

**Inferência em um modelo  
estrutural heterocedástico  
com erros nas variáveis**

Pedro Lucas Cortés Olivares

TESE APRESENTADA  
AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
OBTENÇÃO DO TÍTULO  
DE  
DOUTOR EM CIÊNCIAS

Programa: Estatística

Orientador: Prof. Dr. Heleno Bolfarine

Co-orientador: Prof. Dr. Reinaldo Boris Arellano-Valle

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da  
CAPES/CNPq

São Paulo, agosto de 2011

# Inferência em um modelo estrutural heterocedástico com erros nas variáveis

Esta tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa realizada por Pedro Lucas Cortés Olivares em 29/08/2011. O original encontra-se disponível no Instituto de Matemáticas e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Heleno Bolfarine (Presidente) - IME-USP.
- Prof. Dr. Reinaldo Boris Arellano Valle - PUC-Chile.
- Profa. Dra. Lourdes Coral Contreras Montenegro - UFMG.
- Profa. Dra. Rosângela Helena Loschi - UFMG.
- Prof. Dr. Celso Rômulo Barbosa Cabral - UFAM.

## Agradecimentos

Agradeço, em primeiro lugar, a Deus por me dar forças, saúde e serenidade durante toda a trajetória desse trabalho.

À minha esposa Giglia, companheira de todas as jornadas, pelo suporte emocional, incentivo e confiança e às minhas filhas, Elisa e Fernanda, pelos momentos de alegria e força.

Aos meus orientadores, os professores Heleno Bolfarine e Reinaldo Arellano Valle, realmente pelo privilégio que tive de trabalhar sob a orientação deles e pelo respeito, confiança e compreensão durante todo o desenvolvimento desse trabalho.

À professora Pilar Iglesias Zuazola (*In memoriam*), minha boa amiga, eu realmente agradeço tudo o que ela fez para que eu pudesse continuar com meu aperfeiçoamento profissional.

Ao professor Mário de Castro, pelas sugestões e esclarecimentos nos vários encontros que tivemos no começo do desenvolvimento desse trabalho.

Ao professor Dr Guillermo Mondaca Ortiz Decano da Faculdade de Ciências Básicas da Universidade de Antofagasta - Chile, pelo apoio incondicional durante todo o desenvolvimento desse trabalho, a ajuda dele foi fundamental para a conclusão do doutorado.

Aos meus amigos, antigos e novos Mario Rojas, Cristian Villegas, Diego Gallardo, Carlos Montenegro, Alvaro, Monica, Manuel, Fernanda, Nilton, Marcos, German, Elsa, em geral agradeço a todas as pessoas que direta ou indiretamente contribuíram para este trabalho.

Aos membros da banca examinadora pelas valiosas sugestões e comentários sobre minha tese e pesquisas futuras.

---

## Resumo

Neste trabalho estudamos o modelo de regressão estrutural heterocedástico (MREH) com erros nas variáveis. Diferentes situações para as quais o modelo com erros nas variáveis (MEV) estrutural usual passa para um MREH com erros nas variáveis são estudadas. Primeiro consideramos um modelo estrutural normal com erros de medição heterocedásticos, além disso, estudamos o MEV estrutural normal considerando a variável não observada  $x_i$  como heterocedástica (não abordado na literatura), logo estendemos este modelo para o caso com replicações não-balanceadas na variável explanatória (covariável). Também propusemos um método geral para determinar a matriz de variâncias-covariâncias dos diferentes MREH com erros nas variáveis aqui apresentados. Finalmente, apresentamos uma abordagem Bayesiana simples para o modelo estrutural normal com erros nas variáveis com  $x_i$  heterocedástica. Utilizamos técnicas clássicas tais como o método dos momentos (MM) e máxima verossimilhança (MV) para obter estimadores consistentes e suas distribuições limite (ou assintóticas). Estimativas de máxima verossimilhança são obtidas numericamente através do algoritmo EM. A estimação consistente da variância assintótica dos estimadores de máxima verossimilhança também é discutida. Testes estatísticos são propostos para testar hipóteses de interesse. Simulações de Monte Carlo foram realizadas para estudar o desempenho dos estimadores e testes. Um conjunto de dados reais é analisado usando os métodos propostos.

**Palavras-chave:** Heterocedasticidade na covariável, erros de medida, modelo de regressão estrutural.



---

# Abstract

We studied the heteroscedastic structural regression model (HSRM) with measurement errors. Different situations for which the usual structural errors-in-variable model (EVM) changes to a HSRM with measurement errors are studied. First we consider a normal structural model with heteroscedastic measurement errors, in addition, we studied the normal structural EVM, considering the unobserved variable  $x_i$  as heteroscedastic (not discussed in the literature), we have extended this model to the case with unbalanced replication in the independent variable. Also we proposed a general method to determine the variance-covariance matrix of the HSRM with errors in variables presented here. Finally, we present a simple Bayesian approach for the normal structural model with errors in variables with  $x_i$  heteroscedastic. We use traditional techniques such as the method of moments and maximum likelihood to obtain consistent estimators and their limit distributions. Maximum likelihood estimates are computed numerically via the EM algorithm. Consistent estimation of the asymptotic variance of the maximum likelihood estimators is also discussed. Test statistics are proposed for testing hypotheses of interest. We used Monte Carlo simulations to study the performances of the estimators and the hypotheses tests. A real data set was analyzed according to the methods proposed.

**Keywords:** Heteroscedasticity in the covariate, measurement errors, structural regression model.

# Sumário

Lista de Abreviaturas	viii
Lista de Símbolos	ix
Lista de Figuras	xi
Lista de Tabelas	xii
<b>1</b> Introdução	<b>1</b>
1.1 Considerações Iniciais . . . . .	1
1.2 O modelo com erros nas variáveis (MEV) . . . . .	2
1.3 Modelo Geral . . . . .	4
1.4 Organização do trabalho . . . . .	6
1.5 Conjunto de dados reais . . . . .	6
1.5.1 Dados do projeto WHO MONICA . . . . .	6
<b>2</b> Modelo com Erros de Medição Heterocedásticos	<b>8</b>
2.1 Introdução . . . . .	8
2.2 Definição do modelo . . . . .	8
2.3 Estimadores de momentos (EMM) . . . . .	9
2.3.1 EMM para o parâmetro $\theta = (\alpha, \beta, \mu_x, \sigma_x^2)^T$ . . . . .	12
2.3.2 Distribuição assintótica do estimador $\hat{\theta}_{\text{EMM}}$ . . . . .	13
2.4 Teste de hipóteses . . . . .	16

2.5	Estimação de máxima verossimilhança (EMV) . . . . .	17
2.5.1	O algoritmo EM . . . . .	17
2.5.2	Aplicação do algoritmo EM no MREH . . . . .	18
2.5.3	Matriz de covariâncias assintóticas do EMV de $\theta$ . . . . .	23
2.5.4	Teste de hipóteses . . . . .	28
2.6	Estudos de simulação . . . . .	28
<b>3</b>	<b>MEV Estrutural Heterocedástico na Variável não Observável</b>	<b>32</b>
3.1	Introdução . . . . .	32
3.2	Definição do Modelo . . . . .	32
3.3	Estimadores de Momentos . . . . .	33
3.3.1	EMM supondo $\bar{\sigma}_x^2$ conhecido . . . . .	35
3.3.2	EMM supondo $\bar{a}$ ou $a_i = \frac{\sigma_{x_i}^2}{\sigma_u^2}$ , $i = 1, \dots, n$ conhecidas . . . . .	36
3.4	Distribuição Assintótica dos Estimadores de Momentos . . . . .	36
3.4.1	Distribuição assintótica de $\hat{\theta}_{1\text{EMM}}$ supondo $\bar{\sigma}_x^2$ conhecido . . . . .	37
3.4.2	Distribuição assintótica de $\hat{\theta}_{1\text{EMM}}$ supondo $\bar{a}$ conhecido . . . . .	39
3.5	Estimação de Máxima Verossimilhança . . . . .	42
3.5.1	Função de Log-Verossimilhança . . . . .	43
3.5.2	Estimação de Máxima Verossimilhança usando o Algoritmo EM . . . . .	43
3.6	Matriz de covariâncias assintóticas do EMV de $\theta$ . . . . .	46
3.6.1	Matriz de Informação Observada . . . . .	47
3.6.2	Matriz de informação esperada . . . . .	50
3.7	Teste de Hipóteses . . . . .	52
3.8	Avaliação do Ajuste . . . . .	53
3.9	Uma Visão Unificada . . . . .	54
3.9.1	Modelo Geral . . . . .	54
3.9.2	Função de log-verossimilhança . . . . .	55

<i>SUMÁRIO</i>	vi
3.9.3 Equações de Verossimilhança . . . . .	56
3.9.4 Matriz de Informação de Fisher . . . . .	56
3.9.5 Exemplos . . . . .	59
3.10 Estudo de Simulação . . . . .	63
3.11 Aplicação . . . . .	64
<b>4 Replicações Não-Balanceadas na Variável exploratória</b>	<b>69</b>
4.1 Introdução . . . . .	69
4.2 Estimadores alternativos para os parâmetros $\sigma_u^2$ , $\bar{\sigma}_x^2$ e $\sigma_{x_i}^2$ . . . . .	70
4.3 Estimação de máxima verossimilhança . . . . .	78
4.3.1 Máxima verossimilhança . . . . .	78
4.3.2 Equações de verossimilhança . . . . .	79
4.3.3 Algoritmo EM . . . . .	80
4.4 Matriz de covariância assintótica dos EMV . . . . .	84
4.4.1 Informação observada . . . . .	84
<b>5 Enfoque Bayesiano</b>	<b>87</b>
5.1 Introdução . . . . .	87
5.2 O modelo estrutural sob a ótica bayesiana . . . . .	88
5.2.1 Distribuição <i>a posteriori</i> . . . . .	89
5.2.2 Distribuição preditiva . . . . .	90
5.3 O modelo estrutural identificável com particular <i>priori</i> de componentes independentes . . . . .	92
5.4 Distribuição <i>a posteriori</i> . . . . .	93
5.4.1 Distribuições condicionais completas baseados nos dados observados ( $\mathbf{Z}_{\text{obs}}$ )	94
5.5 O modelo estrutural não identificável com particular <i>priori</i> de componentes independentes . . . . .	98
5.6 Distribuição <i>a posteriori</i> . . . . .	98

5.6.1	Distribuições condicionais completas baseados nos dados observados ( $\mathbf{Z}_{\text{obs}}$ )	99
5.6.2	Distribuições condicionais completas <i>a posteriori</i> baseadas nos dados completos ( $\mathbf{W}_c$ ) . . . . .	104
5.7	Simulação . . . . .	107
<b>6</b>	<b>Contribuições e Estudos futuros</b>	<b>113</b>
6.1	Principais contribuições . . . . .	113
6.2	Comentários finais . . . . .	114
6.3	Sugestões para pesquisas futuras . . . . .	115
<b>A</b>	<b>Informação Observada</b>	<b>116</b>
<b>B</b>	<b>Informação de Fisher e Reparametrizações</b>	<b>119</b>
B.1	Reparametrizações . . . . .	119
B.2	Informação de Fisher . . . . .	119
<b>C</b>	<b>Modelagem Bayesiana</b>	<b>121</b>
C.1	Métodos MCMC . . . . .	121
C.2	Programa WinBUGS . . . . .	122
	<b>Referências Bibliográficas</b>	<b>124</b>

---

## Lista de Abreviaturas

MEV	Modelo com erros nas variáveis.
MREH	Modelo de regressão estrutural heterocedástico.
EMM	Estimador de método dos momentos.
EMV	Estimador de máxima verossimilhança.
EM	Esperança e maximização.
iid	variáveis independentes e identicamente distribuídas.
indep.	variáveis independentes.



## Lista de Símbolos

$\mathbf{A}, \mathbf{B}, \dots$	Matrizes
$\mathbf{a}, \mathbf{b}, \dots$	vetores
$\text{tr}(\mathbf{A})$	Traço da matriz $\mathbf{A}$
$\det(\mathbf{A})$	Determinante da matriz $\mathbf{A}$
$\text{vec}(\cdot)$	Operador <i>vec</i>
$\text{vech}(\cdot)$	Operador <i>vech</i>
$\mathbf{D}_n$	Matriz de duplicação
$\boldsymbol{\theta}$	Vetor de parâmetros desconhecidos
$\hat{\boldsymbol{\theta}}$	Estimador do vetor de parâmetros desconhecidos
$\alpha, \beta, \dots$	parâmetros desconhecidos
$\hat{\alpha}, \hat{\beta}, \dots$	Estimadores dos parâmetros desconhecidos.
$\xrightarrow{D}$	Convergência em distribuição
$\xrightarrow{q.c.}$	Convergência quase certa
$L(\cdot)$	Função de verossimilhança
$l(\cdot)$	O logaritmo da função de verossimilhança. Também chamaremos de função de log-verossimilhança
$l_c(\cdot)$	O logaritmo da função de verossimilhança completa (dados observados + dados não observados). Também chamaremos de função de log-verossimilhança completa
$d(\cdot)$	Operador diferencial



---

## Lista de Figuras

1.1	Dados do projeto WHO MONICA . . . . .	7
2.1	Algoritmo EM no MREH, quando $(\sigma_{e_i}^2, \sigma_{u_i}^2)$ , $i = 1, \dots, n$ são conhecidas . . . . .	22
3.1	Algoritmo EM para modelo (2.12) . . . . .	46
3.2	Linhas estimadas de regressão para os dados do projeto WHO MONICA . . . . .	67
3.3	Gráfico QQ-plot e Envelope para os Homens . . . . .	68
3.4	QQ-plot e Envelope das mulheres . . . . .	68
4.1	Algoritmo EM . . . . .	83
5.1	Programa WinBUGS para o modelo definido em (5.1) e (5.2) . . . . .	109
5.2	"Trace" para os parâmetros, baseados em uma única seqüência de Gibbs de comprimento $T=30,000$ . . . . .	110
5.3	Comportamento histórico das cadeias para os parâmetros, baseados em uma única seqüência de Gibbs de comprimento $T=30,000$ . . . . .	110
5.4	Gráficos das funções de auto-correlação dos parâmetros, baseados em uma única seqüência de Gibbs de comprimento $T=30,000$ . . . . .	111
5.5	Cadeias paralelas para os parâmetros, baseados em uma única seqüência de Gibbs de comprimento $T=10,000$ . . . . .	111
5.6	Gráficos das estimativas das densidades marginais <i>a posteriori</i> para os parâmetros, baseados em uma única seqüência de Gibbs de comprimento $T=10,000$ . . . . .	112

## Lista de Tabelas

2.1	Estimativas (erros-padrão) dos parâmetros de regressão do modelo estrutural com erros de medição heterocedásticos, no caso $\sqrt{\sigma_{e_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 4)$ e $\sqrt{\sigma_{u_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 1.5)$ . . . . .	30
2.2	Taxas de rejeição da hipótese $\mathbf{H}_0 : (\alpha, \beta)^T = (0, 1)^T$ , usando as estatísticas de Wald dadas em (2.18) e (2.33) (ao nível nominal de 5% de significância) no modelo com erros de medição heterocedásticos, para $n = 40, 80$ e $160$ respectivamente. . . . .	31
3.1	Comportamento do DP-IE: Erro padrão usando a matriz de informação de Fisher dada em (3.20) e DP-IG: Erro padrão usando a matriz de informação geral dado em (3.48), quando usamos as estimativas de máxima verossimilhança na sua obtenção. . . . .	63
3.2	Estimativas (erros-padrão) dos parâmetros de regressão do MEV estrutural heterocedástico na variável não observável, no caso $\sqrt{\sigma_{x_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 4)$ , $i = 1, \dots, n$ . EP-IE: Erro padrão usando a matriz de informação esperada, EP-IG: Erro padrão usando a matriz de informação geral dado em (3.43) e EP-IE: Desvio padrão usando a matriz de informação observada. . . . .	65
3.3	Taxas de rejeição da hipótese $\mathbf{H}_0 : (\alpha, \beta)^T = (0, 1)^T$ , usando as estatísticas de Wald dadas em (2.18) e (3.22) (ao nível nominal de 5% de significância) no MEV estrutural heterocedástico na variável não observável, para $n = 40, 80$ e $160$ respectivamente. . . . .	66
3.4	Estimativas dos parâmetros estruturais $\alpha, \beta, \mu_x, \sigma_c^2$ e $\sigma_u^2$ , com seus correspondentes erros padrão (em parênteses) . . . . .	67

5.1 Estimativas do valor esperado, erro padrão, quantis e intervalos de credibilidade para os parâmetros  $\alpha$ ,  $\beta$ ,  $\mu_x$ ,  $\sigma_c^2$  e  $\sigma_u^2$ , baseados em uma única seqüência de Gibbs de comprimento  $T=10,000$ . . . . . 109

---

# Capítulo 1

## Introdução

### 1.1 Considerações Iniciais

Nesta tese são apresentados alguns resultados novos (ainda não discutidos na literatura) sobre o modelo de regressão estrutural heterocedástico (MREH) com erros nas variáveis. Não é de nosso interesse fazer uma revisão extensiva sobre todos os tópicos aqui desenvolvidos, portanto suporemos que o leitor tenha alguma familiaridade com temas tais como: Modelos de regressão com erros nas variáveis, teoria assintótica, implementação de algoritmos tipo MCMC via software WinBUGS, simulações de Monte Carlo, entre outros.

O objetivo geral de nosso trabalho é aperfeiçoar e estender o MEV estrutural usual para um MREH com erros nas variáveis, que nos permita modelar com maior precisão, situações nas quais os modelos de regressão usuais não são adequados. Nosso objetivo final, na medida que for possível (disponibilidade de dados), é buscar uma aplicação do modelo aqui desenvolvido, em áreas importantes, nas quais as variâncias dos erros de medição variam entre as observações, cenário comumente encontrado em exemplos de áreas tais como Química Analítica, Epidemiologia, Botânica, Astronomia e Engenharia Civil.

Os algoritmos computacionais para as simulações foram implementados no programa estatístico R (R. Development Core Team (2011)) e a implementação do método MCMC, na parte Bayesiana, foi feito com o programa computacional WinBUGS (Bayesian inference Using Gibbs Sampler for Windows).

## 1.2 O modelo com erros nas variáveis (MEV)

Na prática existem várias situações em que a variável explanatória (covariável) não pode ser observada diretamente, mas com erros. Isto é comum no caso de observações obtidas por meio de análises laboratoriais resultando em estimativas e não nos verdadeiros valores da variável explanatória. Modelos de regressão em que a variável explanatória está sujeita a erros de medição são denominados modelos com erros de medição ou modelos com erros nas variáveis, desde o século passado este modelo tem sido amplamente discutido na literatura. Sprent (1990) descreve o seu desenvolvimento histórico e uma vasta bibliografia pode ser encontrada em Fuller (1987), Cheng and Van Ness (1999) e Carroll et al. (2006).

O modelo de regressão linear simples com erros nas variáveis clássico está definido pelas seguintes equações

$$Y_i = y_i + e_i, \quad \text{com } y_i = \alpha + \beta x_i \quad (1.1)$$

$$X_i = x_i + u_i, \quad i = 1, \dots, n, \quad (1.2)$$

em que os erros de medição  $e_i$  e  $u_i$  são independentes e identicamente distribuídos seguindo uma distribuição normal com média zero e variância  $\sigma_e^2$  e  $\sigma_u^2$ , respectivamente, isto é

- 1)  $e_i \stackrel{iid}{\sim} N(0, \sigma_e^2)$
- 2)  $u_i \stackrel{iid}{\sim} N(0, \sigma_u^2)$
- 3)  $e_i \perp u_i, i = 1, \dots, n.$

A idéia principal por trás das equações (1.1) e (1.2) é que as verdadeiras observações  $(y_1, x_1), \dots, (y_n, x_n)$  (variáveis latentes) não são observadas diretamente e, portanto, a estimação será baseada em  $(Y_1, X_1), \dots, (Y_n, X_n)$ , que são observados. Se a quantidade  $x_i$  é considerada como uma quantidade fixa (não observada), então o modelo é chamado modelo de regressão funcional. Por outro lado, se a quantidade  $x_i$  é considerada como uma quantidade aleatória, então o modelo é chamado modelo de regressão estrutural, veja Fuller (1987) e Cheng and Van Ness (1999), para mais detalhes. No caso estrutural, a suposição típica segundo Arellano-Valle and Bolfarine (1998) é que

$$x_i \stackrel{iid}{\sim} N(\mu_x, \sigma_x^2), \quad \text{com } x_i \text{ independente de } (e_j, u_j), \forall i, j. \quad (1.3)$$

Como é bem sabido, existem problemas com a estimação dos parâmetros em ambos os casos.



No caso funcional, a estimação do parâmetro  $\beta$  não é consistente (verossimilhança ilimitada). No caso estrutural, surgem problemas de falta de identificabilidade do modelo. Uma suposição geralmente utilizada para contornar o problema de identificação considera que a razão da confiabilidade (ou coeficiente de atenuação),  $k_x = \sigma_x^2 / (\sigma_x^2 + \sigma_u^2)$  é conhecida, veja Fuller (1987).

Nosso trabalho tem como foco principal os modelos de regressão com erros nas variáveis, onde ambas as variáveis ( $Y$  e  $X$ ) estão sujeitas a erros de medição heterocedásticos. A generalização do modelo definido pelas equações (1.1) e (1.2), que não merecia muita atenção na literatura, surge quando os erros de medição são considerados heterocedásticos, isto é, assumindo que

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \stackrel{\text{indep.}}{\sim} N_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{e_i}^2 & 0 \\ 0 & \sigma_{u_i}^2 \end{pmatrix} \right], \quad i = 1, \dots, n. \quad (1.4)$$

O modelo de regressão funcional com a suposição dada em (1.4) é discutido por Fuller (1987), Ripley and Thompson (1987) e Galea-Rojas et al. (2003) sob abordagem da máxima verossimilhança supondo  $\sigma_{e_i}^2$  e  $\sigma_{u_i}^2$  conhecidas e maiores que zero,  $i = 1, \dots, n$ . A falta de interesse no modelo de regressão heterocedástico pode ser devido ao fato de que ambas as variâncias  $\sigma_{e_i}^2$  e  $\sigma_{u_i}^2$  precisam ser conhecidas,  $i = 1, \dots, n$ . No entanto, esta é uma configuração comum em áreas como a Química Analítica (Ripley and Thompson (1987)), Epidemiologia (Kulathinal et al. (2002)), Astronomia (Patriota (2010)), Botânica, entre outras.

Recentemente, modelos funcionais com erros de medição heterocedásticos e modelos estruturais com erro na equação e erros de medição heterocedásticos têm sido objeto de pesquisas (e.g., Kulathinal et al. (2002), Cheng and Riu (2006), Patriota and Bolfarine (2008)), de Castro et al. (2008), Patriota et al. (2009), Patriota (2010)). Em diversas aplicações supõe-se que os erros de medição são não correlacionados e suas variâncias são conhecidas e não negativas. No geral as metodologias encontradas na literatura consideram um modelo funcional heterocedástico com erros nas variáveis supondo variâncias conhecidas, contudo, na prática, as variâncias em geral não são conhecidas.

Neste trabalho concentramos nossa atenção, apenas no modelo de regressão estrutural heterocedástico (MREH) com erros nas variáveis e desenvolvemos uma teoria que nos permitiu fazer inferência sobre os parâmetros estruturais do modelo em estudo. A seguir, apresentamos o modelo geral que foi utilizado em nosso trabalho.

### 1.3 Modelo Geral

Consideremos o seguinte modelo

$$y_i = \alpha + \beta x_i + q_i \quad (1.5)$$

$$X_i = x_i + u_i \quad (1.6)$$

$$Y_i = y_i + e_i \quad i = 1, \dots, n, \quad (1.7)$$

sendo  $Y_i$  a variável de resposta observada,  $X_i$  a variável explanatória observada ambas sujeitas a erros de medição,  $y_i$  a verdadeira variável de resposta não observada e  $x_i$  a verdadeira variável explanatória não observada. O relacionamento entre as verdadeiras variáveis não observadas (latentes)  $y_i$  e  $x_i$  é dado pela equação (1.5), onde  $\alpha$  e  $\beta$  são os parâmetros de regressão desconhecidos e  $q_i$  é conhecido como erro da equação. O erro da equação significa que em algumas situações, as verdadeiras variáveis não estão perfeitamente relacionadas se fatores distintos de  $x_i$  são os responsáveis da variação em  $y_i$ . O erro da equação poderia não existir, nesse caso fazemos  $q_i = 0$  na equação (1.5).

Definindo

$$\mathbf{a} = (\alpha, 0)^T \quad \text{e} \quad \mathbf{b} = (\beta, 1)^T, \quad (1.8)$$

temos que

$$\mathbf{Z}_i = (Y_i, X_i)^T = \mathbf{a} + x_i \mathbf{b} + (e_i + q_i, u_i)^T, \quad i = 1, \dots, n. \quad (1.9)$$

As suposições impostas ao modelo são:

$(e_i, u_i)^T$  e  $(q_j, x_j)^T$  são independentes,  $q_i$  e  $x_i$  são independentes, com

$$q_i \stackrel{i.i.d}{\sim} N(0, \sigma^2), \quad x_i \stackrel{indep.}{\sim} N(\mu_x, \sigma_{x_i}^2) \quad \text{e}$$

$$(e_i, u_i)^T \stackrel{indep.}{\sim} N_2[(0, 0)^T; \text{diag}\{\sigma_{e_i}^2, \sigma_{u_i}^2\}], \quad i, j = 1, \dots, n,$$

onde  $\text{diag}\{\sigma_{e_i}^2, \sigma_{u_i}^2\}$  denota uma matriz diagonal de dimensão 2 com elementos  $\sigma_{e_i}^2$  e  $\sigma_{u_i}^2$  na diagonal.

Das suposições da acima temos de (1.9) que

$$\mathbf{Z}_i \stackrel{indep.}{\sim} N_2(\boldsymbol{\mu}(\boldsymbol{\theta}); \boldsymbol{\Sigma}_i(\boldsymbol{\theta})), \quad i = 1, \dots, n,$$

onde  $\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{a} + \mu_x \mathbf{b}$  e  $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{\Psi}_i + \sigma_{x_i}^2 \mathbf{b}\mathbf{b}^T$ , com  $\boldsymbol{\Psi}_i = \text{diag}(\sigma_{e_i}^2 + \sigma^2, \sigma_{u_i}^2)$ , sendo  $\boldsymbol{\theta}$  o vetor de parâmetros estruturais do modelo. Além disso note que,  $\boldsymbol{\theta}$  vai depender da condição de identificabilidade imposta ao modelo.

