0	-	28	12
2	1981	1	0	29	13
2	1982	0	1	27	11
2	1983	1	0	31	14
3	1980	1	-	28	12
3	1981	0	1	29	13
3	1982	1	0	27	11
3	1983	0	1	31	14
4	1980	1	-	38	16
4	1981	0	1	32	13
4	1982	1	0	25	11
4	1983	0	1	30	14

Um modelo de efeitos fixos ou aleatórios de respostas binárias (Y_{it}) para Dados em Painel com efeitos defasados da variável observada é dado por ¹ :

$$Y_{it}^* = \mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i + U_{it}, \quad i = 1, \dots, N \quad t = 1, \dots, T; \quad (4.1)$$

$$Y_{it} = 1, \text{ se } Y_{it}^* > 0 \quad Y_{it} = 0, \text{ caso contrário,}$$

$$P(Y_{it} = 1 | \mathbf{x}_i, \alpha_i, Y_{i0}, \dots, Y_{i,t-1}) = F_U(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i), \quad (4.2)$$

$$P(Y_{i0} = 1 | \mathbf{x}_i, \alpha_i) = p_0(\mathbf{x}_i, \alpha_i),$$

onde:

- Y_{it} : Variável indicadora da participação do indivíduo i no tempo t ;
- $Y_{i,t-1}$: Variável indicadora da participação do indivíduo i no tempo $t - 1$;
- Y_{i0} : Variável indicadora da participação do indivíduo i no tempo inicial;
- Y_{it}^* : Variável do processo latente que guia a participação na decisão;
- \mathbf{x}_{it} : Vetor $k \times 1$ de determinantes exógenas da participação no tempo t ;
- α_i : Efeito individual específico não observado;
- U_{it} s: Erros transitórios que são assumidos *iid* com função de distribuição $F_U(\cdot)$ satisfazendo $F_U(-a) = F_U(a) \forall a \in \Re$;
- γ e $\boldsymbol{\beta}$: Parâmetros de interesse.

Ao construir um modelo dinâmico de resposta binária para dados em painel, duas questões importantes devem ser abordadas:

- Como tratar as observações iniciais Y_{i0} ;
- Como distinguir uma real dependência intertemporal das decisões de uma dependência meramente espúria.

¹A exposição a seguir é dada por Greene (2003)

A primeira questão importa para a derivação de estimadores consistentes.

A segunda é importante porque a dependência temporal observada, entre os eventos, pode ter vindo do fato de que um dado acontecimento observado ocorreu devido a uma dada experiência passada que alterou um comportamento individual ou devido a uma sucessão de perturbações não observáveis e correlacionadas ao longo do tempo ou, ainda, devido a uma combinação das duas. Para atender essa segunda questão, o valor defasado da variável resposta Y_{it} pode ser colocado como uma variável regressora.

Da expressão (4.1) deduz-se que o processo Y_{it} , para $t = 1, 2, \dots, T$ é uma cadeia de markov com dois estados (0 e 1) com probabilidades de transição não-estacionárias dadas por: ²

$$p_{i,t}(0, 1) = \mathbb{P}(Y_{it} = 1 | Y_{i,t-1} = 0) = \mathbb{P}(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i + U_{it} > 0),$$

Assumindo que U_{it} tenha uma função densidade simétrica em torno de zero, obtém-se:

$$\begin{aligned} p_{i,t}(0, 1) &= \mathbb{P}(U_{it} > -(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i)) = \\ &= \mathbb{P}(U_{it} < \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i) = F(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i). \end{aligned}$$

Da mesma forma:

$$p_{i,t}(1, 0) = \mathbb{P}(Y_{it} = 0 | Y_{i,t-1} = 1) = \mathbb{P}(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma + \alpha_i + U_{it} < 0),$$

$$p_{i,t}(1, 0) = \mathbb{P}(U_{it} < -(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma + \alpha_i)) = \mathbb{P}_{i,t}(U_{it} > \mathbf{x}'_{it}\boldsymbol{\beta} + \gamma + \alpha_i),$$

$$p_{i,t}(1, 0) = 1 - F(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma + \alpha_i).$$

De uma forma geral:

$$\begin{aligned} p(y_{it}, y_{i,t+1}) &= \left[F(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i) \right]^{1-y_{i,t+1}} \left[1 - F(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma + \alpha_i) \right]^{y_{i,t+1}}, \\ &= \left[F((\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i)(2y_{it} - 1)) \right]. \end{aligned} \quad (4.3)$$

²Ver CHAY, K.Y.; HYSLOP, D. (2000)

Para estimar os parâmetros de regressão pelo método da máxima verossimilhança faz-se uso da função densidade conjunta das T observações de Y_{it} para o i -ésimo indivíduo que é dada por:

$$f_Y(y_{i0}, \dots, y_{iT}) = \pi_0(y_{i0}) \cdot p(y_{i0}, y_{i1}) \times \dots \times p(y_{i,t-1}, y_{iT}). \quad (4.4)$$

Nessa expressão, $\pi_0(y_{i0}) = \mathbb{P}(Y_{i0} = y_{i0} | \mathbf{x}_i, \alpha_i)$, os valores de $p(y_{it}, y_{i,t+1})$ foram dados em (4.3) e dependem dos \mathbf{x}'_{it} e dos efeitos individuais α_i , que podem ser fixos ou aleatórios.

Quanto aos valores de $\pi_0(Y_{i0})$, nos trabalhos realizados nas diversas áreas do conhecimento, uma das três hipóteses para as condições iniciais é usualmente considerada:

1. As condições iniciais são verdadeiramente exógenas;

Assumir que as condições iniciais são exógenas significa que Y_0 não foi gerado pelo modelo, logo, ele não tem correlação com o erro transitório U_{it} .

2. O processo no tempo 0 é assumido estar em equilíbrio, com $\pi_i(e)$ a distribuição de equilíbrio de Y_{i0} .

3. Supõe-se que o estado inicial Y_{i0} siga um modelo não-dinâmico (conhecido por “forma reduzida”),³:

$$Y_{i0}^* = \mathbf{x}'_{i0} \Pi_1 + \alpha_i \Pi_2 + U_{i0}, \quad i = 1, \dots, N, \quad \mathbb{P}(U_{i0} < a) = F_0(a), \quad (4.5)$$

$$Y_{i0} = 1(Y_{i0}^* > 0), \quad \mathbb{P}(Y_{i0} = 1 | \mathbf{x}'_{i0}, \alpha_i) = F_0(\mathbf{x}'_{i0} \Pi_1 + \alpha_i \Pi_2),$$

$$\mathbb{P}(Y_{i0} = 0 | \mathbf{x}'_{i0}, \alpha_i) = 1 - F_0(\mathbf{x}'_{i0} \Pi_1 + \alpha_i \Pi_2),$$

$$\text{Logo } \mathbb{P}(Y_{i0} = y_{i0} | \mathbf{x}'_{i0}, \alpha_i) = F_0\{(2y_{i0} - 1)(\mathbf{x}'_{i0} \Pi_1 + \alpha_i \Pi_2)\}. \quad (4.6)$$

Se $F_0(\cdot)$ for a função acumulada da distribuição normal (Φ), tem-se um modelo probito para as condições iniciais, se $F_0(\cdot)$ for a função acumulada da distribuição logística (Λ), tem-se o modelo logito para as condições iniciais.

4.1 Função de Verossimilhança

A função de verossimilhança para uma observação $\mathbf{y}_i = (y_{i0}, \dots, y_{iT})'$ de um modelo dinâmico de resposta binária, é escrita de forma diferente segundo

³Stewart (2006) sugere ver Heckman 1981

as condições iniciais assumidas e as hipóteses sobre os efeitos individuais.⁴ Assumindo que $F_U(\cdot)$ tenha função densidade simétrica, os seguintes casos podem ser considerados:

4.1.1 Modelo de Efeitos Fixos

Caso 1 - Substituindo na expressão (4.4) os valores das probabilidades de transição (4.3) para o modelo dinâmico de escolha binária e supondo **condições iniciais fixas** obtém-se a função de verossimilhança dos parâmetros relativa ao indivíduo i :

$$L_i(\boldsymbol{\beta}, \gamma, \boldsymbol{\alpha}; y) = \prod_{t=1}^T F_U\{(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i) \cdot (2y_{it} - 1)\}. \quad (4.7)$$

Caso 2 - Supondo que no início da observação o processo esteja em **estado de equilíbrio**, a função de verossimilhança dos parâmetros relativa ao indivíduo i dado α_i será:

$$L_i = \prod_{t=1}^T F_U\{(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i) \cdot (2y_{it} - 1)\} [\pi_i^e(1)]^{y_{i0}} \cdot [1 - \pi_i^e(1)]^{1-y_{i0}}. \quad (4.8)$$

Caso 3 - Supondo que o processo ainda não esteja em equilíbrio no tempo $t = 0$, então o estado inicial para o indivíduo i , Y_{i0} pode ser tratado como aleatório. Por Heckman (4.6), a função de verossimilhança da amostra relativa ao indivíduo i é:

$$L_i = \prod_{t=1}^T F_U\{(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i) \cdot (2y_{it} - 1)\} \cdot \mathbb{P}(Y_{i0} = y_{i0} | \mathbf{x}_{i0}, \alpha_i), \quad (4.9)$$

com $\mathbb{P}(Y_{i0} = y_{i0} | \mathbf{x}_{i0}, \alpha_i) = F_0\{(2y_{i0} - 1)(\mathbf{x}'_{i0}\boldsymbol{\Pi}_1 + \alpha_i \boldsymbol{\Pi}_2)\}$.

4.1.2 Modelo de Efeitos Aleatórios

Caso 1 - Se α_i é aleatório, com distribuição $G(\alpha)$, a função de verossimilhança dos parâmetros $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ sob o modelo de efeitos aleatórios com **condições iniciais fixas** relativa a todos os “N” indivíduos do painel é:

$$L(\boldsymbol{\beta}, \gamma, \boldsymbol{\alpha}; y) = \prod_{i=1}^N \int \prod_{t=1}^T \Phi\{(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i) \cdot (2y_{it} - 1)\} dG(\alpha). \quad (4.10)$$

⁴Ver CHAY, K.Y.; HYSLOP, D. (2000)

Caso 2 - Para a 2ª hipótese sobre as condições iniciais, a função de verossimilhança relativa ao indivíduo i é:

$$L_i = \prod_{t=1}^T F_0\{(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i) \cdot (2y_i - 1)\} [\pi_i^e(1)]^{y_{i0}} \cdot [1 - \pi_i^e(1)]^{1-y_{i0}} dG(\alpha) \quad (4.11)$$

Caso 3 - Com as condições iniciais modeladas por Heckman (4.6), a função de verossimilhança da amostra relativa ao indivíduo i é:

$$L_i = \prod_{t=1}^T F_U\{(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i) \cdot (2y_{it} - 1)\} \cdot IP(Y_{i0} = y_{i0} | \mathbf{x}_{i0}, \alpha_i) dG(\alpha), \quad (4.12)$$

onde $IP(Y_{i0} = y_{i0} | \mathbf{x}_{i0}, \alpha_i) = F\{(2y_i - 1)(\mathbf{x}'_{i0}\Pi_1 + \alpha_i\Pi_2)\}$.

4.2 Estimação de alguns Modelos Dinâmicos de Resposta Binária

4.2.1 Modelo de Efeitos Fixos segundo Honoré e Kyriazidou

No modelo (4.1), Honoré e Kyriazidou (2000) consideram o caso em que os efeitos individuais (α_i) são tratados como efeitos fixos.

Na ausência de defasagem no modelo (4.1) ou seja, ($\gamma = 0$), para ($T \geq 2$) com os erros transitórios (U_{it}) *iid* com distribuição logística, usando uma função de verossimilhança condicional na estatística suficiente, $(\sum_{t=1}^T Y_{it})_i$, remove-se os parâmetros incidentais do modelo e obtem-se uma estimativa consistente para $\boldsymbol{\beta}$. (Ver página 34 do presente trabalho).

O modelo Logito dinâmico é dado pela equação (4.1) com as seguintes especificações:

$$IP(Y_{i0} = 1 | \mathbf{x}_i, \alpha_i) = p_0(\mathbf{x}_i, \alpha_i), \quad (4.13)$$

$$IP(Y_{it} = 1 | \mathbf{x}_i, \alpha_i, y_{i0}, \dots, y_{i,t-1}) = \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i)}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i)}, \quad t = 1, 2, 3, \dots, T$$

onde \mathbf{x}_i é a matriz $T \times k$ de variáveis explicativas, $\boldsymbol{\beta}$ é um vetor $k \times 1$ de coeficientes e α_i a heterogeneidade não-observada.

A função de verossimilhança completa de um conjunto de dados em painel sob este modelo é:

$$L = \prod_{i=1}^N \prod_{t=1}^T p_0(\mathbf{x}_i, \alpha_i) \times \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i)}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i)}, \quad t = 1, 2, 3, \dots, T. \quad (4.14)$$

A estimação dos parâmetros $\boldsymbol{\beta}$ através da maximização da função (4.14) é possível mas tem associado o problema de parâmetros incidentais (α_i) , $1 \leq i \leq N$, de forma que os estimadores assim calculados não são consistentes.

Inspirado na idéia de Chamberlain (1980) para o modelo estático logito, Honoré e Kyriazidou (2000) propõem uma verossimilhança condicional para o caso $T = 4$ que é livre dos parâmetros α_i . Para isso consideraram os seguintes eventos dois a dois disjuntos:

$$A_i \equiv \{Y_{i0} = d_{i0}, Y_{i1} = 0, Y_{i2} = 1, Y_{i3} = d_{i3}\},$$

$$B_i \equiv \{Y_{i0} = d_{i0}, Y_{i1} = 1, Y_{i2} = 0, Y_{i3} = d_{i3}\},$$

$$C_i \equiv \{Y_{i0} = d_{i0}, Y_{i1} = 0, Y_{i2} = 0, Y_{i3} = d_{i3}\},$$

$$D_i \equiv \{Y_{i0} = d_{i0}, Y_{i1} = 1, Y_{i2} = 1, Y_{i3} = d_{i3}\}.$$

O espaço amostral S_i relativo ao processo (Y_{it}) , $t = 0, 1, \dots, T$, para o i -ésimo indivíduo pode ser visto como a reunião da partição $(A_i, B_i, C_i$ e $D_i)$ definida acima. Por outro lado, quando se considera o processo relativo a todos os indivíduos, o espaço amostral S é o produto cartesiano dos espaços amostrais S_i relativos a cada indivíduo. Pela suposição subjacente à definição de dados em painel, os N processos (Y_{it}) individuais são independentes.

Kyriazidou e Honoré (2000) sugerem o uso de uma verossimilhança condicional no evento $\cap(A_i \cup B_i, \mathbf{x}_i, \alpha_i, \mathbf{x}_{i2} = \mathbf{x}_{i3})$. A probabilidade condicional de C_i dado $A_i \cup B_i$ é zero, o mesmo ocorre para a probabilidade condicional de D_i dado $A_i \cup B_i$. Os eventos A_i, B_i, C_i e D_i , para um mesmo i , são dependentes.

A_i e B_j são independentes, $\forall i \neq j$;

A_i e A_j são independentes, $\forall i \neq j$;

B_i e B_j são independentes, $\forall i \neq j$;

Probabilidades condicionais:

$$\begin{aligned} \mathbb{P}(A_i|\mathbf{x}_i, \alpha_i) &= p_0(\mathbf{x}_i, \alpha_i)^{d_0} \cdot (1 - p_0(\mathbf{x}_i, \alpha_i))^{1-d_0} \\ &\quad \times \frac{1}{1 + \exp(\mathbf{x}'_{i1}\boldsymbol{\beta} + \gamma d_0 + \alpha_i)} \\ &\times \frac{\exp(\mathbf{x}'_{i2}\boldsymbol{\beta} + \alpha_i)}{1 + \exp(\mathbf{x}'_{i2}\boldsymbol{\beta} + \alpha_i)} \times \frac{\exp(d_3\mathbf{x}'_{i3}\boldsymbol{\beta} + d_3\gamma + d_3\alpha_i)}{1 + \exp(\mathbf{x}'_{i3}\boldsymbol{\beta} + \gamma + \alpha_i)}. \end{aligned} \quad (4.15)$$

$$\begin{aligned} \mathbb{P}(B_i|\mathbf{x}_i, \alpha_i) &= p_0(\mathbf{x}_i, \alpha_i)^{d_0} \cdot (1 - p_0(\mathbf{x}_i, \alpha_i))^{1-d_0} \\ &\times \frac{\exp(\mathbf{x}'_{i1}\boldsymbol{\beta} + \gamma d_0 + \alpha_i)}{1 + \exp(\mathbf{x}'_{i1}\boldsymbol{\beta} + \gamma d_0 + \alpha_i)} \times \frac{1}{1 + \exp(\mathbf{x}'_{i2}\boldsymbol{\beta} + \gamma + \alpha_i)} \\ &\quad \times \frac{\exp(d_3\mathbf{x}'_{i3}\boldsymbol{\beta} + d_3\alpha_i)}{1 + \exp(\mathbf{x}'_{i3}\boldsymbol{\beta} + \alpha_i)}. \end{aligned} \quad (4.16)$$

Honoré e Kyriazidou (2000) mostraram que os efeitos individuais são eliminados, quando se condiciona em $\mathbf{x}_{i2} = \mathbf{x}_{i3}$ e em $A_i \cup B_i$:

$$\mathbb{P}(A_i|\mathbf{x}_i, \alpha_i, A_i \cup B_i, \mathbf{x}_{i2} = \mathbf{x}_{i3}) = \frac{\mathbb{P}(A_i|\cdot)}{\mathbb{P}(A_i|\cdot) + \mathbb{P}(B_i|\cdot)} = \frac{1}{1 + \frac{\mathbb{P}(B_i|\cdot)}{\mathbb{P}(A_i|\cdot)}},$$

Mas

$$\frac{\mathbb{P}(B_i|\cdot)}{\mathbb{P}(A_i|\cdot)} = \frac{e^{(\mathbf{x}'_{i1}\boldsymbol{\beta} + \gamma d_0 + \alpha_i)} \times \frac{1}{1 + e^{(\mathbf{x}'_{i2}\boldsymbol{\beta} + \gamma + \alpha_i)}} \times \frac{e^{(d_3\mathbf{x}'_{i3}\boldsymbol{\beta} + d_3\alpha_i)}}{1 + e^{(\mathbf{x}'_{i3}\boldsymbol{\beta} + \alpha_i)}}}{\frac{e^{(\mathbf{x}'_{i2}\boldsymbol{\beta} + \alpha_i)}}{1 + e^{(\mathbf{x}'_{i2}\boldsymbol{\beta} + \alpha_i)}} \times \frac{e^{(d_3\mathbf{x}'_{i3}\boldsymbol{\beta} + d_3\gamma + d_3\alpha_i)}}{1 + e^{(\mathbf{x}'_{i3}\boldsymbol{\beta} + \gamma + \alpha_i)}}}.$$

Desta forma, como o indivíduo i só contribui para a verossimilhança condicional se sua realização for $\mathbf{x}_{i3} = \mathbf{x}_{i2}$ e cancelando os termos em comum obtém-se:

$$\frac{\mathbb{P}(B_i|\cdot)}{\mathbb{P}(A_i|\cdot)} = \exp((\mathbf{x}'_{i1} - \mathbf{x}'_{i2})\boldsymbol{\beta} + (d_0 - d_3)\gamma),$$

$$\mathbb{P}(A_i|\mathbf{x}_i, \alpha_i, A_i \cup B_i, \mathbf{x}_{i2} = \mathbf{x}_{i3}) = \frac{1}{1 + \exp\{(\mathbf{x}_{i1} - \mathbf{x}_{i2})\boldsymbol{\beta} + \gamma(d_{i0} - d_{i3})\}}. \quad (4.17)$$

Analogamente,

$$\mathbb{P}(B_i|\mathbf{x}_i, \alpha_i, A_i \cup B_i, \mathbf{x}_{i2} = \mathbf{x}_{i3}) = \frac{\exp\{(\mathbf{x}_{i1} - \mathbf{x}_{i2})\boldsymbol{\beta} + \gamma(d_{i0} - d_{i3})\}}{1 + \exp\{(\mathbf{x}_{i1} - \mathbf{x}_{i2})\boldsymbol{\beta} + \gamma(d_{i0} - d_{i3})\}}. \quad (4.18)$$

Observa-se que (4.17) e (4.18) não dependem dos efeitos individuais (α_i).