O modelo definido pelas equações (1.5), (1.6) e (1.7) é bem geral, entretanto esta tese foca a atenção em alguns casos particulares deste modelo, que são mostrados a seguir.

- (1) No capítulo 2, consideramos  $q_i = 0$  e  $\sigma_{x_i}^2 = \sigma_x^2, i = 1, \dots, n$ . Além disso

$$x_i \stackrel{i.i.d.}{\sim} N(\mu_x, \sigma_x^2) \quad \text{e} \quad (e_i, u_i)^T \stackrel{indep.}{\sim} N_2[(0, 0)^T; \text{diag}\{\sigma_{e_i}^2, \sigma_{u_i}^2\}], \quad i = 1, \dots, n.$$

Supondo as variâncias heterocedásticas  $\sigma_{e_i}^2$  e  $\sigma_{u_i}^2$  conhecidas  $\forall i = 1, \dots, n$ , obtemos um MREH com erros nas variáveis identificável. Neste caso o parâmetro estrutural é dado por  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_x^2)^T$ . Este modelo é um caso particular do modelo estudado por Kulathinal et al. (2002) quando  $q_i = 0, i = 1, \dots, n$ . Este é conhecido como MREH com erros nas variáveis sem erro da equação.

- (2) No capítulo 3, consideramos  $q_i = 0, \sigma_{e_i}^2 = \sigma_e^2$  e  $\sigma_{u_i}^2 = \sigma_u^2, i = 1, \dots, n$ . Além disso

$$(e_i, u_i)^T \stackrel{i.i.d.}{\sim} N_2[(0, 0)^T; \text{diag}\{\sigma_e^2, \sigma_u^2\}], \quad \text{e} \quad x_i \stackrel{indep.}{\sim} N(\mu_x, \sigma_{x_i}^2) \quad i = 1, \dots, n.$$

Supondo a variância heterocedástica  $\sigma_{x_i}^2$  conhecida  $\forall i = 1, \dots, n$ , obtemos um MREH com erros nas variáveis no qual a variância da variável não observada  $x_i$  é considerada heterocedástica. Neste caso o parâmetro estrutural é dado por  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2)^T$ . Este modelo ainda não foi abordado na literatura.

- (3) No capítulo 4, consideramos a mesma configuração que no capítulo 3, acrescentando réplicas não balanceadas apenas na variável explanatória. Neste caso temos que a equação (1.6) muda para

$$X_{ij} = x_i + u_{ij} \quad \text{onde} \quad u_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma_u^2) \quad j = 1, \dots, m_i, \quad i = 1, \dots, n.$$

O modelo resultante é um MREH com erros nas variáveis replicado.



---

No capítulo 5 consideramos o mesmo modelo que no capítulo 3, isto é, um MREH com erros nas variáveis, onde a variável não observada  $x_i$  tem distribuição normal com média  $\mu_x$  e variância  $\sigma_{x_i}^2$ ,  $i = 1, \dots, n$ . Este modelo é estudado sob abordagem Bayesiana.

## 1.4 Organização do trabalho

Este trabalho está organizado em seis capítulos mais um apêndice.

No Capítulo 1 apresentamos uma visão geral da forma dos modelos que serão estudados nos capítulos subsequentes. No Capítulo 2 estudamos inferência estatística no MREH com erros nas variáveis (sem erro na equação), quando as variâncias das medidas dos erros variam entre as observações. Este modelo é um caso particular do modelo estudado por Kulathinal et al. (2002). No capítulo 3 fazemos inferência num MREH com erros de medição (sem erro na equação) e heterocedasticidade na variável não observada  $x_i$ . Este modelo ainda não é discutido na literatura. No Capítulo 4 estendemos o modelo estudado no capítulo 3 para o caso de réplicas não balanceadas apenas na variável explanatória. No Capítulo 5 fazemos inferência no modelo estudado no capítulo 3 usando inferência Bayesiana. No Capítulo 6 apresentamos os comentários finais e propostas de pesquisas futuras. Concluimos a tese com o Apêndice, onde apresentamos alguns resultados importantes utilizados no desenvolvimento desta tese.

## 1.5 Conjunto de dados reais

### 1.5.1 Dados do projeto WHO MONICA

O projeto WHO MONICA acompanha o decurso de doenças cardiovasculares. O objetivo principal deste acompanhamento é conseguir poder relacionar os fatores conhecidos de risco com a mortalidade cardiovascular e doenças coronarianas. O conjunto de dados esta composto de 36 países cujos participantes são mulheres de diferentes países, e 38 países cujos participantes são homens de diferentes países. A idade dos participantes flutua entre 35-64 años e são em sua maioria europeus. Para cada indivíduo existe um escore indicando o risco médio e outro escore indicando a ocorrência media do evento. O escore de do risco médio foi definido como uma combinação linear entre status de fumante (fumante ou não fumante), pressão sistólica sanguínea, índice de massa corpórea e o nível de colesterol total. A variável de resposta  $Y$  observada é a tendência estimada da ocorrência do evento. Esta tendência é estimada por uma

regressão linear entre o logaritmo natural do índice médio anual da ocorrência do evento e o ano que o evento ocorreu. A variável explanatória observada,  $X$  é o índice médio anual em % associado ao evento. Estamos interessados em relacionar a variável resposta não observada,  $y$  (o verdadeiro índice médio anual em % da ocorrência do evento) com a verdadeira covariável não observada  $x$  (o verdadeiro índice médio anual em % do risco associado ao evento). As variáveis  $Y$  e  $X$  são observadas no lugar de  $y$  e  $x$  respectivamente. Na Figura 1.1 apresentamos os gráficos de dispersão para os dados dos homens (a) e das mulheres (b) com as barras de erros (observação  $\pm$  erro padrão).

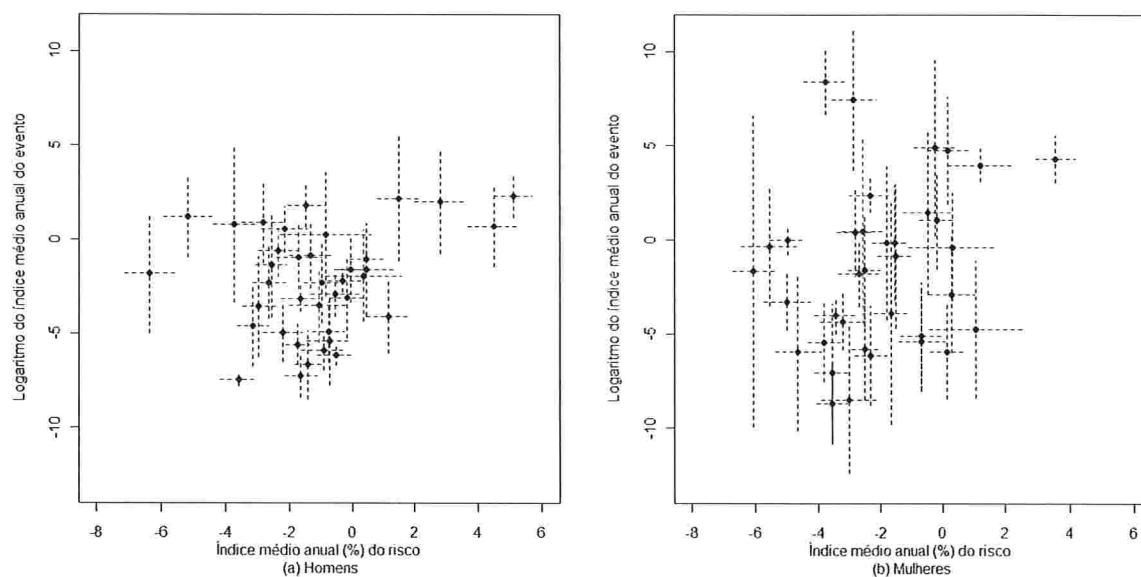


Figura 1.1. Dados do projeto WHO MONICA

## Capítulo 2

# Inferência no Modelo Estrutural Normal com Erros de Medição Heterocedásticos

### 2.1 Introdução

O objetivo deste capítulo é obter os estimadores via os métodos dos momentos e de máxima verossimilhança em um modelo estrutural normal com erros de medição heterocedásticos, isto é, quando as variâncias dos erros de medição  $e_i$  e  $u_i$  variam entre as observações. Também pretende-se determinar a distribuição assintótica destes estimadores. Neste caso as variâncias heterocedásticas dos erros de medição  $\sigma_{e_i}^2$  e  $\sigma_{u_i}^2$  serão consideradas conhecidas. Finalmente um estudo de simulação será feito para avaliar os estimadores encontrados.

### 2.2 Definição do modelo

Sejam

$$\begin{aligned}y_i &= \alpha + \beta x_i \\X_i &= x_i + u_i \\Y_i &= y_i + e_i \quad i = 1, \dots, n,\end{aligned}\tag{2.1}$$

onde  $(X_i, Y_i)$  são medições de  $(x_i, y_i)$ ,  $(u_i, e_i)$  são erros de medida e  $(\alpha, \beta)$  são parâmetros desconhecidos.

Admitimos, aqui, que a variável  $x_i$  e os erros de medição  $e_i, u_i$  são independentes e normalmente distribuídos como segue:

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{indep.}}{\sim} N_3 \left[ \begin{pmatrix} 0 \\ 0 \\ \mu_x \end{pmatrix}; \begin{pmatrix} \sigma_{e_i}^2 & 0 & 0 \\ 0 & \sigma_{u_i}^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right], \quad i = 1, \dots, n. \quad (2.2)$$

Isto leva ao modelo estrutural normal heteroscedástico dado por:

$$\begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix} \stackrel{\text{indep.}}{\sim} N_3 \left[ \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \\ \mu_x \end{pmatrix}; \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_{e_i}^2 & \beta\sigma_x^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_{u_i}^2 & \sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 & \sigma_x^2 \end{pmatrix} \right], \quad i = 1, \dots, n. \quad (2.3)$$

Assumindo  $\sigma_{e_i}^2$  e  $\sigma_{u_i}^2$  conhecidas, o modelo torna-se identificável veja Fuller (1987), Kulathinal et al. (2002), então, o vetor de parâmetros desconhecidos é  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_x^2)^T$ . Como a dimensão do vetor de estatísticas suficientes é igual a 5 e o vetor de parâmetros  $\boldsymbol{\theta}$  tem dimensão 4, temos a situação em que os estimadores de máxima verossimilhança não podem ser obtidos igualando-se as estatísticas suficientes a seus valores esperados como em Cheng and Van Ness (1999). O modelo está super especificado e neste caso o EMV do parâmetro  $\boldsymbol{\theta}$ , tem de ser obtido maximizando diretamente a função de verossimilhança ou através de algum método numérico tipo algoritmo EM.

### 2.3 Estimadores de momentos (EMM)

Definindo o vetor de dados observados por  $\mathbf{Z}_i = (Y_i, X_i)^T$  temos que

$$\mathbf{Z}_i \stackrel{\text{indep.}}{\sim} N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i), \quad i = 1, \dots, n,$$

onde

$$E(\mathbf{Z}_i) = \boldsymbol{\mu} = \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix} \quad \text{e} \quad \text{Var}(\mathbf{Z}_i) = \boldsymbol{\Sigma}_i = \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_{e_i}^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_{u_i}^2 \end{pmatrix} \quad (2.4)$$

Sejam

$$\mathbf{a} \text{ e } \mathbf{b} \text{ como em (1.8) e } \Psi_i = \begin{pmatrix} \sigma_{\epsilon_i}^2 & 0 \\ 0 & \sigma_{u_i}^2 \end{pmatrix}, \quad i = 1, \dots, n, \quad (2.5)$$

então temos que

$$\boldsymbol{\mu} = \mathbf{a} + \mu_x \mathbf{b} \text{ e } \Sigma_i = \sigma_x^2 \mathbf{b} \mathbf{b}^T + \Psi_i, \quad i = 1, \dots, n. \quad (2.6)$$

O seguinte resultado fornece os estimadores de momentos do parâmetro  $\boldsymbol{\theta}$ .

**Proposição 2.1.** Sejam  $\boldsymbol{\mu}$  e  $\Sigma_i$ , o primeiro e segundo momentos do vetor aleatório  $\mathbf{Z}_i$  e sejam

$$\bar{\mathbf{Z}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \text{ e } \mathbf{S}_{\mathbf{Z}} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T, \quad (2.7)$$

o vetor de médias amostrais e a matriz de covariâncias amostrais de  $\mathbf{Z}_i$  respectivamente.

Então,

$$\mathbb{E}(\bar{\mathbf{Z}}) = \boldsymbol{\mu} \quad \text{e} \quad \mathbb{E}(\mathbf{S}_{\mathbf{Z}}) = \sigma_x^2 \mathbf{b} \mathbf{b}^T + \bar{\Psi},$$

$$\text{onde } \bar{\Psi} = \begin{pmatrix} \bar{\sigma}_\epsilon^2 & 0 \\ 0 & \bar{\sigma}_u^2 \end{pmatrix}, \quad \text{com } \bar{\sigma}_\epsilon^2 = \frac{1}{n} \sum_{i=1}^n \sigma_{\epsilon_i}^2 \text{ e } \bar{\sigma}_u^2 = \frac{1}{n} \sum_{i=1}^n \sigma_{u_i}^2.$$

**Prova:** A primeira parte é direta, ou seja,

$$\mathbb{E}(\bar{\mathbf{Z}}) = \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(\mathbf{Z}_i) = \boldsymbol{\mu}.$$

Para a segunda parte notemos em primeiro lugar que:

$$\begin{aligned}
 \text{Var}(\mathbf{Z}_i - \bar{\mathbf{Z}}) &= \text{Var}(\mathbf{Z}_i) + \text{Var}(\bar{\mathbf{Z}}) - 2\text{Cov}(\mathbf{Z}_i, \bar{\mathbf{Z}}) \\
 &= \Sigma_i + \frac{1}{n^2} \sum_{i=1}^n \Sigma_i - \frac{2}{n} \Sigma_i \\
 &= \frac{(n-2)\Sigma_i}{n} + \frac{\bar{\Sigma}}{n} \\
 &= \frac{(n-2)(\sigma_x^2 \mathbf{b}\mathbf{b}^T + \Psi_i)}{n} + \frac{\sigma_x^2 \mathbf{b}\mathbf{b}^T + \bar{\Psi}}{n} \\
 &= \frac{(n-1)\sigma_x^2 \mathbf{b}\mathbf{b}^T}{n} + \frac{(n-2)\Psi_i + \bar{\Psi}}{n},
 \end{aligned}$$

onde  $\bar{\Sigma} = \frac{1}{n} \sum_{i=1}^n \Sigma_i$ . Levando em conta que  $E(\mathbf{Z}_i - \bar{\mathbf{Z}}) = \mathbf{0}$ , temos que

$$\text{Var}(\mathbf{Z}_i - \bar{\mathbf{Z}}) = E\{(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T\}.$$

Assim,

$$\begin{aligned}
 E(\mathbf{S}_Z) &= \frac{1}{n-1} \sum_{i=1}^n E\{(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T\} \\
 &= \frac{1}{n-1} \sum_{i=1}^n \text{Var}(\mathbf{Z}_i - \bar{\mathbf{Z}}) \\
 &= \frac{1}{n-1} \sum_{i=1}^n \left( \frac{(n-1)}{n} \sigma_x^2 \mathbf{b}\mathbf{b}^T + \frac{(n-2)\Psi_i + \bar{\Psi}}{n} \right) \\
 &= \sigma_x^2 \mathbf{b}\mathbf{b}^T + \frac{(n-2)}{n-1} \bar{\Psi} + \frac{1}{n-1} \bar{\Psi} \\
 &= \sigma_x^2 \mathbf{b}\mathbf{b}^T + \bar{\Psi}.
 \end{aligned}$$

□

### 2.3.1 EMM para o parâmetro $\theta = (\alpha, \beta, \mu_x, \sigma_x^2)^\top$

Para o cálculo dos estimadores de momentos, definimos as seguintes quantidades amostrais:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (2.8)$$

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{e} \quad S_{YX} = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})X_i \quad (2.9)$$

Assim, podemos definir o vetor de médias e matriz de covariâncias amostrais tais como

$$\bar{\mathbf{Z}} = \begin{pmatrix} \bar{Y} \\ \bar{X} \end{pmatrix} \quad \text{e} \quad \mathbf{S}_Z = \begin{pmatrix} S_Y^2 & S_{YX} \\ S_{YX} & S_X^2 \end{pmatrix}. \quad (2.10)$$

Então, os estimadores de momentos podem ser obtidos usando a Proposição 2.1, isto é, temos que resolver as seguintes equações de momentos:

$$\boldsymbol{\mu}(\hat{\theta}) = \bar{\mathbf{Z}}, \quad \sigma_x^2(\hat{\theta})\mathbf{b}(\hat{\theta})\mathbf{b}(\hat{\theta})^\top + \bar{\boldsymbol{\Psi}} = \mathbf{S}_Z. \quad (2.11)$$

Desenvolvendo estas equações para as componentes de  $\hat{\theta}$ , obtemos:

$$\hat{\alpha} + \hat{\beta}\hat{\mu}_x = \bar{Y} \quad \Longrightarrow \quad \hat{\alpha} = \bar{Y} - \hat{\beta}\hat{\mu}_x \quad ; \quad \hat{\mu}_x = \bar{X},$$

$$\hat{\sigma}_x^2 + \bar{\sigma}_u^2 = S_X^2 \quad \Longrightarrow \quad \hat{\sigma}_x^2 = S_X^2 - \bar{\sigma}_u^2, \quad \text{se} \quad S_X^2 - \bar{\sigma}_u^2 > 0,$$

$$\hat{\beta}\hat{\sigma}_x^2 = S_{YX} \quad \Longrightarrow \quad \hat{\beta} = \frac{S_{YX}}{S_X^2 - \bar{\sigma}_u^2}, \quad \text{se} \quad S_X^2 - \bar{\sigma}_u^2 > 0.$$

Portanto, assumindo  $\bar{\sigma}_u^2$  conhecido, os estimadores de momentos são dados por:

$$\hat{\mu}_{x\text{EMM}} = \bar{X}, \quad \hat{\alpha}_{\text{EMM}} = \bar{Y} - \hat{\beta}_{\text{EMM}}\bar{X}, \quad (2.12)$$

$$\hat{\beta}_{\text{EMM}} = \frac{S_{YX}}{S_X^2 - \bar{\sigma}_u^2}, \quad \hat{\sigma}_{x\text{EMM}}^2 = S_X^2 - \bar{\sigma}_u^2, \quad \text{se} \quad S_X^2 - \bar{\sigma}_u^2 > 0. \quad (2.13)$$



Note que os estimadores de momentos obtidos em (2.12) e (2.13) não são únicos, pois a equação  $\beta^2 \sigma_x^2 + \bar{\sigma}_e^2 = S_Y^2$ , que também é fornecida pelas equações de momentos, não foi utilizada na sua obtenção. Isto é, neste caso a relação  $(\mu, \bar{\Sigma}) \rightarrow \theta$  não é 1:1. Estes estimadores de momentos também são propostos em Kulathinal et al. (2002). Contudo, se supusermos que  $\bar{\sigma}_e^2$  fosse conhecido ao em vez de  $\bar{\sigma}_u^2$ , então esta equação teria que ser usada. Em tal caso, os EMM dos parâmetros  $\beta$  e  $\sigma_x^2$  seriam dados por:

$$\hat{\beta}_{\text{EMM}} = \frac{S_Y^2 - \bar{\sigma}_e^2}{S_{XY}}, \quad \text{se } S_{YX} \neq 0 \quad \text{e} \quad \hat{\sigma}_{x_{\text{EMM}}}^2 = \frac{S_{YX}^2}{S_Y^2 - \bar{\sigma}_e^2} \quad \text{se } S_Y^2 - \bar{\sigma}_e^2 > 0, \quad (2.14)$$

e os estimadores dos parâmetros  $\alpha$  e  $\mu_x$  são dados como em (2.12).

Note também, que se no caso da suposição de identificabilidade  $\bar{\sigma}_u^2$  conhecido fazemos  $\sigma_{u_i}^2 = \sigma_u^2, i = \dots, n$  (caso homoscedástico), o EMM do parâmetro  $\beta$  dado em (2.13) coincide com o EMV do parâmetro  $\beta$  dado em Cheng and Van Ness (1999), pag. 18 caso (c). De maneira semelhante, no caso da suposição  $\bar{\sigma}_e^2$  conhecido fazendo  $\sigma_{e_i}^2 = \sigma_e^2, i = \dots, n$ , temos que o EMM do parâmetro  $\beta$  dado em (2.14) é igual ao EMV de  $\beta$  dado em Cheng and Van Ness (1999), pag. 18 caso (d).

### 2.3.2 Distribuição assintótica do estimador $\hat{\theta}_{\text{EMM}}$

Como os estimadores de momentos tem solução explícita, podemos usar o método delta para obter sua distribuição assintótica. Para isto, precisamos das seguintes suposições adicionais de regularidade.

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma_{e_i}^2 &= \nu_e^*, & \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (\sigma_{e_i}^2)^2 &= \nu_e^{**}, & \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma_{e_i}^2 \sigma_{u_i}^2 &= \nu_{eu}^{**}, \\ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma_{u_i}^2 &= \nu_u^*, & \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (\sigma_{u_i}^2)^2 &= \nu_u^{**}. \end{aligned} \quad (2.15)$$

O próximo resultado fornece a distribuição assintótica do estimador  $\hat{\theta}_{\text{EMM}}$ .



**Proposição 2.2.** Seja  $\hat{\theta}_{\text{EMM}}$  o estimador de momentos do parâmetro  $\theta$  dado em (2.12) e (2.13). Se as suposições em (2.15) são satisfeitas, então

$$\sqrt{n}(\hat{\theta}_{\text{EMM}} - \theta) \xrightarrow{D} N_4[\mathbf{0}_4, \Upsilon(\theta)], \text{ quando } n \rightarrow \infty,$$

onde

$$\Upsilon(\theta) = \begin{bmatrix} \sigma_x^2 + \nu_u^* & -\beta\nu_u^* & 0 & 0 \\ -\beta\nu_u^* & \Upsilon_{22} & \Upsilon_{23} & \frac{2\beta\mu_x}{\sigma_x^2}(\sigma_x^2\nu_u^* + \nu_u^{**}) \\ 0 & \Upsilon_{32} & \Upsilon_{33} & \frac{-2\beta}{\sigma_x^2}(\sigma_x^2\nu_u^* + \nu_u^{**}) \\ 0 & \frac{2\beta\mu_x}{\sigma_x^2}(\sigma_x^2\nu_u^* + \nu_u^{**}) & \frac{-2\beta}{\sigma_x^2}(\sigma_x^2\nu_u^* + \nu_u^{**}) & \Upsilon_{44} \end{bmatrix}$$

com

$$\begin{aligned} \Upsilon_{22} &= \frac{1}{\sigma_x^4}(\beta^2\sigma_x^4\nu_u^* + \beta^2\mu_x^2\sigma_x^2\nu_u^* + \mu_x^2\sigma_x^2\nu_\epsilon^* + \sigma_x^4\nu_\epsilon^* + 2\beta^2\mu_x^2\nu_u^{**} + \mu_x^2\nu_{eu}^{**}) \\ \Upsilon_{23} &= \Upsilon_{32} = \frac{-\mu_x}{\sigma_x^4}(\beta^2\sigma_x^2\nu_u^* + \sigma_x^2\nu_\epsilon^* + 2\beta^2\nu_u^{**} + \nu_{eu}^{**}) \\ \Upsilon_{33} &= \frac{1}{\sigma_x^4}(\beta^2\sigma_x^2\nu_u^* + \sigma_x^2\nu_\epsilon^* + 2\beta^2\nu_u^{**} + \nu_{eu}^{**}) \\ \Upsilon_{44} &= 2\sigma_x^4 + 4\sigma_x^2\nu_u^* + 2\nu_u^{**}. \end{aligned}$$

**Prova:** A prova pode ser feita com o auxílio de dois resultados assintóticos, a saber, Teoremas 3.1 e 3.2 em Arellano-Valle et al. (2002). Estes teoremas garantem que sob as suposições feitas,

$$\bar{\mathbf{Z}} \xrightarrow{q.c.} \boldsymbol{\mu}, \quad \mathbf{S}_{\mathbf{Z}} \xrightarrow{q.c.} \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_*, \quad (2.16)$$

e

$$\begin{pmatrix} \sqrt{n}(\bar{\mathbf{Z}} - \boldsymbol{\mu}) \\ \sqrt{n} \text{vech}(\mathbf{S}_{\mathbf{Z}} - \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_*) \end{pmatrix} \xrightarrow{D} N_{\frac{1}{2}p(p+3)} \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Gamma} + \boldsymbol{\Gamma}_* \\ \boldsymbol{\Gamma}' + \boldsymbol{\Gamma}'_* & \boldsymbol{\Lambda} + \boldsymbol{\Lambda}_* \end{pmatrix} \right), \quad (2.17)$$

quando  $n \rightarrow \infty$ , onde o operador  $\text{vech}(\cdot)$  é definido como em Magnus and Neudecker (2007), pag. 49.

No nosso caso,  $\bar{\mathbf{Z}}$  e  $\mathbf{S}_{\mathbf{Z}}$  são dadas em (2.10) e além disso,  $\boldsymbol{\Sigma}_* = \boldsymbol{\Gamma} = \boldsymbol{\Gamma}_* = \boldsymbol{\Lambda}_* = \mathbf{0}$ ,  $\boldsymbol{\Sigma}$  é dada

por

$$\Sigma = \begin{pmatrix} \beta^2 \sigma_x^2 + \nu_e^* & \beta \sigma_x^2 \\ \beta \sigma_x^2 & \sigma_x^2 + \nu_u^* \end{pmatrix},$$

e  $\Lambda$  é dada por

$$\Lambda = \begin{bmatrix} 2\beta^4 \sigma_x^4 + 4\beta^2 \sigma_x^2 \nu_e^* + 2\nu_e^{**} & 2\beta^2 \sigma_x^4 & 2\beta \sigma_x^2 (\beta^2 \sigma_x^2 + \nu_e^*) \\ 2\beta^2 \sigma_x^4 & 2\sigma_x^4 + 4\sigma_x^2 \nu_u^* + 2\nu_u^{**} & 2\beta \sigma_x^2 (\sigma_x^2 + \nu_u^*) \\ 2\beta \sigma_x^2 (\beta^2 \sigma_x^2 + \nu_e^*) & 2\beta \sigma_x^2 (\sigma_x^2 + \nu_u^*) & \beta^2 \sigma_x^2 (2\sigma_x^2 + \nu_u^*) + \sigma_x^2 \nu_e^* + \nu_{eu}^{**} \end{bmatrix}.$$

Assim, aplicando (2.16) e (2.17) resulta que

$$\sqrt{n} \begin{pmatrix} \bar{\mathbf{Z}} - \boldsymbol{\mu} \\ \text{vech}(\mathbf{S}_Z - \Sigma) \end{pmatrix} = \sqrt{n} \begin{bmatrix} \bar{Y} - \alpha - \beta \mu_x \\ \bar{X} - \mu_x \\ S_Y^2 - \beta^2 \sigma_x^2 - \nu_e^* \\ S_X^2 - \sigma_x^2 - \nu_u^* \\ S_{YX} - \beta \sigma_x^2 \end{bmatrix} \xrightarrow{D} N_5 \left[ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix} \right].$$

Agora, seja  $\boldsymbol{\phi} = (\alpha + \beta \mu_x, \mu_x, \sigma_x^2, \beta \sigma_x^2)^T$  e definamos a função  $g: \mathbb{R}^4 \rightarrow \mathbb{R}^4$ , tal que:  $g(\mathbf{x}) = g(x_1, x_2, x_3, x_4) = (x_2, x_1 - \frac{x_1}{x_3} x_2, \frac{x_1}{x_3}, x_3)^T$  para todo  $\mathbf{x} \in \mathbb{R}^4$ . Assim,

$$\hat{\boldsymbol{\theta}}_{\text{EMM}} = g(\bar{Y}, \bar{X}, S_X^2 - \bar{\sigma}_u^2, S_{YX}) = \begin{bmatrix} \bar{X} \\ \bar{Y} - \frac{S_{YX}}{S_X^2 - \bar{\sigma}_u^2} \bar{X} \\ \frac{S_{YX}}{S_X^2 - \bar{\sigma}_u^2} \\ S_X^2 - \bar{\sigma}_u^2 \end{bmatrix} = \begin{bmatrix} \hat{\mu}_{\text{EMM}} \\ \hat{\alpha}_{\text{EMM}} \\ \hat{\beta}_{\text{EMM}} \\ \hat{\sigma}_{x_{\text{EMM}}}^2 \end{bmatrix}$$

e

$$g(\phi) = g(\alpha + \beta\mu_x, \mu_x, \sigma_x^2, \beta\sigma_x^2) = \begin{bmatrix} \mu_x \\ \alpha \\ \beta \\ \sigma_x^2 \end{bmatrix}.$$

Portanto,

$$\dot{\mathbf{G}}(\phi) = \frac{\partial}{\partial \mathbf{x}^T} \mathbf{g}(\mathbf{x}) \Big|_{\mathbf{x}=\phi} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & -\beta & \frac{\beta\mu_x}{\sigma_x^2} & \frac{-\mu_x}{\sigma_x^2} \\ 0 & 0 & \frac{-\beta}{\sigma_x^2} & \frac{1}{\sigma_x^2} \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Então pelo método delta, veja Sen and Singer (1993), Teorema 3.4.6, temos que:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{EMM}} - \boldsymbol{\theta}) \xrightarrow{D} N_4[\mathbf{0}_4, \boldsymbol{\Upsilon}(\boldsymbol{\theta})], \text{ quando } n \rightarrow \infty, \text{ onde}$$

$$\boldsymbol{\Upsilon}(\boldsymbol{\theta}) = \dot{\mathbf{G}}(\phi) \boldsymbol{\Upsilon}_1 \dot{\mathbf{G}}(\phi)^T \quad \text{e} \quad \boldsymbol{\Upsilon}_1 = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda} \end{pmatrix}.$$

Finalmente fazendo o produto  $\dot{\mathbf{G}}(\phi) \boldsymbol{\Upsilon}_1 \dot{\mathbf{G}}(\phi)^T$ , algumas manipulações algébricas conduzem à expressão dada para  $\boldsymbol{\Upsilon}(\boldsymbol{\theta})$ .  $\square$

Um resultado similar pode ser obtido se usamos a restrição de identificabilidade  $\bar{\sigma}_e^2$  conhecido.

## 2.4 Teste de hipóteses

Para testar a hipótese  $H_0 : \mathbf{G}\boldsymbol{\theta} = \mathbf{d}$ , podemos utilizar a estatística de Wald dada por:

$$\mathcal{W}_{\text{EMM}} = n \left( \mathbf{G}\hat{\boldsymbol{\theta}}_{\text{EMM}} - \mathbf{d} \right)^T \left[ \mathbf{G}\boldsymbol{\Upsilon}(\hat{\boldsymbol{\theta}}_{\text{EMM}})\mathbf{G}^T \right]^{-1} \left( \mathbf{G}\hat{\boldsymbol{\theta}}_{\text{EMM}} - \mathbf{d} \right), \quad (2.18)$$

que, sob  $H_0$ , converge em distribuição para uma distribuição  $\chi^2_{(q)}$ , em que  $q$  é o posto da matriz  $\mathbf{G}$ . Este resultado é uma consequência imediata do fato que  $\hat{\boldsymbol{\theta}}_{\text{EMM}}$  é um estimador consistente para  $\boldsymbol{\theta}$  e que  $\hat{\boldsymbol{\theta}}_{\text{EMM}}$  é aproximadamente normal com média  $\boldsymbol{\theta}$  e variância  $\Upsilon(\boldsymbol{\theta})$  em grandes amostras.

## 2.5 Estimação de máxima verossimilhança (EMV)

### 2.5.1 O algoritmo EM

O algoritmo EM, desenvolvido por Dempster et al. (1977), tem sido utilizado há bastante tempo como forma de obter aproximações numéricas para o estimador de máxima verossimilhança (EMV) em modelos complexos. O objetivo deste método iterativo é determinar o valor do vetor de parâmetros desconhecidos  $\boldsymbol{\theta}$ , ( $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ ), que maximize  $L(\mathbf{z}|\boldsymbol{\theta})$  a verossimilhança dos dados observados (ou incompletos). Fazendo uma escolha apropriada da função de verossimilhança dos dados completos, denotada por  $L_c(\mathbf{w}|\boldsymbol{\theta})$ , temos que

$$L(\mathbf{z}|\boldsymbol{\theta}) = \int_{\mathcal{W}(\mathbf{z})} L_c(\mathbf{w}|\boldsymbol{\theta}) d\mathbf{w}, \quad \text{onde } \mathcal{W}(\mathbf{z}) = \{\mathbf{w} : \mathbf{z} = \mathbf{z}(\mathbf{w})\}.$$

O correspondente  $\mathbf{w} \in \mathcal{W}$  não é observada diretamente, mas apenas indiretamente, através de  $\mathbf{z}$ , isto é,  $\mathbf{w}$  só é conhecida por estar em  $\mathcal{W}(\mathbf{z})$ . Aqui  $\mathbf{w}$  e  $\mathbf{z}$  denotam os vetores de dados completos e incompletos, respectivamente.

O algoritmo EM, na sua forma geral, é como segue:

**Alvo:** Encontrar o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$ , isto é, gostaríamos de obter  $\hat{\boldsymbol{\theta}}$  tal que

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \{L(\mathbf{z}|\boldsymbol{\theta})\}.$$

Definindo  $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = E(\log\{L_c(\mathbf{w}|\boldsymbol{\theta})|\mathbf{Z}, \hat{\boldsymbol{\theta}}\})$ , o algoritmo EM definirá a passagem da iteração  $\boldsymbol{\theta}^p$ , obtida pelo algoritmo no passo  $p$  para a iteração  $\boldsymbol{\theta}^{p+1}$ , obtida pelo algoritmo no passo  $(p+1)$ , da seguinte forma:

**Etapa E:** Calcule  $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$

**Etapa M:** Escolha  $\boldsymbol{\theta}^{p+1} \in \Theta$ , que maximize  $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ .

**Parada:** Continue até que  $|\boldsymbol{\theta}^p - \boldsymbol{\theta}^{p+1}| < \epsilon$  ou  $|L(\mathbf{z}|\boldsymbol{\theta}^p) - L(\mathbf{z}|\boldsymbol{\theta}^{p+1})| < \epsilon$ , para algum  $\epsilon > 0$ .

A idéia intuitiva por trás do algoritmo EM é a seguinte: gostaríamos de maximizar  $L_c(\mathbf{w}|\boldsymbol{\theta})$ , mas não a conhecemos ou é muito complexa, ao invés disso maximizamos  $E(\log\{L_c(\mathbf{w}|\boldsymbol{\theta})|\mathbf{Z}, \boldsymbol{\theta}^p\})$  que no geral é mais simples, em que  $\boldsymbol{\theta}^p$  é a estimativa de  $\boldsymbol{\theta}$  obtida na etapa M da iteração anterior.

As condições que garantem a convergência do algoritmo EM, isto é, que a seqüência gerada pelas etapas E e M atingem um valor  $\hat{\boldsymbol{\theta}}$  que maximiza a função  $\log\{L(\mathbf{z}|\boldsymbol{\theta})\}$ , isto é, que o limite desta seqüência é o EMV de  $\boldsymbol{\theta}$ , são dadas nos artigos de Dempster et al. (1977), Wu (1983) e Boyles (1983). Um resultado importante deste algoritmo é que cada iteração do algoritmo EM aumenta a verossimilhança dos dados observados, isto é,  $\log\{L(\mathbf{z}|\boldsymbol{\theta}^p)\} \leq \log\{L(\mathbf{z}|\boldsymbol{\theta}^{p+1})\}$ .

### 2.5.2 Aplicação do algoritmo EM no MREH

O algoritmo EM pode ser usado no modelo estrutural com erros de medição heteroscedásticos, ao considerar as verdadeiras variáveis  $x_i$  como dados perdidos.

Definimos o vetor de dados completos por  $\mathbf{W}_i = (Y_i, X_i, x_i)^T$ ,  $i = 1, \dots, n$ . Com as suposições feitas para este modelo em (2.2) e com as restrições de identificabilidade  $(\sigma_{c_i}^2, \sigma_{u_i}^2)$ ,  $i = 1, \dots, n$  conhecidas, segue que a distribuição conjunta de  $\mathbf{W}_i$  é dada por

$$\mathbf{W}_i \stackrel{indep.}{\sim} N_3(\mathbf{v}, \boldsymbol{\Omega}_i), \quad i = 1, \dots, n, \quad (2.19)$$

onde  $\mathbf{v}$  e  $\boldsymbol{\Omega}_i$  são dadas como em (2.3), isto é

$$\mathbf{v} = \begin{pmatrix} \boldsymbol{\mu} \\ \mu_x \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Omega}_i = \begin{pmatrix} \Sigma_i & \sigma_x^2 \mathbf{b} \\ \sigma_x^2 \mathbf{b}^T & \sigma_x^2 \end{pmatrix}, \quad i = 1, \dots, n, \quad (2.20)$$

com  $\boldsymbol{\mu}$  e  $\Sigma_i$  como em (2.4). Isto leva a seguinte função de verossimilhança:

$$L_c(\boldsymbol{\theta}) = \prod_{i=1}^n \frac{1}{(2\pi)^{3/2} |\boldsymbol{\Omega}_i|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{w}_i - \mathbf{v})^T \boldsymbol{\Omega}_i^{-1} (\mathbf{w}_i - \mathbf{v}) \right].$$

Logo, o logaritmo da verossimilhança dos dados completos pode ser escrito como:

$$l_c(\boldsymbol{\theta}) = -\frac{3n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log |\det \boldsymbol{\Omega}_i| - \frac{1}{2} \sum_{i=1}^n (\mathbf{w}_i - \mathbf{v})^T \boldsymbol{\Omega}_i^{-1} (\mathbf{w}_i - \mathbf{v}). \quad (2.21)$$

Usando os resultados em Muirhead (1982) (Teoremas A.5.2 e A.5.3), obtemos:

$$\det(\Omega_i) = \sigma_x^2 \sigma_{e_i}^2 \sigma_{u_i}^2, \quad (2.22)$$

e

$$\Omega_i^{-1} = \begin{pmatrix} \frac{1}{\sigma_{e_i}^2} & 0 & \frac{-\beta}{\sigma_{e_i}^2} \\ 0 & \frac{1}{\sigma_{u_i}^2} & \frac{-1}{\sigma_{u_i}^2} \\ \frac{-\beta}{\sigma_{e_i}^2} & \frac{-1}{\sigma_{u_i}^2} & \frac{1}{\sigma_x^2} + \frac{\beta^2}{\sigma_{e_i}^2} + \frac{1}{\sigma_{u_i}^2} \end{pmatrix}, \quad i = 1, \dots, n. \quad (2.23)$$

Fazendo algumas manipulações algébricas, temos que:

$$\begin{aligned} (\mathbf{w}_i - \mathbf{v})^T \Omega_i^{-1} (\mathbf{w}_i - \mathbf{v}) &= \frac{1}{\sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2} \left( \sigma_x^2 \sigma_{u_i}^2 Y_i^2 + \mu_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 + \sigma_x^2 \sigma_{e_i}^2 x_i^2 + \sigma_x^2 \sigma_{e_i}^2 X_i^2 \right. \\ &\quad + \sigma_{u_i}^2 \sigma_{e_i}^2 x_i^2 - 2\alpha \sigma_x^2 \sigma_{u_i}^2 Y_i - 2\sigma_x^2 \sigma_{e_i}^2 X_i x_i + \beta^2 \sigma_x^2 \sigma_{u_i}^2 x_i^2 \\ &\quad \left. - 2\mu \sigma_{u_i}^2 \sigma_{e_i}^2 x_i + \alpha^2 \sigma_x^2 \sigma_{u_i}^2 - 2\beta \sigma_x^2 \sigma_{u_i}^2 Y_i x_i + 2\beta \alpha \sigma_x^2 \sigma_{u_i}^2 x_i \right), \quad (2.24) \end{aligned}$$

e

$$\log[\det(\Omega_i)] = \log(\sigma_x^2 \sigma_{e_i}^2 \sigma_{u_i}^2) = \log(\sigma_x^2) + \log(\sigma_{e_i}^2) + \log(\sigma_{u_i}^2), \quad i = 1, \dots, n. \quad (2.25)$$

Substituindo (2.24) e (2.25) em (2.21), obtemos a seguinte expressão:

$$\begin{aligned} l_c(\theta) &= -\frac{3n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log(\sigma_{e_i}^2 \sigma_{u_i}^2) - \frac{n}{2} \log(\sigma_x^2) - \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2} \left( \sigma_x^2 \sigma_{u_i}^2 Y_i^2 \right. \\ &\quad + \mu_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 + \sigma_x^2 \sigma_{e_i}^2 x_i^2 + \sigma_x^2 \sigma_{e_i}^2 X_i^2 + \sigma_{u_i}^2 \sigma_{e_i}^2 x_i^2 - 2\alpha \sigma_x^2 \sigma_{u_i}^2 Y_i - 2\sigma_x^2 \sigma_{e_i}^2 X_i x_i \\ &\quad \left. + \beta^2 \sigma_x^2 \sigma_{u_i}^2 x_i^2 - 2\mu \sigma_{u_i}^2 \sigma_{e_i}^2 x_i + \alpha^2 \sigma_x^2 \sigma_{u_i}^2 - 2\beta \sigma_x^2 \sigma_{u_i}^2 Y_i x_i + 2\beta \alpha \sigma_x^2 \sigma_{u_i}^2 x_i \right). \quad (2.26) \end{aligned}$$

Como descrito anteriormente, o algoritmo EM, definirá a passagem  $\theta^p \rightarrow \theta^{p+1}$ , da seguinte forma:

Etapa E (passo  $p$ )

Calcular:  $Q(\theta|\theta^{(p)}) = E(l_c(\theta|\mathbf{W})|Y, X, \theta^{(p)})$

Etapa M (passo  $(p + 1)$ )

Achar o ponto de máximo  $\hat{\theta}$  de  $Q(\theta|\theta^{(p)})$  com relação a  $\theta$ .

Fazer  $\theta^{(p+1)} = \hat{\theta}$  e  $p = p + 1$ .

Na etapa E (passo  $p$ ), precisamos conhecer a distribuição condicional dos dados não observados com respeito aos dados observados, isto é, precisamos da distribuição da variável aleatória  $(x_i|Y_i, X_i)$ .

Segundo Mardia et al. (1989) (Teorema 3.2.4, pag. 63) temos que:

$(x_i|Y_i, X_i) \sim N(E(x_i|Y_i, X_i), \text{Var}(x_i|Y_i, X_i))$ , onde

$$E(x_i|Y_i, X_i) = \frac{\sigma_{e_i}^2 \sigma_{u_i}^2 \mu_x + \beta \sigma_x^2 \sigma_{u_i}^2 (Y_i - \alpha) + \sigma_x^2 \sigma_{e_i}^2 X_i}{(\beta^2 \sigma_x^2 + \sigma_{e_i}^2) \sigma_{u_i}^2 + \sigma_{e_i}^2 \sigma_x^2}$$

$$v(x_i) = \text{Var}(x_i|Y_i, X_i) = \frac{\sigma_x^2 \sigma_{e_i}^2 \sigma_{u_i}^2}{(\beta^2 \sigma_x^2 + \sigma_{e_i}^2) \sigma_{u_i}^2 + \sigma_{e_i}^2 \sigma_x^2}$$

Fazendo:

$$\hat{x}_i = E(x_i|Y_i, X_i) \quad \text{e} \quad \hat{x}_i^2 = v(x_i) + (\hat{x}_i)^2, \quad i = 1, \dots, n,$$

segue que:

$$Q(\theta|\theta^{(p)}) \propto -\frac{n}{2} \log(\sigma_x^{2(p)}) - \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_x^{2(p)} \sigma_{u_i}^{2(p)} \sigma_{e_i}^{2(p)}} \left( \sigma_x^{2(p)} \sigma_{u_i}^{2(p)} Y_i^2 + \mu_x^{2(p)} \sigma_{u_i}^{2(p)} \sigma_{e_i}^{2(p)} \right. \\ \left. + \sigma_x^{2(p)} \sigma_{e_i}^{2(p)} \hat{x}_i^{2(p)} + \sigma_x^{2(p)} \sigma_{e_i}^{2(p)} X_i^2 + \sigma_{u_i}^{2(p)} \sigma_{e_i}^{2(p)} \hat{x}_i^{2(p)} - 2\alpha^{(p)} \sigma_x^{2(p)} \sigma_{u_i}^{2(p)} Y_i \right. \\ \left. - 2\sigma_x^{2(p)} \sigma_{e_i}^{2(p)} X_i \hat{x}_i^{(p)} + \beta^{2(p)} \sigma_x^{2(p)} \sigma_{u_i}^{2(p)} \hat{x}_i^{2(p)} - 2\mu_x^{(p)} \sigma_{u_i}^{2(p)} \sigma_{e_i}^{2(p)} \hat{x}_i^{(p)} \right. \\ \left. + \alpha^{2(p)} \sigma_x^{2(p)} \sigma_{u_i}^{2(p)} - 2\beta^{(p)} \sigma_x^{2(p)} \sigma_{u_i}^{2(p)} Y_i \hat{x}_i^{(p)} + 2\beta^{(p)} \alpha^{(p)} \sigma_x^{2(p)} \sigma_{u_i}^{2(p)} \hat{x}_i^{(p)} \right), \quad (2.27)$$

em que:

$$\hat{x}_i^{(p)} = \frac{\sigma_{e_i}^{2(p)} \sigma_{u_i}^{2(p)} \mu_x^{(p)} + \beta^{(p)} \sigma_x^{2(p)} \sigma_{u_i}^{2(p)} (Y_i - \alpha^{(p)}) + \sigma_x^{2(p)} \sigma_{e_i}^{2(p)} X_i}{(\beta^{2(p)} \sigma_x^{2(p)} + \sigma_{e_i}^{2(p)}) \sigma_{u_i}^{2(p)} + \sigma_{e_i}^{2(p)} \sigma_x^{2(p)}},$$



$$\text{Var}(x_i|Y_i, X_i)^{(p)} = \frac{\sigma_x^{2(p)} \sigma_{e_i}^{2(p)} \sigma_{u_i}^{2(p)}}{(\beta^{2(p)} \sigma_x^{2(p)} + \sigma_{e_i}^{2(p)}) \sigma_{u_i}^{2(p)} + \sigma_{e_i}^{2(p)} \sigma_x^{2(p)}},$$

$$\widehat{x}_i^{2(p)} = \text{Var}(x_i|Y_i, X_i)^{(p)} + \left(\widehat{x}_i^{(p)}\right)^2, \quad i = 1, \dots, n.$$

Na etapa M (passo  $(p + 1)$ ), temos que maximizar (2.27). Resolvendo as equações

$$\frac{\partial}{\partial \mu_x^{(p)}} [Q(\theta|\theta^{(p)})] = -\frac{n\mu_x^{(p)}}{\sigma_x^{2(p)}} + \sum_{i=1}^n \frac{\widehat{x}_i^{(p)}}{\sigma_x^{2(p)}} = 0,$$

$$\frac{\partial}{\partial \sigma_x^{2(p)}} [Q(\theta|\theta^{(p)})] = \frac{1}{2\sigma_x^{4(p)}} [n\mu_x^{2(p)} - n\sigma_x^{2(p)} + \sum_{i=1}^n (\widehat{x}_i^{2(p)} - 2\mu_x^{(p)}\widehat{x}_i^{(p)})] = 0,$$

$$\frac{\partial}{\partial \alpha^{(p)}} [Q(\theta|\theta^{(p)})] = -\sum_{i=1}^n \left( \frac{\alpha^{(p)} - Y_i + \beta^{(p)}\widehat{x}_i^{(p)}}{\sigma_{e_i}^2} \right) = 0,$$

$$\frac{\partial}{\partial \beta^{(p)}} [Q(\theta|\theta^{(p)})] = -\sum_{i=1}^n \left( \frac{\beta^{(p)}\widehat{x}_i^{2(p)} - Y_i\widehat{x}_i^{(p)} + \alpha^{(p)}\widehat{x}_i^{(p)}}{\sigma_{e_i}^2} \right) = 0,$$

obtemos os estimadores de máxima verossimilhança para  $\beta$ ,  $\alpha$ ,  $\mu_x$  e  $\sigma_x^2$  respectivamente.