Para a estimação dos parâmetros α e β , maximiza-se a função de log-verossimilhança condicional (4.19):

$$l(\beta, \gamma; y | \cap_{j=1}^N A_j \cup B_j) = \sum_{i=1}^N 1\{Y_{i1} + Y_{i2} = 1\} \cdot 1\{\mathbf{x}_{i2} - \mathbf{x}_{i3} = 0\} \\ \times \ln \left\{ \frac{\exp((\mathbf{x}_{i1} - \mathbf{x}_{i2})\beta + g(y_{i0} - y_{i3}))^{y_{i1}}}{1 + \exp((\mathbf{x}_{i1} - \mathbf{x}_{i2})\beta + g(y_{i0} - y_{i3}))} \right\}. \quad (4.19)$$

Na expressão acima $1(A)$ é função indicadora do conjunto A e $g(y_{i0} - y_{i3}) = \gamma(d_{i0} - d_{i3})$. O estimador resultante terá todas as propriedades usuais (será consistente e \sqrt{N} -assintoticamente normal).

A estimação a partir da log-verossimilhança acima é um pouco restritiva na prática. Por exemplo, no campo da economia, dificilmente teremos uma situação onde $\mathbf{x}_{i1} = \mathbf{x}_{i2}$ (salvo o caso particular em que a distribuição de \mathbf{x}_i seja controlada). Por conta disso Honoré e Kyriazidou sugeriram substituir $1\{\mathbf{x}_{i2} - \mathbf{x}_{i3} = 0\}$ na log-verossimilhança acima por uma função que dê pesos às diferenças $\mathbf{x}_{i1} - \mathbf{x}_{i2}$, procurando tornar mais relevantes aquelas que estiverem mais próximas de zero.

A extensão da proposta de Honoré e Kyriazidou é fazer a estimação utilizando a “log-verossimilhança” condicional abaixo:

$$l(\beta, \gamma; y | \cap_{j=1}^N A_j \cup B_j) = \sum_{i=1}^N 1\{Y_{i1} + Y_{i2} = 1\} \cdot k \left\{ \frac{\mathbf{x}_{i2} - \mathbf{x}_{i3}}{\sigma_N} \right\} \\ \times \ln \left\{ \frac{\exp((\mathbf{x}_{i1} - \mathbf{x}_{i2})\beta + g(y_{i0} - y_{i3}))^{y_{i1}}}{1 + \exp((\mathbf{x}_{i1} - \mathbf{x}_{i2})\beta + g(y_{i0} - y_{i3}))} \right\}. \quad (4.20)$$

Na expressão (4.20) $k(\cdot)$ é o núcleo responsável por dar o peso apropriado à observação i , enquanto que σ_N é um parâmetro de suavização que diminui a medida que N cresce. Este parâmetro pode ser por exemplo $\sigma_N = \sigma \times N^{-1/5}$, com $\sigma = 0,5, 1$ ou 3 . Para o uso desta teoria $k(\cdot)$ deve obedecer a algumas condições de regularidade padrão, por exemplo $k(\cdot) \rightarrow 0$.

4.2.2 Modelo Probito Dinâmico de Heckman com Erros Autocorrelacionados

Nesta seção será apresentada a estimação de um modelo probito dinâmico de efeitos aleatórios dado por Heckman (1981) com erros autocorrelacionados pelo

método da Máxima Verossimilhança Simulada (MVS), segundo Stewart (2006). O uso desse método de estimação é mais conveniente quando os termos do erro da equação latente são serialmente correlacionados no tempo.

Será considerado o seguinte modelo de regressão para a variável latente:

$$Y_{it}^* = \mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i + U_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T; \quad (4.21)$$

O modelo ainda especifica que:

- i) os erros U_{it} ($t = 1, \dots, T$ e i fixo) têm uma distribuição normal multivariada com estrutura de autocorrelação arbitrária
- ii) α_i , o efeito aleatório específico do indivíduo é independente de \mathbf{x}_{it} e U_{it} , $\forall t$.
- iii)

$$Y_{it} = \begin{cases} 1 & \text{se } Y_{it}^* \geq 0 \quad t = 0, 1, \dots, T. \\ 0 & \text{caso contrário.} \end{cases} \quad (4.22)$$

- iv) as condições iniciais são modeladas segundo Heckman por: ⁵

$$Y_{i0}^* = \mathbf{x}'_{i0}\Pi_1 + \alpha_i\Pi_2 + U_{i0}, \quad i = 1, \dots, N$$

O termo genérico de variância de Y_{i0}^* será denotado por V_{i0} . Desta forma $V_{i0} = \alpha_i\Pi_2 + U_{i0}$ e $V_{it} = \alpha_i + U_{it}$ para $t \geq 1$, onde \mathbf{x}_{it} é um vetor de variáveis explicativas, em particular $U_{it} \sim N(0, \sigma_u^2)$. Dois modelos com autocorrelação de 1ª ordem são considerados aqui: AR(1) e MA(1). No primeiro deles U_{it} é especificado como um processo autoregressivo de primeira ordem:

$$U_{it} = \rho U_{it-1} + \varepsilon_{it} \quad \text{AR}(1), \quad (4.23)$$

com $-1 < \rho < 1$ e $t \geq 1$. Para este modelo a matriz de variância e covariância (Ω) dos erros do modelo V_{it} , com $\sigma_u^2 = 1$ é dada por:

⁵Stewart (2006) sugere ver Heckman (1981).

$$\Omega = \begin{bmatrix} \Pi_2^2 \sigma_\alpha^2 + 1 & & & & \\ \Pi_2 \sigma_\alpha^2 + \rho & \sigma_\alpha^2 + 1 & & & \\ \Pi_2 \sigma_\alpha^2 + \rho^2 & \sigma_\alpha^2 + \rho & \sigma_\alpha^2 + 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \Pi_2 \sigma_\alpha^2 + \rho^{T-1} & \sigma_\alpha^2 + \rho^{T-2} & \sigma_\alpha^2 + \rho^{T-3} & \dots & \sigma_\alpha^2 + 1 \end{bmatrix} = \text{cov}(V_{it}).$$

No segundo modelo U_{it} é especificado como um processo de médias móveis de 1ª ordem:

$$U_{it} = \varepsilon_{it} - \mu \varepsilon_{it-1} \quad \text{MA}(1), \quad (4.24)$$

Neste caso a matriz de variância e covariância de V_{it} , com $\sigma_u^2 = 1$ é dada por:

$$\Omega = \begin{bmatrix} \Pi_2^2 \sigma_\alpha^2 + 1 & & & & \\ \Pi_2 \sigma_\alpha^2 - \omega & \sigma_\alpha^2 + 1 & & & \\ \Pi_2 \sigma_\alpha^2 & \sigma_\alpha^2 - \omega & \sigma_\alpha^2 + 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \Pi_2 \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 + 1 \end{bmatrix} = \text{cov}(V_{it}),$$

onde $\omega = \mu/(1 - \mu^2)$.

Adotando a notação: $(L_{i0}, S_{i0}) = (-\mathbf{x}'_{i0}\boldsymbol{\beta}, \infty)$ se $Y_{i0} = 1$ ou $(-\infty, -\mathbf{x}'_{i0}\boldsymbol{\beta})$ se $Y_{i0} = 0$ e $(L_{it}, S_{it}) = (-\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1}, \infty)$ se $Y_{it} = 1$ ou $(-\infty, -\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1})$ se $Y_{it} = 0$ para $t \geq 1$, a função densidade conjunta das observações (Y_{it}) , $t = 0, 1, \dots, T$, para um indivíduo i , pode ser escrita da seguinte maneira:

$$f_Y(y_{i0}, y_{i1}, \dots, y_{it}) = \mathbb{P} \left[0 < (2y_{i0} - 1)(\mathbf{x}'_{i0}\boldsymbol{\beta} + \Pi_2 \alpha_i + U_{i0}) < \infty; \right. \\ \left. 0 < (2y_{i1} - 1)(\mathbf{x}'_{i1}\boldsymbol{\beta} + \gamma y_{i0} + \alpha_i + U_{i1}) < \infty; \right. \\ \left. \dots ; 0 < (2y_{it} - 1)(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i + U_{it}) < \infty \right]. \quad (4.25)$$

Sejam $V_{i0} = \Pi_2 \alpha_i + U_{i0}$ e $V_{it} = \alpha_i + U_{it}$, $t \geq 1$. A função densidade (4.25) acima pode ainda ser escrita como:

$$f = \mathbb{P} \left[-(2y_{i0} - 1)\mathbf{x}'_{i0}\boldsymbol{\beta} < (2y_{i0} - 1)V_{i0} < \infty; \right. \\ \left. -(2y_{i1} - 1)(\mathbf{x}'_{i1}\boldsymbol{\beta} + \gamma y_{i0}) < (2y_{i1} - 1)V_{i1} < \infty; \right. \\ \left. \dots ; -(2y_{it} - 1)(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1}) < (2y_{it} - 1)V_{it} < \infty \right]. \quad (4.26)$$

A função de verossimilhança (4.26), pode ser vista como:

$$f = \mathbb{P} [L_{i0} < V_{i0} < S_{i0}, \dots, L_{it} < V_{it} < S_{it}]. \quad (4.27)$$

Essa função densidade conjunta é escrita através de uma integral múltipla.

A seguir será apresentado o método de estimação por máxima verossimilhança simulada (MVS) que utiliza o Simulador GHK desenvolvido independentemente por Geweke, Hajivassiliou e Keane (Keane, 1994).

Método da Estimação MVS - Estimador GHK

Considere o modelo dado por 4.21 com os erros U_{it} autocorrelacionados. Os estimadores de máxima verossimilhança de tal modelo podem ser encontrados via simulação. Um estimador de máxima verossimilhança simulada (MVS) baseado no algoritmo de GHK Geweke, Hajivassiliou e Keane ⁶ será apresentado nesta sub-seção.

Stewart (2006) cita em seu artigo que Geweke (1989), Borsh - Saupan e Hajivassilou (1993) propuseram um simulador chamado de “condicionante recursivo suave” (SRC). Em artigos mais recentes o simulador SRC foi rotulado como Simulador Geweke-Hajivassilou-Keane (GHK).

O simulador GHK calcula probabilidades do tipo:

$$\mathbb{P} [\mathbf{h} < \mathbf{v} < \mathbf{b}] = \int_{\mathbf{h}}^{\mathbf{b}} g(\mathbf{v}) d\mathbf{v}. \quad (4.28)$$

onde \mathbf{v} é um vetor aleatório k -dimensional com distribuição $N(\mu, \Omega)$. Para o cálculo de (4.28) será feita a seguinte mudança: $\mathbf{v} = \mathbf{C}\boldsymbol{\eta}_i$, onde $\boldsymbol{\eta} \sim N(0, I)$ e \mathbf{C} é a decomposição triangular inferior de Cholesky de Ω (isto é, tal que $\mathbf{C}\mathbf{C}' = \Omega$).

$$\mathbb{P} [\mathbf{h} < \mathbf{v} < \mathbf{b}] = \mathbb{P} [\mathbf{h} < \mathbf{C}\boldsymbol{\eta} < \mathbf{b}]. \quad (4.29)$$

Observe que $\mathbf{v} = \mathbf{C}\boldsymbol{\eta}$ é dado por:

$$\mathbf{v} = \begin{bmatrix} c_{11} \\ c_{21} & c_{22} \\ \vdots & \vdots & \ddots \\ c_{T1} & c_{T2} & \dots & c_{TN} \end{bmatrix} \times \begin{bmatrix} \eta_{i0} \\ \eta_{i1} \\ \vdots \\ \eta_{iT} \end{bmatrix} = \begin{bmatrix} c_{11}\eta_{i0} \\ c_{21}\eta_{i0} & c_{22}\eta_{i1} \\ \vdots & \vdots & \ddots \\ c_{T1}\eta_{i0} & c_{T2}\eta_{i1} & \dots & c_{TN}\eta_{iT} \end{bmatrix} \quad (4.30)$$

⁶Keane, 1994

Pode-se representar o evento ($\mathbf{h} < \mathbf{C}\boldsymbol{\eta} < \mathbf{b}$) da seguinte forma:

$$\begin{aligned} h_1 &< c_{11}\eta_{i0} < b_1 \\ h_2 &< c_{21}\eta_{i0} + c_{22}\eta_{i1} < b_2 \\ &\vdots \\ h_N &< c_{T1}\eta_{i0} + \cdots + c_{TN}\eta_{iT} < b_N \end{aligned}$$

onde c_{ij} , h_i e b_i são elementos de \mathbf{C} , \mathbf{h} e \mathbf{b} , respectivamente e $\boldsymbol{\eta}_{ij} \sim N(0, I)$. Isolando os η'_{ij} s tem-se que ($\mathbf{h} < \mathbf{C}\boldsymbol{\eta} < \mathbf{b}$) pode ser representado como a intersecção dos seguintes eventos (A_j):

$$\frac{h_1}{c_{11}} < \eta_{i0} < \frac{b_1}{c_{11}} \quad (A_0), \quad (4.31)$$

$$\frac{h_2 - c_{21}\eta_{i0}}{c_{22}} < \eta_{i1} < \frac{b_2 - c_{21}\eta_{i0}}{c_{22}} \quad (A_1),$$

⋮

$$\frac{h_N - c_{T1}\eta_{i0} - c_{T2}\eta_{i1} - \cdots - c_{T,t-1,N}\eta_{i,t-1}}{c_{TN}} < \eta_{iT} < \frac{b_N - c_{T1}\eta_{i0} - c_{T2}\eta_{i1} - \cdots - c_{T,t-1,N}\eta_{i,t-1}}{c_{TN}} \quad (A_T).$$

Pelas expressões acima, a probabilidade de interesse pode ser dada por:

$$\mathbb{P}[h < \mathbf{C}\boldsymbol{\eta} < b] = \mathbb{P}(A_0)\mathbb{P}(A_1|A_0) \times \cdots \times \mathbb{P}(A_t|A_0 \cdots, A_{t-1}). \quad (4.32)$$

Geweke, Hajivassilou e Keane, propuseram a estimação por simulação de cada um dos fatores presentes em (4.32). O fundamento da estimação por simulação de probabilidade de um evento A está no Teorema (4.1) abaixo e seu corolário. A seguir, mostra-se que cada fator da expressão (4.32) é uma probabilidade condicional dado um evento, (A_i , por exemplo.) que por sua vez pode ser estimada por simulação em 2 passos: simula-se η_{i0}^* , que tem distribuição normal truncada, e simula-se η_{i1} , observando se ocorre ou não o evento A_1 . O número médio de ocorrências de A_1 em N simulações será um estimador de $\mathbb{P}(A_1|A_0)$.

Teorema 4.1 *Seja $(\Omega, \mathcal{A}, \mathbb{P})$ um espaço de probabilidade e A um evento em \mathcal{A} . A variável $I_A : \Omega \rightarrow \mathbb{R}$ definida por $I_A(\omega) = 1$ se $\omega \in A$ e $I_A(\omega) = 0$ caso contrário, é tal que $\mathbb{P}(A) = \mathbb{E}(I_A(\omega))$ e $\text{Var}(I_A(\omega)) = \mathbb{P}(A)(1 - \mathbb{P}(A))$. No espaço*

produto (Ω^N) com a medida produto, seja $\omega \in \Omega^N$. Então a variável $S_A(\omega) = \sum_{j=1}^N I_A(\omega_j)$ tem esperança $\mathbb{E}(S_A(\omega)) = N\mathbb{P}(A)$ e variância $\text{Var}(S_A(\omega)) = N\mathbb{P}(A)(1-\mathbb{P}(A))$. Desta forma a variável $\frac{S_A(\omega)}{N}$ tem esperança $\mathbb{P}(A)$ e variância $\frac{\mathbb{P}(A)(1-\mathbb{P}(A))}{N}$, e quando $N \rightarrow \infty$, $\frac{S_A(\omega)}{N}$ converge quase-certamente para $\mathbb{P}(A)$.

Corolário 4.1 Princípio da Simulação de $\mathbf{P}(A)$: Em vista do resultado acima um estimador de $\mathbb{P}(A)$ pode ser obtido realizando N vezes o experimento cujo espaço amostral é Ω , e calculando $\frac{\sum_{j=1}^N I_A(\omega_j)}{N}$ que é a frequência relativa de ocorrências do evento A , entre as N observações.

Teorema 4.2 : $\mathbb{P}(A_1|A_0) = \mathbb{E}(\mathbb{P}(A_1|\eta_{i0}^*))$, onde η_{i0}^* tem distribuição normal truncada em (a_0, b_0) .

Prova:

Pelos eventos (4.31) temos:

$$A_0 : a_0 < \eta_{i0} < b_0 \quad \text{com} \quad (a_0, b_0) = \left(\frac{h_1}{c_{11}}, \frac{b_1}{c_{11}} \right);$$

$$A_1 : a_1 < \eta_{i1} < b_1 \quad \text{com} \quad (a_1, b_1) = \left(\frac{h_2 - c_{21}\eta_{i0}}{c_{22}}, \frac{b_2 - c_{21}\eta_{i0}}{c_{22}} \right);$$

Temos que,

$$\mathbb{P}(A_1|A_0) = \frac{\mathbb{P}(A_1, A_0)}{\mathbb{P}(A_0)},$$

com $\mathbb{P}(A_0) = \int_{a_0}^{b_0} f(\eta_{i0})d\eta_{i0}$ e $\mathbb{P}(A_0, A_1) = \int_{a_1}^{b_1} \int_{a_0}^{b_0} f(\eta_{i0}, \eta_{i1})d\eta_{i0}d\eta_{i1}$,

$$\mathbb{P}(A_1|A_0) = \frac{\int_{a_1}^{b_1} \int_{a_0}^{b_0} f(\eta_{i0}, \eta_{i1})d\eta_{i0}d\eta_{i1}}{\int_{a_0}^{b_0} f(\eta_{i0})d\eta_{i0}},$$

$$\mathbb{P}(A_1|A_0) = \int_{a_1}^{b_1} \int_{a_0}^{b_0} \frac{f(\eta_{i0}, \eta_{i1})d\eta_{i0}d\eta_{i1}}{f(\eta_{i0})} \times \frac{f(\eta_{i0})}{\int_{a_0}^{b_0} f(\eta_{i0})d\eta_{i0}},$$

$$\mathbb{P}(A_1|A_0) = \int_{a_0}^{b_0} \left[\int_{a_1}^{b_1} f(\eta_{i1}|\eta_{i0})d\eta_{i1} \right] \times \left[\frac{f(\eta_{i0})}{\int_{a_0}^{b_0} f(\eta_{i0})d\eta_{i0}} \right] d\eta_{i0},$$

Logo,

$$\mathbb{P}(A_1|A_0) = \mathbb{E} \left[\int_{a_1}^{b_1} f(\eta_{i1}|\eta_{i0}) d\eta_1 \right] \quad \text{em relação a} \quad \frac{f(\eta_{i0})}{\int_{a_0}^{b_0} f(\eta_{i0}) d\eta_{i0}},$$

como

$$\int_{a_1}^{b_1} f(\eta_{i1}|\eta_{i0}) d\eta_1 = \mathbb{P} \left\{ \left(\frac{h_2 - c_{21}\eta_{i0}}{c_{22}} < \eta_{i1} < \frac{b_2 - c_{21}\eta_{i0}}{c_{22}} \right) | \eta_{i0}^* \right\},$$

temos,

$$\mathbb{P}(A_1|A_0) = \mathbb{E} \left[\mathbb{P} \left\{ \left(\frac{h_2 - c_{21}\eta_{i0}}{c_{22}} < \eta_{i1} < \frac{b_2 - c_{21}\eta_{i0}}{c_{22}} \right) | \eta_{i0}^* \right\} \right],$$

$$\mathbb{P}(A_1|A_0) = \mathbb{E} (\mathbb{P}(A_1|\eta_{i0}^*)).$$

η_{i0}^* tem uma distribuição normal truncada em (a_0, b_0) . ■

Comentários:

Como η_{i0}^* e η_{i1} são independentes, podemos simular uma amostra onde:

$$\mathbb{P}(A_1|\eta_{i0}^*) = \mathbb{E}_{\eta_{i0}}(I(A_1)|\eta_{i0}^*) = \begin{cases} 1 & \text{se } a_1 < \eta_{i1} < b_1 \\ 0 & \text{caso contrário.} \end{cases}$$

$$\mathbb{P}(A_1|A_0) = \mathbb{E}(\mathbb{E}(I(A_1)|\eta_{i0}^*)).$$

Como no início da prova foi designado que $a_0 < \eta_{i0} < b_0$, assim como η_{i0}^* , pode-se escrever:

$$\mathbb{P}(A_1|A_0) = \mathbb{E}_{\eta_{i0}}(\mathbb{P}(A_1|\eta_{i0})).$$

Observe que a probabilidade condicional $\mathbb{P}(A_1|A_0)$ pode ser vista como uma média dos indicadores $I((A_1)|\eta_{i0}) = \mathbb{P}(A_1|\eta_{i0})$, via simulação. Essas probabilidades condicionais $\mathbb{P}(A_1|A_0)$ estão em função dos η' s, que dependem de covariáveis e da matriz de Cholesky C , que depende de Ω .

Teorema 4.3 Um Estimador não-viciado de $\mathbb{P}[\mathbf{h} < \mathbf{C}\boldsymbol{\eta} < \mathbf{b}]$ (Ver (4.32)), é dado por:

$$= \frac{1}{R} \left\{ \sum_{r=1}^R \mathbb{P}(A_0) \mathbb{P}(A_1 | \eta_{r0}) \times \mathbb{P}(A_2 | \eta_{r0}, \eta_{r1}) \times \cdots \times \mathbb{P}(A_t | \eta_{r0}, \cdots, \eta_{r,t-1}) \right\}, \quad (4.33)$$

onde os $\eta'_{r,i}$ s são seqüências independentes de normais padrão truncadas e R é o número de simulações. (Geweke, Hajivassilou e Keane)⁷.

Essas probabilidades condicionais são dadas por:

$$\mathbb{P}(A_i | \eta_{r1}, \cdots, \eta_{r,t-1}) = \Phi \left(\frac{b_i - c_{t1}\eta_{r1} - \cdots - c_{t-1,i}\eta_{r,t-1}}{c_{ti}} \right) - \Phi \left(\frac{h_i - c_{t1}\eta_{r1} - \cdots - c_{t-1,i}\eta_{r,t-1}}{c_{ti}} \right), \quad (4.34)$$

onde Φ é a função de distribuição acumulada da distribuição normal padrão.

Ajuste do Modelo Probito Dinâmico utilizando a MVS

Para o caso do modelo dinâmico de resposta binária os limites de integração para os η' s são obtidos a partir de uma normal truncada, que faz uso das equações latentes para o período de tempo inicial ($t = 0$) e para $t \geq 1$.

Para o início do processo temos:

$$Y_{i0}^* = \mathbf{x}'_{i0} \Pi_1 + \alpha_i \Pi_2 + U_{i0}, \quad i = 1, \cdots, N.$$

Seja $\mu_{i0} = \mathbf{x}'_{i0} \Pi_1$ e $V_{i0} = \alpha_i \Pi_2 + U_{i0}$, logo:

$$Y_{i0}^* = \mu_{i0} + V_{i0},$$

$$\mathbb{P}(Y_{i0} = 1) = \mathbb{P}(Y_{i0}^* > 0) = \mathbb{P}(V_{i0} > -\mathbf{x}'_{i0} \Pi_1),$$

Pela primeira linha da matriz dada em (4.30), V_{i0} será agora dado por:

$$\begin{aligned} &= \mathbb{P}(c_{11}\eta_{i0} > -\mu_{i0}) = \mathbb{P}(\eta_{i0} > -\mu_{i0}/c_{11}), \\ &= \Phi \left(\frac{\mu_{i0}}{c_{11}} \right) = \Phi(a_{i0}). \end{aligned}$$

⁷Stewart (2006) sugere ver por exemplo Gourieroux e Monfort, 1966, e Cameron e Trivedi, 2005

Para o tempo 1, temos:

$$Y_{i1}^* = \mathbf{x}'_{i1}\boldsymbol{\beta} + \gamma y_{i0} + V_{i1},$$

com $V_{i1} = c_{21}\eta_{i0} + c_{22}\eta_{i1}$ e $\mu_{i1} = \mathbf{x}'_{i1}\boldsymbol{\beta} + \gamma y_{i0}$, logo:

$$\begin{aligned} \mathbb{P}(Y_{i1} = 1) &= \mathbb{P}(Y_{i1}^* > 0) = \mathbb{P}(\mu_{i1} + V_{i1} > 0), \\ &= \mathbb{P}(c_{21}\eta_{i0} + c_{22}\eta_{i1} > -\mu_{i1}), \\ &= \Phi\left(\frac{\mu_{i1} + c_{21}\eta_{i0}}{c_{22}}\right) = \Phi(a_{i1}). \end{aligned}$$

Usando a decomposição de Cholesky, a equação latente pode ser escrita como:

$$Y_{it}^* = \mu_{it} + \sum_{j=1}^T C_{jt}\eta_{ij}, \quad (4.35)$$

onde $\mu_{it} = \gamma y_{i,t-1} + \mathbf{x}'_{it}\boldsymbol{\beta}$ para $t \geq 1$ e $\mu_{i0} = \mathbf{x}'_{i0}\boldsymbol{\beta}$. A probabilidade de uma seqüência observada de Y 's é dada por:

$$\begin{aligned} \mathbb{P}_i &= \Phi((2y_{i0} - 1)a_{i0}) \times \int_{L_{i1}}^{S_{i1}} \Phi((2y_{i1} - 1)a_{i1})\phi(\eta_{i0})d\eta_{i0} \\ &\times \int_{L_{i1}}^{S_{i1}} \int_{L_{i2}}^{S_{i2}} \Phi((2y_{i2} - 1)a_{i2})\phi(\eta_{i0})\phi(\eta_{i1})d\eta_{i1}d\eta_{i0} \\ &\times \dots, \end{aligned} \quad (4.36)$$

onde $a_{i0} = \mu_{i0}/c_{11}$, $a_{i1} = (\mu_{i1} + c_{21}\eta_{i0})/c_{22}$, $a_{i2} = (\mu_{i2} + c_{31}\eta_{i0} + c_{32}\eta_{i1})/c_{33}$, etc. e $(L_{it}, S_{it}) = (-a_{it}, \infty)$ se $Y_{it} = 1$ ou $(-\infty, -a_{it})$ se $Y_{it} = 0$.

Como essas probabilidades são difíceis de ser calculadas, podemos simulá-las. O simulador GHK de $\mathbb{P}[\mathbf{h} < \mathbf{C}\boldsymbol{\eta} < \mathbf{b}]$ foi dado pelo Teorema (4.3).

A função densidade e a função de distribuição acumulada de uma Normal Truncada em $[L_{it}, S_{it}]$ são dadas por:

$$f(x) = \frac{\phi(x)}{\Phi(S_{it}) - \Phi(L_{it})}.$$

$$F(x) = \frac{\Phi(x) - \Phi(L_{it})}{\Phi(S_{it}) - \Phi(L_{it})}.$$

Para retirar uma amostra aleatória desta distribuição, pode ser usada a técnica de transformação de uma variável com distribuição Uniforme, pela função acumulada

inversa da distribuição Normal Truncada. Para isso usa-se um gerador de números aleatórios seguindo uma distribuição Uniforme $[0,1]$, obtendo-se um valor ξ . Um número aleatório x , resultante de uma distribuição Normal Truncada em $[L_{it}, S_{it}]$, será obtido fazendo:

$$\xi = F(x) = \frac{\Phi(x) - \Phi(L_{it})}{\Phi(S_{it}) - \Phi(L_{it})},$$

$$\xi [\Phi(S_{it}) - \Phi(L_{it})] = \Phi(x) - \Phi(L_{it}),$$

$$\Phi(L_{it}) + \xi\Phi(S_{it}) - \xi\Phi(L_{it}) = \Phi(x),$$

$$\Phi(x) = (1 - \xi)\Phi(L_{it}) + \xi\Phi(S_{it}),$$

$$x = \Phi^{-1} [(1 - \xi)\Phi(L_{it}) + \xi\Phi(S_{it})],$$

onde x representa a variável η_{it} . Desta forma, os passos no simulador GHK para a r -ésima retirada relativa ao indivíduo i para este modelo são:

1. Calcular $a_{i0} = \mu_{i0}/c_{11}$.
2. Retirar ξ_{i0}^r de uma uniforme padrão e calcule $\eta_{i0}^r = \Phi^{-1} [(1 - \xi_{i0}^r)\Phi(L_{i0}) + \xi_{i0}^r\Phi(S_{i0})]$ e $a_{i1}^r = (\mu_{i1} + c_{21}\eta_{i0}^r)/c_{22}$.
3. Retirar ξ_{i1}^r de uma uniforme padrão e calcule $\eta_{i1}^r = \Phi^{-1} [(1 - \xi_{i1}^r)\Phi(L_{i1}) + \xi_{i1}^r\Phi(S_{i1})]$ e $a_{i2}^r = (\mu_{i2} + c_{31}\eta_{i0}^r + c_{32}\eta_{i1}^r)/c_{33}$.
4. Repetir os passos sucessivamente para os períodos restantes.

A função de verossimilhança simulada será dada por:

$$L^* = \prod_{i=1}^n \left\{ \frac{1}{R} \sum_{r=1}^R \left[\Phi((2y_{i0} - 1)a_{i0}) \prod_{t=1}^T \Phi((2y_{it} - 1)a_{it}^r) \right] \right\}, \quad (4.37)$$

onde a_{it}^r representa o valor de a_{it} na r -ésima simulação.

Nesta seção foi apresentado o método de estimação por máxima verossimilhança simulada (MVS) para o modelo probito dinâmico (MPD) de efeitos aleatórios com os erros autocorrelacionados. A seguir será apresentado o modelo probito dinâmico com erros não-correlacionados definido por Wooldridge (2005) e sua sugestão para o ajuste de tal modelo que permite o uso de pacotes computacionais que ajustam modelos probitos não-dinâmicos.

4.2.3 Modelo Probito Dinâmico para Dados em Painel com Erros Não-Correlacionados, segundo Wooldridge

Nesta seção será apresentada a proposta de Wooldridge (2005) para (4.1) que sugere modelar a distribuição dos efeitos não-observados α_i condicionando-a ao valor inicial e às variáveis explicativas exógenas. Essa sugestão já havia sido feita antes para alguns modelos particulares como para o modelo linear AR(1) sem covariáveis ⁸.

Para o modelo dinâmico de resposta binária, Wooldridge cita Arellano e Carrasco(2003) que estudaram um estimador de máxima verossimilhança condicionado em $(\mathbf{x}_{i0}, y_{i0})$, assumindo que a distribuição condicional do efeito não-observado é discreta.

Wooldridge mostra que sua proposta para modelos não-lineares dinâmicos é mais flexível que as propostas já apresentadas por outros autores e mais simples computacionalmente: No caso dos modelos logito ou probito, os softwares que ajustam modelos não-dinâmicos com efeitos aleatórios podem ser usados para estimar os parâmetros desse modelo dinâmico.

Assumindo que o par $(\mathbf{x}_{it}, Y_{it})$ é observado nos tempos $t = 1, \dots, T$, um Modelo Probito Dinâmico com efeitos não-observados pode ser definido por (4.38) e (4.39):

$$P(Y_{it} = 1 | y_{i,t-1}, \dots, y_{i0}, \mathbf{x}'_i, \alpha_i) = \Phi(\mathbf{x}'_{it}\boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i). \quad (4.38)$$

α_i tem uma função densidade de probabilidade condicional dada por:

$$h(\alpha_i | y_{i0}, \mathbf{x}_i; \boldsymbol{\delta}) \quad (4.39)$$

Considerações sobre o MPD de Wooldridge:

Em (4.38), supõe-se que a probabilidade da variável resposta Y_{it} assumir o valor 1 depende somente do valor da resposta imediatamente anterior $Y_{i,t-1}$, dados os valores de $Y_{i,t-1}, \dots, Y_{i0}, \mathbf{x}_{it}$ e α_i e que \mathbf{x}_{it} satisfaz a hipótese de exogeneidade estrita, ou seja, somente \mathbf{x}_{it} aparece no lado direito, embora $\mathbf{x}_i = (\mathbf{x}_{i0}, \dots, \mathbf{x}_{iT})$ apareça no conjunto condicionante ao lado esquerdo.

⁸Wooldridge (2005) cita Chamberlain (1980)

Estimação do MPD de Wooldridge por MV condicional

Wooldridge (2005), propõe usar a função densidade condicional de (Y_{i1}, \dots, Y_{iT}) dado (y_{i0}, \mathbf{x}) para fazer estimação dos parâmetros. Essa densidade pode ser obtida multiplicando a densidade de (Y_{i1}, \dots, Y_{iT}) dado $(y_{i0}, \mathbf{x}_i, \alpha_i)$ pela densidade de α_i dado (y_{i0}, \mathbf{x}_i) .