A figura 2.1 contém um esboço do esquema iterativo do algoritmo EM.



início

- Fornecer valores iniciais  $\theta^{(0)} = (\alpha^{(0)}, \beta^{(0)}, \mu_x^{(0)}, \sigma_x^{2(0)}, \sigma_{e_i}^2, \sigma_{u_i}^2)$ ;
- passo E: Avaliar

$$\hat{x}_i^{(p)} = \frac{\sigma_{e_i}^{2(p)} \sigma_{u_i}^{2(p)} \mu_x^{(p)} + \beta^{(p)} \sigma_x^{2(p)} \sigma_{u_i}^{2(p)} (Y_i - \alpha^{(p)}) + \sigma_x^{2(p)} \sigma_{e_i}^{2(p)} X_i}{(\beta^{2(p)} \sigma_x^{2(p)} + \sigma_{e_i}^{2(p)}) \sigma_{u_i}^{2(p)} + \sigma_{e_i}^{2(p)} \sigma_x^{2(p)}}$$

$$V(x_i)^{(p)} = \frac{\sigma_x^{2(p)} \sigma_{e_i}^{2(p)} \sigma_{u_i}^{2(p)}}{(\beta^{2(p)} \sigma_x^{2(p)} + \sigma_{e_i}^{2(p)}) \sigma_{u_i}^{2(p)} + \sigma_{e_i}^{2(p)} \sigma_x^{2(p)}}$$

$$\hat{x}_i^{2(p)} = V x_i^{(p)} + \left( \hat{x}_i^{(p)} \right)^2, \quad i = 1, \dots, n,$$

- passo M: Avaliar

$$\beta^{(p+1)} = \frac{\left( \sum_{i=1}^n \frac{1}{\sigma_{e_i}^2} \right) \left( \sum_{i=1}^n \frac{Y_i \hat{x}_i^{(p)}}{\sigma_{e_i}^2} \right) - \left( \sum_{i=1}^n \frac{Y_i}{\sigma_{e_i}^2} \right) \left( \sum_{i=1}^n \frac{\hat{x}_i^{(p)}}{\sigma_{e_i}^2} \right)}{\left( \sum_{i=1}^n \frac{1}{\sigma_{e_i}^2} \right) \left( \sum_{i=1}^n \frac{\hat{x}_i^{2(p)}}{\sigma_{e_i}^2} \right) - \left( \sum_{i=1}^n \frac{\hat{x}_i^{(p)}}{\sigma_{e_i}^2} \right)^2},$$

$$\alpha^{(p+1)} = \frac{\sum_{i=1}^n \frac{Y_i}{\sigma_{e_i}^2} - \beta^{(p+1)} \left( \sum_{i=1}^n \frac{1}{\sigma_{e_i}^2} \right) \left( \frac{1}{n} \sum_{i=1}^n \hat{x}_i^{(p)} \right)}{\sum_{i=1}^n \frac{1}{\sigma_{e_i}^2}},$$

$$\mu_x^{(p+1)} = \frac{1}{n} \sum_{i=1}^n \hat{x}_i^{(p)},$$

$$\sigma_x^{2(p+1)} = \frac{1}{n} \sum_{i=1}^n \hat{x}_i^{2(p)} - \left( \frac{1}{n} \sum_{i=1}^n \hat{x}_i^{(p)} \right)^2,$$

- repetir passos E e M até que  $|\theta^{(p)} - \theta^{(p+1)}| < \epsilon$ , para algum  $\epsilon > 0$ ,

fim

Figura 2.1. Algoritmo EM no MREH, quando  $(\sigma_{e_i}^2, \sigma_{u_i}^2)$ ,  $i = 1, \dots, n$  são conhecidas

### 2.5.3 Matriz de covariâncias assintóticas do EMV de $\theta$

A matriz de variâncias-covariâncias assintótica do EMV do vetor de parâmetros  $\theta = (\alpha, \beta, \mu_x, \sigma_x^2)^T$  é dada pela inversa da matriz de informação de Fisher ( $\mathbf{I}_F$ ), onde

$$\mathbf{I}_F = -\mathbf{E} \left[ \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \right],$$

com  $l(\theta) = \log[\mathbf{L}(\theta)]$ , sendo  $\mathbf{L}(\theta)$  a função de verossimilhança dos dados observados  $\mathbf{Z}_i$ . Assim dada uma amostra de tamanho  $n$  de  $\mathbf{Z}$  temos:

$$L(\theta) = \prod_{i=1}^n f_{\mathbf{Z}_i}(\mathbf{z}_i) = \prod_{i=1}^n \frac{1}{2\pi |\Sigma_i|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu})^T \Sigma_i^{-1} (\mathbf{z}_i - \boldsymbol{\mu}) \right].$$

Logo, a log-verossimilhança pode ser escrita como:

$$l(\theta) = -n \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log |\det \Sigma_i| - \frac{1}{2} \sum_{i=1}^n (\mathbf{z}_i - \boldsymbol{\mu})^T \Sigma_i^{-1} (\mathbf{z}_i - \boldsymbol{\mu}).$$

Após várias manipulações algébricas, obtemos a seguinte expressão para o log-verossimilhança:

$$\begin{aligned} l(\theta) = & -\frac{1}{2} \sum_{i=1}^n \log(A_i(\beta, \sigma_x^2)) - \frac{1}{2} \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} \left( \sigma_x^2 Y_i^2 - 2\sigma_x^2 \alpha Y_i \right. \\ & + \sigma_{u_i}^2 Y_i^2 - 2\alpha \sigma_{u_i}^2 Y_i - 2\beta \mu_x \sigma_{u_i}^2 Y_i + \alpha^2 \sigma_x^2 + \alpha^2 \sigma_{u_i}^2 \\ & + 2\beta \alpha \mu_x \sigma_{u_i}^2 + \beta^2 \mu_x^2 \sigma_{u_i}^2 - 2\beta \sigma_x^2 X_i Y_i + 2\beta \alpha \sigma_x^2 X_i + \sigma_{e_i}^2 X_i^2 \\ & \left. + \beta^2 \sigma_x^2 X_i^2 - 2\mu_x \sigma_{e_i}^2 X_i + \mu_x^2 \sigma_{e_i}^2 \right), \end{aligned} \quad (2.28)$$

em que

$$A_i(\beta, \sigma_x^2) = \beta^2 \sigma_x^2 \sigma_{u_i}^2 + \sigma_x^2 \sigma_{e_i}^2 + \sigma_{e_i}^2 \sigma_{u_i}^2, \quad i = 1, \dots, n.$$

**Proposição 2.3.** A matriz de informação de Fisher para o parâmetro  $\theta$  em um modelo estrutural normal com erros nas variáveis, em que as variâncias dos erros de medição são heterocedásticas é dada por

$$\mathbf{I}_F(\theta) = \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\beta} & I_{\alpha\mu} & 0 \\ I_{\beta\alpha} & I_{\beta\beta} & I_{\beta\mu} & I_{\beta\sigma} \\ I_{\mu\alpha} & I_{\mu\beta} & I_{\mu\mu} & 0 \\ 0 & I_{\sigma\beta} & 0 & I_{\sigma\sigma} \end{pmatrix},$$

onde

$$I_{\alpha\alpha} = \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} (\sigma_x^2 + \sigma_{u_i}^2),$$

$$I_{\alpha\beta} = I_{\beta\alpha} = \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} (\mu_x \sigma_x^2 + \mu_x \sigma_{u_i}^2),$$

$$I_{\alpha\mu} = I_{\mu\alpha} = \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} (\beta \sigma_{u_i}^2),$$

$$\begin{aligned} I_{\beta\beta} = \sum_{i=1}^n \frac{1}{A_i^3(\beta, \sigma_x^2)} & \left( \mu_x^2 \sigma_{u_i}^6 \sigma_{e_i}^4 - \sigma_x^2 \sigma_{u_i}^6 \sigma_{e_i}^4 - 2\sigma_x^4 \sigma_{u_i}^4 \sigma_{e_i}^4 + 2\beta^4 \sigma_x^6 \sigma_{u_i}^6 \right. \\ & + \mu_x^2 \sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^4 + 5\beta^2 \sigma_x^8 \sigma_{u_i}^2 \sigma_{e_i}^2 + \beta^4 \sigma_x^8 \sigma_{u_i}^4 - \sigma_x^6 \sigma_{u_i}^2 \sigma_{e_i}^4 + 5\beta^2 \sigma_x^4 \sigma_{u_i}^6 \sigma_{e_i}^2 \\ & + 10\beta^2 \sigma_x^6 \sigma_{u_i}^4 \sigma_{e_i}^2 + \beta^4 \mu_x^2 \sigma_x^6 \sigma_{u_i}^4 + \beta^4 \mu_x^2 \sigma_x^4 \sigma_{u_i}^6 + 2\mu_x^2 \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^4 \\ & \left. + 2\beta^2 \mu_x^2 \sigma_x^2 \sigma_{u_i}^6 \sigma_{e_i}^2 + 5\beta^2 \mu_x^2 \sigma_x^6 \sigma_{u_i}^2 \sigma_{e_i}^2 + 7\beta^2 \mu_x^2 \sigma_x^4 \sigma_{u_i}^4 \sigma_{e_i}^2 \right), \end{aligned}$$

$$I_{\beta\mu} = I_{\mu\beta} = \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} (\beta \mu_x \sigma_{u_i}^2),$$

$$\begin{aligned} I_{\beta\sigma} = I_{\sigma\beta} = \sum_{i=1}^n \frac{1}{A_i^3(\beta, \sigma_x^2)} & \left( \beta^5 \sigma_x^4 \sigma_{u_i}^6 - \beta \sigma_{u_i}^6 \sigma_{e_i}^4 + \beta^3 \mu_x^2 \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 - \beta \mu_x^2 \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^4 \right. \\ & \left. + 3\beta^3 \sigma_x^4 \sigma_{u_i}^4 \sigma_{e_i}^2 - \beta \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^4 - \beta \mu_x^2 \sigma_{u_i}^4 \sigma_{e_i}^4 + 2\beta^3 \sigma_x^2 \sigma_{u_i}^6 \sigma_{e_i}^2 \right), \end{aligned}$$

$$I_{\mu\mu} = \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} (\beta^2 \sigma_{u_i}^2 + \sigma_{e_i}^2),$$

$$\begin{aligned} I_{\sigma\sigma} = \sum_{i=1}^n \frac{1}{2A_i^3(\beta, \sigma_x^2)} & \left( 3\beta^4 \sigma_{u_i}^6 \sigma_{e_i}^2 + \beta^6 \sigma_x^2 \sigma_{u_i}^6 - 8\beta^3 \mu_x \sigma_{u_i}^4 \sigma_{e_i}^2 + \sigma_x^2 \sigma_{e_i}^6 \right. \\ & + 5\beta^4 \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 + 5\beta^2 \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^4 + 8\beta^3 \mu_x \sigma_{u_i}^4 \sigma_{e_i}^2 + 4\beta^2 \sigma_{u_i}^4 \sigma_{e_i}^4 \\ & \left. + \sigma_{u_i}^2 \sigma_{e_i}^6 + 10\beta^4 \mu_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 + 2\beta^2 \mu_x^2 \sigma_{u_i}^2 \sigma_{e_i}^4 \right). \end{aligned}$$

**Prova:** A prova será feita com o auxílio do software matemático Maple, versão 10.

Derivando (2.28) em relação a  $\alpha$ ,  $\beta$ ,  $\mu_x$  e  $\sigma_x^2$  respectivamente, obtemos as derivadas de primeira ordem, resultando:

$$\frac{\partial l(\theta)}{\partial \alpha} = \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} (\sigma_x^2 Y_i + \sigma_{u_i}^2 Y_i - \alpha \sigma_x^2 - \alpha \sigma_{u_i}^2 - \beta \mu_x \sigma_{u_i}^2 - \beta \sigma_x^2 X_i) \quad (2.29)$$

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \beta} = \sum_{i=1}^n \frac{-1}{A_i^2(\beta, \sigma_x^2)} & \left( \beta^3 \sigma_x^4 \sigma_{u_i}^4 + \alpha \mu_x \sigma_{u_i}^4 \sigma_{e_i}^2 - \mu_x \sigma_{u_i}^4 \sigma_{e_i}^2 Y_i + \beta \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 \right. \\ & + \beta \sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^2 + \beta \mu_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 - \sigma_x^4 \sigma_{e_i}^2 X_i Y_i + \alpha \sigma_x^4 \sigma_{e_i}^2 X_i - \beta \sigma_x^4 \sigma_{u_i}^2 Y_i^2 \\ & - \beta \alpha^2 \sigma_x^2 \sigma_{u_i}^4 - \beta \alpha^2 \sigma_x^4 \sigma_{u_i}^2 - \beta \sigma_x^2 \sigma_{u_i}^4 Y_i^2 - \beta^2 \alpha \mu_x \sigma_x^2 \sigma_{u_i}^4 + \alpha \mu_x \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 \\ & + \beta^2 \sigma_x^4 \sigma_{u_i}^2 X_i Y_i - \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i Y_i - \beta^2 \alpha \sigma_x^4 \sigma_{u_i}^2 X_i + \alpha \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i \\ & + 2\beta \alpha \sigma_x^4 \sigma_{u_i}^2 Y_i + 2\beta \alpha \sigma_x^4 \sigma_{u_i}^4 Y_i - \beta \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i^2 + 2\beta \mu_x \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i \\ & \left. + \beta^2 \mu_x \sigma_x^2 \sigma_{u_i}^4 Y_i - \mu_x \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 Y_i \right), \end{aligned} \quad (2.30)$$

$$\frac{\partial l(\theta)}{\partial \mu_x} = \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} (\beta \sigma_{u_i}^2 Y_i - \beta \alpha \sigma_{u_i}^2 - \beta^2 \mu_x \sigma_{u_i}^2 + \sigma_{e_i}^2 X_i - \mu_x \sigma_{e_i}^2), \quad (2.31)$$

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \sigma_x^2} = \sum_{i=1}^n \frac{-1}{2A_i^2(\beta, \sigma_x^2)} & \left( 2\mu_x \sigma_{e_i}^4 X_i - \beta^4 \mu_x^2 \sigma_{u_i}^4 - \beta^2 \alpha^2 \sigma_{u_i}^4 - \beta^2 \sigma_{u_i}^4 Y_i^2 \right. \\ & + \beta^2 \sigma_{u_i}^4 \sigma_{e_i}^2 + \beta^4 \sigma_x^2 \sigma_{u_i}^4 + 2\beta^2 \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 + 2\beta^2 \mu_x \sigma_{u_i}^2 \sigma_{e_i}^2 X_i \\ & + 2\beta \alpha \sigma_{u_i}^2 \sigma_{e_i}^2 X_i - \beta^2 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i^2 - 2\beta^3 \mu_x \alpha \sigma_{u_i}^4 + 2\beta^3 \mu_x \sigma_{u_i}^4 Y_i \\ & + 2\beta^2 \alpha \sigma_{u_i}^4 Y_i - 2\beta \sigma_{u_i}^2 \sigma_{e_i}^2 X_i Y_i - 2\beta \alpha \mu_x \sigma_{u_i}^2 \sigma_{e_i}^2 + 2\beta \mu_x \sigma_{u_i}^2 \sigma_{e_i}^2 Y_i \\ & \left. - 2\beta^2 \mu_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 + \sigma_x^2 \sigma_{e_i}^4 - \mu_x^2 \sigma_{e_i}^4 - \sigma_{e_i}^4 X_i^2 + \sigma_{u_i}^2 \sigma_{e_i}^4 \right). \end{aligned} \quad (2.32)$$

Agora derivando, (2.29) em relação a  $\alpha$ ,  $\beta$ ,  $\mu_x$ ,  $\sigma_x^2$ , (2.30) em relação a  $\beta$ ,  $\mu_x$ ,  $\sigma_x^2$ , (2.31) em relação a  $\mu_x$ ,  $\sigma_x^2$  e (2.32) em relação a  $\sigma_x^2$ , obtemos as derivadas de segunda ordem dadas por:

$$\frac{\partial^2 l(\theta)}{\partial \alpha^2} = - \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} (\sigma_x^2 + \sigma_{u_i}^2),$$

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha \partial \beta} &= - \sum_{i=1}^n \frac{1}{A_i^2(\beta, \sigma_x^2)} \left( -\beta^2 \mu_x \sigma_x^2 \sigma_{u_i}^4 + \mu_x \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 + \mu_x \sigma_{u_i}^4 \sigma_{e_i}^2 \right. \\ &\quad - \beta^2 \sigma_x^4 \sigma_{u_i}^2 X_i + \sigma_x^4 \sigma_{e_i}^2 X_i + \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i + 2\beta \sigma_x^4 \sigma_{u_i}^2 Y_i \\ &\quad \left. + 2\beta \sigma_x^2 \sigma_{u_i}^4 Y_i - 2\beta \alpha \sigma_x^4 \sigma_{u_i}^2 - 2\beta \alpha \sigma_x^2 \sigma_{u_i}^4 \right),\end{aligned}$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha \partial \mu_x} = - \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} \beta \sigma_{u_i}^2,$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha \partial \sigma_x^2} = - \sum_{i=1}^n \frac{1}{A_i^2(\beta, \sigma_x^2)} \beta \sigma_{u_i}^2 \left( \beta \sigma_{u_i}^2 Y_i - \beta \alpha \sigma_{u_i}^2 - \beta^2 \mu_x \sigma_{u_i}^2 + \sigma_{e_i}^2 X_i - \mu_x \sigma_{e_i}^2 \right),$$

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta^2} &= - \sum_{i=1}^n \frac{1}{A_i^3(\beta, \sigma_x^2)} \left[ \sigma_{u_i}^2 \left( \mu_x^2 \sigma_{u_i}^4 \sigma_{e_i}^4 + \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^4 + 2\sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^4 - \beta^4 \sigma_x^6 \sigma_{u_i}^4 \right. \right. \\ &\quad - \sigma_x^4 \sigma_{e_i}^4 X_i^2 - \sigma_x^6 \alpha^2 \sigma_{e_i}^2 - \sigma_x^6 \sigma_{e_i}^2 Y_i^2 + \sigma_x^6 \sigma_{e_i}^4 + \mu_x^2 \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^4 + 3\beta^2 \sigma_x^6 \sigma_{u_i}^2 Y_i^2 \\ &\quad - 2\sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^2 Y_i^2 + 3\beta^2 \alpha^2 \sigma_x^4 \sigma_{u_i}^4 - 2\alpha^2 \sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^2 - \alpha^2 \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 \\ &\quad + 3\beta^2 \alpha^2 \sigma_x^6 \sigma_{u_i}^2 + 3\beta^2 \sigma_x^4 \sigma_{u_i}^2 Y_i^2 - \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 Y_i^2 + 2\alpha \sigma_x^6 \sigma_{e_i}^2 Y_i - \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^4 X_i^2 \\ &\quad + 2\mu_x \sigma_x^4 \sigma_{e_i}^4 X_i - 3\beta^2 \mu_x^2 \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 + 2\beta^3 \alpha \mu_x \sigma_x^4 \sigma_{u_i}^4 - 6\beta \alpha \mu_x \sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^2 \\ &\quad - 6\beta \alpha \mu_x \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 - 2\beta^3 \sigma_x^6 \sigma_{u_i}^2 X_i Y_i + 6\beta \sigma_x^6 \sigma_{e_i}^2 X_i Y_i + 6\beta \sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i Y_i \\ &\quad + 2\beta^3 \alpha \sigma_x^6 \sigma_{u_i}^2 X_i - 6\beta \alpha \sigma_x^6 \sigma_{e_i}^2 X_i - 6\beta \alpha \sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i - 6\beta^2 \alpha \sigma_x^6 \sigma_{u_i}^2 Y_i \\ &\quad + 4\alpha \sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^2 Y_i - 6\beta^2 \alpha \sigma_x^4 \sigma_{u_i}^4 Y_i + 2\alpha \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 Y_i + 3\beta^2 \sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i^2 \\ &\quad - 6\beta^2 \mu_x \sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i + 2\mu_x \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^4 X_i - 2\beta^3 \mu_x \sigma_x^4 \sigma_{u_i}^4 Y_i \\ &\quad \left. \left. + 6\beta \mu_x \sigma_x^4 \sigma_{u_i}^2 \sigma_{e_i}^2 Y_i + 6\beta \mu_x \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 Y_i \right) \right],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta \partial \mu_x} &= - \sum_{i=1}^n \frac{1}{A_i^2(\beta, \sigma_x^2)} \left[ \sigma_{u_i}^2 \left( \alpha \sigma_{u_i}^2 \sigma_{e_i}^2 - \sigma_{u_i}^2 \sigma_{e_i}^2 Y_i + 2\beta \mu_x \sigma_{u_i}^2 \sigma_{e_i}^2 - \beta^2 \alpha \sigma_x^2 \sigma_{u_i}^2 \right. \right. \\ &\quad \left. \left. + \alpha \sigma_x^2 \sigma_{e_i}^2 + 2\beta \sigma_x^2 \sigma_{e_i}^2 X_i + \beta^2 \sigma_x^2 \sigma_{u_i}^2 Y_i - \sigma_x^2 \sigma_{e_i}^2 Y_i \right) \right],\end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta \partial \sigma_x^2} &= - \sum_{i=1}^n \frac{1}{A_i^3(\beta, \sigma_x^2)} \left[ \sigma_{u_i}^2 \left( \beta \sigma_{u_i}^4 \sigma_{e_i}^4 + \beta \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^4 - 2\beta \mu_x^2 \sigma_{u_i}^2 \sigma_{e_i}^4 - \alpha \mu_x \sigma_{u_i}^2 \sigma_{e_i}^4 \right. \right. \\ &+ \mu_x \sigma_{u_i}^2 \sigma_{e_i}^4 Y_i + \beta^3 \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 + \beta^3 \alpha^2 \sigma_x^2 \sigma_{u_i}^4 - \beta \alpha^2 \sigma_{u_i}^4 \sigma_{e_i}^2 + \beta^3 \sigma_x^2 \sigma_{u_i}^4 Y_i^2 \\ &- \beta \sigma_{u_i}^4 \sigma_{e_i}^2 Y_i^2 - \sigma_{u_i}^2 \sigma_{e_i}^4 X_i Y_i + \alpha \sigma_{u_i}^2 \sigma_{e_i}^4 X_i - \beta \sigma_{u_i}^2 \sigma_{e_i}^4 X_i^2 - 2\beta^3 \mu_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 \\ &- \alpha \mu_x \sigma_x^2 \sigma_{e_i}^4 + \alpha \sigma_x^2 \sigma_{e_i}^4 X_i + \mu_x \sigma_x^2 \sigma_{e_i}^4 Y_i - \sigma_x^2 \sigma_{e_i}^4 X_i Y_i + \beta \sigma_x^2 \sigma_{e_i}^4 X_i^2 \\ &- \beta \alpha^2 \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 - \beta \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 Y_i^2 - 2\beta^3 \alpha \sigma_x^2 \sigma_{u_i}^4 Y_i - 2\beta \mu_x \sigma_x^2 \sigma_{e_i}^4 X_i \\ &+ 2\beta \alpha \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 Y_i + 2\beta \alpha \sigma_{u_i}^4 \sigma_{e_i}^2 Y_i + \beta^3 \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i^2 - 2\beta^3 \mu_x \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i \\ &+ 2\beta \mu_x \sigma_{u_i}^2 \sigma_{e_i}^4 X_i - \beta^4 \mu_x \sigma_x^2 \sigma_{u_i}^4 Y_i + 3\beta^2 \mu_x \sigma_{u_i}^4 \sigma_{e_i}^2 Y_i + \beta^4 \alpha \mu_x \sigma_x^2 \sigma_{u_i}^4 \\ &\left. \left. - 3\beta^2 \alpha \mu_x \sigma_{u_i}^4 \sigma_{e_i}^2 + 3\beta^2 \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i Y_i - 3\beta^2 \alpha \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^2 X_i \right) \right], \end{aligned}$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \mu_x^2} = - \sum_{i=1}^n \frac{1}{A_i(\beta, \sigma_x^2)} \left( \beta^2 \sigma_{u_i}^2 + \sigma_{e_i}^2 \right),$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \mu_x \partial \sigma_x^2} &= - \sum_{i=1}^n \frac{1}{A_i^2(\beta, \sigma_x^2)} \left( \beta \sigma_{u_i}^2 Y_i - \beta \alpha \sigma_{u_i}^2 - \beta^2 \mu_x \sigma_{u_i}^2 + \sigma_{e_i}^2 X_i - \mu_x \sigma_{e_i}^2 \right) \\ &\left( \beta^2 \sigma_{u_i}^2 + \sigma_{e_i}^2 \right), \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial (\sigma_x^2)^2} &= \sum_{i=1}^n \frac{1}{2A_i^3(\beta, \sigma_x^2)} \left( \beta^4 \sigma_{u_i}^6 \sigma_{e_i}^2 + \beta^6 \sigma_x^2 \sigma_{u_i}^6 - 2\beta^4 \sigma_{u_i}^6 Y_i^2 - 2\beta^4 \alpha^2 \sigma_{u_i}^6 \right. \\ &- 2\beta^6 \mu_x^2 \sigma_{u_i}^6 + 4\mu_x \sigma_{e_i}^6 X_i + 4\beta^4 \alpha \sigma_{u_i}^6 Y_i + 4\beta^5 \mu_x \sigma_{u_i}^6 Y_i - 4\beta^5 \alpha \mu_x \sigma_{u_i}^6 \\ &- 2\beta^4 \sigma_{u_i}^4 \sigma_{e_i}^2 X_i^2 + \sigma_x^2 \sigma_{e_i}^6 - 4\beta \alpha \mu_x \sigma_{u_i}^2 \sigma_{e_i}^4 - 2\sigma_{e_i}^6 X_i^2 - 4\beta^3 \sigma_{u_i}^4 \sigma_{e_i}^2 X_i Y_i \\ &+ 4\beta^3 \alpha \sigma_{u_i}^4 \sigma_{e_i}^2 X_i + 4\beta^4 \mu_x \sigma_{u_i}^4 \sigma_{e_i}^2 X_i - 2\mu_x^2 \sigma_{e_i}^6 + 4\beta \mu_x \sigma_{u_i}^2 \sigma_{e_i}^4 Y_i \\ &+ 3\beta^4 \sigma_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 + 3\beta^2 \sigma_x^2 \sigma_{u_i}^2 \sigma_{e_i}^4 - 4\beta \sigma_{u_i}^2 \sigma_{e_i}^4 X_i Y_i + 4\beta \alpha \sigma_{u_i}^2 \sigma_{e_i}^4 X_i \\ &+ 4\beta^2 \alpha \sigma_{u_i}^4 \sigma_{e_i}^2 Y_i + 8\beta^3 \mu_x \sigma_{u_i}^4 \sigma_{e_i}^2 Y_i - 8\beta^3 \alpha \mu_x \sigma_{u_i}^4 \sigma_{e_i}^2 + 8\beta^2 \mu_x \sigma_{u_i}^2 \sigma_{e_i}^4 X_i \\ &+ 2\beta^2 \sigma_{u_i}^4 \sigma_{e_i}^4 + \sigma_{u_i}^2 \sigma_{e_i}^6 - 2\beta^2 \sigma_{u_i}^4 \sigma_{e_i}^2 Y_i^2 - 2\beta^2 \alpha^2 \sigma_{u_i}^4 \sigma_{e_i}^2 - 6\beta^4 \mu_x^2 \sigma_{u_i}^4 \sigma_{e_i}^2 \\ &\left. \left. - 4\beta^2 \sigma_{u_i}^2 \sigma_{e_i}^4 X_i^2 - 6\beta^2 \mu_x^2 \sigma_{u_i}^2 \sigma_{e_i}^4 \right) \right). \end{aligned}$$

Finalmente tomando esperança a cada uma das derivadas de segunda ordem obtidas acima, obtemos o resultado pretendido.  $\square$

O Teorema 5.2.2 em Sen and Singer (1993) (generalização do Teorema 5.2.1), pode ser estendido para o caso em que as variáveis aleatórias sejam apenas independentes, desde que



se justifique o uso de alguma versão da Lei Forte (Fracca) dos Grandes Números para mostrar que  $V_n(\boldsymbol{\theta}) \rightarrow -\mathbf{I}^*(\boldsymbol{\theta})$  quase certamente ou em probabilidade (onde  $\mathbf{I}^*(\boldsymbol{\theta})$  é a média da informação sobre  $\boldsymbol{\theta}$ ) e alguma versão do Teorema do Limite Central para mostrar que  $U_n(\boldsymbol{\theta})$  convenientemente normalizado converge em distribuição para uma variável aleatória normal. Assim temos que

$$\sqrt{n} \left( \hat{\boldsymbol{\theta}}_{\text{EMV}} - \boldsymbol{\theta} \right) \xrightarrow{D} N_4 \left[ \mathbf{0}_4, \Phi^{-1}(\boldsymbol{\theta}) \right], \quad \text{onde} \quad \Phi(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{I}_F(\boldsymbol{\theta}),$$

e dizemos que  $\hat{\boldsymbol{\theta}}_{\text{EMV}}$  é distribuída assintoticamente normal com vetor de média  $\boldsymbol{\theta}$  e matriz de covariância  $\mathbf{I}_F^{-1}(\boldsymbol{\theta})$ . Isto é denotado por

$$\hat{\boldsymbol{\theta}}_{\text{EMV}} \xrightarrow{a} N_4 \left[ \boldsymbol{\theta}, \mathbf{I}_F^{-1}(\boldsymbol{\theta}) \right].$$

Uma versão alternativa à apresentada acima para  $\mathbf{I}_F$  é apresentada em Kulathinal et al. (2002).