A densidade conjunta de (Y_{i1}, \dots, Y_{iT}) dado $(Y_{i0} = y_{i0}, \mathbf{x}_i = \mathbf{x}_i, \alpha_i = \alpha)$ indicada por $g(y_{i1}, \dots, y_{iT}, \mathbf{x}, \alpha_i, y_{i0})$ pode ser fatorada da seguinte forma:

$$g(y_{i1}, \dots, y_{iT} | \mathbf{x}_i, \alpha_i, y_{i0}) =$$

$$g(y_{i1} | y_{i0}, \mathbf{x}_i, \alpha_i) \cdot g(y_{i2} | y_{i1}, y_{i0}, \mathbf{x}, \alpha) \times \dots \times g(y_{iT} | y_{i,t-1}, y_{i,t-2}, \dots, y_{i,1}, y_{i0}, \mathbf{x}, \alpha),$$

Pela expressão (4.38) tem-se que no máximo uma defasagem do Y_{it} aparece na distribuição condicionada aos resultados anteriores de Y_{it} desde o período de tempo inicial Y_{i0} , logo:

$$g(y_{i1}, \dots, y_{iT} | \mathbf{x}_i, \alpha_i, y_{i0}) =$$

$$\begin{aligned} &g(y_{i1}, \dots, y_{iT} | \mathbf{x}_i, \alpha_i) \cdot g(y_{i2} | y_{i1}, \mathbf{x}, \alpha) \cdot g(y_{i3} | y_{i2}, \mathbf{x}, \alpha) \cdot \dots \times g(y_{iT} | y_{i,t-1}, \mathbf{x}, \alpha) \\ &= \prod_{t=1}^T g_t(y_{it} | y_{i,t-1}, \mathbf{x}, \alpha). \end{aligned} \quad (4.40)$$

Pela Equação (4.38)

$$\begin{aligned} g(y_{i1}, \dots, y_{iT} | \mathbf{x}, \alpha, y_{i0}) &= \left[\Phi(\mathbf{x}'_{it} \boldsymbol{\beta} + \gamma y_{i0} + \alpha_i) \right]^{y_{i0}} \cdot \left[1 - \Phi(\mathbf{x}'_{it} \boldsymbol{\beta} + \gamma y_{i0} + \alpha_i) \right]^{1-y_{i0}} \\ &\times \dots \times \left[\Phi(\mathbf{x}'_{it} \boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i) \right]^{y_{i,t-1}} \cdot \left[1 - \Phi(\mathbf{x}'_{it} \boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i) \right]^{1-y_{i,t-1}} \\ &= \prod_{t=1}^T \left(\left[\Phi(\mathbf{x}'_{it} \boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i) \right]^{y_{it}} \left[1 - \Phi(\mathbf{x}'_{it} \boldsymbol{\beta} + \gamma y_{i,t-1} + \alpha_i) \right]^{1-y_{it}} \right). \end{aligned} \quad (4.41)$$

A densidade de (Y_{i1}, \dots, Y_{iT}) dado (y_{i0}, \mathbf{x}_i) é obtida integrando (4.41) em relação a $h(\alpha | Y_0, \mathbf{x}; \boldsymbol{\delta})$:

$$\int_{\mathbb{R}^J} \left(\prod_{t=1}^T g_t(y_{it} | \mathbf{x}_t, y_{i,t-1}, \alpha; \boldsymbol{\beta}, \boldsymbol{\gamma}) \right) \cdot h(\alpha | y_{i0}, \mathbf{x}_i; \boldsymbol{\delta}) d\alpha. \quad (4.42)$$

Portanto, a função de log-verossimilhança dados Y_{i0}, \mathbf{x}_i para uma observação i :

$$l_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) = \log \left[\int_{\mathbb{R}^J} \left(\prod_{t=1}^T g_t(y_{it} | \mathbf{x}_{it}, y_{i,t-1}, \alpha; \boldsymbol{\beta}, \boldsymbol{\gamma}) \right) \cdot h(\alpha | y_{i0}, \mathbf{x}_i; \boldsymbol{\delta}) d\alpha \right]. \quad (4.43)$$

Na equação (4.39), suponha-se que:

$$\alpha_i|y_{i0}, \mathbf{x}'_i \sim \text{Normal}(\omega_0 + \omega_1 y_{i0} + \mathbf{x}'_i \boldsymbol{\omega}_2, \sigma_a^2), \quad (4.44)$$

sendo então $\boldsymbol{\delta} = (\omega_1, \omega_2)$. Equivalentemente, (4.44) pode ser escrita como:

$$\alpha_i = \omega_0 + \omega_1 y_{i0} + \mathbf{x}'_i \boldsymbol{\omega}_2 + a_i, \quad (4.45)$$

onde $a_i|(y_{i0}, \mathbf{x}'_i) \sim \text{Normal}(0, \sigma_a^2)$

Substituindo α_i da expressão (4.45), o modelo para Y_{it}^* é:

$$Y_{it}^* = \mathbf{x}'_{it} \boldsymbol{\beta} + \gamma Y_{i,t-1} + \omega_0 + \omega_1 Y_{i0} + \mathbf{x}_i \boldsymbol{\omega}_2 + a_i + U_{it}, \quad (4.46)$$

onde $U_{it} | (\mathbf{x}_i, Y_{i,t-1}, \dots, Y_{i0}, a_i) \sim N(0,1)$. Assim a função densidade conjunta de (Y_{i1}, \dots, Y_{iT}) dado $(Y_{i0} = y_{i0}, \mathbf{x}_i = \mathbf{x}, a_i = a_i)$ é:

$$\begin{aligned} & \prod_{t=1}^T \{ \Phi(\mathbf{x}'_t \boldsymbol{\beta} + \gamma y_{i,t-1} + \omega_0 + \omega_1 y_{i0} + \mathbf{x}'_t \boldsymbol{\omega}_2 + a_i)^{y_{it}} \\ & \times [1 - \Phi(\mathbf{x}'_t \boldsymbol{\beta} + \gamma y_{i,t-1} + \omega_0 + \omega_1 y_{i0} + \mathbf{x}'_t \boldsymbol{\omega}_2 + a_i)]^{1-y_{it}} \}. \end{aligned} \quad (4.47)$$

Integrando a equação(4.47) com relação à densidade $\text{Normal}(0, \sigma_a^2)$ obtém-se a densidade de (Y_{i1}, \dots, Y_{iT}) dado $(Y_{i0} = y_{i0}, \mathbf{x}'_i = \mathbf{x}'_i)$:

$$\begin{aligned} & \int_{\mathbb{R}} \left(\prod_{t=1}^T \{ \Phi(\mathbf{x}'_t \boldsymbol{\beta} + \gamma y_{i,t-1} + \omega_0 + \omega_1 y_{i0} + \mathbf{x}'_t \boldsymbol{\omega}_2 + a_i)^{y_{it}} \right. \\ & \left. \times [1 - \Phi(\mathbf{x}'_t \boldsymbol{\beta} + \gamma y_{i,t-1} + \omega_0 + \omega_1 y_{i0} + \mathbf{x}'_t \boldsymbol{\omega}_2 + a_i)]^{1-y_{it}} \right) (1/\sigma_a) \phi(a/\sigma_a) da. \end{aligned} \quad (4.48)$$

A verossimilhança anterior pode ser escrita da seguinte forma:

$$\begin{aligned} L_i &= \frac{1}{\sigma_a} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{a_i}{\sigma_a})^2} \\ & \times \prod_{t=1}^T \Phi(q_{it}(\mathbf{x}'_t \boldsymbol{\beta} + \gamma y_{i,t-1} + \omega_0 + \omega_1 y_{i0} + \mathbf{x}'_t \boldsymbol{\omega}_2 + a_i)) da_i, \end{aligned} \quad (4.49)$$

onde $q_{it} = 2y_{it} - 1$. Seja $r_i = a_i/\sigma_a\sqrt{2} \Rightarrow a_i = (\sigma_a\sqrt{2})r_i \Rightarrow a = \theta r_i$.

Logo $\theta = \sigma_a\sqrt{2}$ e $da = \theta dr_i$.

Com estas substituições a equação (4.49) resulta em:

$$L_i = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-r_i^2} \times \prod_{t=1}^T \Phi(q_{it}(\mathbf{x}'_t \boldsymbol{\beta} + \gamma y_{i,t-1} + \omega_0 + \omega_1 y_{i0} + \mathbf{x}'_t \boldsymbol{\omega}_2 + \theta r_i)) dr_i. \quad (4.50)$$

Observe que as verossimilhanças nas equações (4.49) e (4.50) têm a mesma estrutura da função de verossimilhança do **modelo Probito estático de efeitos aleatórios** dado na pág.37 ((3.33) e (3.34)), com variáveis explicativas no período de tempo t , dadas por:

$$\mathbf{z}'_{it} \equiv (\mathbf{x}'_{it}, y_{i,t-1}, 1, y_{i0}, \mathbf{x}'_i). \quad (4.51)$$

Somando as logverossimilhanças dadas por (4.50), para $i = 1, \dots, N$, a logverossimilhança conjunta para todo painel é obtida. As integrais podem ser aproximadas usando a quadratura de Gauss-Hermite, de modo que uma expressão aproximada para a verossimilhança conjunta, denotada por l_H , é:

$$l_H = \sum_{i=1}^N \left\{ \left[\frac{1}{\sqrt{\pi}} \sum_{h=1}^H \prod_{t=1}^T p_h \phi(q_{it}(\mathbf{x}'_t \boldsymbol{\beta} + \gamma y_{i,t-1} + w_0 + w_1 y_{i0} + \mathbf{x}'_t \mathbf{w}_2 + \theta z_h)) \right] \right\} \quad (4.52)$$

onde H é o número de pontos para a quadratura, p_h e z_h são os pesos e nós para a quadratura respectivamente. Usando a expressão aproximada para l_H dada por (4.50), procura-se os valores dos parâmetros que maximizam a função l_H . Isso pode ser feito usando um software para ajuste de um modelo estático de efeitos aleatórios, apesar do modelo ser dinâmico. Neste ajuste, são considerados parâmetros os coeficientes dos elementos de \mathbf{z}_{it} dado em (4.51) que aparecem em (4.50), e ainda σ_a ou equivalentemente $\theta = \sigma_a \sqrt{2}$. Desta forma os parâmetros a serem estimados serão: $\boldsymbol{\beta}, \gamma, w_0, w_1, \mathbf{w}_2$ e σ_a^2 .

Peters (2005) sugere utilizar as covariáveis dadas por (4.53) para tentar diminuir o problema da colinearidade entre as variáveis:

$$\mathbf{z}'_{it} \equiv (\mathbf{x}'_{it}, y_{i,t-1}, 1, y_{i0}, \bar{\mathbf{x}}'_i), \quad (4.53)$$

$$l_H = \sum_{i=1}^N \left\{ \left[\frac{1}{\sqrt{\pi}} \sum_{h=1}^H \prod_{t=1}^T p_h \phi(q_{it}(\mathbf{x}'_t \boldsymbol{\beta} + \gamma y_{i,t-1} + w_0 + w_1 y_{i0} + \bar{\mathbf{x}}'_t \mathbf{w}_2 + \theta z_h)) \right] \right\} \quad (4.54)$$

Logo pode-se utilizar um software destinado a estimar os parâmetros do modelo probito estático para estimar os parâmetros do modelo probito dinâmico de Wooldridge. Stata, Limdep e R são exemplos de pacotes que podem ser utilizados.

Capítulo 5

Resultados

Foram propostos como objetivos desta dissertação a criação de programas em R para ajuste de alguns dos modelos de resposta binária pelo método da máxima verossimilhança, a realização de um estudo por simulação sobre propriedades dos estimadores obtidos por tais ajustes e ainda a verificação de propriedades relativas à distribuição conjunta dos estimadores de um dos parâmetros em uma seqüência de análises intermediárias do painel. Os resultados obtidos serão apresentados nas diversas seções deste capítulo.

5.1 Programa para ajuste de um Modelo Probit Estático (MPE)

Um primeiro resultado do estudo feito foi a elaboração de um programa no R para o ajuste pelo método de máxima verossimilhança de um modelo estático de resposta binária com efeitos aleatórios, utilizando o recurso de quadratura para o cálculo de integrais. Este programa está no Apêndice B e a ausência de um programa com tal objetivo justifica sua inclusão neste trabalho. A seguir, a validade do programa foi verificada a partir da comparação das estimativas calculadas e os valores reais dos parâmetros e da comparação das estimativas calculadas pelo programa R com as fornecidas por softwares comerciais. Foram simulados painéis de dados segundo um modelo probito estático de coeficientes

de regressão conhecidos para alguns valores de N e T e a seguir, foram estimados os coeficientes a partir dos dados simulados usando o programa em R e dois softwares comerciais, o Stata e o Limdep. Os valores reais dos coeficientes estão na Tabela 5.1 e os valores das estimativas desses coeficientes que foram iguais nos três programas, são mostrados na Tabela 5.2. Comparando as referidas tabelas, pode-se julgar quão próximas estão os valores reais dos coeficientes e suas estimativas pelo programa R.

O processo gerador de dados referentes a um indivíduo i para o Modelo Estático de Efeitos Aleatórios foi :

$$Y_{it}^* = \mathbf{x}'_{it}\boldsymbol{\beta} + U_{it} + \alpha_i, \quad i = 1, \dots, N \quad t = 1, \dots, T, \quad (5.1)$$

$$Y_{it} = 1, \text{ se } Y_{it}^* > 0 \quad 0, \text{ caso contrário,}$$

As observações referentes a cada indivíduo foram geradas independentemente das observações referentes aos demais indivíduos. Na geração de observações do modelo, foram consideradas apenas duas covariáveis, X_1 e X_2 , a primeira com distribuição de Bernoulli e a segunda com distribuição Normal. Para cada i , e $t = 1, \dots, T$, as covariáveis X_1 e X_2 , os erros aleatórios e a heterogeneidade individual foram todos gerados independentemente. A heterogeneidade individual não-observada, α_i foi gerada de uma Normal(0,1) e os erros U_{it} , foram gerados de uma normal multivariada de média igual ao vetor nulo de dimensão T e matriz de covariância identidade $T \times T$. As T observações de X_1 e X_2 em cada indivíduo i ($i = 1, \dots, N$), nos diversos tempos t ($t = 1, \dots, T$) foram geradas independentemente. Os dados foram gerados no R com semente 123. A Tabela 5.1 mostra os valores reais atribuídos aos parâmetros. Conforme explicado nas

TABELA 5.1: *Valores atribuídos aos parâmetros do MPE em uma simulação.*

	β_1	β_2	θ	σ_α	ρ
Valores reais	1	-0,7	1,41	1	0,5

páginas 35 e 37:

$$\rho = \frac{\sigma_\alpha^2}{1 + \sigma_\alpha^2}$$

$$\theta = \sigma_\alpha \sqrt{2}$$

Foi feito o ajuste por MV do modelo estático de efeitos aleatórios ao painel de dados gerados no software R, no Limdep e no software Stata, com 12 pontos de quadratura. A Tabela 5.2 mostra as estimativas dos parâmetros, que foram iguais para os três pacotes.

TABELA 5.2: *Estimativas por MV dos parâmetros de um painel simulado segundo um MPE e $\ln L^*$ via R, Limdep e Stata.*

N	T	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\alpha$	$\hat{\rho}$	$\ln L$
100	4	1,56	-0,96	2,26	0,69	-94,55
200	4	0,92	-0,64	0,94	0,47	-193,70
300	5	1,00	-0,88	0,95	0,47	-323,27
400	5	1,12	-0,75	0,98	0,49	-401,11
500	6	1,04	-0,57	1,09	0,54	-665,89
600	6	1,05	-0,70	1,11	0,55	-789,78

* $\ln L$ = log-verossimilhança

Como exposto acima, o programa em R para o ajuste do Modelo Probit Estático fornece exatamente as mesmas estimativas que os pacotes comerciais utilizados. Os resultados obtidos estão bem próximos dos verdadeiros valores, que se encontram na Tabela 5.1. À medida que aumentamos o N e o T as estimativas ficam mais próximas dos verdadeiros valores.

5.2 Ajuste de um Modelo Probit Dinâmico-(MPD) segundo Wooldridge usando o programa para (MPE)

O processo gerador de dados de um indivíduo i segundo um Modelo Probit Dinâmico de Efeitos Aleatórios de Wooldridge é:

$$Y_{i0} \sim \text{Bernoulli}(p) \quad (5.2)$$

$$\alpha_i = \omega_0 + \omega_1 Y_{i0} + \mathbf{x}'_i \boldsymbol{\omega}_2 + a_i, \quad (5.3)$$

onde $a_i | (Y_{i0}, \mathbf{x}'_i) \sim \text{Normal}(0, \sigma_a^2)$

$$Y_{it}^* = \mathbf{x}'_{it} \boldsymbol{\beta} + \gamma Y_{i,t-1} + \alpha_i + U_{it}, \quad t = 1, \dots, T, \quad (5.4)$$

onde $U_{it} | (\mathbf{x}_i, Y_{i,t-1}, \dots, Y_{i0}, a_i) \sim N(\mathbf{0}, \mathbf{I}_{(T \times T)})$, $t = 1, \dots, T$.

$$Y_{it} = 1, \text{ se } Y_{it}^* > 0 \quad 0, \text{ caso contrário,}$$

Wooldridge (2005) mostrou que o ajuste do modelo acima por verossimilhança condicional (aos valores de (\mathbf{x}_i, y_{i0}) , $i = 1, \dots, N$) é igual ao ajuste por máxima verossimilhança de um modelo probito estático com covariáveis \mathbf{z}'_{it} dadas por $\mathbf{z}'_{it} \equiv (\mathbf{x}'_{it}, Y_{i,t-1}, 1, Y_{i0}, \mathbf{x}'_i)$. Peters (2006), sugere usar $\mathbf{z}'_{it} \equiv (\mathbf{x}'_{it}, Y_{i,t-1}, 1, Y_{i0}, \bar{\mathbf{x}}_i)$. Os coeficientes de regressão destes componentes serão indicados por: w_0 (intercepto), w_1 (coeficiente de Y_{i0}) e w_{2j} (coeficientes de $\bar{\mathbf{x}}_j$, $j = 1, 2$).

Nesta seção será verificado essa afirmativa, usando um software qualquer que ajuste um modelo probito estático para ajustar um modelo Probito dinâmico de Wooldridge. O modelo (5.4) foi gerado usando duas covariáveis, em conseqüência, para usar a idéia de Wooldridge será ajustado um modelo probito estático para as respostas y_{it} em termos das variáveis $\mathbf{x}_{i1t}, \mathbf{x}_{i2t}, Y_{i,t-1}, 1, Y_{i0}, \bar{x}_{i1}$ e \bar{x}_{i2} com coeficientes $\beta_1, \beta_2, \gamma, w_0, w_1, w_{21}$ e w_{22} , respectivamente. Ressalta-se que o interesse maior reside na estimação dos três primeiros parâmetros.

Na geração de observações do modelo, foram consideradas apenas duas covariáveis, X_1 e X_2 , a primeira com distribuição de Bernoulli e a segunda com distribuição Normal. As observações referentes a cada indivíduo foram geradas independentemente das observações referentes aos demais indivíduos. Para cada i , e $t = 1, \dots, T$, as covariáveis X_1 e X_2 , os erros aleatórios e a heterogeneidade individual foram todos gerados independentemente entre si. A heterogeneidade individual não-observada, α_i foi gerada condicionalmente ao valor de Y_{i0} segundo (5.3) onde nessa equação $a_i \sim N(0, 1)$. Os erros U_{it} , foram gerados de uma normal multivariada de média igual ao vetor nulo de dimensão T e matriz de covariância identidade $T \times T$. As T observações de X_1 e X_2 em cada indivíduo i ($i = 1, \dots, N$), nos diversos tempos t ($t = 1, \dots, T$) foram geradas independentemente.