#### 2.5.4 Teste de hipóteses

Para testar a hipótese  $H_0 : \mathbf{G}\boldsymbol{\theta} = \mathbf{d}$ , podemos utilizar a estatística de Wald dada por:

$$\mathcal{W}_{\text{EMV}} = n \left( \mathbf{G}\hat{\boldsymbol{\theta}}_{\text{EMV}} - \mathbf{d} \right)^T \left[ \mathbf{G} \hat{\Phi}(\hat{\boldsymbol{\theta}}_{\text{EMV}}) \mathbf{G}^T \right] \left( \mathbf{G}\hat{\boldsymbol{\theta}}_{\text{EMV}} - \mathbf{d} \right) \quad (2.33)$$

$$= \left( \mathbf{G}\hat{\boldsymbol{\theta}}_{\text{EMV}} - \mathbf{d} \right)^T \left[ \mathbf{G} \mathbf{I}_F(\hat{\boldsymbol{\theta}}_{\text{EMV}}) \mathbf{G}^T \right] \left( \mathbf{G}\hat{\boldsymbol{\theta}}_{\text{EMV}} - \mathbf{d} \right) \quad (2.34)$$

onde  $\hat{\Phi}(\hat{\boldsymbol{\theta}}_{\text{EMV}}) = \frac{1}{n} \mathbf{I}_F(\hat{\boldsymbol{\theta}}_{\text{EMV}})$  e  $\mathcal{W}_{\text{EMV}}$  sob  $H_0$  converge em distribuição para uma distribuição  $\chi_{(q)}^2$ , em que  $q$  é o posto da matriz  $\mathbf{G}$ .

## 2.6 Estudos de simulação

Nesta seção estudos de simulação, usando diferentes tamanhos amostrais, são feitos a fim de avaliar o comportamento das estatísticas do tipo Wald dadas em (2.18), (2.33) e dos estimadores dos parâmetros de regressão do MREH com erros de medição proposto em (2.1) com as suposições em (2.2), assumindo a condição de identificabilidade  $\sigma_{e_i}^2$  e  $\sigma_{u_i}^2$ ,  $i = 1, \dots, n$ ,



conhecidas.

Para calcular as estimativas dos parâmetros do modelo consideramos simulações de Monte Carlo. Para gerar os dados  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , usamos o mesmo conjunto de valores para os parâmetros que foi usado em de Castro et al. (2008), isto é:

$$\alpha = -1, \quad \beta = 1, \quad \text{e} \quad x_i \stackrel{i.i.d.}{\sim} N(-2, 4), \quad i = 1, \dots, n, \quad (2.35)$$

além disso as variâncias  $\sigma_{e_i}^2$  e  $\sigma_{u_i}^2$  foram geradas pelas distribuições uniformes  $\sqrt{\sigma_{e_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 4)$  e  $\sqrt{\sigma_{u_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 1.5)$ ,  $i = 1, \dots, n$  respectivamente e são mantidas fixas em todas as simulações. Os tamanhos amostrais utilizados nas simulações foram  $n = 40, 80$  e  $160$ . Foram realizadas 10.000 simulações Monte Carlo para cada rodada e em cada rodada foram calculadas as estimativas dos estimadores de momentos propostos em (2.12) e (2.13) com seus respectivos erros padrões, estas estimativas de momentos são usadas para iniciar o algoritmo EM proposto na Figura 2.1 com  $\epsilon = 1e - 0.7$ , este algoritmo permite obter as estimativas dos estimadores de máxima verossimilhança com seus erros padrões correspondentes. Se de fato o procedimento de estimação é adequado, espera-se que as estimativas dos parâmetros estejam próximas dos valores dos parâmetros que foram usados na geração dos dados.

Os resultados da simulação são apresentados na Tabela 2.1, onde observamos, que tanto o EMM como o EMV produzem resultados coerentes para os diferentes tamanhos amostrais, isto é, as estimativas dos parâmetros obtidas pelo método dos momentos e de máxima verossimilhança ficam muito perto dos valores verdadeiros utilizados na simulação. Além disso, é possível perceber que os erros padrões (em parênteses) das estimativas diminuem quando o tamanho amostral cresce, além disso, mais importante ainda, o erro padrão das estimativas de máxima verossimilhança é menor que o erro padrão das estimativas de momentos.

Para examinar o comportamento das estatísticas do tipo Wald dadas em (2.18), (2.33), foram feitas simulações de Monte Carlo para calcular as taxas de rejeição do teste utilizando estas estatísticas ao nível nominal de 5%. A configuração para os parâmetros foi a seguinte:

$$x_i \stackrel{i.i.d.}{\sim} N(-2, 4), \quad \sqrt{\sigma_{e_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 4) \quad \text{e} \quad \sqrt{\sigma_{u_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 1.5), \quad i = 1, \dots, n. \quad (2.36)$$

A hipótese nula é  $\mathbf{H}_0 : (\alpha, \beta)^T = (0, 1)^T$  e os valores para  $(\alpha, \beta)$  utilizados para calcular as taxas de rejeição, foi uma grade de pontos igualmente espaçados em torno de  $(0, 1)$  inclusive, a saber:  $(-1, 0.6)$ ,  $(-1, 1)$ ,  $(-1, 1.4)$ ,  $(0, 0.6)$ ,  $(0, 1)$ ,  $(0, 1.4)$ ,  $(1, 0.6)$ ,  $(1, 1)$ ,  $(1, 1.4)$ . Com esta grade de pontos é possível monitorar não apenas o tamanho empírico, mas também o

Tabela 2.1. Estimativas (erros-padrão) dos parâmetros de regressão do modelo estrutural com erros de medição heterocedásticos, no caso  $\sqrt{\sigma_{e_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 4)$  e  $\sqrt{\sigma_{u_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 1.5)$ .

n	Método	$\alpha$	$\beta$	$\mu_x$	$\sigma_x^2$
40	EMM	-0.968 (0.648)	1.016 (0.249)	-1.994 (0.355)	4.009 (1.151)
	EMV	-1.002 (0.415)	0.997 (0.143)	-1.994 (0.3470)	3.903 (0.687)
	Verdadeiro	-1.000	1.000	-2.000	4.000
80	EMM	-0.982 (0.445)	1.009 (0.169)	-1.999 (0.251)	4.002 (0.805)
	EMV	-1.001 (0.298)	1.000 (0.105)	-1.999 (0.246)	3.947 (0.495)
	Verdadeiro	-1.000	1.000	-2.000	4.000
160	EMM	-0.999 (0.325)	1.006 (0.124)	-1.997 (0.179)	3.997 (0.578)
	EMV	-1.000 (0.233)	1.001 (0.083)	-1.997 (0.176)	3.969 (0.358)
	Verdadeiro	-1.000	1.000	-2.000	4.000

poder do teste. Os tamanhos amostrais utilizados nas simulações foram  $n = 40, 80, 160$  e para cada tripla  $(\alpha, \beta, n)$  foram realizadas 10,000 simulações de Monte Carlo, em cada rodada as estatísticas de Wald dadas em (2.18), (2.33) foram usadas para verificar as evidências contra a hipótese nula ao nível nominal de 5%. Se de fato o procedimento de estimação é adequado, espera-se que sob a hipótese nula, as taxas de rejeição sejam próximas de 5%.

A Tabela 2.2 apresenta os resultados da simulação e observamos que as taxas de rejeição sob a hipótese nula tendem a 0.05 (5%) e as taxas de rejeição para valores de  $(\alpha, \beta)$  em torno da hipótese nula tendem a 1.00 (100%) à medida que o tamanho amostral cresce, como esperado.

Tabela 2.2. Taxas de rejeição da hipótese  $\mathbf{H}_0 : (\alpha, \beta)^T = (0, 1)^T$ , usando as estatísticas de Wald dadas em (2.18) e (2.33) (ao nível nominal de 5% de significância) no modelo com erros de medição heterocedásticos, para  $n = 40, 80$  e  $160$  respectivamente.

n	$\alpha$	$\mathcal{W}_{\text{EMM}}$			$\mathcal{W}_{\text{EMV}}$		
		$\beta$			$\beta$		
		0.6	1.0	1.4	0.6	1.0	1.4
40	-1	0.374	0.552	0.992	0.734	0.909	1.000
	0	0.743	<b>0.067</b>	0.545	0.943	<b>0.076</b>	0.933
	1	0.990	0.553	0.285	0.998	0.885	0.603
80	-1	0.642	0.885	0.999	0.974	0.996	1.000
	0	0.896	<b>0.061</b>	0.854	0.997	<b>0.059</b>	0.989
	1	1.000	0.850	0.491	1.000	0.996	0.877
160	-1	0.926	0.987	1.000	0.975	0.999	1.000
	0	0.998	<b>0.054</b>	0.996	1.000	<b>0.054</b>	1.000
	1	1.000	0.992	0.839	1.000	1.000	0.998

## Capítulo 3

# Inferência no Modelo Estrutural Normal com Erros nas Variáveis com $x_i$ Heterocedástica

### 3.1 Introdução

O objetivo deste capítulo será obter estimadores de momentos e estimadores de máxima verossimilhança (usando algoritmo EM), bem como as correspondentes variâncias assintóticas dos estimadores obtidos, num modelo estrutural normal onde a variável não observada  $x_i$  tem distribuição normal heteroscedástica, isto é,  $x_i \stackrel{indep.}{\sim} N(\mu_x, \sigma_{x_i}^2)$ ,  $i = 1, \dots, n$ . Nós consideramos a possibilidade de heterogeneidade na variância da variável não observada  $x_i$ , de modo que temos um modelo de regressão estrutural heteroscedástico (MREH) com erros nas variáveis. Para a estimação dos parâmetros, algumas suposições adicionais devem ser consideradas, como no caso de um modelo de regressão estrutural típico. Nós suporemos uma das seguintes condições de identificabilidade:

- i)  $\sigma_{x_i}^2$  conhecido,  $i = 1, \dots, n$ ;
- ii)  $k_i = \frac{\sigma_{x_i}^2}{\sigma_{x_i}^2 + \sigma_u^2}$  ou, equivalente,  $a_i = \sigma_{x_i}^2 / \sigma_u^2$  conhecido,  $i = 1, \dots, n$ .

### 3.2 Definição do Modelo

Considere o modelo (2.1), onde admitimos, que a variável latente  $x_i$  e o vetor de erros de medição  $(e_i, u_i)^T$  são independentes e normalmente distribuídos, isto é:

$$x_i \stackrel{indep.}{\sim} N(\mu_x, \sigma_{x_i}^2) \text{ e } \begin{pmatrix} e_i \\ u_i \end{pmatrix} \stackrel{i.i.d.}{\sim} N_2 \left[ \mathbf{0}, \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right], i = 1, \dots, n. \quad (3.1)$$

Isto é equivalente a supor que o vetor aleatório  $(e_i, u_i, x_i)^T$ , tem a seguinte distribuição:

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{indep.}}{\sim} N_3 \left[ \begin{pmatrix} 0 \\ 0 \\ \mu_x \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_{x_i}^2 \end{pmatrix} \right], \quad (3.2)$$

o que leva ao modelo estrutural heteroscedástico dado por:

$$\begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix} \stackrel{\text{indep.}}{\sim} N_3 \left[ \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \\ \mu_x \end{pmatrix}; \begin{pmatrix} \beta^2\sigma_{x_i}^2 + \sigma_e^2 & \beta\sigma_{x_i}^2 & \beta\sigma_{x_i}^2 \\ \beta\sigma_{x_i}^2 & \sigma_{x_i}^2 + \sigma_u^2 & \sigma_{x_i}^2 \\ \beta\sigma_{x_i}^2 & \sigma_{x_i}^2 & \sigma_{x_i}^2 \end{pmatrix} \right], \quad i = 1, \dots, n. \quad (3.3)$$

Se as variâncias  $\sigma_{x_i}^2$ ,  $i = 1, \dots, n$  são conhecidas, então o modelo torna-se identificável e o vetor de parâmetros desconhecidos é dado por  $\theta = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2)^T$ .

### 3.3 Estimadores de Momentos

A distribuição conjunta dos dados observados  $\mathbf{Z}_i = (Y_i, X_i)^T$ ,  $i = 1, \dots, n$ , é dada por:  $\mathbf{Z}_i \stackrel{\text{indep.}}{\sim} N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$ ,  $i = 1, \dots, n$ , onde

$$\boldsymbol{\mu} = \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix} = \mathbf{a} + \mathbf{b}\mu_x, \quad (3.4)$$

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \beta^2\sigma_{x_i}^2 + \sigma_e^2 & \beta\sigma_{x_i}^2 \\ \beta\sigma_{x_i}^2 & \sigma_{x_i}^2 + \sigma_u^2 \end{pmatrix} = \sigma_{x_i}^2 \mathbf{b}\mathbf{b}^T + \boldsymbol{\Psi}, \quad (3.5)$$

com

$$\mathbf{a} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \beta \\ 1 \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Psi} = \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}. \quad (3.6)$$

O seguinte resultado fornece o estimador de momentos do parâmetro  $\theta$ .

**Proposição 3.1.** Sejam  $\bar{\mathbf{Z}}$  e  $\mathbf{S}_Z$  o vetor de médias e a matriz de covariâncias amostrais, isto é,

$$\bar{\mathbf{Z}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \quad \text{e} \quad \mathbf{S}_Z = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T. \quad \text{Então,}$$

$$E(\bar{\mathbf{Z}}) = \boldsymbol{\mu} \quad \text{e} \quad E(\mathbf{S}) = \bar{\sigma}_x^2 \mathbf{b}\mathbf{b}^T + \boldsymbol{\Psi}, \quad \text{onde} \quad \bar{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^n \sigma_{x_i}^2.$$

**Prova:** A demonstração é similar à da Proposição 2.1.

A primeira parte é direta, pois:

$$E(\bar{\mathbf{Z}}) = E\left(\frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i\right) = \frac{1}{n} \sum_{i=1}^n E(\mathbf{Z}_i) = \boldsymbol{\mu}.$$

Para a segunda parte notemos em primeiro lugar que:

$$\begin{aligned} \text{Var}(\mathbf{Z}_i - \bar{\mathbf{Z}}) &= \text{Var}(\mathbf{Z}_i) + \text{Var}(\bar{\mathbf{Z}}) - 2\text{Cov}(\mathbf{Z}_i, \bar{\mathbf{Z}}) \\ &= \boldsymbol{\Sigma}_i + \frac{1}{n^2} \sum_{i=1}^n \boldsymbol{\Sigma}_i - \frac{2}{n} \boldsymbol{\Sigma}_i \\ &= \frac{(n-2)\boldsymbol{\Sigma}_i}{n} + \frac{\bar{\boldsymbol{\Sigma}}}{n} \\ &= \frac{(n-2)(\sigma_{x_i}^2 \mathbf{b}\mathbf{b}^T + \boldsymbol{\Psi})}{n} + \frac{(\bar{\sigma}_x^2 \mathbf{b}\mathbf{b}^T + \boldsymbol{\Psi})}{n} \\ &= \frac{((n-2)\sigma_{x_i}^2 + \bar{\sigma}_x^2) \mathbf{b}\mathbf{b}^T}{n} + \frac{(n-1)\boldsymbol{\Psi}}{n} \end{aligned}$$

sendo que  $\bar{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\Sigma}_i = \bar{\sigma}_x^2 \mathbf{b}\mathbf{b}^T + \boldsymbol{\Psi}$ . Levando em conta que  $E(\mathbf{Z}_i - \bar{\mathbf{Z}}) = 0$ , temos que

$$\text{Var}(\mathbf{Z}_i - \bar{\mathbf{Z}}) = E[(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T].$$

Assim,

$$\begin{aligned} E[\mathbf{S}_{\mathbf{Z}}] &= \frac{1}{n-1} \sum_{i=1}^n E[(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T] \\ &= \frac{1}{n-1} \sum_{i=1}^n \text{Var}(\mathbf{Z}_i - \bar{\mathbf{Z}}) \\ &= \frac{1}{n-1} \sum_{i=1}^n \left( \frac{((n-2)\sigma_{x_i}^2 + \bar{\sigma}_x^2) \mathbf{b}\mathbf{b}^T}{n} + \frac{(n-1)\boldsymbol{\Psi}}{n} \right) \\ &= \frac{(n-2)\bar{\sigma}_x^2 \mathbf{b}\mathbf{b}^T}{n-1} + \frac{\bar{\sigma}_x^2 \mathbf{b}\mathbf{b}^T}{n-1} + \boldsymbol{\Psi} \\ &= \bar{\sigma}_x^2 \mathbf{b}\mathbf{b}^T + \boldsymbol{\Psi}, \end{aligned}$$

□



concluindo a prova.

### 3.3.1 EMM supondo $\bar{\sigma}_x^2$ conhecido

Sob a restrição de identificabilidade  $\bar{\sigma}_x^2$  conhecida ou  $\sigma_{x_i}^2$ ,  $i = 1, \dots, n$  conhecidas, os estimadores de momentos para  $\alpha$ ,  $\beta$ ,  $\mu_x$ ,  $\sigma_e^2$  e  $\sigma_u^2$  podem ser obtidos usando a Proposição 3.1, com  $\bar{\mathbf{Z}}$  e  $\mathbf{S}_Z$  definidos em (2.10).

Neste caso temos as seguintes equações de momentos:

$$\mu(\hat{\boldsymbol{\theta}}) = \bar{\mathbf{Z}}, \quad \bar{\sigma}_x^2 \mathbf{b}(\hat{\boldsymbol{\theta}}) \mathbf{b}^T(\hat{\boldsymbol{\theta}}) + \boldsymbol{\Psi}(\hat{\boldsymbol{\theta}}) = \mathbf{S}_Z.$$

Das equações acima temos:

$$\hat{\alpha} + \hat{\beta} \hat{\mu}_x = \bar{Y} \implies \hat{\alpha} = \bar{Y} - \hat{\beta} \hat{\mu}_x \quad ; \quad \hat{\mu}_x = \bar{X},$$

$$\hat{\beta}^2 \bar{\sigma}_x^2 + \hat{\sigma}_e^2 = S_Y^2 \implies \hat{\sigma}_e^2 = S_Y^2 - \hat{\beta}^2 \bar{\sigma}_x^2, \quad \text{se } S_Y^2 - \hat{\beta}^2 \bar{\sigma}_x^2 > 0,$$

$$\bar{\sigma}_x^2 + \hat{\sigma}_u^2 = S_X^2 \implies \hat{\sigma}_u^2 = S_X^2 - \bar{\sigma}_x^2, \quad \text{se } S_X^2 - \bar{\sigma}_x^2 > 0,$$

$$\hat{\beta} \bar{\sigma}_x^2 = S_{YX} \implies \hat{\beta} = \frac{S_{YX}}{\bar{\sigma}_x^2},$$

de modo que os seguintes estimadores de momentos são obtidos:

$$\hat{\mu}_{x\text{EMM}} = \bar{X}, \quad \hat{\alpha}_{\text{EMM}} = \bar{Y} - \hat{\beta}_{\text{EMM}} \bar{X}, \quad \hat{\beta}_{\text{EMM}} = \frac{S_{YX}}{\bar{\sigma}_x^2},$$

$$\hat{\sigma}_{u\text{EMM}}^2 = S_X^2 - \bar{\sigma}_x^2, \quad \text{se } S_X^2 - \bar{\sigma}_x^2 > 0,$$

$$\hat{\sigma}_{e\text{EMM}}^2 = S_Y^2 - \frac{S_{YX}^2}{\bar{\sigma}_x^2}, \quad \text{se } S_Y^2 - \frac{S_{YX}^2}{\bar{\sigma}_x^2} > 0.$$



### 3.3.2 EMM supondo $\bar{a}$ ou $a_i = \frac{\sigma_{x_i}^2}{\sigma_u^2}$ , $i = 1, \dots, n$ conhecidas

Notemos que

$$\frac{1}{n} \sum_{i=1}^n a_i = \frac{1}{n\sigma_u^2} \sum_{i=1}^n \sigma_{x_i}^2 \implies \bar{a} = \frac{1}{\sigma_u^2} \bar{\sigma}_x^2 \implies \bar{\sigma}_x^2 = \bar{a}\sigma_u^2.$$

Substituindo  $\bar{\sigma}_x^2 = \bar{a}\sigma_u^2$  na Proposição 3.1, as equações de momentos ficam dadas por:

$$\mu(\hat{\theta}) = \bar{\mathbf{Z}}, \quad \bar{a}\hat{\sigma}_u^2 \mathbf{b}(\hat{\theta}) \mathbf{b}^T(\hat{\theta}) + \Psi(\hat{\theta}) = \mathbf{S}_Z.$$

Das equações acima temos:

$$\hat{\alpha} + \hat{\beta}\hat{\mu}_x = \bar{Y} \implies \hat{\alpha} = \bar{Y} - \hat{\beta}\hat{\mu}_x \quad ; \quad \hat{\mu}_x = \bar{X}$$

$$\hat{\beta}^2 \bar{a} \hat{\sigma}_u^2 + \hat{\sigma}_e^2 = S_Y^2 \implies \hat{\sigma}_e^2 = S_Y^2 - \hat{\beta}^2 \bar{a} \hat{\sigma}_u^2, \quad \text{se } S_Y^2 - \hat{\beta}^2 \bar{a} \hat{\sigma}_u^2 > 0$$

$$\bar{a} \hat{\sigma}_u^2 + \hat{\sigma}_u^2 = S_X^2 \implies \hat{\sigma}_u^2 = \frac{S_X^2}{(1 + \bar{a})}$$

$$\hat{\beta} \bar{a} \hat{\sigma}_u^2 = S_{YX} \implies \hat{\beta} = \frac{S_{YX}}{\bar{a} \hat{\sigma}_u^2}.$$

de modo que os seguintes estimadores de momentos são obtidos:

$$\hat{\mu}_{x_{\text{EMM}}} = \bar{X}, \quad \hat{\alpha}_{\text{EMM}} = \bar{Y} - \hat{\beta}_{\text{EMM}} \bar{X}, \quad \hat{\beta}_{\text{EMM}} = \frac{(1 + \bar{a}) S_{YX}}{\bar{a} S_X^2},$$

$$\hat{\sigma}_{u_{\text{EMM}}}^2 = \frac{S_X^2}{(1 + \bar{a})}, \quad \hat{\sigma}_{e_{\text{EMM}}}^2 = S_Y^2 - \frac{(1 + \bar{a}) S_{YX}^2}{\bar{a} S_x^2}, \quad \text{se } S_Y^2 > \frac{(1 + \bar{a}) S_{YX}^2}{\bar{a} S_x^2}.$$

## 3.4 Distribuição Assintótica dos Estimadores de Momentos

Como os estimadores de momentos tem solução explícita podemos usar o método delta para obter a distribuição assintótica destes estimadores. Neste caso, nosso interesse será concentrado nos parâmetros  $\alpha$  e  $\beta$ . Denotaremos por  $\hat{\theta}_{1_{\text{EMM}}} = (\hat{\alpha}_{\text{EMM}}, \hat{\beta}_{\text{EMM}})^T$  o estimador de momentos de nosso parâmetro de interesse  $\theta_1 = (\alpha, \beta)^T$ , subconjunto de  $\theta$ .

### 3.4.1 Distribuição assintótica de $\hat{\theta}_{1\text{EMM}}$ supondo $\bar{\sigma}_x^2$ conhecido

Suposições adicionais:

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{\sigma_{x_i}^2}{n} = t^*, \quad \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{(\sigma_{x_i}^2)^2}{n} = t^{**}. \quad (3.7)$$

O próximo resultado fornece a distribuição assintótica do estimador  $\hat{\theta}_{1\text{EMM}}$ .