$\rho = \frac{\sigma_a^2}{1 + \sigma_a^2}$ é o coeficiente de correlação entre V_{it} e V_{is} , para $t \neq s$ com $V_{it} = U_{it} + \alpha_i$, $\forall i$ e t e $\theta = \sigma_a \sqrt{2}$.

Foi feito o ajuste por MVC do modelo probito dinâmico de efeitos aleatórios de

TABELA 5.3: *Valores atribuídos aos parâmetros do MPD de Wooldridge em uma simulação.*

Parâmetros	β_1	β_2	γ	w_0	w_1	w_{21}	w_{22}	θ	ρ
Valores reais	1	-0,7	1,2	-0,5	1	0,5	0,5	1,42	0,5

Wooldridge ao painel de dados gerados no software R, no Limdep e no software Stata, com 12 pontos de quadratura. A Tabela 5.4 mostra as estimativas dos parâmetros, que foram iguais para os três pacotes.

TABELA 5.4: *Estimativas por MVC dos parâmetros de seis painéis de dados simulados segundo o MPD de Wooldridge, lnL* via R, Limdep e Stata.*

Parâmetros	Variável	T=4		T=5		T=6	
		N=100	N=200	N=300	N=400	N=500	N=600
β_1	X_1	1,29	0,90	0,94	1,17	1,08	1,23
		(0,43)	(0,13)	(0,10)	(0,13)	(0,08)	(0,09)
β_2	X_2	-0,36	-0,50	-0,73	-0,71	-0,89	-0,83
		(0,42)	(0,24)	(0,18)	(0,18)	(0,13)	(0,14)
γ	γ	1,68	0,96	1,26	1,18	1,07	1,40
		(0,60)	(0,25)	(0,19)	(0,19)	(0,14)	(0,16)
w_0	1	0,31	-0,83	-1,07	-0,59	-0,68	-0,54
		(0,61)	(0,30)	(0,25)	(0,27)	(0,22)	(0,24)
w_1	Y_{i0}	0,84	0,78	1,00	1,11	1,22	1,02
		(0,60)	(0,29)	(0,22)	(0,25)	(0,18)	(0,18)
w_{21}	\bar{X}_1	0,83	0,51	0,47	0,64	0,56	0,61
		(0,33)	(0,11)	(0,08)	(0,10)	(0,07)	(0,08)
w_{22}	\bar{X}_2	-1,75	1,18	1,41	1,07	1,11	0,89
		(1,23)	(0,53)	(0,44)	(0,47)	(0,38)	(0,42)
θ		1,46	0,79	0,79	1,07	1,07	1,30
		(0,70)	(0,28)	(0,19)	(0,21)	(0,14)	(0,16)
σ_a		1,03	0,55	0,55	0,76	0,76	0,92
ρ		0,68	0,39	0,38	0,53	0,53	0,63
		(0,24)	(0,19)	(0,11)	(0,11)	(0,07)	(0,06)
lnL		-76	-146	-262	-295	-537	-589

*lnL = log-verossimilhança

O desvio padrão de cada estimativa está entre parênteses.

Analisando a Tabela 5.4, verifica-se que as estimativas dos três primeiros parâmetros estão próximos dos verdadeiros valores dados na Tabela 5.3.

Para avaliar o efeito do número de pontos utilizados para a integração numérica em (4.52), sobre as estimativas obtidas, foram gerados 4 painéis com $T = 6$; Escolheu-se $N = 500$ e $N = 600$ e duas sementes (214 e 222) diferentes. Os valores atribuídos aos

parâmetros foram os mesmos que no exemplo anterior e estão novamente descritos na Tabela 5.5. Na Tabela 5.6 estão apresentadas as estimativas dos parâmetros obtidas com 10 pontos de quadratura.

TABELA 5.5: *Valores atribuídos aos parâmetros do MPD de Wooldridge em uma simulação.*

Parâmetros	β_1	β_2	γ	w_0	w_1	w_{21}	w_{22}	θ	ρ
Valores reais	1	-0,7	1,2	-0,5	1	0,5	0,5	1,42	0,5

TABELA 5.6: *Estimativas por MVC dos parâmetros e $\ln L^*$ via R, para 4 painéis de dados simulados segundo Wooldridge usando 10 pontos de quadratura.*

Parâmetros	Variáveis	Semente 222		Semente 214	
		N=500	N=600	N=500	N=600
β_1	X_1	0,94	0,89	0,97	0,90
β_2	X_2	-0,67	-0,79	-0,53	-0,61
γ	$Y_{i,t-1}$	0,83	0,85	0,71	0,78
w_0	1	0,26	0,28	-0,85	-0,04
w_1	Y_{i0}	1,10	1,10	0,80	1,1
w_{21}	\bar{X}_1	0,56	0,48	0,58	0,54
w_{22}	\bar{X}_1	0,58	0,61	0,78	0,94
$\theta = \sigma_a \sqrt{2}$		1,08	1,08	1,32	1,09
σ_a		0,76	0,76	0,94	0,77
$\rho = \sigma_a^2 / (1 + \sigma_a^2)$		0,37	0,37	0,47	0,37
$\ln L$		-876	-1093	-907	-1082

* $\ln L$ = log-verossimilhança

As estimativas dos parâmetros pelo programa em R estão bem próximas dos respectivos valores reais, como é possível observar nas Tabelas 5.5 e 5.6, comprovando que o modelo probito dinâmico de Wooldridge (2005) pode realmente ser ajustado usando um pacote computacional qualquer que ajuste um modelo probito estático. Na próxima seção serão mostrados os resultados de simulação de 100 painéis de mesma dimensão para um MPD, o que permitirá avaliar propriedades dos estimadores propostos.

5.3 Estudo por simulação das propriedades dos estimadores por máxima verossimilhança do MPD de Wooldridge

Nesta seção apresenta-se o estudo por simulação de propriedades do estimador dos coeficientes do **Modelo Probito Dinâmico** (MPD) de efeitos aleatórios apresentado na seção 4.2.3 no capítulo 4 definido por Wooldridge. Foram gerados segundo o modelo da seção 5.2 $n = 100$ painéis de observações referentes a uma variável resposta binária Y e duas covariáveis X_1 e X_2 , a primeira com distribuição de Bernoulli e a segunda com distribuição Normal. As observações referentes a cada indivíduo foram geradas independentemente das observações referentes aos demais indivíduos. Para cada i , e $t = 1, \dots, T$, as covariáveis X_1 e X_2 , os erros aleatórios e a heterogeneidade individual foram todos gerados independentemente entre si. A heterogeneidade individual não-observada, α_i foi gerada condicionalmente ao valor de Y_{i0} segundo (5.3) onde nessa equação $a_i \sim N(0, 4)$. Os erros U_{it} , foram gerados de uma normal multivariada de média igual ao vetor nulo de dimensão T e matriz de covariância identidade $T \times T$. As T observações de X_1 e X_2 em cada indivíduo i ($i = 1, \dots, N$), nos diversos tempos t ($t = 1, \dots, T$) foram geradas independentemente. Os dados foram gerados no R 2.6.0. A Tabela 5.7 apresenta as médias das estimativas obtidas em 100 simulações do modelo dado em (5.2), com $N = 500$ e $T = 5$.

TABELA 5.7: Médias das estimativas por MVC dos coeficientes e EQM em 100 simulações de um MPD de Wooldridge, via Stata.

Variáveis	Coefficientes	Valores Estimados	Valores Reais	EQM
X_1	β_1	0,97	1	0,0109
X_2	β_2	-0,67	-0,70	0,0209
Y_{t-1}	γ	1,02	1,20	0,0624
Y_0	w_1	-0,09	0,70	0,5321
\bar{X}_1	w_{21}	0,57	0,50	0,0249
\bar{X}_2	w_{22}	0,61	0,5	1,1421

O interesse do pesquisador quando usa um modelo que considera a Esperança (possivelmente condicional) de uma variável resposta Y_{it} como função do estado anterior

$(Y_{i,t-1})$ e de covariáveis X_1 e X_2 , é nos coeficientes dessas variáveis. Os coeficientes relativos à modelagem feita por Wooldridge para a heterogeneidade individual α_i em termos de y_{i0} e $\bar{\mathbf{x}}_i$ foram introduzidos apenas para levar a uma simplificação no ajuste de modelos dinâmicos e não são de interesse maior. Nesse sentido, a Tabela 5.7 mostra estimativas satisfatórias para os coeficientes de interesse que são γ , β_1 e β_2 . Os valores dos erros quadráticos médios estão na coluna 4 e observa-se que são erros bem pequenos. Estes resultados indicam que possivelmente os estimadores dos coeficientes são consistentes.

5.4 Estimando o MPD de Heckman usando a sugestão de Wooldridge

Conforme visto na seção 4.1.3, o ajuste de um MPD de Heckman por máxima verossimilhança condicional exige esforço computacional muito grande, ao usar, por exemplo, o método de máxima verossimilhança simulada (Stewart,2005). Dessa forma, é interessante tentar outras formas de obter estimadores consistentes para os parâmetros. Nesta seção será apresentado um estudo por simulação das propriedades do estimador por máxima verossimilhança condicional de um modelo segundo Wooldridge para os coeficientes de um **Modelo Probito Dinâmico - MPD** de Efeitos Aleatórios segundo Heckman (1964). O objetivo é detectar se introduzindo covariáveis adicionais, como em (4.53), e ajustando um MPE (como fez Wooldridge em seu MPD) obtém-se estimadores do modelo de Heckman com boas propriedades quanto a consistência, ausência de vício, etc.

O processo gerador de dados do modelo de Heckman é:

$$Y_{it}^* = \mathbf{x}'_{it}\boldsymbol{\beta} + \gamma Y_{i,t-1} + \alpha_i + U_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T; \quad (5.5)$$

- i) os erros U_{it} ($t = 1, \dots, T$ e i fixo) têm uma distribuição normal multivariada com estrutura de autocorrelação arbitrária;
- ii) α_i , o efeito aleatório específico do indivíduo é independente de \mathbf{x}_{it} e U_{it} , $\forall t$.
- iv) as condições iniciais são modeladas por: ¹

$$Y_{i0}^* = \mathbf{x}'_{i0}\Pi_1 + \alpha_i\Pi_2 + U_{i0}, \quad i = 1, \dots, N, \quad (5.6)$$

¹Stewart (2006) sugere ver Heckman (1981).

iii)

$$Y_{it} = \begin{cases} 1 & \text{se } Y_{it}^* \geq 0 \quad t = 0, 1, \dots, T. \\ 0 & \text{caso contrário.} \end{cases} \quad (5.7)$$

Para cada i , e $t = 1, \dots, T$, as covariáveis X_1 e X_2 , os erros aleatórios (U_{it}) e a heterogeneidade individual (α_i) foram todos gerados independentemente entre si. Y_{i0} foi gerado segundo (5.6), onde nessa equação $\alpha_i \sim N(0, 1)$, $X_2 \sim \text{Bernoulli}(p)$ e $X_1 \sim \text{Normal}$, $t = 0, 1, \dots, T$. Os erros U_{it} , foram gerados de uma normal multivariada de média igual ao vetor nulo de dimensão T e matriz de covariância identidade $T \times T$. As T observações de X_1 e X_2 em cada indivíduo i ($i = 1, \dots, N$), nos diversos tempos t ($t = 0, \dots, T$) foram geradas independentemente. Ressalta-se que no modelo de Heckman, Y_{i0} depende da heterogeneidade α_i e um erro independente U_{i0} .

Foram gerados no R 2.6.0 $n = 100$ painéis de $N = 600$ firmas que foram observadas por $T = 6$ tempos. Os resultados obtidos em média no R para 100 simulações foram:

TABELA 5.8: *Médias das estimativas por MVC dos coeficientes e EQM em 100 simulações de painéis de dados gerados segundo um MPD de Heckman usando a sugestão de Wooldridge.*

Variável	Coefficientes	Valores Estimados	Valores Reais	EQM
X_1	β_1	0,9774	1,00	0,0034
X_2	β_2	-0,6881	-0,70	0,0112
Y_{t-1}	γ	0,9604	1,20	0,0699
Y_0	w_1	-0,1697	1,00	1,3783
\bar{X}_1	w_{21}	0,0286	0,50	0,2241
\bar{X}_2	w_{22}	0,0048	0,50	0,3341

Nesse sentido, a Tabela 5.8 mostra que as médias das estimativas dos coeficientes β_1 e β_2 estão bem próximas dos respectivos verdadeiros valores. A média das estimativas de γ já não está tão próxima do verdadeiro valor. Os valores dos erros quadráticos médios estão na coluna 5 e observa-se que são erros bem pequenos. Apesar dos dados não terem sido gerados segundo o MPD de Wooldridge, foi tentado estimar os parâmetros β e γ conforme proposto por Wooldridge. Esses resultados comprovam

que pode ser utilizada a sugestão de Wooldridge para o ajuste de um MPD com as especificações de Heckman utilizando um pacote que ajusta um MPE.

Concluimos que parece ser razoável estimar os parâmetros de um MPD de Heckman como se fosse um MPD de Wooldridge. Ressalta-se que esta afirmação deve ser comprovada em outros estudos por simulação para outras covariáveis e/ou outros valores de parâmetros do modelo de regressão. Em aplicações práticas, dois modelos que tem sido mais usados são os de Wooldridge e Heckman. A adequabilidade desses modelos pode ser objeto de estudo teórico, com a criação de uma metodologia de análise de resíduos.

5.5 Análises intermediárias em um modelo probito dinâmico: estudo via *bootstrap*

No seguimento de um corte transversal no decorrer do tempo, é interessante fazer análises dos dados em diversos instantes entre o tempo $t = 0$ e $t = T$. Pretende-se nesta seção realizar um estudo empírico sobre a distribuição conjunta dos estimadores dos coeficientes de um modelo de regressão probito dinâmico segundo Wooldridge nos diversos momentos de análise. Mais precisamente pretendeu-se iniciar este estudo analisando a estrutura de covariância dos estimadores de um parâmetro, em três momentos sucessivos do tempo. Para um painel com $T = 10$ tempos, e $N = 200$ indivíduos, foram feitas três análises intermediárias, isto é, fez-se inicialmente a estimação dos coeficientes do modelo de regressão com os dados dos indivíduos que foram coletados até o tempo $T = 4$, a seguir, procedeu-se à estimação dos coeficientes com os dados coletados sobre os mesmos indivíduos até o tempo $T = 7$ e depois, fez-se a análise com os dados coletados até o tempo $T = 10$. Foi usada a técnica de *bootstrap* para estimar a matriz de covariância dos estimadores de um específico coeficiente.

Nesta seção serão apresentados os resultados de 345 amostras de *bootstraps* de um painel de dados com $T = 10$ referentes a $N = 200$ indivíduos. Foi gerado um painel de dados com resposta binária e duas variáveis explicativas, com $T = 10$ e $N = 200$, $\mathbf{x}_1 \sim N(0, 4)$ e $\mathbf{x}_2 \sim \text{Bernoulli}(0, 5)$. Para a heterogeneidade individual não-observada α_i e para os erros U_i , foi atribuída uma distribuição Normal padrão. Foram retiradas 345 amostras *bootstrap* deste painel de dados. O ajuste do modelo MPD de Wooldridge

a cada reamostra do painel foi feito usando o programa descrito na seção anterior, usando quadratura de Gauss Hermite com 12 pontos. As médias das estimativas de cada parâmetro nas 345 reamostras nas três análises intermediárias estão na Tabela 5.9.

TABELA 5.9: Médias das estimativas por MVC dos parâmetros em 345 amostras bootstrap de um painel gerado segundo um MPD de Wooldridge em 3 análises intermediárias.

Parâmetros	Valores Reais	$T = 4$	$T = 7$	$T = 10$
β_1	1,0	1,0456	1,0289	1,0423
β_2	-0,7	-0,7157	-0,7167	-0,7107
γ	1,2	1,2492	1,2177	1,2279
w_0	-0,5	-0,5497	-0,5823	-0,5658
w_1	1,0	1,0320	1,0327	1,0327
w_{21}	0,5	0,5203	0,5095	0,5238
w_{22}	0,5	0,4990	0,5769	0,5312
θ	$\sqrt{2}$	0,9268	0,8830	0,9366

Os resultados apresentados na Tabela 5.9 para as 3 análises intermediárias estão bem próximos dos verdadeiros valores.

A seguir serão apresentadas as matrizes de covariância empíricas dos estimadores de cada coeficiente nas 3 análises intermediárias.

1. Estimativa da matriz de covariância dos estimadores do coeficiente de X_1 em três análises.

$$\widehat{\text{Cov}}(\hat{\beta}_1) = \begin{bmatrix} 0,01776745 & 0,001459321 & -9,534828e^{-05} \\ 0,001459321 & 0,034935708 & 0,001798766 \\ -9,534828e^{-05} & 0,001798766 & 0,02391871 \end{bmatrix},$$

2. Estimativa da matriz de covariância dos estimadores do coeficiente de X_2 em três análises.

$$\widehat{\text{Cov}}(\hat{\beta}_2) = \begin{bmatrix} 0,038516193 & 0,006862199 & 0,005218810 \\ 0,006862199 & 0,055977855 & 0,003613077 \\ 0,005218810 & 0,003613077 & 0,036931599 \end{bmatrix},$$

3. Estimativa da matriz de covariância dos estimadores do coeficiente de $Y_{i,t-1}$ em três análises.

$$\widehat{\text{Cov}}(\hat{\gamma}) = \begin{bmatrix} 0,066111039 & 0,005534907 & -0,001120226 \\ 0,005534907 & 0,053344967 & -0,001928603 \\ -0,001120226 & -0,001928603 & 0,052077510 \end{bmatrix},$$

4. Estimativa da matriz de covariância dos estimadores do Intercepto em três análises.

$$\widehat{\text{Cov}}(\hat{w}_0) = \begin{bmatrix} 0,14754814 & 0,04116892 & 0,03128483 \\ 0,04116892 & 0,14322801 & 0,02988888 \\ 0,03128483 & 0,02988888 & 0,14824947 \end{bmatrix},$$

5. Estimativa da matriz de covariância dos estimadores do coeficiente de Y_0 em três análises.

$$\widehat{\text{Cov}}(\hat{w}_1) = \begin{bmatrix} 0,084862709 & 0,01652189 & 0,008405749 \\ 0,016521888 & 0,11751074 & 0,026383118 \\ 0,008405749 & 0,02638312 & 0,090033945 \end{bmatrix},$$

6. Estimativa da matriz de covariância dos estimadores do coeficiente de \bar{x}_1 em três análises.

$$\widehat{\text{Cov}}(\hat{w}_{21}) = \begin{bmatrix} 0,0126068643 & 0,0003092687 & 0,001092491 \\ 0,0003092687 & 0,0178201333 & 0,001799488 \\ 0,0010924912 & 0,0017994879 & 0,014660628 \end{bmatrix},$$

7. Estimativa da matriz de covariância dos estimadores do coeficiente de \bar{x}_2 em três análises.

$$\widehat{\text{Cov}}(\hat{w}_{22}) = \begin{bmatrix} 0,3862544 & 0,11442409 & 0,09293150 \\ 0,1144241 & 0,40588280 & 0,08998835 \\ 0,0929315 & 0,08998835 & 0,45667169 \end{bmatrix},$$

8. Estimativa da matriz de covariância dos estimadores de $\theta = \sigma_a\sqrt{2}$ em três análises.

$$\widehat{\text{Cov}}(\hat{\theta}) = \begin{bmatrix} 0,05843021 & 0,002065360 & 2,642168e^{-05} \\ 0,002065360 & 0,092892544 & 0,005155809 \\ 2,642168e^{-05} & 0,005155809 & 0,07841238 \end{bmatrix}.$$

Para a conclusão deste estudo, pretende-se examinar se a matriz de covariância dos estimadores de um especificado parâmetro em k análises tem estrutura conhecida. Especialmente, deseja-se verificar se essa estrutura é a de “acréscimos independentes” em seqüências de somas $S_m = U_1 + \dots + U_m$, $m = 1, \dots, R$, onde as variáveis U_i , $i = 1, \dots, N$, são independentes e identicamente distribuídas. A razão por trás disso é que a presença de tal estrutura de covariância, garante a possibilidade de fazer testes de hipóteses sobre um dos parâmetros de regressão ao longo da coleta de dados, sob um nível de significância pré-fixado α . Isso permite que a tomada de decisões a respeito dos parâmetros possa ocorrer antes do encerramento da coleta dos dados de todos os dados do painel.

Ressalta-se que este tipo de análise sequencial é muito utilizado na área de Ensaio Clínico de Medicamentos: grupos de N indivíduos são observados ao longo do tempo quanto a uma variável resposta Y_{it} e diversas variáveis explicativas. Nesta área de aplicação, usualmente deseja-se rejeitar H_0 , se for esse o caso, ao menor período de tempo, por razões éticas, econômicas etc. Foi desenvolvido então a técnica de testes sequenciais em grupo, que permite construir as regiões de rejeição de H_0 em cada momento da análise, de forma que no teste global o nível de significância seja igual a um determinado valor α . Pretendemos aperfeiçoar a metodologia usada de amostragem *bootstrap*, de modo a obter os resultados em tempo menor, tendo em vista que é necessário aumentar o número de amostras *bootstraps*, e também modificar o número de análises e parâmetros.

Capítulo 6

Considerações Finais

Neste trabalho, o objetivo foi o estudo de modelos de regressão dinâmicos com variável resposta binária para dados em painel. Para um melhor entendimento do assunto, primeiramente foi feito no capítulo 2 um estudo sobre os principais resultados acerca dos modelos de regressão estáticos com efeitos aleatórios para variáveis respostas contínuas. Em seguida, no capítulo 3, foram introduzidos os modelos de regressão de resposta binária, tanto para observações independentes das variáveis em um corte transversal da população como para observações em um acompanhamento longitudinal de um corte transversal da população. Neste último caso, foram abordados os modelos de regressão estáticos de resposta binária para dados em painel, com ênfase nos modelos Logito e Probit.

Finalmente, no capítulo 4, foi feita a apresentação dos modelos de regressão dinâmicos de resposta binária para dados em painel. Nesse capítulo foram abordados os principais resultados desenvolvidos nos últimos anos acerca deste assunto. A primeira proposta estudada foi a de Honoré e Kyriazidou (2000) que desenvolve um modelo Logito dinâmico condicional de efeitos fixos. Para a estimação dos parâmetros desse modelo os referidos autores utilizaram uma verossimilhança condicional que é livre dos parâmetros incidentais individuais α_i . Para que esses efeitos individuais α_i fossem eliminados da verossimilhança foi necessário fazer um condicionamento no evento que é a união das observações de indivíduos que apresentam mudança na variável resposta do tempo 1 para o tempo 2 e igualdade nas covariáveis nos tempos 2 e 3. A estimação a partir dessa verossimilhança condicional é um pouco restritiva na prática: no campo da economia, por exemplo, dificilmente se observa no painel um número razoável de casos

onde $\mathbf{x}_{i2} = \mathbf{x}_{i3}$. Por conta disso Honoré e Kyriazidou sugerem uma verossimilhança condicional modificada: a verossimilhança relativa a um indivíduo do painel é multiplicada por uma função “kernel” que dá maiores pesos às observações individuais em que as diferenças $\mathbf{x}_{i2} - \mathbf{x}_{i3}$, estejam mais próximas do zero. Por conta de todas essas restrições práticas não foi desenvolvida nesta dissertação nenhuma adaptação computacional para o ajuste desse modelo mas é de interesse desenvolver algo no futuro para a solução deste problema.

Ainda no capítulo 4 é feita a apresentação de um Modelo Probit Dinâmico de efeitos aleatórios com erros autocorrelacionados. Nesta subseção foi mostrada a abordagem dada por Stewart (2006) para o ajuste de um MPD de Heckman quando os erros do modelo são serialmente autocorrelacionados. Stewart criou um programa em Stata, o *redpace*, que é baseado no algoritmo de simulação GHK para o cálculo de integrais de uma densidade normal multivariada em regiões cujos limites, para a k -ésima componente do vetor aleatório de dimensão $n > k > 1$, são dados recursivamente em termos da $(k - 1)$ -ésima componente. Stewart propôs a estimação dos parâmetros do MPD por Máxima Verossimilhança Simulada, que consiste em otimizar a função de verossimilhança do MPD usando o simulador GHK. O uso desse programa requer, no entanto, um esforço computacional muito grande, pois para ajustar um conjunto de dados de painel com $N = 799$ e $T = 6$ Stewart (2006) reporta que foram necessárias 15 horas de execução do programa até a convergência.

São ainda apresentados no capítulo 4, dois MPD de efeitos aleatórios com erros não correlacionados: o mais tradicional, que se baseia em uma proposta de Heckman (1981) para $IP(Y_{i0} = 1|\alpha_i)$ e que possibilita uma estimação dos parâmetros por máxima verossimilhança e um recentemente proposto por Wooldridge que não faz uso da modelagem das condições iniciais para estimação dos coeficientes de um modelo dinâmico, mas possibilita a consideração de uma verossimilhança condicional em y_{i0} .

Após o levantamento da literatura recente sobre o tema proposto, ficou constatada a ausência de programas em R para o ajuste de vários dos modelos de regressão de resposta binária. Por esse motivo, foi redigido e testado um programa para o ajuste do Modelo Probit Estático (MPE) de efeitos aleatórios. A partir desse programa, ficou possível o ajuste de um modelo probito dinâmico de Wooldridge(2005), indicado por MPDW, simplesmente acrescentando ao conjunto de covariáveis, a variável resposta defasada, um intercepto e o vetor de médias em todos os tempos das covariáveis.

A seguir, foi feito um estudo por simulação de propriedades dos estimadores dos parâmetros de um MPDW a partir do artifício descrito acima. Foi feito também um estudo por simulação de propriedades dos estimadores dos parâmetros de um modelo probito dinâmico de Heckman com erros não correlacionados, indicado por MPDH, a partir do artifício proposto por Wooldridge. Para realizar esta tarefa foram gerados no R vários conjuntos de dados em painel com as especificações do MPDW, no primeiro caso, e do MPDH no segundo caso. O ajuste desses modelos aos dados foi feito usando o programa em R desenvolvido para o MPDW. Os resultados obtidos, apresentados no capítulo 5, indicam que o vício do estimador de Wooldridge para cada parâmetro destes modelos é pequeno, assim como o erro quadrático médio, para $N = 600$. Isso justifica em parte, a análise de um painel de dados reais usando o ajuste sugerido por Wooldridge.

Foi ainda abordada a questão do comportamento dos estimadores de um MPDW com dimensão de tempo T , ao longo de K análises intermediárias feitas em tempos $T_1 < T_2 < \dots < T_K$. Este estudo foi realizado usando a técnica de bootstrap, reamostrando com reposição um painel simulado segundo um MPDW com $N = 200$ indivíduos e $T = 3$ tempos, estimando os parâmetros dos painéis reamostrados nos tempos $T_1 = 4 < T_2 = 7 < T_3 = 10$. Este estudo mostrou que as estimativas dos parâmetros estão, em média, bastante próximos dos valores escolhidos para os parâmetros. O estudo prosseguirá no futuro, avaliando se poderão ser usadas as técnicas de testes sequenciais em grupo para análises intermediárias de um painel, sob a hipótese de um MPDW. Essa avaliação decorrerá de uma análise das matrizes de covariâncias dos estimadores e requer um prosseguimento do trabalho, o que se pretende fazer no futuro.

Referências Bibliográficas

- [1] CHAMBERLAIN, G. “Analysis of covariance with Qualitative Data”. Review of Economic Studies, **47**, p. 225-238, 1980.
- [2] CHAY, K.Y.; HYSLOP, D. “Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence using Alternative Approaches”. *Working paper*, UC Berkeley, 2000.
- [3] HSIAO, C. “Analysis of Panel Data”. Cambridge: Cambridge University Press, 1986. 246p.
- [4] DARGAY, J.M.; VYTHOULKAS, P.C. “Estimation of a dynamic Car ownership model for France using a pseudo-panel approach.” Journal of Transport Economics and Policy. **33** 3, p. 238-302, 1999.
- [5] GREENE, W. H. “Econometric analysis”. 5.ed. New Jersey: Prentice-Hall, 2003. 1026p.
- [6] FULLER, W.A; BATTESE, G.E. “Estimation of linear models with crossed-error structure”, Journal of Econometrics, **2**, p. 67-78, 1974.
- [7] HECKMAN, J.J. “The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process,” in C.F. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, 1981, p.179-195. MIT Press: Cambridge,MA.
- [8] HONORÉ, BO E.; KYRIAZIDOU, E. “Panel Data Discrete Choice Models with Lagged Dependent Variables”, *Econometrica*, **68**, N°4 - p. 839-874, 2000.
- [9] JUDGE, G. G. ...et al. “Introduction to the practice of econometrics”. 2 ed. New York: Wiley, 1988. 1024 p.
- [10] KEANE, M. “A computationally practical simulation estimator for panel data”, *Econometrica*, **62**, p.95-116, 1994.
- [11] MARQUES, L.D. “Modelos dinâmicos com Dados em Painel: revisão de literatura”. Disponível em: <<http://www.fep.up.pt/investigacao/workingpapers/wp100.PDF>>. Acesso em: 09 abr. 2006.
- [12] PETERS, B. “Persistence of Innovation: Stylised Facts and Panel Data Evidence ”. *Discussion Paper N° 05-81*, 2005. Disponível em: <<http://opus.zbw-kiel.de/volltexte/2006/4564/pdf/dp0581.pdf>> . Acesso em: 07 set. 2007.

-
- [13] SANTOS, M.J; KASSOUF, A.L. “Uma Investigação Econômica da Influência do Mercado de Drogas Ilícitas Sobre a Criminalidade Brasileira”. *Economia, Brasília(DF)*, n°2, p.187-210, 2007.
- [14] SOBREIRA, R.; CAMPOS, B.C. “Avaliação dos Resultados do FUNDEF” <http://www.boletimdeconjuntura.ufpr.br/Textosdiscussao_pdf/Texto%2026%202005.pdf>. Acesso em: 06 fev. 2008.
- [15] STEWART, M.B. “Maximum Simulated Likelihood Estimation of Random Effects Dynamic Probit Models with Autocorrelated Errors”, *The Stata Journal*, **6**, p.256-272, 2006.
- [16] WOOLDRIDGE, J.M. “Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity”, *Journal of Applied Econometrics*, **20**, p.39-54, 2005.

Apêndice A

Matrizes Subdivididas

Suponhamos que uma matriz $p \times p$ A possa ser partida em blocos, isto é, que A possa ser escrita da seguinte maneira:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

sendo A_{11} e A_{22} matrizes quadradas de ordem k e $(p - k)$, respectivamente. Então se A for inversível, a decomposição em blocos de A^{-1} pode ser obtida fazendo uma partição de A^{-1} . Seja a inversa de A representada pela seguinte matriz:

$$A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}$$

onde:

$$A^{11} = [A_{11} - A_{12}A_{22}^{-1}A_{21}]^{-1}$$

$$A^{21} = -A_{22}^{-1}A_{21}A^{11}$$

$$A^{12} = -A^{11}A_{12}A_{22}^{-1}$$

$$A^{22} = A_{22}^{-1} - A^{21}A_{12}A_{22}^{-1}$$

Estes resultados só podem ser utilizados se A for quadrada, todas as matrizes da diagonal principal devem ser quadradas e por último A_{11} e $(A_{22} - A_{21}A_{11}^{-1}A_{12})$ devem ser não-singulares(invertíveis).

Para obter os resultados acima é necessário multiplicar simbolicamente A por A^{-1} , e forçar o produto ser igual a $I_{p \times p}$, com I = matriz identidade.

Escrevendo as quatro condições para que ocorra $A \times A^{-1} = I_{p \times p}$

$$1) A_{11} \times A^{11} + A_{12} \times A^{21} = I_{\{k \times k\}}$$

- 2) $A_{11} \times A^{12} + A_{12} \times A^{22} = 0_{\{k \times (p-k)\}}$
 3) $A_{21} \times A^{11} + A_{22} \times A^{21} = 0_{\{(p-k) \times k\}}$
 4) $A_{21} \times A^{12} + A_{22} \times A^{22} = I_{\{(p-k) \times (p-k)\}}$

Da equação 3) obtém-se:

$$A^{21} = -(A_{22})^{-1} \times A_{21} \times A^{11}$$

Substituindo na equação 1), temos:

$$1) A_{11} \times A^{11} + A_{12} \times [-(A_{22})^{-1} \times A_{21} \times A^{11}] = I_{\{k \times k\}}$$

Isto é o mesmo que:

$$[A_{11} - A_{12} \times (A_{22})^{-1} \times A_{21}] A^{11} = I_{\{k \times k\}}$$

Desta última igualdade decorre que:

$$A^{11} = [A_{11} - A_{12} \times (A_{22})^{-1} \times A_{21}]^{-1}$$

Desta forma já foi encontrado A^{11} e A^{21} escritos em termos de blocos da matriz A . Para obter os outros dois blocos de A^{-1} , pode-se repetir o artifício acima usado, forçando o produto de $A^{-1} \times A$ ser igual a $I_{\{(n-p) \times (n-p)\}}$, resultando nas seguintes equações:

- 1*) $A^{11} \times A_{11} + A^{12} \times A_{21} = I_{\{k \times k\}}$
 2*) $A^{11} \times A_{12} + A^{12} \times A_{22} = 0_{\{k \times (p-k)\}}$
 3*) $A^{21} \times A_{11} + A^{22} \times A_{21} = 0_{\{(p-k) \times k\}}$
 4*) $A^{21} \times A_{12} + A^{22} \times A_{22} = I_{\{(p-k) \times (p-k)\}}$

Da equação 2*), obtém-se:

$$A^{12} = -A^{11} \times A_{12} (A_{22})^{-1}$$

Falta apenas A^{22}

Da equação 4*) obtém-se:

$$A^{22} A_{22} = I_{(p-k) \times (p-k)} - A^{21} \times A_{12}$$

$$A^{22} = A_{22}^{-1} - A^{21} \times A_{12} (A_{22}^{-1})$$

Mas substituindo A_{21} , temos:

$$A^{22} = A_{22}^{-1} + [(A_{22}^{-1}) A_{21} \times A^{11}] A_{12} (A_{22}^{-1})$$

Em resumo, obtivemos:

$$A^{-1} = \begin{bmatrix} A^{11} & -A^{11} \times A_{12} \times A_{22}^{-1} \\ -A_{22}^{-1} A_{21} A_{11} & A_{22}^{-1} + [(A_{22}^{-1}) A_{21} \times A^{11}] A_{12} (A_{22}^{-1}) \end{bmatrix}$$

onde:

$$A^{11} = [A_{11} - A_{12} \times (A_{22}^{-1}) A_{21}]^{-1}$$

Esta é apenas uma das diversas representações que podem ser obtidas para A^{-1} .