**Proposição 3.2.** Seja  $\hat{\theta}_{1\text{EMM}} = (\bar{Y} - \hat{\beta}_{\text{EMM}} \bar{X}, \frac{S_{YX}}{\bar{\sigma}_x^2})^T$  o estimador de momentos do parâmetro  $\theta_1 = (\alpha, \beta)^T$ . Se as suposições em (3.7) são satisfeitas, então

$$\sqrt{n}(\hat{\theta}_{1\text{EMM}} - \theta_1) \xrightarrow{D} N_2[\mathbf{0}_2, \Delta(\theta)], \text{ quando } n \rightarrow \infty,$$

onde

$$\Delta(\theta) = \begin{bmatrix} \Delta_{11} & -\mu_x(2\beta^4 t^{**} + 4\beta^2 \sigma_e^2 t^* + 2\sigma_e^4) \\ -\mu_x(2\beta^4 t^{**} + 4\beta^2 \sigma_e^2 t^* + 2\sigma_e^4) & 2\beta^4 t^{**} + 4\beta^2 \sigma_e^2 t^* + 2\sigma_e^4 \end{bmatrix},$$

em que

$$\Delta_{11} = 2\beta^4 \mu_x^2 t^{**} + 4\beta^2 \mu_x^2 \sigma_e^2 t^* + 2\mu_x^2 \sigma_e^4 + \beta^2 \sigma_u^2 + \sigma_e^2.$$

**Prova:** Usando os resultados assintóticos dados em (2.16) e (2.17) (Arellano-Valle et al. (2002)) temos que

$$\Sigma_* = \Gamma = \Gamma_* = \Lambda_* = \mathbf{0},$$

$$\Sigma = \begin{pmatrix} \beta^2 t^* + \sigma_e^2 & \beta t^* \\ \beta t^* & \sigma_u^2 + t^* \end{pmatrix}$$

$$\Lambda = \begin{bmatrix} 2\beta^4 t^{**} + 4\beta^2 \sigma_e^2 t^* + 2\sigma_e^4 & 2\beta^2 t^{**} & 2\beta^3 t^{**} + 2\beta \sigma_e^2 t^* \\ 2\beta^2 t^{**} & 2t^{**} + 4\sigma_u^2 t^* + 2\sigma_u^4 & 2\beta t^{**} + 2\beta \sigma_u^2 t^* \\ 2\beta^3 t^{**} + 2\beta \sigma_e^2 t^* & 2\beta t^{**} + 2\beta \sigma_u^2 t^* & 2\beta^2 t^{**} + \beta^2 \sigma_u^2 t^* + \sigma_e^2 t^* + \sigma_e^2 \sigma_u^2 \end{bmatrix}.$$

Assim,

$$\sqrt{n} \begin{pmatrix} \bar{Z} - \mu \\ \text{vech}(\mathbf{S}_Z - \Sigma) \end{pmatrix} = \sqrt{n} \begin{bmatrix} \bar{Y} - \alpha - \beta \mu_x \\ \bar{X} - \mu_x \\ S_Y^2 - \beta^2 t^* - \sigma_e^2 \\ S_X^2 - \sigma_u^2 - t^* \\ S_{YX} - \beta t^* \end{bmatrix} \xrightarrow{D} N_5 \left[ \begin{pmatrix} \mathbf{0}_2 \\ \mathbf{0}_3 \end{pmatrix}, \begin{pmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix} \right].$$

Agora, seja  $\phi = (\alpha + \beta \mu_x, \mu_x, \beta, \sigma_u^2, \sigma_e^2)^T$  e definamos a função  $g : \mathbb{R}^5 \rightarrow \mathbb{R}^2$  tal que:  $g(\mathbf{x}) = g(x_1, x_2, x_3, x_4, x_5) = (x_1 - x_3 x_2, x_3)^T$  para todo  $\mathbf{x} \in \mathbb{R}^5$ . Assim,

$$\hat{\theta}_{1\text{EMM}} = g(\bar{Y}, \bar{X}, \frac{S_{YX}}{\bar{\sigma}_x^2}, S_X^2, S_Y^2)^T = \begin{bmatrix} \bar{Y} - \frac{S_{YX}}{\bar{\sigma}_x^2} \bar{X} \\ \frac{S_{YX}}{\bar{\sigma}_x^2} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_{\text{EMM}} \\ \hat{\beta}_{\text{EMM}} \end{bmatrix}$$

e

$$g(\phi) = g(\alpha + \beta \mu_x, \mu_x, \beta, \sigma_u^2, \sigma_e^2)^T = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

Então

$$\dot{G}(\phi) = \frac{\partial}{\partial \mathbf{x}^T} \mathbf{g}(\mathbf{x}) \Big|_{\mathbf{x}=\phi} = \begin{bmatrix} 1 & -\beta & -\mu_x & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Então pelo método delta, veja Sen and Singer (1993), Teorema 3.4.6 temos que:

$$\sqrt{n}(\hat{\theta}_{1\text{EMM}} - \theta_1) \xrightarrow{D} N_2[\mathbf{0}_2, \Delta(\theta)], \text{ quando } n \rightarrow \infty, \text{ onde}$$

$$\Delta(\theta) = \dot{G}(\phi)\Delta_1\dot{G}(\phi)^T, \quad \text{onde} \quad \Delta_1 = \begin{pmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix}.$$

Finalmente fazendo o produto  $\dot{G}(\phi)\Delta_1\dot{G}(\phi)^T$ , algumas manipulações algébricas conduzem à expressão dada para  $\Delta(\theta)$ .  $\square$

Da Proposição 3.2 temos que quando o tamanho da amostra tende ao infinito ( $n \rightarrow \infty$ ),  $\hat{\theta}_{1_{EMM}}$  é distribuída assintoticamente normal com vetor de médias  $\theta_1$  e matriz de covariância  $\frac{1}{n}\Delta(\theta)$ , isto é

$$\hat{\theta}_{1_{EMM}} \xrightarrow{a} N_2[\theta_1, \text{VA}(\hat{\theta}_{1_{EMM}})], \quad \text{onde} \quad \text{VA}(\hat{\theta}_{1_{EMM}}) = \frac{1}{n}\Delta(\theta).$$

Então, para os componentes do vetor  $\theta_1$ , temos que:

$$\text{VA}(\hat{\alpha}_{EMM}) = \frac{1}{n}(2\beta^4\mu_x^2t^{**} + 4\beta^2\mu_x^2\sigma_e^2t^* + 2\mu_x^2\sigma_e^4 + \beta^2\sigma_u^2 + \sigma_e^2) \quad e$$

$$\text{VA}(\hat{\beta}_{EMM}) = \frac{1}{n}(2\beta^4t^{**} + 4\beta^2\sigma_e^2t^* + 2\sigma_e^4).$$

### 3.4.2 Distribuição assintótica de $\hat{\theta}_{1_{EMM}}$ supondo $\bar{a}$ conhecido

Além das suposições já feitas anteriormente, vamos supor que:

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{a_i}{n} = \lambda^*, \quad \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{a_i^2}{n} = \lambda^{**}. \quad (3.8)$$

O próximo resultado fornece a distribuição assintótica do estimador  $\hat{\theta}_{1_{EMM}}$ .

**Proposição 3.3.** Seja  $\hat{\theta}_{1_{EMM}} = (\bar{Y} - \hat{\beta}_{EMM}\bar{X}, \frac{(1+\bar{a})S_{YX}}{\bar{a}S_X^2})^T$  o estimador de momentos do parâmetro  $\theta_1 = (\alpha, \beta)^T$ . Se as suposições em (3.8) são satisfeitas, então

$$\sqrt{n}(\hat{\theta}_{1_{EMM}} - \theta_1) \xrightarrow{D} N_2[\mathbf{0}_2, \Upsilon(\theta)], \quad \text{quando } n \rightarrow \infty,$$

onde

$$\Upsilon(\theta) = \begin{bmatrix} \Upsilon_{11} & \Upsilon_{12} \\ \Upsilon_{21} & \Upsilon_{22} \end{bmatrix},$$

com

$$\begin{aligned} \Upsilon_{11} &= \frac{1}{\sigma_u^4} \left[ \sigma_e^2 \sigma_u^4 + \beta^2 \sigma_u^6 + 2\beta^6 \mu_x^2 \sigma_u^4 \lambda^{**} + 4\beta^4 \mu_x^2 \sigma_u^2 \sigma_e^2 \lambda^* + 2\beta^2 \mu_x^2 \sigma_e^4 \right. \\ &\quad \left. - 4\beta^3 \mu_x^2 \sigma_u^4 \lambda^{**} + 2\sigma_u^4 \mu_x^2 \lambda^{**} + 4\sigma_u^4 \mu_x^2 \lambda^* + 2\sigma_u^4 \mu_x^2 \right], \\ \Upsilon_{12} = \Upsilon_{21} &= \frac{-2\mu_x}{\sigma_u^4} \left[ \beta^6 \sigma_u^4 \lambda^{**} + 2\beta^4 \sigma_u^2 \sigma_e^2 \lambda^* + \beta^2 \sigma_e^4 - \beta^3 \sigma_u^4 \lambda^{**} + \sigma_u^4 \lambda^{**} \right. \\ &\quad \left. + 2\sigma_u^4 \lambda^* + \sigma_u^4 \right], \\ \Upsilon_{22} &= \frac{2}{\sigma_u^4} \left[ \beta^6 \sigma_u^4 \lambda^{**} + 2\beta^4 \sigma_u^2 \sigma_e^2 \lambda^* + \beta^2 \sigma_e^4 - 2\beta^3 \sigma_u^4 \lambda^{**} + \sigma_u^4 \lambda^{**} \right. \\ &\quad \left. + 2\sigma_u^4 \lambda^* + \sigma_u^4 \right]. \end{aligned}$$

**Prova:** De modo similar ao caso anterior, usando os resultados assintóticos dados em (2.16) e (2.17) temos que:  $\Sigma_* = \Gamma = \Gamma_* = \Lambda_* = \mathbf{0}$ ,

$$\Sigma = \begin{bmatrix} \beta^2 \sigma_u^2 \lambda^* + \sigma_e^2 & \beta \sigma_u^2 \lambda^* \\ \beta \sigma_u^2 \lambda^* & (1 + \lambda^*) \sigma_u^2 \end{bmatrix}$$

e

$$\Lambda = \begin{bmatrix} \Lambda_{11} & 2\beta^2 \sigma_u^2 \lambda^{**} & 2\beta^3 \sigma_u^4 \lambda^{**} + 2\beta \sigma_e^2 \sigma_u^2 \lambda^* \\ 2\beta^2 \sigma_u^2 \lambda^{**} & 2\sigma_u^2 \lambda^{**} + 4\sigma_u^4 \lambda^* + 2\sigma_e^4 & 2\beta \sigma_u^4 \lambda^{**} + 2\beta \sigma_u^4 \lambda^* \\ 2\beta^3 \sigma_u^4 \lambda^{**} + 2\beta \sigma_e^2 \sigma_u^2 \lambda^* & 2\beta \sigma_u^4 \lambda^{**} + 2\beta \sigma_u^4 \lambda^* & \Lambda_{22} \end{bmatrix},$$

com

$$\Lambda_{11} = 2\beta^4 \sigma_u^2 \lambda^{**} + 4\beta^2 \sigma_e^2 \sigma_u^2 \lambda^* + 2\sigma_e^4$$

e

$$\Lambda_{22} = 2\beta^2 \sigma_u^4 \lambda^{**} + \beta^2 \sigma_u^4 \lambda^* + \sigma_c^2 \sigma_u^2 \lambda^* + \sigma_c^2 \sigma_u^2.$$

Assim,

$$\sqrt{n} \begin{pmatrix} \bar{\mathbf{Z}} - \boldsymbol{\mu} \\ \text{vech}(\mathbf{S}_Z - \boldsymbol{\Sigma}) \end{pmatrix} = \sqrt{n} \begin{bmatrix} \bar{Y} - \alpha - \beta \mu_x \\ \bar{X} - \mu_x \\ S_Y^2 - \beta^2 \sigma_u^2 \lambda^* - \sigma_c^2 \\ S_X^2 - (1 + \lambda^*) \sigma_u^2 \\ S_{YX} - \beta \sigma_u^2 \lambda^* \end{bmatrix} \xrightarrow{D} N_5 \left[ \begin{pmatrix} \mathbf{0}_2 \\ \mathbf{0}_3 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda} \end{pmatrix} \right].$$

Agora, seja  $\boldsymbol{\phi} = (\alpha + \beta \mu_x, \mu_x, \sigma_u^2, \beta \sigma_u^2, \sigma_c^2)^T$  e definamos a função  $g : \mathbb{R}^5 \rightarrow \mathbb{R}^2$  tal que:  $g(\mathbf{x}) = g(x_1, x_2, x_3, x_4, x_5) = (x_1 - \frac{x_1}{x_3} x_2, \frac{x_1}{x_3})^T$  para todo  $\mathbf{x} \in \mathbb{R}^5$ .

Assim

$$\hat{\boldsymbol{\theta}}_{\text{EMM}} = g(\bar{Y}, \bar{X}, \bar{a} S_X^2, (1 + \bar{a}) S_{YX}, S_Y^2)^T = \begin{bmatrix} \bar{Y} - \frac{(1+\bar{a})}{\bar{a}} \frac{S_{YX}}{S_X^2} \bar{X} \\ \frac{(1+\bar{a})}{\bar{a}} \frac{S_{YX}}{S_X^2} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_{\text{EMM}} \\ \hat{\beta}_{\text{EMM}} \end{bmatrix}$$

e

$$g(\boldsymbol{\phi}) = g(\alpha + \beta \mu_x, \mu_x, \sigma_u^2, \beta \sigma_u^2, \sigma_c^2)^T = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

Então

$$\dot{\mathbf{G}}(\boldsymbol{\phi}) = \frac{\partial}{\partial \mathbf{x}^T} \mathbf{g}(\mathbf{x}) \Big|_{\mathbf{x}=\boldsymbol{\phi}} = \begin{bmatrix} 1 & -\beta & \frac{\beta \mu_x}{\sigma_u^2} & \frac{-\mu_x}{\sigma_u^2} & 0 \\ 0 & 0 & \frac{-\beta}{\sigma_u^2} & \frac{1}{\sigma_u^2} & 0 \end{bmatrix}.$$

Então, pelo método delta, veja Sen and Singer (1993), Teorema 3.4.6, temos que:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{EMM}} - \boldsymbol{\theta}_1) \xrightarrow{D} N_2[\mathbf{0}_2, \boldsymbol{\Upsilon}(\boldsymbol{\theta})], \text{ quando } n \rightarrow \infty, \text{ onde}$$

$$\boldsymbol{\Upsilon}(\boldsymbol{\theta}) = \dot{\mathbf{G}}(\boldsymbol{\phi}) \boldsymbol{\Upsilon}_1 \dot{\mathbf{G}}(\boldsymbol{\phi})^T, \text{ onde } \boldsymbol{\Upsilon}_1 = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda} \end{pmatrix}.$$

Finalmente fazendo o produto  $\dot{\mathbf{G}}(\phi)\Upsilon_1\dot{\mathbf{G}}(\phi)^T$ , algumas manipulações algébricas conduzem à expressão dada para  $\Upsilon(\theta)$ .  $\square$

Da Proposição 3.3 temos que quando o tamanho da amostra tende ao infinito ( $n \rightarrow \infty$ ),  $\hat{\theta}_{1\text{EMM}}$  é distribuída assintoticamente normal com vetor de média  $\theta_1$  e matriz de covariância  $\frac{1}{n}\Upsilon(\theta)$ , isto é,

$$\hat{\theta}_{1\text{EMM}} \xrightarrow{a} N_2[\theta_1, \text{VA}(\hat{\theta}_{1\text{EMM}})], \quad \text{onde} \quad \text{VA}(\hat{\theta}_{1\text{EMM}}) = \frac{1}{n}\Upsilon(\theta). \text{ Então,}$$

$$\begin{aligned} \text{VA}(\hat{\alpha}_{\text{EMM}}) &= \frac{1}{n\sigma_u^4} \left[ \sigma_e^2\sigma_u^4 + \beta^2\sigma_u^6 + 2\beta^6\mu_x^2\sigma_u^4\lambda^{**} + 4\beta^4\mu_x^2\sigma_u^2\sigma_e^2\lambda^* + 2\beta^2\mu_x^2\sigma_e^4 \right. \\ &\quad \left. - 4\beta^3\mu_x^2\sigma_u^4\lambda^{**} + 2\sigma_u^4\mu_x^2\lambda^{**} + 4\sigma_u^4\mu_x^2\lambda^* + 2\sigma_u^4\mu_x^2 \right]. \end{aligned}$$

$$\begin{aligned} \text{VA}(\hat{\beta}_{\text{EMM}}) &= \frac{2}{n\sigma_u^4} \left[ \beta^6\sigma_u^4\lambda^{**} + 2\beta^4\sigma_u^2\sigma_e^2\lambda^* + \beta^2\sigma_e^4 - 2\beta^3\sigma_u^4\lambda^{**} + \sigma_u^4\lambda^{**} \right. \\ &\quad \left. + 2\sigma_u^4\lambda^* + \sigma_u^4 \right]. \end{aligned}$$

### 3.5 Estimação de Máxima Verossimilhança

Sob as suposições em (3.1), (3.2) e assumindo  $\sigma_{x_1}^2, \dots, \sigma_{x_n}^2$  conhecidas, a distribuição conjunta do vetor de dados observados,  $\mathbf{Z}_i = \mathbf{a} + x_i\mathbf{b} + (e_i, u_i)^T$ , é dada por:

$$\mathbf{Z}_i \stackrel{\text{indep.}}{\sim} N_2(\mathbf{a} + \mu_x\mathbf{b}, \Sigma_i), \quad i = 1, \dots, n, \quad (3.9)$$

onde  $\Sigma_i$  e os vetores  $\mathbf{a}$  e  $\mathbf{b}$  são definidos em (3.5) e (3.6) respectivamente.

O determinante e inversa da matriz de covariâncias  $\Sigma_i$ , são dados por

$$|\Sigma_i| = |\Psi|(1 + \sigma_{x_i}^2\mathbf{b}^T\Psi^{-1}\mathbf{b}) = \sigma_e^2\sigma_u^2\sigma_{x_i}^2 c_i \quad \text{e} \quad \Sigma_i^{-1} = \Psi^{-1} - c_i^{-1}\Psi^{-1}\mathbf{b}\mathbf{b}^T\Psi^{-1},$$



onde

$$c_i = \sigma_{x_i}^{-2}(1 + \sigma_{x_i}^2 \mathbf{b}^T \Psi^{-1} \mathbf{b}), \quad i = 1, \dots, n,$$

com  $\Psi$  definida em (3.6).

### 3.5.1 Função de Log-Verossimilhança

Seja  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2)^T$  nosso parâmetro de interesse. A função de log-verossimilhança correspondente ao modelo definido por (3.9) pode ser escrita como

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n l_i(\boldsymbol{\theta}), \quad \text{onde}$$

$$l_i(\boldsymbol{\theta}) = \text{constante} - \frac{1}{2} \log |\Sigma_i| - \frac{1}{2} (\mathbf{Z}_i - \mathbf{a} - \mu_x \mathbf{b})^T \Sigma_i^{-1} (\mathbf{Z}_i - \mathbf{a} - \mu_x \mathbf{b}).$$

Após várias manipulações algébricas obtemos a seguinte expressão para o log-verossimilhança correspondente à  $i$ -ésima observação

$$l_i(\boldsymbol{\theta}) \propto -\frac{1}{2} \log |\Sigma_i| - \frac{1}{2} [t_i - c_i^{-1} h_i^2], \quad (3.10)$$

onde

$$t_i = \sigma_e^{-2} (Y_i - \alpha - \beta \mu_x)^2 + \sigma_u^{-2} (X_i - \mu_x)^2$$

$$h_i = \beta \sigma_e^{-2} (Y_i - \alpha - \beta \mu_x) + \sigma_u^{-2} (X_i - \mu_x), \quad i = 1, \dots, n.$$

### 3.5.2 Estimação de Máxima Verossimilhança usando o Algoritmo EM

Das suposições feitas para este modelo em (3.1) e (3.2) segue que a distribuição conjunta do vetor de dados completos  $\mathbf{W}_i$ , é dada por

$$\mathbf{W}_i \stackrel{\text{indep.}}{\sim} N_3(\boldsymbol{\nu}, \boldsymbol{\Omega}_i), \quad i = 1, \dots, n, \quad (3.11)$$

onde  $\mathbf{v}$  e  $\Omega_i$  são dadas como em (3.3), isto é,

$$\mathbf{v} = \begin{pmatrix} \mathbf{a} + \mathbf{b}\mu_x \\ \mu_x \end{pmatrix} \quad \text{e} \quad \Omega_i = \begin{pmatrix} \sigma_{x_i}^2 \mathbf{b}\mathbf{b}^T + \Psi & \sigma_{x_i}^2 \mathbf{b} \\ \sigma_{x_i}^2 \mathbf{b}^T & \sigma_{x_i}^2 \end{pmatrix}, \quad i = 1, \dots, n. \quad (3.12)$$

Usando os resultados em Muirhead (1982) (Teoremas A.5.2 e A.5.3), o determinante e inversa da matriz de covariâncias  $\Omega_i$ , são dados por

$$|\Omega_i| = \sigma_{x_i}^2 \sigma_e^2 \sigma_u^2 \quad \text{e} \quad \Omega_i^{-1} = \begin{pmatrix} \Psi^{-1} & -\Psi^{-1}\mathbf{b} \\ -\mathbf{b}^T\Psi^{-1} & c_i \end{pmatrix}. \quad (3.13)$$

A função de log-verossimilhança correspondente ao modelo definido em (3.11) pode ser escrita como

$$l_c(\boldsymbol{\theta}) = \sum_{i=1}^n l_i(\boldsymbol{\theta}), \quad \text{onde}$$

$$l_{c_i}(\boldsymbol{\theta}) = \text{constante} - \frac{1}{2} \log|\Omega_i| - \frac{1}{2} (\mathbf{w}_i - \mathbf{v})^T \Omega_i^{-1} (\mathbf{w}_i - \mathbf{v}),$$

substituindo (3.13) em  $l_{c_i}(\boldsymbol{\theta})$  e, após várias manipulações algébricas obtemos,

$$\begin{aligned} l_{c_i}(\boldsymbol{\theta}) \propto & -\frac{1}{2} \log(\sigma_e^2) - \frac{1}{2} \log(\sigma_u^2) - \frac{1}{2} \log(\sigma_{x_i}^2) - \frac{1}{2} \sigma_e^{-2} (Y_i - \alpha - \beta\mu_x)^2 \\ & + \beta \sigma_e^{-2} (Y_i - \alpha - \beta\mu_x) x_i - \beta \sigma_e^{-2} \mu_x (Y_i - \alpha - \beta\mu_x) \\ & - \frac{1}{2} \sigma_u^{-2} (X_i - \mu_x)^2 + \sigma_u^{-2} (X_i - \mu_x) x_i - \sigma_u^{-2} \mu_x (X_i - \mu_x) \\ & - \frac{1}{2} c_i x_i^2 + c_i \mu_x x_i - \frac{1}{2} c_i \mu_x^2. \end{aligned}$$

Então, na etapa E do algoritmo EM temos que

$$Q_i(\boldsymbol{\theta}) = E(l_{c_i}(\boldsymbol{\theta} | \mathbf{W}_i) | Y_i, X_i) = E(l_{c_i}(\boldsymbol{\theta} | x_i) | Y_i, X_i),$$

é dada por

$$\begin{aligned}
Q_i(\boldsymbol{\theta}) \propto & -\frac{1}{2} \log(\sigma_e^2) - \frac{1}{2} \log(\sigma_u^2) - \frac{1}{2} \log(\sigma_{x_i}^2) - \frac{1}{2} \sigma_e^{-2} (Y_i - \alpha - \beta \mu_x)^2 \\
& + \beta \sigma_e^{-2} (Y_i - \alpha - \beta \mu_x) \widehat{x}_i - \beta \sigma_e^{-2} \mu_x (Y_i - \alpha - \beta \mu_x) \\
& - \frac{1}{2} \sigma_u^{-2} (X_i - \mu_x)^2 + \sigma_u^{-2} (X_i - \mu_x) \widehat{x}_i - \sigma_u^{-2} \mu_x (X_i - \mu_x) \\
& - \frac{1}{2} c_i \widehat{x}_i^2 + c_i \mu_x \widehat{x}_i - \frac{1}{2} c_i \mu_x^2,
\end{aligned} \tag{3.14}$$

onde

$$\widehat{x}_i = E(x_i | Y_i, X_i) \quad \text{e} \quad \widehat{x}_i^2 = v(x_i) + (\widehat{x}_i)^2,$$

com

$$\begin{aligned}
E(x_i | Y_i, X_i) &= \mu_x + c_i^{-1} h_i \\
V(x_i) &= \text{Var}(x_i | Y_i, X_i) = c_i^{-1} = \frac{\sigma_{x_i}^2 \sigma_e^2 \sigma_u^2}{(\beta^2 \sigma_{x_i}^2 + \sigma_e^2) \sigma_u^2 + \sigma_e^2 \sigma_{x_i}^2}, \quad i = 1, \dots, n.
\end{aligned}$$

Na etapa M, temos que maximizar  $Q(\boldsymbol{\theta}) = \sum_{i=1}^n Q_i(\boldsymbol{\theta})$ , com  $Q_i(\boldsymbol{\theta})$  dada em (3.14). Resolvendo as equações

$$\begin{aligned}
\frac{\partial}{\partial \mu_x} [Q(\boldsymbol{\theta})] &= 0, & \frac{\partial}{\partial \alpha} [Q(\boldsymbol{\theta})] &= 0, & \frac{\partial}{\partial \beta} [Q(\boldsymbol{\theta})] &= 0 \\
\frac{\partial}{\partial \sigma_u^2} [Q(\boldsymbol{\theta})] &= 0, & \frac{\partial}{\partial \sigma_e^2} [Q(\boldsymbol{\theta})] &= 0,
\end{aligned}$$

obtemos os estimadores de máxima verossimilhança para  $\alpha$ ,  $\beta$ ,  $\mu_x$ ,  $\sigma_e^2$  e  $\sigma_u^2$ , respectivamente.

A figura 3.1 contém um esboço do esquema iterativo do algoritmo EM.

início

- Fornecer valores iniciais  $\theta^{(0)} = (\alpha^{(0)}, \beta^{(0)}, \mu_x^{(0)}, \sigma_{x_i}^2, \sigma_e^{2(0)}, \sigma_u^{2(0)})$ ,
- passo E: Avaliar

$$h_i^{(p)} = \beta^{(p)} \sigma_e^{-2(p)} (Y_i - \alpha^{(p)} - \beta^{(p)} \mu_x^{(p)}) + \sigma_u^{-2(p)} (X_i - \mu_x^{(p)})$$

$$c_i^{-1(p)} = \frac{\sigma_{x_i}^{2(p)} \sigma_e^{2(p)} \sigma_u^{2(p)}}{\beta^{2(p)} \sigma_{x_i}^{2(p)} \sigma_u^{2(p)} + \sigma_{x_i}^{2(p)} \sigma_e^{2(p)} + \sigma_e^{2(p)} \sigma_u^{2(p)}}$$

$$\hat{x}_i^{(p)} = \mu_x^{(p)} + c_i^{-1(p)} h_i^{(p)}, \quad V(x_i)^{(p)} = c_i^{-1(p)}$$

$$\hat{x}_i^{2(p)} = V(x_i)^{(p)} + \left(\hat{x}_i^{(p)}\right)^2, \quad i = 1, \dots, n,$$

- passo M: Avaliar

$$\mu_x^{(p+1)} = \frac{\sum_{i=1}^n \frac{\hat{x}_i^{(p)}}{\sigma_{x_i}^{2(p)}}}{\sum_{i=1}^n \frac{1}{\sigma_{x_i}^{2(p)}}}, \quad \beta^{(p+1)} = \frac{\sum_{i=1}^n Y_i \hat{x}_i^{(p)} - n \bar{Y} \bar{\hat{x}}^{(p)}}{\sum_{i=1}^n \hat{x}_i^{2(p)} - n (\bar{\hat{x}}^{(p)})^2},$$

$$\alpha^{(p+1)} = \bar{Y} - \bar{\hat{x}}^{(p)} \beta^{(p+1)}$$

$$\sigma_u^{2(p+1)} = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{x}_i^{(p)})^2$$

$$\sigma_e^{2(p+1)} = \frac{1}{n} \sum_{i=1}^n \left[ (Y_i - \alpha^{(p+1)}) - \beta^{(p+1)} \hat{x}_i^{(p)} \right]^2$$

- repetir passos E e M até que  $|\theta^{(p)} - \theta^{(p+1)}| < \epsilon$ , para algum  $\epsilon > 0$ ,

fim

Figura 3.1. Algoritmo EM para modelo (2.12)

### 3.6 Matriz de covariâncias assintóticas do EMV de $\theta$

A matriz de variâncias-covariâncias assintótica do EMV do parâmetro  $\theta = (\alpha, \beta, \mu_x, \sigma_c^2, \sigma_u^2)^T$  é dada pela inversa da matriz de informação de Fisher denotada por

$\mathbf{I}_F(\boldsymbol{\theta})$ , onde

$$\mathbf{I}_F(\boldsymbol{\theta}) = -\mathbb{E} \left[ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right],$$

com  $l(\boldsymbol{\theta}) = \log[L(\boldsymbol{\theta})]$ , sendo  $L(\boldsymbol{\theta})$  a função de verossimilhança dos dados observados  $(\mathbf{Z}_i)$ .

### 3.6.1 Matriz de Informação Observada

A matriz de informação observada denotada por  $\mathbf{K}(\boldsymbol{\theta})$ , é dada por:

$$\mathbf{K}(\boldsymbol{\theta}) = - \sum_{i=1}^n \mathbf{K}_i(\boldsymbol{\theta}), \text{ onde } \mathbf{K}_i(\boldsymbol{\theta}) = \frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}. \quad (3.15)$$

A matriz de derivadas segundas  $\mathbf{K}_i(\boldsymbol{\theta})$  pode ser obtida como segue.

Derivando (3.10) em relação a  $\psi$ , obtemos a seguinte expressão para as derivadas de primeiro:

$$\frac{\partial l_i(\boldsymbol{\theta})}{\partial \psi} = -\frac{1}{2} \frac{\partial}{\partial \psi} \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} \frac{\partial t_i}{\partial \psi} + c_i^{-1} h_i \frac{\partial h_i}{\partial \psi} - \frac{1}{2} h_i^2 c_i^{-2} \frac{\partial c_i}{\partial \psi}, \quad (3.16)$$

onde  $\psi = \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2$

Agora derivando (3.16) em relação a  $\gamma$ , obtemos uma expressão para as derivadas de segunda ordem dada por:

$$\begin{aligned} \frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \gamma \partial \psi} = & -\frac{1}{2} \frac{\partial^2}{\partial \gamma \partial \psi} \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} \frac{\partial^2 t_i}{\partial \gamma \partial \psi} + c_i^{-1} \left[ \frac{\partial h_i}{\partial \gamma} \frac{\partial h_i}{\partial \psi} + h_i \frac{\partial^2 h_i}{\partial \gamma \partial \psi} \right] \\ & - h_i c_i^{-2} \left[ \frac{\partial c_i}{\partial \gamma} \frac{\partial h_i}{\partial \psi} + \frac{\partial h_i}{\partial \gamma} \frac{\partial c_i}{\partial \psi} \right] + h_i^2 c_i^{-2} \left[ c_i^{-1} \frac{\partial c_i}{\partial \gamma} \frac{\partial c_i}{\partial \psi} - \frac{1}{2} \frac{\partial^2 c_i}{\partial \gamma \partial \psi} \right], \end{aligned} \quad (3.17)$$

onde  $\gamma, \psi = \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2$ .

Usando (3.16), (3.17) e após várias manipulações algébricas, a matriz de derivadas de segunda ordem de  $l_i(\boldsymbol{\theta})$  em relação a  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2)^T$  é dada por:

$$\mathbf{K}_i(\boldsymbol{\theta}) = \begin{pmatrix} K_{i\alpha\alpha} & K_{i\alpha\beta} & K_{i\alpha\mu_x} & K_{i\alpha\sigma_e^2} & K_{i\alpha\sigma_u^2} \\ K_{i\beta\alpha} & K_{i\beta\beta} & K_{i\beta\mu_x} & K_{i\beta\sigma_e^2} & K_{i\beta\sigma_u^2} \\ K_{i\mu_x\alpha} & K_{i\mu_x\beta} & K_{i\mu_x\mu_x} & K_{i\mu_x\sigma_e^2} & K_{i\mu_x\sigma_u^2} \\ K_{i\sigma_e^2\alpha} & K_{i\sigma_e^2\beta} & K_{i\sigma_e^2\mu_x} & K_{i\sigma_e^2\sigma_e^2} & K_{i\sigma_e^2\sigma_u^2} \\ K_{i\sigma_u^2\alpha} & K_{i\sigma_u^2\beta} & K_{i\sigma_u^2\mu_x} & K_{i\sigma_u^2\sigma_e^2} & K_{i\sigma_u^2\sigma_u^2} \end{pmatrix}, \quad i = 1, \dots, n, \quad (3.18)$$

onde

$$K_{i\alpha\alpha} = -\sigma_e^{-2} + c_i^{-1}\beta^2\sigma_e^{-4}$$

$$\begin{aligned} K_{i\alpha\beta} &= K_{i\beta\alpha} \\ &= -\mu_x\sigma_e^{-2} - \beta\sigma_e^{-4}c_i^{-1}(Y_i - \alpha - 2\beta\mu_x) - h_i\sigma_e^{-2}c_i^{-1} \\ &\quad + 2\beta^2h_i\sigma_e^{-4}c_i^{-2} \end{aligned}$$

$$\begin{aligned} K_{i\alpha\mu_x} &= K_{i\mu_x\alpha} \\ &= -\beta\sigma_e^{-2} + c_i^{-1}(\beta^3\sigma_e^{-4} + \beta\sigma_u^{-2}\sigma_e^{-2}) \end{aligned}$$

$$\begin{aligned} K_{i\alpha\sigma_e^2} &= K_{i\sigma_e^2\alpha} \\ &= -\sigma_e^{-4}(Y_i - \alpha - \beta\mu_x) \\ &\quad + c_i^{-i}[\beta^2\sigma_e^{-6}(Y_i - \alpha - \beta\mu_x) + \beta h_i\sigma_e^{-4}] - \beta^3h_i\sigma_e^{-6}c_i^{-2} \end{aligned}$$

$$\begin{aligned} K_{i\alpha\sigma_u^2} &= K_{i\sigma_u^2\alpha} \\ &= c_i^{-1}\beta\sigma_e^{-2}\sigma_u^{-4}(X_i - \mu_x) - \beta h_i\sigma_e^{-2}\sigma_u^{-4}c_i^{-2} \end{aligned}$$

$$\begin{aligned} K_{i\beta\beta} &= -\sigma_e^{-2}c_i^{-1} - \mu_x^2\sigma_e^{-2} + c_i^{-1}[\sigma_e^{-4}(Y_i - \alpha - 2\beta\mu_x)^2 - 2h_i\mu_x\sigma_e^{-2}] \\ &\quad - 4\beta h_i\sigma_e^{-4}c_i^{-2}(Y_i - \alpha - 2\beta\mu_x) + h_i^2c_i^{-2}(4\beta^2\sigma_e^{-4}c_i^{-1} - \sigma_e^{-2}) + 2\beta^2c_i^{-2}\sigma_e^{-4} \end{aligned}$$

$$\begin{aligned} K_{i\beta\mu_x} &= K_{i\mu_x\beta} \\ &= -c_i^{-1}[\sigma_e^{-2}(\beta^2\sigma_e^{-2} + \sigma_u^{-2})(Y_i - \alpha - 2\beta\mu_x) + 2\beta h_i\sigma_e^{-2}] \\ &\quad + \sigma_e^{-2}(Y_i - \alpha - 2\beta\mu_x) + 2\beta h_i\sigma_e^{-2}c_i^{-2}(\beta^2\sigma_e^{-2} + \sigma_u^{-2}) \end{aligned}$$

$$\begin{aligned}
K_{i_{\beta\sigma_e^2}} &= K_{i_{\sigma_e^2\beta}} \\
&= \beta\sigma_e^{-4}c_i^{-1} - \beta^3\sigma_e^{-6}c_i^{-2} - \mu_x\sigma_e^{-4}(Y_i - \alpha - \beta\mu_x) \\
&\quad - c_i^{-1} [\beta\sigma_e^{-6}(Y_i - \alpha - \beta\mu_x)(Y_i - \alpha - 2\beta\mu_x) + h_i\sigma_e^{-4}(Y_i - \alpha - 2\beta\mu_x)] \\
&\quad + h_i c_i^{-2} [\beta^2\sigma_e^{-6}(Y_i - \alpha - 2\beta\mu_x) + 2\beta^2\sigma_e^{-6}(Y_i - \alpha - \beta\mu_x)] \\
&\quad - h_i^2 c_i^{-2} (2c_i^{-1}\beta^3\sigma_e^{-6} - \beta\sigma_e^{-4})
\end{aligned}$$

$$\begin{aligned}
K_{i_{\beta\sigma_u^2}} &= K_{i_{\sigma_u^2\beta}} \\
&= -\beta\sigma_e^{-2}\sigma_u^{-4}c_i^{-2} - \sigma_e^{-2}\sigma_u^{-4}c_i^{-1}(Y_i - \alpha - 2\beta\mu_x)(X_i - \mu_x) \\
&\quad + h_i\sigma_u^{-4}c_i^{-2} [2\beta\sigma_e^{-2}(X_i - \mu_x) + \sigma_e^{-2}(Y_i - \alpha - 2\beta\mu_x)] \\
&\quad - 2\beta h_i^2\sigma_e^{-2}\sigma_u^{-4}c_i^{-3}
\end{aligned}$$

$$K_{i_{\mu_x\mu_x}} = -(\beta^2\sigma_e^{-2} + \sigma_u^{-2}) + c_i^{-1}(\beta^2\sigma_e^{-2} + \sigma_u^{-2})^2$$

$$\begin{aligned}
K_{i_{\mu_x\sigma_e^2}} &= K_{i_{\sigma_e^2\mu_x}} \\
&= -\beta\sigma_e^{-4}(Y_i - \alpha - \beta\mu_x) + c_i^{-1}[\beta\sigma_e^{-4}(Y_i - \alpha - \beta\mu_x)(\beta^2\sigma_e^{-2} + \sigma_u^{-2}) \\
&\quad + \beta^2 h_i\sigma_e^{-4}] - \beta^2 h_i\sigma_e^{-4}c_i^{-2}(\beta^2\sigma_e^{-2} + \sigma_u^{-2})
\end{aligned}$$

$$\begin{aligned}
K_{i_{\mu_x\sigma_u^2}} &= K_{i_{\sigma_u^2\mu_x}} \\
&= -\sigma_u^{-4}(X_i - \mu_x) + c_i^{-1} [\sigma_u^{-4}(\beta^2\sigma_e^{-2} + \sigma_u^{-2})(X_i - \mu_x) + h_i\sigma_u^{-4}] \\
&\quad - h_i c_i^{-2}\sigma_u^{-4}(\beta^2\sigma_e^{-2} + \sigma_u^{-2})
\end{aligned}$$

$$\begin{aligned}
K_{i_{\sigma_e^2\sigma_e^2}} &= -\frac{1}{2}(-\sigma_e^{-4} + 2\beta^2\sigma_e^{-6}c_i^{-1} - \beta^4\sigma_e^{-8}c_i^{-2}) - \sigma_e^{-6}(Y_i - \alpha - \beta\mu_x)^2 \\
&\quad + c_i^{-1} [\beta^2\sigma_e^{-8}(Y_i - \alpha - \beta\mu_x)^2 + 2\beta h_i\sigma_e^{-6}(Y_i - \alpha - \beta\mu_x)] \\
&\quad - 2h_i\beta^3\sigma_e^{-8}c_i^{-2}(Y_i - \alpha - \beta\mu_x) + h_i^2 c_i^{-2} (c_i^{-1}\beta^4\sigma_e^{-8} - \beta^2\sigma_e^{-6})
\end{aligned}$$

$$\begin{aligned}
K_{i_{\sigma_e^2\sigma_u^2}} &= K_{i_{\sigma_u^2\sigma_e^2}} \\
&= \frac{1}{2}\beta^2\sigma_e^{-4}\sigma_u^{-4}c_i^{-2} + \beta\sigma_e^{-4}\sigma_u^{-4}c_i^{-1}(Y_i - \alpha - \beta\mu_x)(X_i - \mu_x) \\
&\quad - h_i c_i^{-2} [\beta^2\sigma_e^{-4}\sigma_u^{-4}(X_i - \mu_x) + \beta\sigma_e^{-4}\sigma_u^{-4}(Y_i - \alpha - \beta\mu_x)] \\
&\quad + h_i^2 c_i^{-3}\beta^2\sigma_e^{-4}\sigma_u^{-4}
\end{aligned}$$

$$\begin{aligned}
K_{i_{\sigma_u^2\sigma_u^2}} &= \frac{1}{2}(\sigma_u^{-4} - 2\sigma_u^{-6}c_i^{-1} + \sigma_u^{-8}c_i^{-2}) - \sigma_u^{-6}(X_i - \mu_x)^2 \\
&\quad + c_i^{-1} [\sigma_u^{-8}(X_i - \mu_x)^2 + 2h_i\sigma_u^{-6}(X_i - \mu_x)] \\
&\quad - 2h_i\sigma_u^{-8}c_i^{-2}(X_i - \mu_x) + h_i^2 c_i^{-2}(\sigma_u^{-8}c_i^{-1} - \sigma_u^{-6}).
\end{aligned}$$



### 3.6.2 Matriz de informação esperada

Nesta seção, nós fornecemos a matriz de informação esperada do vetor de parâmetros  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2)^T$ . Usamos uma parametrização similar à encontrada em Kulathinal et al. (2002).

Adotando a parametrização  $\boldsymbol{\theta}^* = (\mu_1, \mu_2, \sigma_{11}, \sigma_{22}, \tau)^T$ , onde  $\mu_1 = \alpha + \beta\mu_x$ ,  $\mu_2 = \mu_x$ ,  $\sigma_{11} = \sigma_e^2$ ,  $\sigma_{22} = \sigma_u^2$ ,  $\tau = \beta$  e fazendo  $\boldsymbol{\theta}^* = (\boldsymbol{\mu}, \boldsymbol{\sigma})$ , onde  $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$  e  $\boldsymbol{\sigma} = (\sigma_{11}, \sigma_{22}, \tau)^T$ , pode ser mostrado em algumas passagens (em um modelo estrutural normal heteroscedástico com erros nas variáveis) que:

$$E \left( \frac{\partial^2 l_i(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\mu} \partial \boldsymbol{\sigma}} \right) = \mathbf{0} \quad \text{e} \quad -E \left( \frac{\partial^2 l_i(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T} \right) = \boldsymbol{\Sigma}_i^{-1}, \quad i = 1, \dots, n,$$

o que implica que a matriz de informação do vetor de parâmetros  $\boldsymbol{\theta}^*$  pode ser representada como uma matriz diagonal em blocos na forma

$$\mathbf{I}(\boldsymbol{\theta}^*) = \begin{pmatrix} \mathbf{I}(\boldsymbol{\mu}) & \mathbf{0} \\ \mathbf{0}^T & \mathbf{I}(\boldsymbol{\sigma}) \end{pmatrix}, \quad (3.19)$$

onde  $\mathbf{I}(\boldsymbol{\mu})$  e  $\mathbf{I}(\boldsymbol{\sigma})$  são as matrizes de informação associadas aos vetores de parâmetros  $\boldsymbol{\mu}$  e  $\boldsymbol{\sigma}$ , respectivamente. Neste caso  $\mathbf{I}(\boldsymbol{\mu}) = \sum_{i=1}^n \boldsymbol{\Sigma}_i^{-1}$ .

Para obter  $\mathbf{I}(\boldsymbol{\sigma})$  usaremos a distribuição da variável observada  $\mathbf{Z}_i$  e a seguinte fórmula provada por Dolby (1976):

$$-E \left[ \frac{\partial^2 l_i(\boldsymbol{\theta}^*)}{\partial \psi \partial \phi} \right] = \frac{1}{2} \text{tr} (\boldsymbol{\Sigma}_i^{-1} \boldsymbol{\Sigma}_{i\psi} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\Sigma}_{i\phi}) + \mathbf{d}_{i\psi}^T \boldsymbol{\Sigma}_i^{-1} \mathbf{d}_{i\phi},$$

em que  $\mathbf{d}_i = (\mathbf{Z}_i - \boldsymbol{\mu})$ ,  $\mathbf{d}_{ik} = \frac{\partial \mathbf{d}_i}{\partial k}$  e  $\boldsymbol{\Sigma}_{ik} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial k}$  para  $k = \psi, \phi$ . Aqui  $\psi$  e  $\phi$  são os parâmetros do modelo, neste caso  $\psi, \phi = \sigma_{11} + \tau^2 \sigma_{x_i}^2, \sigma_{22} + \sigma_{x_i}^2, \rho_i$ , onde  $\rho_i$  é o coeficiente de correlação entre  $(Y_i, X_i)$  dado por

$$\rho_i = \frac{\tau \sigma_{x_i}^2}{\sqrt{(\sigma_{11} + \tau^2 \sigma_{x_i}^2)(\sigma_{22} + \sigma_{x_i}^2)}}, \quad i = 1, \dots, n.$$

Daqui segue, que a matriz de informação  $\mathbf{I}_i(\boldsymbol{\theta}_i)$  para  $\boldsymbol{\theta}_i = (\sigma_{11} + \tau^2 \sigma_{x_i}^2, \sigma_{22} + \sigma_{x_i}^2, \rho_i)$  é uma matriz  $3 \times 3$  com elementos

$$\mathbf{I}_i(\boldsymbol{\theta}_i) = \begin{pmatrix} \frac{2-\rho_i^2}{4(1-\rho_i^2)(\sigma_{11}+\tau^2\sigma_{x_i}^2)^2} & \frac{-\rho_i^2}{4(1-\rho_i^2)(\sigma_{11}+\tau^2\sigma_{x_i}^2)(\sigma_{22}+\sigma_{x_i}^2)} & \frac{-\rho_i}{2(1-\rho_i^2)(\sigma_{11}+\tau^2\sigma_{x_i}^2)} \\ \frac{-\rho_i^2}{4(1-\rho_i^2)(\sigma_{11}+\tau^2\sigma_{x_i}^2)(\sigma_{22}+\sigma_{x_i}^2)} & \frac{2-\rho_i^2}{4(1-\rho_i^2)(\sigma_{22}+\sigma_{x_i}^2)^2} & \frac{-\rho_i}{2(1-\rho_i^2)(\sigma_{22}+\sigma_{x_i}^2)} \\ \frac{-\rho_i}{2(1-\rho_i^2)(\sigma_{11}+\tau^2\sigma_{x_i}^2)} & \frac{-\rho_i}{2(1-\rho_i^2)(\sigma_{22}+\sigma_{x_i}^2)} & \frac{1+\rho_i^2}{(1-\rho_i^2)^2} \end{pmatrix}.$$

Note-se que  $\sigma_{x_i}^2$  é constante. A informação sobre os parâmetros  $(\sigma_{11}, \sigma_{22}, \tau)$ , disponível na distribuição de  $\mathbf{Z}_i$  pode ser facilmente obtida (veja Apêndice B) usando  $\mathbf{I}_i(\boldsymbol{\theta}_i)$  e o Jacobiano da transformação denotado por  $\mathbf{J}_i$ , que é uma matriz  $3 \times 3$  dada por

$$\mathbf{J}_i(\boldsymbol{\theta}_i) = \begin{pmatrix} 1 & 0 & \frac{-\tau\sigma_{x_i}^2}{2(\sigma_{11}+\tau^2\sigma_{x_i}^2)\sqrt{(\sigma_{11}+\tau^2\sigma_{x_i}^2)(\sigma_{22}+\sigma_{x_i}^2)}} \\ 0 & 1 & \frac{-\tau\sigma_{x_i}^2}{2(\sigma_{22}+\sigma_{x_i}^2)\sqrt{(\sigma_{11}+\tau^2\sigma_{x_i}^2)(\sigma_{22}+\sigma_{x_i}^2)}} \\ 2\tau\sigma_{x_i}^2 & 0 & \frac{\sigma_{11}\sigma_{x_i}^2}{(\sigma_{11}+\tau^2\sigma_{x_i}^2)\sqrt{(\sigma_{11}+\tau^2\sigma_{x_i}^2)(\sigma_{22}+\sigma_{x_i}^2)}} \end{pmatrix},$$

segue que, a matriz de informação  $\mathbf{I}_i(\sigma_{11}, \sigma_{22}, \tau)$  é dada por

$$\mathbf{I}_i(\sigma_{11}, \sigma_{22}, \tau) = \mathbf{J}_i(\boldsymbol{\theta}_i)\mathbf{I}_i(\boldsymbol{\theta}_i)\mathbf{J}_i(\boldsymbol{\theta}_i)^T.$$

A informação total sobre  $\boldsymbol{\sigma} = (\sigma_{11}, \sigma_{22}, \tau)$  é a soma das informações disponíveis de cada  $\mathbf{Z}_i$  e é denotada por

$$\mathbf{I}(\boldsymbol{\sigma}) = \sum_{i=1}^n \mathbf{I}_i(\sigma_{11}, \sigma_{22}, \tau).$$

Voltando à nossa parametrização, temos agora que a matriz de informação de Fisher para o vetor de parâmetros  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_\epsilon^2, \sigma_u^2)^T$  é dada por

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta}^*)\mathbf{I}(\boldsymbol{\theta}^*)\mathbf{J}(\boldsymbol{\theta}^*)^T, \quad (3.20)$$

onde o Jacobiano é dado por

$$\mathbf{J}(\boldsymbol{\theta}^*) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \mu_2 & 0 & 0 & 0 & 1 \\ \tau & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Sob certas condições de regularidade, pode ser mostrado que a distribuição do estimador de máxima verossimilhança denotado por  $\hat{\boldsymbol{\theta}}_{\text{EMV}}$  se aproxima de uma distribuição normal multivariada quando o tamanho da amostra tende ao infinito ( $n \rightarrow \infty$ ). Assim, a distribuição aproximada de  $\hat{\boldsymbol{\theta}}_{\text{EMV}}$  em grandes amostras é  $N_5(\boldsymbol{\theta}, \mathbf{I}(\boldsymbol{\theta})^{-1})$ . Regiões de confiança e testes de hipóteses para os parâmetros podem ser construídos a partir dos resultados assintóticos. A matriz de informação esperada  $\mathbf{I}(\boldsymbol{\theta})$  pode ser estimada por  $\mathbf{I}(\hat{\boldsymbol{\theta}}_{\text{EMV}})$  ou, alternativamente, por  $\mathbf{K}(\hat{\boldsymbol{\theta}}_{\text{EMV}})$ , com  $\mathbf{K}(\boldsymbol{\theta})$  dado em (3.15).

### 3.7 Teste de Hipóteses

Podemos testar a hipótese

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}, \quad \boldsymbol{\theta}_2 \text{ não especificado}, \quad (3.21)$$

onde  $\boldsymbol{\theta}_1$  é um subconjunto de  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ , com  $p_1$  parâmetros  $1 \leq p_1 \leq 5$ . Fazendo uma partição adequada da matriz de informação esperada e usando o teorema A.5.2 em Muirhead (1982), obtemos:

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_{11}(\boldsymbol{\theta}) & \mathbf{I}_{12}(\boldsymbol{\theta}) \\ \mathbf{I}_{21}(\boldsymbol{\theta}) & \mathbf{I}_{22}(\boldsymbol{\theta}) \end{pmatrix}, \quad \text{e} \quad \mathbf{I}(\boldsymbol{\theta})^{-1} = \begin{pmatrix} \mathbf{I}^{11}(\boldsymbol{\theta}) & \mathbf{I}^{12}(\boldsymbol{\theta}) \\ \mathbf{I}^{21}(\boldsymbol{\theta}) & \mathbf{I}^{22}(\boldsymbol{\theta}) \end{pmatrix},$$

onde  $\mathbf{I}_{11}(\boldsymbol{\theta})$  e  $\mathbf{I}^{11}(\boldsymbol{\theta})$  são matrizes de dimensão  $p_1 \times p_1$ ,  $\mathbf{I}_{12}(\boldsymbol{\theta})$  e  $\mathbf{I}^{12}(\boldsymbol{\theta})$  de dimensão  $p_1 \times (p - p_1)$ ,  $\mathbf{I}_{21}(\boldsymbol{\theta})$  e  $\mathbf{I}^{21}(\boldsymbol{\theta})$  de dimensão  $(p - p_1) \times p_1$  e  $\mathbf{I}_{22}(\boldsymbol{\theta})$  e  $\mathbf{I}^{22}(\boldsymbol{\theta})$  de dimensão  $(p - p_1) \times (p - p_1)$  respectivamente. Assumimos também a não singularidade das matrizes  $\mathbf{I}_{11}(\boldsymbol{\theta})$  e  $\mathbf{I}_{22}(\boldsymbol{\theta})$ .

Definindo

$$\mathbf{I}_{11.2}(\boldsymbol{\theta}) = \mathbf{I}_{11}(\boldsymbol{\theta}) - \mathbf{I}_{12}(\boldsymbol{\theta})\mathbf{I}_{22}^{-1}(\boldsymbol{\theta})\mathbf{I}_{21}(\boldsymbol{\theta}),$$

$$\mathbf{I}_{22.1}(\boldsymbol{\theta}) = \mathbf{I}_{22}(\boldsymbol{\theta}) - \mathbf{I}_{21}(\boldsymbol{\theta})\mathbf{I}_{11}^{-1}(\boldsymbol{\theta})\mathbf{I}_{12}(\boldsymbol{\theta}),$$

temos que os elementos de  $\mathbf{I}(\boldsymbol{\theta})^{-1}$  são tais que

$$\mathbf{I}^{11}(\boldsymbol{\theta}) = \mathbf{I}_{11.2}^{-1}(\boldsymbol{\theta}), \quad \mathbf{I}^{22}(\boldsymbol{\theta}) = \mathbf{I}_{22.1}^{-1}(\boldsymbol{\theta}), \quad \mathbf{I}^{12}(\boldsymbol{\theta}) = -\mathbf{I}_{11}^{-1}(\boldsymbol{\theta})\mathbf{I}_{12}(\boldsymbol{\theta})\mathbf{I}_{22.1}^{-1}(\boldsymbol{\theta}),$$

$$\mathbf{I}^{21}(\boldsymbol{\theta}) = -\mathbf{I}_{22}^{-1}(\boldsymbol{\theta})\mathbf{I}_{21}(\boldsymbol{\theta})\mathbf{I}_{11.2}^{-1}(\boldsymbol{\theta}).$$

Então a hipóteses (3.21) pode ser testada usando o seguinte estatística de Wald

$$\mathcal{W}_{EMV} = (\hat{\boldsymbol{\theta}}_{EMV} - \boldsymbol{\theta}_{10})^T \mathbf{I}_{11.2}(\hat{\boldsymbol{\theta}}_{EMV})(\hat{\boldsymbol{\theta}}_{EMV} - \boldsymbol{\theta}_{10}). \quad (3.22)$$

Sob (3.21), a estatística em (3.22) converge em distribuição para uma distribuição qui-quadrado com  $p_1$  graus de liberdade, quando o tamanho da amostra tende ao infinito, isto é  $\mathcal{W}_{EMV} \xrightarrow{d} \chi_{p_1}^2$ , quando  $n \rightarrow \infty$ .