Apêndice B

Programas

B.1 Programa para o Ajuste de um MPE

Estimação dos parâmetros β_1 , β_2 e θ de um modelo probito estático de efeitos aleatórios por máxima verossimilhança utilizando Quadratura de Gauss-Hermite com 12 pontos:

```
#-----  
#Carregando pacote:  
library(statmod)  
#-----  
#G2 é uma função que calcula para uma pessoa i, em todos os tempos,  
#os diversos valores de probabilidades acumuladas segundo o  
#parâmetro a[k]-(nós da quadratura)  
#-----  
#out1 é o vetor de pesos e nós da quadratura;  
#-----  
G2=function(beta,Y,x1,x2,a,i,theta)  
{  
  out1<- gauss.quad.prob(12,"normal")  
  theta = sqrt(2)  
  p=length(a)  
  T=ncol(Y)
```

```

N=nrow(x1)
U=cbind(c(rep(1,p)))
for(k in 1:p)
{
for (j in 1:T)
{
#-----
#Função de verossimilhança:
U[k]=U[k]*(1 - pnorm((x1[i,j]*beta[1]+x2[i,j]*beta[2]+
sqrt(2)*a[k]),0,1))^(Y[i,j])*(pnorm((x1[i,j]*beta[1]+
x2[i,j]*beta[2]+ sqrt(2)*a[k]),0,1))^(1-Y[i,j])}
}
Lik = out1$weights %*%U
Lik
}
#-----
#A função a seguir exhibe valores de G2:
ExibG=function(beta,Y,x1,x2,a,K)
{
Liky=(G2(beta,Y,x1,x2,a,1, 1))
cat(Liky, "\n")
for (i in 2:K)
{
Liky=( G2(beta,Y,x1,x2,a,i, 1))
cat(Liky, "\n")
}
Liky = -Liky
Liky
}
#-----
#Função que calcula a log-verossimilhança para todos os indivíduos;
Max3 = function(beta,Y,x1,x2,a,K)
{
Liky=0

```

```

    if(G2(beta,Y,x1,x2,a,1,1)>0)
Liky=log(G2(beta,Y,x1,x2,a,1,1))
for (i in 2:K)
{
if(G2(beta,Y,x1,x2,a,i,1)>0)
{
Liky = Liky+log(G2(beta,Y,x1,x2,a,i,1))
}
else if((G2(beta,Y,x1,x2,a,i,1)==0))
{
Liky = - Liky
}
}
Liky
}
#-----
#Função para estimar os parâmetros;
optim(cbind(c(2,1)), Max3, NULL, method = "BFGS",
hessian = TRUE, Y=y,x1=X1,x2=X2,a=cbind(out1$nodes),K=N)
#-----

```

B.2 Programa para o Ajuste de um MPD de Wooldridge

Estimação dos parâmetros β_1 , β_2 , γ , y_0 , w_{21} , w_{22} , w_0 , θ , σ_a e ρ de um modelo probito dinâmico de efeitos aleatórios segundo Wooldridge por máxima verossimilhança utilizando Quadratura de Gauss-Hermite:

```

#-----
#Carregando pacote:
library(statmod)
#-----
#G2 é uma função que calcula para uma pessoa i, em todos os tempos,
#os diversos valores de probabilidades acumuladas segundo o

```

```

#parâmetro a[k]-(nós da quadratura)
#-----
#out1 é o vetor de pesos e nós da quadratura;
#-----
G2=function(delta,y,X1,X2,X1m,X2m,y0,h,i)
{
  out1<- gauss.quad.prob(h,"normal")
  a=cbind(out1$nodes)
  p = length(a)
  T=ncol(y)
  N=nrow(X1)
  U=cbind(c(rep(1,p)))
  #-----
  #Cálculo de alpha:
  alpha[i] = delta[4] + y0*delta[5] + X1m[i]*delta[6] + X2m[i]*delta[7]
  #-----
  #Função de verossimilhança:
  for(k in 1:p)
  {
    U[k] = pnorm((2*y[i,1] -1)*(X1[i,1]*delta[1]+X2[i,1]*delta[2]
    + y0*delta[3] + alpha[i]+ delta[8]*a[k]),0,1)
    for (t in 2:T)
    {
      U[k]=U[k]*pnorm((2*y[i,t] -1)*(X1[i,t]*delta[1]+ X2[i,t]*delta[2]
      + y[i,(t-1)]*delta[3] + alpha[i] + delta[8]*a[k]),0,1)
    }
  }
  Lik = out1$weights%%U
  Lik
}
#-----
#Função de log-verossimilhança:
MaX5 = function(delta, y0, y ,X1,X2,X1m,X2m,h,K)
{

```

```

Liky=0
for (i in 1:K)
{
if(G2(delta,y,X1,X2,X1m,X2m,y0,h,i)!=0)
{
Liky =Liky+log(G2(delta,y,X1,X2,X1m,X2m,y0,h,i))
}
}
Liky
}
#-----
#Função para estimar os parâmetros;
optim(cbind(c(1.05,-0.68,1.25,-0.55,1.10,0.48,0.45,1.21)),
MaX5, NULL, method = "BFGS", hessian = TRUE, control=list(fnscale=-1),
y=y,X1=X1,X2=X2, X1m= X1m, X2m= X2m, y0=y0, h=12,K=N)

```

B.3 Programa para gerar um painel de dados segundo o MPD de Wooldridge

Este programa gera os dados segundo um modelo probito dinâmico de efeitos aleatórios com erros não-correlacionados de Wooldridge:

```

rm(list=ls(all=TRUE)) # removendo algoritmos anteriores;
set.seed(214) #Semente;
N = 500 # N° de elementos;
T = 6 #Tempos observados;
#-----
#Coeficientes
beta1 = 1 ; beta2 = -0.7 ; gama = 1.2
w1 = 1 ; w0 = -0.5 ; w21 = 0.5 ; w22 = 0.5
#-----
#covariáveis
options("digits"=2) #N° de dígitos após a vírgula.
X1 = matrix(rnorm(N*T,1,4),N,T)

```

```

X2 = matrix(rbinom(N*T,1,0.5),N,T)
X1m = numeric()
X2m = numeric()
#-----
alpha = numeric()
y.novo = matrix(0,N,T)
y = matrix(0,N,T)
Y0 = matrix(0,N*T,1)
#-----
#Gerando as variáveis:
for (i in 1:N){
a = rnorm(1, mean=0, sd = 1)
u = rnorm(T, mean=0,sd = 1)
y0 = rbinom(1,1,0.5)
Y0[((i-1)*T +1):(i*T),1]= y0
#-----
#média das covariáveis
X1m[i] = mean(X1[i,])
X2m[i] = mean(X2[i,])
#-----
alpha[i] = w0 + w1*y0 +w21*X1m[i] + w22*X2m[i] + a
#-----
#Cálculo da variável latente:
y.novo[i,1] = beta1*X1[i,1] + beta2*X2[i,1] +
gama*y0 + alpha[i] + u[1]
if (y.novo[i,1]>0) y[i,1] = 1 else y[i,1] = 0
for (t in 2:T){
y.novo[i,t] = beta1*X1[i,t] + + beta2*X2[i,t] +
gama*y[i,t-1] + alpha[i] + u[t]
#-----
#Cálculo da variável resposta:
if (y.novo[i,t]>0) y[i,t] = 1 else y[i,t] = 0
}
}

```

```

#-----
#Montando a matriz de dados para o ajuste:
d1 = matrix(1:T,T*N,1)
d2 = matrix(t(matrix(1:N,N,T)),T*N,1)
d3 = matrix(t(y),T*N,1)
d4=Y0
d5 = matrix(t(X1),T*N,1)
d6 = matrix(t(X2),T*N,1)
d7 = matrix(t(matrix(X1m,N,T)),T*N,1)
d8 = matrix(t(matrix(X2m,N,T)),T*N,1)
d9 = matrix(0,T*N,1)

for (i in 1:(T*N)) if (d1[i] == 1) d9[i] = d4[i] else d9[i] = d3[i-1]

dados = cbind(d1,d2,d3,d4,d5,d6,d7,d8,d9)
colnames(dados)=c("year","firm","y" ,"y0" ,"x1" ,
"x2", "X1m","X2m", "ylag")
mean(y)
mean(Y0)
res = c(w1, gama, beta1, beta2, w21, w22, w0)

```

B.4 Programa para gerar um painel de dados segundo o MPD de Heckman

```

rm(list=ls(all=TRUE)) # removendo algoritmos anteriores;
set.seed(123)#Semente;
N = 600 #N° de elementos;
T = 6 # Tempos observados;
#-----
# Coeficientes
#-----
beta1 = 1 ; beta2 = -0.7 ; gama = 1.2 ;
w1 = 1 ; w0 = -0.5 ; w21 = 0.5;

```

```
w22 = 0.5 ; pi1 = 0.6 ; pi2 = 0.8
#-----
#Geração das variáveis explanatórias:
#-----
options("digits"=2)
X1 = matrix(rnorm(N*T,1,4),N,T)
X2 = matrix(rbinom(N*T,1,0.5),N,T)
x0 = matrix(rbinom(N,1,0.7),N,1)
#-----
X1m = numeric()
X2m = numeric()
alpha = numeric()
y.novo = matrix(0,N,T)
y0.novo = matrix(0,N,1)
y = matrix(0,N,T)
y0 = matrix(0,N,1)
#-----
for (i in 1:N){
a = rnorm(1, mean=0, sd = 1)
u = rnorm(T, mean=0,sd = 1)
u0 = rnorm(N, mean=0,sd = 1)
#-----
#Geração da variável resposta Y0:
#-----
y0.novo[i] = x0[i]*pi1 + a*pi2 + u0[i]
if (y0.novo[i]>0) y0[i] = 1 else y0[i] = 0
#-----
#Vetor de médias das variáveis explicativas:
#-----
X1m[i] = mean(X1[i,])
X2m[i] = mean(X2[i,])
#-----
#Geração da variável resposta Y:
#-----
```

```

y.novo[i,1] = beta1*X1[i,1] + beta2*X2[i,1] + gama*y0[i] + a + u[1]
if (y.novo[i,1]>0) y[i,1] = 1 else y[i,1] = 0

for (t in 2:T){
y.novo[i,t] = beta1*X1[i,t] + + beta2*X2[i,t] + gama*y[i,(t-1)] + a + u[t]
if (y.novo[i,t]>0) y[i,t] = 1 else y[i,t] = 0
}}

#-----
#Criando a matriz de dados:
#-----
d1 = matrix(1:T,T*N,1)
d2 = matrix(t(matrix(1:N,N,T)),T*N,1)
d3 = matrix(t(y),T*N,1)
d4 = matrix(t(y0),T*N,1)
d5 = matrix(t(X1),T*N,1)
d6 = matrix(t(X2),T*N,1)
d7 = matrix(t(matrix(X1m,N,T)),T*N,1)
d8 = matrix(t(matrix(X2m,N,T)),T*N,1)
d9 = matrix(0,T*N,1)

for (i in 1:(T*N)) if (d1[i] == 1) d9[i] = d4[i] else d9[i] = d3[i-1]

dados = cbind(d1,d2,d3,d4,d5,d6,d7,d8,d9)
colnames(dados)=c("year","firm","y" ,"y0" ,"x1" ,
"x2", "X1m","X2m", "ylag")
#-----
mean(y) ; mean(y0)
#-----
#Guardando as matrizes de dados:
require(MASS)
write.matrix(round(x0,2),file="c:/x0.txt",sep = " ")
write.matrix(round(dados,2),file="c:/heckman1.txt",sep = " ")

```

B.5 Programa em STATA para ajustar o MPD de Wooldridge

Abaixo segue o programa em Stata para ajustar o modelo probito dinâmico (MPD) de efeitos aleatórios segundo Wooldridge.

```
clear
set memory 100m
set more off

u c:\quali1.dta
tsset firm year, yearly
xtprobit y y0 ylag x1 x2 x1m x2m
```

B.6 Programa para reamostragem *bootstrap* de uma análise de dados em painel com análises intermediárias

```
#Carregando os pacotes:
library(statmod)
require(bootstrap)
#-----
set.seed(123) # Semente;
options("digits"=4)
N = 200 #Nºde amostras
#-----
# Vetor de tempos:
tv =cbind(c(4,7,10))
k=nrow(tv)
#-----
# Coeficientes:
beta1 = 1 ; beta2 = -0.7 ; gama = 1.2
```

```

w1 = 1 ; w0 = -0.5 ; w21 = 0.5 ; w22 = 0.5
#-----
# Geração das variáveis explicativas:
x1 = matrix(rnorm(N*tv[k],1,4),N, tv[k])
x2 = matrix(rbinom(N*tv[k],1,0.5),N, tv[k])
#-----
# Geração de Y0:
yoo=rbinom(N,1,0.5)
# Erros aleatórios:
uu0= rnorm(N*tv[k],0,1)
# heterogeneidade
a0=rnorm(N,0,1)
#-----
# Bootstrap do uu0 com M replicações
#-----
M=15 # N° de replicações;
R= 145 # Tamanho de cada replicação;
n1<-N*tv[k] # Tamanho do painel;
Bu1<-matrix(0,n1,R)
  for(j in 1:R)
{
  id<-sample(1:n1,n1,replace=T)
  Bu1[,j]<-uu0[id]
}
  #Bu1
#-----
# Bootstrap do a0 (tem um p/ cada indivíduo)
#-----
n2<-N
Ba1<-matrix(0,n2,R)
  for(j in 1:R)
{
  id<-sample(1:n2,n2,replace=T)
  Ba1[,j]<-a0[id] }

```

```

#Ba1
#-----
# Gerar o painel relativo aos dados originais e estimar os coeficientes.
#-----
veja0=cirl(N,tv,cbind(uu0),x1,x2,yoo,cbind(a0),beta1,
beta2,gama,w1,w0,w21,w22)
d1=veja0[[1]] [[1]] [[1]]
d2=veja0 [[1]] [[2]]
d3=veja0 [[2]]
#-----
# Estimação dos coeficientes para o painel até o tempo T=4:
#-----
d=d1
Op1=optim(cbind(c(1.05,-0.68,1.25,-0.55, 1.05,0.51,0.48,1.31)),
MaX5, NULL, method = "BFGS", hessian = TRUE, control=list
(fnscale=-1), y=d[,3], X1=d[,5],X2=d[,6], X3=d[,9], X5=d[,4],
X6= d[,7], X7= d[,8], T=tv[1], K=N)
#-----
Op01=cbind(Op1$par) # Vetor dos coeficientes estimados;
Oph1=as.matrix(Op1$hessian) # Matriz Hessiana;
#-----
# Estimação dos coeficientes para o painel até o tempo T=7:
#-----
d=d2
Op2=optim(cbind(c(1.05,-0.68,1.25,-0.55, 1.05,0.51,0.48,1.31)),
MaX5, NULL, method = "BFGS", hessian = TRUE, control=list
(fnscale=-1), y=d[,3], X1=d[,5],X2=d[,6], X3=d[,9], X5=d[,4],
X6= d[,7], X7= d[,8], T=tv[2], K=N)
#-----
Op02=cbind(Op2$par) # Vetor dos coeficientes estimados;
Oph2=as.matrix(Op2$hessian) # Matriz Hessiana;
#-----
# Estimação dos coeficientes para o painel até o tempo T=10:
#-----

```

```

d=d3
Op3=optim(cbind(c(1.05,-0.68,1.25,-0.55, 1.05,0.51,0.48,1.31)),
MaX5, NULL, method = "BFGS", hessian = TRUE,
control=list(fnscale=-1), y=d[,3], X1=d[,5],X2=d[,6],
X3=d[,9], X5=d[,4], X6= d[,7], X7= d[,8], T=tv[3], K=N)
#-----
Op03=cbind(Op2$par) # Vetor dos coeficientes estimados;
Oph3=as.matrix(Op2$hessian) # Matriz Hessiana;
#-----
# Vetor dos coeficientes estimados nas 3 análises intermediárias:
Op0=cbind(Op01,Op02,Op03)
# Vetor das matrizes hessianas:
Oph=cbind(Oph1,Oph2,oph3)
#-----
# Guardando os resultados:
require(MASS)
write.matrix(as.matrix(Op0),file="c:/Par0.txt",sep = " ")
write.matrix(as.matrix(Oph), file="c:/Hessiana0.txt",
sep = " ")
#-----
# M: N° de amostras bootstrap em uma rodada do computador
# amostra 0: linhas de 1 a 25;
# amostra 1: linhas da 26 até 40;
# amostra 2: linhas da 41 até 55;
# amostra 3: linhas da 25+ (m-1)*15+1 até 25+m*15;
# ...      ...      ...      ...      ...      ...
# amostra 8: linhas da 25+ (m-1)*15+1 até 25+m*15;
# Para a amostra m, selecionar as colunas
# de 25 + (m-1)*15+1 até 25+m*15
# Para isso, na simulação , escolher o m
# e em seguida, redefinir Ba e Bu ;
#-----
#Definir m
#-----

```

```

Ba = cbind(Ba1[, (25+ (m-1)*15+1) : (25+m*15)])
Bu = cbind(Bu1[, (25+ (m-1)*15+1) : (25+m*15)])
#####
# 2ª PARTE:
#####
# Geração de um painel, de T observações no tempo
Novaff=function(N,T,UU, X1,X2,Yoo,A,beta1,
beta2,gama,w1,w0,w21,w22)
{
#-----
# Matriz de médias:
#-----
X1m = matrix(0,N,T)
X2m = matrix(0,N,T)
#-----
# Heterogeneidade não-observada:
alpha = matrix(0,N,1)
#-----
# Variáveis respostas:
y.novo = matrix(0,N,T)
y = matrix(0,N,T)
Y0 = matrix(0,N*T,1)
#-----
for (i in 1:N)
{
y0 = Yoo[i]
Y0[((i-1)*T +1):(i*T),1]= y0
#-----
# Média das variáveis respostas:
#-----
X1m[i,1:T] = mean(X1[i,1:T])
X2m[i,1:T] = mean(X2[i,1:T])
#-----
alpha[i] = w0 + w1*y0 +w21*X1m[i] + w22*X2m[i] + A[i]

```

```

#-----
# Variável resposta:
y.novo[i,1] = beta1*X1[i,1] + beta2*X2[i,1] + gama*y0 +
alpha[i] + UU[((i-1)*T+1),1]
if (y.novo[i,1]>0) y[i,1] = 1 else y[i,1] = 0
for (t in 2:T){
y.novo[i,t] = beta1*X1[i,t] + + beta2*X2[i,t] + gama*y[i,t-1]
+ alpha[i] + UU[((i-1)*T+t),1]
if (y.novo[i,t]>0) y[i,t] = 1 else y[i,t] = 0
}}
#-----
# Criação de matriz de dados:
d1 = matrix(1:T,T*N,1)
d2 = matrix(t(matrix(1:N,N,T)),T*N,1)
d3 = matrix(t(y),T*N,1)
d4 = Y0
d5 = matrix(t(X1[,1:T]),T*N,1)
d6 = matrix(t(X2[,1:T]),T*N,1)
d7 = matrix(t(matrix(X1m,N,T)),T*N,1)
d8 = matrix(t(matrix(X2m,N,T)),T*N,1)
d9 = matrix(0,T*N,1)
for (i in 1:(T*N)) if (d1[i] == 1) d9[i] = d4[i] else d9[i] = d3[i-1]
dados1= cbind(d1,d2,d3,d4,d5,d6,d7,d8,d9)
dados1 }
#-----
# cir1: segmentação do painel
cir1=function(N,TV,UU,X1,X2,Yoo,A,beta1,
beta2,gama,w1,w0,w21,w22)
{
TV=cbind(TV)
L=list(Novaff(N,(TV[1]),UU, X1,X2,Yoo,A,beta1,
beta2,gama,w1,w0,w21,w22))
for(s in 2:nrow(TV))
L=list(L, Novaff(N,(TV[s]),UU, X1,X2,Yoo,A,beta1,

```

```

beta2,gama,w1,w0,w21,w22))
L
}
#-----
# Análise de dados por MV:
#-----
library(statmod)
G2=function(delta,y,X1,X2,X3,X5,X6, X7,T,i)
{
  out1<- gauss.quad.prob(12,"normal")
  a = cbind(out1$nodes)
  U=cbind(c(rep(1,12)))
for(k in 1:12)
  {
for (t in 1:T)
  {
U[k]=U[k]*pnorm((2*y[T*(i-1)+t] -1)*(X1[T*(i-1)+t]*delta[1]+
X2[T*(i-1)+t] *delta[2]+ X3[T*(i-1)+t]*delta[3]+ delta[4] +
X5[T*(i-1)+t]*delta[5]+ X6[T*(i-1)+t]*delta[6]+
X7[T*(i-1)+t]*delta[7]+ delta[8]*a[k]),0,1)
}}
Lik = out1$weights%%U
Lik
}
# Log-verossimilhança:
MaX5 = function(delta,y,X1,X2,X3,X5,X6, X7,T,K)
{
Liky=0
for (i in 1:K)
{
if(G2(delta,y,X1,X2,X3,X5,X6, X7,T,i)!=0)
{
Liky = Liky + log(G2(delta,y,X1,X2,X3,X5,X6, X7,T,i)) }
}
}

```

```

Liky
}
#-----
# Gerar os painéis relativos a cada bootstrap dos erros u e a.
#(M conjuntos de painéis)
#-----
Par=matrix(0,8,3*M)
for( j in 1:15)
{
veja=cirl(N,tv,cbind(Bu[,j]),x1,x2,yoo,cbind(Ba[,j]),
beta1,beta2,gama,w1,w0,w21,w22)
d1=veja[[1]] [[1]] [[1]]
d2=veja [[1]] [[2]]
d3=veja [[2]]
#-----
# Estimação dos coeficientes:
#-----
# Para o painel até o tempo T=4:
#-----
d=d1
Op1 = optim(cbind(c(1.05,-0.68,1.25,-0.55, 1.05,0.51,0.48,1.31)),
MaX5, NULL, method = "BFGS", hessian = TRUE,
control=list(fnscale=-1), y=d[,3], X1=d[,5],X2=d[,6], X3=d[,9],
X5=d[,4], X6= d[,7], X7= d[,8], T=tv[1], K=N)
#-----
Par[ ,(j-1)*3+1]=cbind(Op1$par) # Estimativas dos coeficientes;
#-----
oph=as.matrix(Op1$hessian) # Matriz Hessiana
#-----
# Para o painel até o tempo T=7:
#-----
d=d2
Op2=optim(cbind(c(1.05,-0.68,1.25,-0.55, 1.05,0.51,0.48,1.31)),
MaX5,NULL, method = "BFGS", hessian = TRUE, control=list(fnscale=-1),

```

```

y=d[,3], X1=d[,5],X2=d[,6], X3=d[,9],X5=d[,4], X6= d[,7], X7= d[,8],
T=tv[2], K=N)
#-----
Par[ ,(j-1)*3+2]=cbind(Op2$par) # Estimativas dos coeficientes;
#-----
# Para o painel até o tempo T=10:
#-----
d=d3
Op3=optim(cbind(c(1.05,-0.68,1.25,-0.55, 1.05,0.51,0.48,1.31)),
Max5, NULL, method = "BFGS", hessian = TRUE,
control=list(fnscale=-1), y=d[,3], X1=d[,5],X2=d[,6], X3=d[,9],
X5=d[,4], X6= d[,7], X7= d[,8], T=tv[3], K=N)
#-----
Par[ ,(j-1)*3+3]=cbind(Op3$par) # Estimativas dos coeficientes
}
#-----
write.matrix(Par,file="c:/Amostra5.txt") # Guardando a amostra;
#-----
#####
# 3ª PARTE
#####
# Para as análises iremos abrir os arquivos que foram guardados na parte 2:
#-----
Am1_120=scan(file="c:\\Amostra1.txt")
Am2_120=scan(file="c:\\Amostra2.txt")
Am3_120=scan(file="c:\\Amostra3.txt")
Am4_120=scan(file="c:\\Amostra4.txt")
Am5_120=scan(file="c:\\Amostra5.txt")
Am6_120=scan(file="c:\\Amostra6.txt")
Am7_120=scan(file="c:\\Amostra7.txt")
Am8_120=scan(file="c:\\Amostra8.txt")
#-----
Am1_225=scan(file="c:\\Amostra1_225.txt")
Am2_225=scan(file="c:\\Amostra2_225.txt")

```

```
Am3_225=scan(file="c:\\Amostra3_225.txt")
Am4_225=scan(file="c:\\Amostra4_225.txt")
Am5_225=scan(file="c:\\Amostra5_225.txt")
Am6_225=scan(file="c:\\Amostra6_225.txt")
Am7_225=scan(file="c:\\Amostra7_225.txt")
Am8_225=scan(file="c:\\Amostra8_225.txt")
Am9_225=scan(file="c:\\Amostra9_225.txt")
Am10_225=scan(file="c:\\Amostra10_225.txt")
Am11_225=scan(file="c:\\Amostra11_225.txt")
Am12_225=scan(file="c:\\Amostra12_225.txt")
Am13_225=scan(file="c:\\Amostra13_225.txt")
Am14_225=scan(file="c:\\Amostra14_225.txt")
Am15_225=scan(file="c:\\Amostra15_225.txt")
#-----
# Transformar em matriz:
#-----
Am1_120=matrix(Am1_120,nrow=8,ncol=45,byrow=TRUE)
Am2_120=matrix(Am2_120,nrow=8,ncol=45,byrow=TRUE)
Am3_120=matrix(Am3_120,nrow=8,ncol=45,byrow=TRUE)
Am4_120=matrix(Am4_120,nrow=8,ncol=45,byrow=TRUE)
Am5_120=matrix(Am5_120,nrow=8,ncol=45,byrow=TRUE)
Am6_120=matrix(Am6_120,nrow=8,ncol=45,byrow=TRUE)
Am7_120=matrix(Am7_120,nrow=8,ncol=45,byrow=TRUE)
Am8_120=matrix(Am8_120,nrow=8,ncol=45,byrow=TRUE)
#-----
Am1_225=matrix(Am1_225,nrow=8,ncol=45,byrow=TRUE)
Am2_225=matrix(Am2_225,nrow=8,ncol=45,byrow=TRUE)
Am3_225=matrix(Am3_225,nrow=8,ncol=45,byrow=TRUE)
Am4_225=matrix(Am4_225,nrow=8,ncol=45,byrow=TRUE)
Am5_225=matrix(Am5_225,nrow=8,ncol=45,byrow=TRUE)
Am6_225=matrix(Am6_225,nrow=8,ncol=45,byrow=TRUE)
Am7_225=matrix(Am7_225,nrow=8,ncol=45,byrow=TRUE)
Am8_225=matrix(Am8_225,nrow=8,ncol=45,byrow=TRUE)
Am9_225=matrix(Am9_225,nrow=8,ncol=45,byrow=TRUE)
```

```
Am10_225=matrix(Am10_225,nrow=8,ncol=45,byrow=TRUE)
Am11_225=matrix(Am11_225,nrow=8,ncol=45,byrow=TRUE)
Am12_225=matrix(Am12_225,nrow=8,ncol=45,byrow=TRUE)
Am13_225=matrix(Am13_225,nrow=8,ncol=45,byrow=TRUE)
Am14_225=matrix(Am14_225,nrow=8,ncol=45,byrow=TRUE)
Am15_225=matrix(Am15_225,nrow=8,ncol=45,byrow=TRUE)
#-----
# Definir uma nova matriz transposta:
#-----
Am1t_120 = t(Am1_120)
Am2t_120 = t(Am2_120)
Am3t_120 = t(Am3_120)
Am4t_120 = t(Am4_120)
Am5t_120 = t(Am5_120)
Am6t_120 = t(Am6_120)
Am7t_120 = t(Am7_120)
Am8t_120 = t(Am8_120)
#-----
Am1t_225 = t(Am1_225)
Am2t_225 = t(Am2_225)
Am3t_225 = t(Am3_225)
Am4t_225 = t(Am4_225)
Am5t_225 = t(Am5_225)
Am6t_225 = t(Am6_225)
Am7t_225 = t(Am7_225)
Am8t_225 = t(Am8_225)
Am9t_225 = t(Am9_225)
Am10t_225 = t(Am10_225)
Am11t_225 = t(Am11_225)
Am12t_225 = t(Am12_225)
Am13t_225 = t(Am13_225)
Am14t_225 = t(Am14_225)
Am15t_225 = t(Am15_225)
#-----
```

```
# Salvar essa matriz no C:
#-----
require(MASS)
write.matrix(round(Am1t_120,4),file="c:/Am1t_120.txt",sep = " ")
write.matrix(round(Am2t_120,4),file="c:/Am2t_120.txt",sep = " ")
write.matrix(round(Am3t_120,4),file="c:/Am3t_120.txt",sep = " ")
write.matrix(round(Am4t_120,4),file="c:/Am4t_120.txt",sep = " ")
write.matrix(round(Am5t_120,4),file="c:/Am5t_120.txt",sep = " ")
write.matrix(round(Am6t_120,4),file="c:/Am6t_120.txt",sep = " ")
write.matrix(round(Am7t_120,4),file="c:/Am7t_120.txt",sep = " ")
write.matrix(round(Am8t_120,4),file="c:/Am8t_120.txt",sep = " ")
#-----
write.matrix(round(Am1t_225,4),file="c:/Am1t_225.txt",sep = " ")
write.matrix(round(Am2t_225,4),file="c:/Am2t_225.txt",sep = " ")
write.matrix(round(Am3t_225,4),file="c:/Am3t_225.txt",sep = " ")
write.matrix(round(Am4t_225,4),file="c:/Am4t_225.txt",sep = " ")
write.matrix(round(Am5t_225,4),file="c:/Am5t_225.txt",sep = " ")
write.matrix(round(Am6t_225,4),file="c:/Am6t_225.txt",sep = " ")
write.matrix(round(Am7t_225,4),file="c:/Am7t_225.txt",sep = " ")
write.matrix(round(Am8t_225,4),file="c:/Am8t_225.txt",sep = " ")
write.matrix(round(Am9t_225,4),file="c:/Am9t_225.txt",sep = " ")
write.matrix(round(Am10t_225,4),file="c:/Am10t_225.txt",sep = " ")
write.matrix(round(Am11t_225,4),file="c:/Am11t_225.txt",sep = " ")
write.matrix(round(Am12t_225,4),file="c:/Am12t_225.txt",sep = " ")
write.matrix(round(Am13t_225,4),file="c:/Am13t_225.txt",sep = " ")
write.matrix(round(Am14t_225,4),file="c:/Am14t_225.txt",sep = " ")
write.matrix(round(Am15t_225,4),file="c:/Am15t_225.txt",sep = " ")
#-----
# Juntando as matrizes acima em uma só:
#-----
# AM tem 360 linhas da 1ª rodada e 675 da segunda = 1035
AM = rbind(AM = rbind(Am1t_120, Am2t_120, Am3t_120, Am4t_120, Am5t_120,
  Am6t_120, Am7t_120, Am8t_120, Am1t_225, Am2t_225, Am3t_225, Am4t_225,
  Am5t_225, Am6t_225, Am7t_225, Am8t_225, Am9t_225, Am10t_225, Am11t_225,
```

```

Am12t_225, Am13t_225, Am14t_225, Am15t_225)
#-----
# Criar uma matriz que junte todas as estimativas de b1:
#-----
# Sem outliers
#-----
b1g = matrix(c(AM[,1]),nrow=345,ncol=3,byrow=TRUE)
#Esta matriz vai ter todas as estimativas de b1 na primeira análise intermediária #
#-----
# Criar uma matriz que junte todas as estimativas de b2:
b2g = matrix(c(AM[,2]),nrow=345,ncol=3,byrow=TRUE)
#-----
# Criar uma matriz que junte todas as estimativas de gama:
gama = matrix(c(AM[,3]),nrow=345,ncol=3,byrow=TRUE)
#-----
# Criar matrizes que junte todas as estimativas dos outros 5 coeficientes:
w0 = matrix(c(AM[,4]),nrow=345,ncol=3,byrow=TRUE)
w1 = matrix(c(AM[,5]),nrow=345,ncol=3,byrow=TRUE)
w21 = matrix(c(AM[,6]),nrow=345,ncol=3,byrow=TRUE)
theta = matrix(c(AM[,7]),nrow=345,ncol=3,byrow=TRUE)
theta = matrix(c(AM[,8]),nrow=345,ncol=3,byrow=TRUE)
##*****
###*****Médias dos coeficientes para cada painel: *****
##*****
b1m = matrix(0,3,1)
for (i in 1:ncol(b1g))
b1m[i]=mean(b1g[,i])
#-----
b2m = matrix(0,3,1)
for (i in 1:ncol(b2g))
b2m[i]=mean(b2g[,i])
#-----
gamaM = matrix(0,3,1)
for (i in 1:ncol(gama))

```

```

gamaM[i]=mean(gama[,i])
#-----
w0M = matrix(0,3,1)
for (i in 1:ncol(w0))
w0M[i]=mean(w0[,i])
#-----
w1M = matrix(0,3,1)
for (i in 1:ncol(w1))
w1M[i]=mean(w1[,i])
#-----
w21M = matrix(0,3,1)
for (i in 1:ncol(w21))
w21M[i]=mean(w21[,i])
#-----
w22M = matrix(0,3,1)
for (i in 1:ncol(w22))
w22M[i]=mean(w22[,i])
#-----
thetaM = matrix(0,3,1)
for (i in 1:ncol(theta))
thetaM[i]=mean(theta[,i])
#-----
# Valores reais
# beta1 = 1, beta2 = -0.7, gama = 1.2, w1 = 1, w0 = -0.5,
w21 = 0.5, w22 = 0.5, theta=1,41
#-----
# Matriz de médias:
medias = cbind(b1m,b2m,gamaM,w0M,w1M,w21M,w22M,thetaM)
colnames(medias)=c("b1Medio","b2Medio","gamaMedio" ,
"w0_Medio" ,"w1_Medio" ,"w21_Medio", "w22_Medio","theta_Medio")
#-----
# Cálculo das Covariâncias:
#-----
# Matriz de covariância:

```

```
#b1:
covb1 <- matrix(0,3,3)
covb1[1,1] <- varb1[1,]
covb1[1,2] <- cov(b1g[,1],b1g[,2])
covb1[1,3] <- cov(b1g[,1],b1g[,3])
covb1[2,1] <- cov(b1g[,2],b1g[,1])
covb1[2,2] <- varb1[2,]
covb1[2,3] <- cov(b1g[,2],b1g[,3])
covb1[3,1] <- cov(b1g[,3],b1g[,1])
covb1[3,2] <- cov(b1g[,3],b1g[,2])
covb1[3,3] <- varb1[3,]

### b2:
covb2 <- matrix(0,3,3)
covb2[1,1] <- varb2[1,]
covb2[1,2] <- cov(b2g[,1],b2g[,2])
covb2[1,3] <- cov(b2g[,1],b2g[,3])
covb2[2,1] <- cov(b2g[,2],b2g[,1])
covb2[2,2] <- varb2[2,]
covb2[2,3] <- cov(b2g[,2],b2g[,3])
covb2[3,1] <- cov(b2g[,3],b2g[,1])
covb2[3,2] <- cov(b2g[,3],b2g[,2])
covb2[3,3] <- varb2[3,]

### gama:
covgama <- matrix(0,3,3)
covgama[1,1] = vargama[1,]
covgama[1,2] = cov(gama[,1],gama[,2])
covgama[1,3] = cov(gama[,1],gama[,3])
covgama[2,1] = cov(gama[,2],gama[,1])
covgama[2,2] = vargama[2,]
covgama[2,3] = cov(gama[,2],gama[,3])
covgama[3,1] = cov(gama[,3],gama[,1])
covgama[3,2] = cov(gama[,3],gama[,2])
covgama[3,3] = vargama[3,]

### w0:
```

```
covw0 <- matrix(0,3,3)
covw0[1,1] = varw0[1,]
covw0[1,2] = cov(w0[,1],w0[,2])
covw0[1,3] = cov(w0[,1],w0[,3])
covw0[2,1] = cov(w0[,2],w0[,1])
covw0[2,2] = varw0[2,]
covw0[2,3] = cov(w0[,2],w0[,3])
covw0[3,1] = cov(w0[,3],w0[,1])
covw0[3,2] = cov(w0[,3],w0[,2])
covw0[3,3] = varw0[3,]
### w1:
covw1 <- matrix(0,3,3)
covw1[1,1] = varw1[1,]
covw1[1,2] = cov(w1[,1],w1[,2])
covw1[1,3] = cov(w1[,1],w1[,3])
covw1[2,1] = cov(w1[,2],w1[,1])
covw1[2,2] = varw1[2,]
covw1[2,3] = cov(w1[,2],w1[,3])
covw1[3,1] = cov(w1[,3],w1[,1])
covw1[3,2] = cov(w1[,3],w1[,2])
covw1[3,3] = varw1[3,]
### w21:
covw21 <- matrix(0,3,3)
covw21[1,1] = varw21[1,]
covw21[1,2] = cov(w21[,1],w21[,2])
covw21[1,3] = cov(w21[,1],w21[,3])
covw21[2,1] = cov(w21[,2],w21[,1])
covw21[2,2] = varw21[2,]
covw21[2,3] = cov(w21[,2],w21[,3])
covw21[3,1] = cov(w21[,3],w21[,1])
covw21[3,2] = cov(w21[,3],w21[,2])
covw21[3,3] = varw21[3,]
### w22:
covw22 <- matrix(0,3,3)
```

```
covw22[1,1] = varw22[1,]
covw22[1,2] = cov(w22[,1],w22[,2])
covw22[1,3] = cov(w22[,1],w22[,3])
covw22[2,1] = cov(w22[,2],w22[,1])
covw22[2,2] = varw22[2,]
covw22[2,3] = cov(w22[,2],w22[,3])
covw22[3,1] = cov(w22[,3],w22[,1])
covw22[3,2] = cov(w22[,3],w22[,2])
covw22[3,3] = varw22[3,]

### theta:
covtheta <- matrix(0,3,3)
covtheta[1,1] = vartheta[1,]
covtheta[1,2] = cov(theta[,1],theta[,2])
covtheta[1,3] = cov(theta[,1],theta[,3])
covtheta[2,1] = cov(theta[,2],theta[,1])
covtheta[2,2] = vartheta[2,]
covtheta[2,3] = cov(theta[,2],theta[,3])
covtheta[3,1] = cov(theta[,3],theta[,1])
covtheta[3,2] = cov(theta[,3],theta[,2])
covtheta[3,3] = vartheta[3,]
```