### 3.8 Avaliação do Ajuste

Para avaliar a qualidade do ajuste do MREH normal com erro nas variáveis, podemos construir um gráfico das distâncias transformadas, veja Lange et al. (1989).

Lembremos que

$$\delta_i(\boldsymbol{\theta}) = (\mathbf{Z}_i - \mathbf{a} - \mu_x \mathbf{b})^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Z}_i - \mathbf{a} - \mu_x \mathbf{b}), \quad i = 1, \dots, n, \quad (3.23)$$

é a distância de Mahalanobis. Sob normalidade,  $\delta_i(\boldsymbol{\theta})$  tem uma distribuição qui-quadrado com dois graus de liberdade ( $\chi_2^2$ ),  $i = 1, \dots, n$ . Além disso,  $\delta_i(\hat{\boldsymbol{\theta}})$  tem assintoticamente a mesma distribuição qui-quadrado ( $\chi_2^2$ ). Então usando a aproximação de Wilson -Hilferty (Wilson and Hilferty (1931), Johnson et al. (1994)) temos que

$$r_i = 3 \left\{ \left[ \delta_i(\hat{\boldsymbol{\theta}})/2 \right]^{1/3} - 8/9 \right\}, \quad i = 1, \dots, n, \quad (3.24)$$

tem, aproximadamente, uma distribuição normal padrão, isto é,  $r_i \stackrel{i.i.d.}{\sim} N(0, 1)$ ,  $i = 1, \dots, n$ ,

aproximadamente. Então um Q-Q plot das distâncias transformadas  $r_i$ ,  $i = 1, \dots, n$ , dadas em (3.24) pode ser utilizado para avaliar o ajuste do modelo MREH normal com erro nas variáveis na prática.

### 3.9 Uma Visão Unificada

Nesta seção apresentamos uma visão geral do modelo estrutural normal heteroscedástico com erros nas variáveis, para as diferentes situações em que as variâncias dos erros de medição e a variância da variável não observada  $x_i$ , variam entre as observações.

#### 3.9.1 Modelo Geral

Consideremos o seguinte modelo

$$\begin{aligned} y_i &= \alpha + \beta x_i + q_i \\ X_i &= x_i + u_i \\ Y_i &= y_i + e_i \quad i = 1, \dots, n. \end{aligned} \quad (3.25)$$

Definindo  $\mathbf{a}$  e  $\mathbf{b}$  como em (1.8), temos que

$$\mathbf{Z}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix} = \mathbf{a} + x_i \mathbf{b} + \begin{pmatrix} e_i + q_i \\ u_i \end{pmatrix}. \quad (3.26)$$

As suposições impostas ao modelo são:

$(e_i, u_i)^T$  e  $(q_j, x_j)^T$  são independentes,  $q_i$  e  $x_i$  são independentes, com

$$\begin{aligned} q_i &\stackrel{i.i.d}{\sim} N(0, \sigma^2), \quad x_i \stackrel{indep.}{\sim} N(\mu_x, \sigma_{x_i}^2) \quad e \\ \begin{pmatrix} e_i \\ u_i \end{pmatrix} &\stackrel{indep.}{\sim} N_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_{e_i}^2 & 0 \\ 0 & \sigma_{u_i}^2 \end{pmatrix} \right], \quad i, j = 1, \dots, n. \end{aligned} \quad (3.27)$$

As suposições dadas acima implicam que

$$\mathbf{Z}_i \stackrel{\text{indep.}}{\sim} N_2(\boldsymbol{\mu}; \boldsymbol{\Sigma}_i), \quad i = 1, \dots, n,$$

onde  $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$  e  $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ , sendo  $\boldsymbol{\theta}$  o vetor de parâmetros estruturais do modelo, além disso,

$$\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{\Psi}_i + \sigma_{x_i}^2 \mathbf{b}\mathbf{b}^T = \begin{pmatrix} \beta^2 \sigma_{x_i}^2 + \sigma_{e_i}^2 + \sigma^2 & \beta \sigma_{x_i}^2 \\ \beta \sigma_{x_i}^2 & \sigma_{x_i}^2 + \sigma_{u_i}^2 \end{pmatrix}, \quad (3.28)$$

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{a} + \mu_x \mathbf{b} = \begin{pmatrix} \alpha + \beta \mu_x \\ \mu_x \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Psi}_i = \begin{pmatrix} \sigma_{e_i}^2 + \sigma^2 & 0 \\ 0 & \sigma_{u_i}^2 \end{pmatrix}, \quad (3.29)$$

Note que,  $\boldsymbol{\theta}$  o vetor de parâmetros estruturais do modelo, vai depender da condição de identificabilidade imposta ao modelo.

### 3.9.2 Função de log-verossimilhança

A função de log-verossimilhança é dada por

$$\begin{aligned} L(\boldsymbol{\theta}) &= \sum_{i=1}^n l_i(\boldsymbol{\theta}), \quad \text{onde} \\ l_i(\boldsymbol{\theta}) &= -\frac{1}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} (\mathbf{Z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}), \end{aligned} \quad (3.30)$$

em que

$$|\boldsymbol{\Sigma}_i| = |\boldsymbol{\Psi}_i| (1 + \sigma_{x_i}^2 \mathbf{b}^T \boldsymbol{\Psi}_i^{-1} \mathbf{b}) \quad \text{e} \quad \boldsymbol{\Sigma}_i^{-1} = \boldsymbol{\Psi}_i^{-1} - \lambda_i \boldsymbol{\Psi}_i^{-1} \mathbf{b}\mathbf{b}^T \boldsymbol{\Psi}_i^{-1}, \quad (3.31)$$

com

$$\lambda_i = \frac{\sigma_{x_i}^2 (\sigma_{e_i} + \sigma^2) \sigma_{u_i}^2}{(\beta^2 \sigma_{x_i}^2 + \sigma_{e_i}^2 + \sigma^2) \sigma_{u_i}^2 + \sigma_{x_i}^2 (\sigma_{e_i} + \sigma^2)}, \quad i = 1, \dots, n. \quad (3.32)$$

### 3.9.3 Equações de Verossimilhança

De (3.30) temos

$$\begin{aligned} l_i(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i) &\propto -\frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}(\mathbf{Z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}) \\ &\propto -\frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\text{tr} \left[ \boldsymbol{\Sigma}_i^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}) (\mathbf{Z}_i - \boldsymbol{\mu})^T \right]. \end{aligned} \quad (3.33)$$

Diferenciando (3.33) (como em Magnus and Neudecker (2007) ( pag. 81)) em relação a  $\boldsymbol{\mu}$  e  $\boldsymbol{\Sigma}_i$  temos,

$$\begin{aligned} dl_i(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i) &= -\frac{1}{2}d\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\text{tr} \left[ (d\boldsymbol{\Sigma}_i^{-1}) (\mathbf{Z}_i - \boldsymbol{\mu}) (\mathbf{Z}_i - \boldsymbol{\mu})^T \right] \\ &\quad - \frac{1}{2}\text{tr} \left[ \boldsymbol{\Sigma}_i^{-1} d(\mathbf{Z}_i - \boldsymbol{\mu}) (\mathbf{Z}_i - \boldsymbol{\mu})^T \right] \\ &= -\frac{1}{2}\text{tr} \left[ \boldsymbol{\Sigma}_i^{-1} (d\boldsymbol{\Sigma}_i) \right] + \frac{1}{2}\text{tr} \left[ \boldsymbol{\Sigma}_i^{-1} (d\boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}) (\mathbf{Z}_i - \boldsymbol{\mu})^T \right] \\ &\quad + \frac{1}{2}\text{tr} \left( \boldsymbol{\Sigma}_i^{-1} \left[ (\mathbf{Z}_i - \boldsymbol{\mu}) (d\boldsymbol{\mu})^T + (d\boldsymbol{\mu}) (\mathbf{Z}_i - \boldsymbol{\mu})^T \right] \right) \\ &= \frac{1}{2}\text{tr} \left\{ (d\boldsymbol{\Sigma}_i) \left( \boldsymbol{\Sigma}_i^{-1} \left[ (\mathbf{Z}_i - \boldsymbol{\mu}) (\mathbf{Z}_i - \boldsymbol{\mu})^T - \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \right) \right\} \\ &\quad + (d\boldsymbol{\mu})^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}). \end{aligned} \quad (3.34)$$

Se ignoramos a restrição de simetria sobre  $\boldsymbol{\Sigma}_i$ , obtemos

$$\mathbf{S}_\mu = \sum_{i=1}^n \boldsymbol{\Sigma}_i^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}) \quad (3.35)$$

$$\mathbf{S}_\Sigma = \sum_{i=1}^n \mathbf{S}_{\Sigma_i}, \quad (3.36)$$

onde

$$\mathbf{S}_{\Sigma_i} = \boldsymbol{\Sigma}_i^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}) (\mathbf{Z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1}, \quad i = 1, \dots, n. \quad (3.37)$$

### 3.9.4 Matriz de Informação de Fisher

Para obter a matriz de informação temos que levar em conta a simetria da matriz  $\boldsymbol{\Sigma}_i$ , implícita ou explicitamente. Nós preferimos o tratamento implícito, usando para tal efeito a



matriz de duplicação  $D_n$ , definida como em Magnus and Neudecker (2007) ( pag. 48).

**Proposição 3.4.** Sejam  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$  vetores aleatórios independentes de dimensão  $m \times 1$ , tais que

$$\mathbf{Z}_i \stackrel{indep.}{\sim} N_m(\boldsymbol{\mu}; \boldsymbol{\Sigma}_i), \quad i = 1, \dots, n,$$

onde  $\boldsymbol{\Sigma}_i$  é definida positiva, além disso  $n \geq m + 1$ . Então a informação sobre  $\boldsymbol{\mu}$  e  $\text{vech}(\boldsymbol{\Sigma}_i)$  contida na  $i$ -ésima observação  $\mathbf{Z}_i$ , denotada por  $\mathcal{F}_i$  é a matriz de dimensão  $\frac{1}{2}m(m+3) \times \frac{1}{2}m(m+3)$  dada por

$$\mathcal{F}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2} \mathbf{D}_m^T (\boldsymbol{\Sigma}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1}) \mathbf{D}_m \end{pmatrix}, \quad i = 1, \dots, n. \quad (3.38)$$

**Prova:** Como  $\boldsymbol{\Sigma}_i$  é uma função lineal de  $\text{vech}(\boldsymbol{\Sigma}_i)$ , temos que la diferencial de segunda ordem de  $\boldsymbol{\Sigma}_i$  é nula, isto é,  $d^2\boldsymbol{\Sigma}_i = d(d\boldsymbol{\Sigma}_i) = 0$ , assim a diferencial de segunda ordem de  $l_i(\boldsymbol{\mu}, \text{vech}(\boldsymbol{\Sigma}_i))$  em relação a  $\boldsymbol{\mu}$  e  $\boldsymbol{\Sigma}_i$  pode ser obtida diferenciando (3.34), isto é

$$\begin{aligned} d^2 l_i(\boldsymbol{\mu}, \text{vech}(\boldsymbol{\Sigma}_i)) &= \frac{1}{2} \text{tr} \left\{ (d\boldsymbol{\Sigma}_i) (d\boldsymbol{\Sigma}_i^{-1}) \left( \left[ (\mathbf{Z}_i - \boldsymbol{\mu})(\mathbf{Z}_i - \boldsymbol{\mu})^T - \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \right) \right\} \\ &\quad + \frac{1}{2} \text{tr} \left\{ (d\boldsymbol{\Sigma}_i) (\boldsymbol{\Sigma}_i^{-1}) \left( \left[ (\mathbf{Z}_i - \boldsymbol{\mu})(\mathbf{Z}_i - \boldsymbol{\mu})^T - \boldsymbol{\Sigma}_i \right] d\boldsymbol{\Sigma}_i^{-1} \right) \right\} \\ &\quad + \frac{1}{2} \text{tr} \left\{ (d\boldsymbol{\Sigma}_i) (\boldsymbol{\Sigma}_i^{-1}) \left( \left[ (\mathbf{Z}_i - \boldsymbol{\mu})(\mathbf{Z}_i - \boldsymbol{\mu})^T - d\boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \right) \right\} \\ &\quad + (d\boldsymbol{\mu})^T (d\boldsymbol{\Sigma}_i^{-1}) (\mathbf{Z}_i - \boldsymbol{\mu}) - (d\boldsymbol{\mu})^T \boldsymbol{\Sigma}_i^{-1} d\boldsymbol{\mu}. \end{aligned} \quad (3.39)$$

Notar que não é necessário desenvolver a diferencial segunda completamente em termos de  $d\boldsymbol{\mu}$  e  $d\text{vech}(\boldsymbol{\Sigma}_i)$ , já que tomando esperança de (3.39) e levando em conta que  $E(\mathbf{Z}_i - \boldsymbol{\mu}) = 0$  e  $E(\mathbf{Z}_i - \boldsymbol{\mu})(\mathbf{Z}_i - \boldsymbol{\mu})^T = \boldsymbol{\Sigma}_i$ ,  $i = 1, \dots, n$ , segue que

$$\begin{aligned} -E[d^2 l_i] &= \frac{1}{2} \text{tr} \left\{ (d\boldsymbol{\Sigma}_i) (d\boldsymbol{\Sigma}_i^{-1}) \left( \left[ E(\mathbf{Z}_i - \boldsymbol{\mu})(\mathbf{Z}_i - \boldsymbol{\mu})^T - \boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \right) \right\} \\ &\quad + \frac{1}{2} \text{tr} \left\{ (d\boldsymbol{\Sigma}_i) (\boldsymbol{\Sigma}_i^{-1}) \left( \left[ E(\mathbf{Z}_i - \boldsymbol{\mu})(\mathbf{Z}_i - \boldsymbol{\mu})^T - \boldsymbol{\Sigma}_i \right] d\boldsymbol{\Sigma}_i^{-1} \right) \right\} \\ &\quad + \frac{1}{2} \text{tr} \left\{ (d\boldsymbol{\Sigma}_i) (\boldsymbol{\Sigma}_i^{-1}) \left( \left[ E(\mathbf{Z}_i - \boldsymbol{\mu})(\mathbf{Z}_i - \boldsymbol{\mu})^T - d\boldsymbol{\Sigma}_i \right] \boldsymbol{\Sigma}_i^{-1} \right) \right\} \\ &\quad + (d\boldsymbol{\mu})^T (d\boldsymbol{\Sigma}_i^{-1}) E(\mathbf{Z}_i - \boldsymbol{\mu}) - (d\boldsymbol{\mu})^T \boldsymbol{\Sigma}_i^{-1} d\boldsymbol{\mu} \\ &= \frac{1}{2} \text{tr} \left\{ (d\boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} (d\boldsymbol{\Sigma}_i) \boldsymbol{\Sigma}_i^{-1} \right\} + (d\boldsymbol{\mu})^T \boldsymbol{\Sigma}_i^{-1} d\boldsymbol{\mu}. \end{aligned} \quad (3.40)$$

Agora vetorizando (3.40) (com o operador  $vec(\cdot)$  definido como em Magnus and Neudecker (2007) ( pag. 30)) e usando a matriz de duplicação obtemos

$$\begin{aligned}
- E [d^2 l_i(\boldsymbol{\mu}, \text{vech}(\boldsymbol{\Sigma}_i))] &= \frac{1}{2} (\text{vec } d\boldsymbol{\Sigma}_i)^T (\boldsymbol{\Sigma}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1}) \text{vec } d\boldsymbol{\Sigma}_i + (d\boldsymbol{\mu})^T \boldsymbol{\Sigma}_i^{-1} d\boldsymbol{\mu} \\
&= \frac{1}{2} (d\text{vech}(\boldsymbol{\Sigma}_i))^T \mathbf{D}_m^T (\boldsymbol{\Sigma}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1}) \mathbf{D}_m d\text{vech}(\boldsymbol{\Sigma}_i) \\
&\quad + (d\boldsymbol{\mu})^T \boldsymbol{\Sigma}_i^{-1} d\boldsymbol{\mu}.
\end{aligned} \tag{3.41}$$

De (3.41) temos que a informação sobre  $\boldsymbol{\mu}$  e  $\text{vech}(\boldsymbol{\Sigma}_i)$  contida na  $i$ -ésima observação  $\mathbf{Z}_i$  é dada por

$$\mathcal{F}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2} \mathbf{D}_m^T (\boldsymbol{\Sigma}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1}) \mathbf{D}_m \end{pmatrix}, \quad i = 1, \dots, n. \tag{3.42}$$

□

**Observação:** Se adotamos a parametrização  $\boldsymbol{\theta}^* = (\boldsymbol{\mu}, \boldsymbol{\sigma})$ , temos que a matriz de informação esperada para  $\boldsymbol{\theta}^*$  pode ser expressa como uma matriz diagonal em blocos da forma

$$\mathcal{F}(\boldsymbol{\theta}^*) = \begin{pmatrix} \mathbf{I}(\boldsymbol{\mu}) & \mathbf{0} \\ \mathbf{0}^T & \mathbf{I}(\boldsymbol{\sigma}) \end{pmatrix},$$

onde  $\mathbf{I}(\boldsymbol{\mu}) = \sum_{i=1}^n \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_i)$  e  $\mathbf{I}(\boldsymbol{\sigma}) = \sum_{i=1}^n \mathbf{I}_i(\boldsymbol{\sigma})$ , com  $\mathbf{I}_i(\boldsymbol{\sigma}) = \mathbf{J}_i \mathbf{I}_i(\boldsymbol{\theta}_i) \mathbf{J}_i^T$ , onde

$$\mathbf{I}_i(\boldsymbol{\theta}_i) = \frac{1}{2} \mathbf{D}_m^T (\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_i) \otimes \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_i)) \mathbf{D}_m.$$

Assim para uma reparametrização específica  $\boldsymbol{\theta} = g(\boldsymbol{\theta}^*)$ , temos que a informação de Fisher associada ao novo vetor de parâmetros  $\boldsymbol{\theta}$  é dada por: (veja Apêndice B)

$$\begin{aligned}
\mathcal{F}(\boldsymbol{\theta}) &= \mathbf{J}(\boldsymbol{\theta}^*) \mathcal{F}(\boldsymbol{\theta}^*) \mathbf{J}^T(\boldsymbol{\theta}^*), \text{ onde } \mathcal{F}(\boldsymbol{\theta}^*) = \sum_{i=1}^n \mathcal{F}_i(\boldsymbol{\theta}^*), \text{ com} \\
\mathcal{F}_i(\boldsymbol{\theta}^*) &= \mathbf{J}_i \mathcal{F}_i(\boldsymbol{\theta}_i) \mathbf{J}_i^T, \quad i = 1, \dots, n,
\end{aligned} \tag{3.43}$$

em que  $\mathcal{F}_i(\boldsymbol{\theta}_i)$  é dada em (3.42) e  $\mathbf{J}_i$  é o Jacobiano associado à transformação.

### 3.9.5 Exemplos

**Exemplo 3.1.** Consideremos o seguinte modelo estrutural heteroscedástico com erros nas variáveis com erro de equação (modelo desenvolvido por Kulathinal et al., 2002)

$$\begin{aligned} y_i &= \alpha + \beta x_i + q_i \\ X_i &= x_i + u_i \\ Y_i &= y_i + e_i \quad i = 1, \dots, n, \end{aligned} \quad (3.44)$$

em que  $(e_i, u_i)^T$  e  $(q_j, x_j)^T$  são independentes,  $q_i$  e  $x_j$  são independentes,

$$q_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2), \quad x_i \stackrel{i.i.d.}{\sim} N(\mu_x, \sigma_x^2) \quad e$$

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \stackrel{indcp.}{\sim} N_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_{e_i}^2 & 0 \\ 0 & \sigma_{u_i}^2 \end{pmatrix} \right], \quad i, j = 1, \dots, n. \quad (3.45)$$

Note que este modelo corresponde ao modelo definido em (3.25) com  $\sigma_{x_i}^2 = \sigma_x^2$ ,  $i = 1, \dots, n$ . Para fazer o modelo identificável vamos supor que as variâncias dos erros de medida  $\sigma_{e_i}^2$  e  $\sigma_{u_i}^2$  são conhecidas e, portanto, o vetor de parâmetros desconhecidos é  $\theta = (\alpha, \beta, \mu_x, \sigma_x^2, \sigma^2)^T$ . Então a matriz de informação esperada para o parâmetro  $\theta$  pode ser obtida como segue.

Consideremos a seguinte parametrização  $\theta^* = (\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})^T$ , onde  $\mu_1 = \alpha + \beta\mu_x$ ,  $\mu_2 = \mu_x$ ,  $\sigma_{11} = \beta^2\sigma_x^2 + \sigma^2$ ,  $\sigma_{12} = \beta\sigma_x^2$ ,  $\sigma_{22} = \sigma_x^2$  e seja  $\theta_i = (\sigma_{11} + \sigma_{e_i}^2, \rho_i, \sigma_{22} + \sigma_{u_i}^2)^T$ , onde  $\rho_i$  é o coeficiente de correlação entre  $(Y_i, X_i)$  dado por

$$\rho_i = \frac{\sigma_{12}}{\sqrt{(\sigma_{11} + \sigma_{e_i}^2)(\sigma_{22} + \sigma_{u_i}^2)}}, \quad i = 1, \dots, n,$$

de forma que

$$\begin{aligned} \Sigma_i(\theta_i) &= \begin{pmatrix} \beta^2\sigma_x^2 + \sigma^2 + \sigma_{e_i}^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_{u_i}^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{11} + \sigma_{e_i}^2 & \rho_i \sqrt{(\sigma_{11} + \sigma_{e_i}^2)(\sigma_{22} + \sigma_{u_i}^2)} \\ \rho_i \sqrt{(\sigma_{11} + \sigma_{e_i}^2)(\sigma_{22} + \sigma_{u_i}^2)} & \sigma_{22} + \sigma_{u_i}^2 \end{pmatrix}. \end{aligned}$$

Então  $\mathcal{F}_i(\sigma_{11}, \sigma_{12}, \sigma_{22}) = J_i(\boldsymbol{\theta}_i)\mathcal{F}_i(\boldsymbol{\theta}_i)J_i^T(\boldsymbol{\theta}_i)$ , onde

$$J_i(\boldsymbol{\theta}_i) = \begin{pmatrix} 1 & -\frac{\sigma_{12}}{2(\sigma_{11} + \sigma_{e_i}^2)\sqrt{(\sigma_{11} + \sigma_{e_i}^2)(\sigma_{22} + \sigma_{u_i}^2)}} & 0 \\ 0 & \frac{1}{\sqrt{(\sigma_{11} + \sigma_{e_i}^2)(\sigma_{22} + \sigma_{u_i}^2)}} & 0 \\ 0 & -\frac{\sigma_{12}}{2(\sigma_{11} + \sigma_{e_i}^2)\sqrt{(\sigma_{22} + \sigma_{u_i}^2)(\sigma_{22} + \sigma_{u_i}^2)}} & 1 \end{pmatrix},$$

$$\mathcal{F}_i(\boldsymbol{\theta}_i) = \frac{1}{2}\mathbf{D}_2^T (\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_i) \otimes \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_i)) \mathbf{D}_2, \text{ com } \mathbf{D}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

segue que

$$\mathcal{F}(\boldsymbol{\theta}^*) = \begin{pmatrix} \sum_{i=1}^n \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_i) & \mathbf{0} \\ \mathbf{0}^T & \sum_{i=1}^n J_i(\boldsymbol{\theta}_i)\mathcal{F}_i(\boldsymbol{\theta}_i)J_i^T(\boldsymbol{\theta}_i) \end{pmatrix}.$$

Finalmente a matriz de informação do parâmetro  $\boldsymbol{\theta}$  é dada por  $\mathcal{F}(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta}^*)\mathcal{F}(\boldsymbol{\theta}^*)\mathbf{J}^T(\boldsymbol{\theta}^*)$ , onde

$$\mathbf{J}(\boldsymbol{\theta}^*) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \mu_2 & 0 & 2\sigma_{12} & \sigma_{22} & 0 \\ \frac{\sigma_{12}}{\sigma_{22}} & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma_{12}^2}{\sigma_{22}^2} & \frac{\sigma_{12}}{\sigma_{22}} & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

**Exemplo 3.2.** Consideremos agora o seguinte modelo estrutural heteroscedástico com erros nas variáveis

$$\begin{aligned} y_i &= \alpha + \beta x_i \\ X_i &= x_i + u_i \\ Y_i &= y_i + e_i \quad i = 1, \dots, n, \end{aligned} \tag{3.46}$$

em que  $e_i$  e  $x_j$  são independentes,  $u_i$  e  $x_j$  são independentes,

$$x_i \stackrel{\text{indep.}}{\sim} N(\mu_x, \sigma_{x_i}^2) \quad \text{e}$$

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right], \quad i, j = 1, \dots, n. \quad (3.47)$$

Note que este modelo corresponde ao modelo definido em (3.25) com  $q_i = 0$ ,  $\sigma_{e_i}^2 = \sigma_e^2$ ,  $\sigma_{u_i}^2 = \sigma_u^2$   $i = 1, \dots, n$ .

Para fazer o modelo identificável vamos supor que a variância da variável  $x_i$ ,  $\sigma_{x_i}^2$  é conhecida e, portanto, o vetor de parâmetros desconhecidos é  $\theta = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2)^\top$ .

De forma análoga ao exemplo anterior, sejam  $\theta^* = (\mu_1, \mu_2, \sigma_{11}, \tau, \sigma_{22})^\top$ , onde  $\mu_1 = \alpha + \beta\mu_x$ ,  $\mu_2 = \mu_x$ ,  $\sigma_{11} = \sigma_e^2$ ,  $\tau = \beta$ ,  $\sigma_{22} = \sigma_u^2$  e seja  $\theta_i = (\sigma_{11} + \tau^2\sigma_{x_i}^2, \rho_i, \sigma_{22} + \sigma_{x_i}^2)^\top$ , onde  $\rho_i$  é o coeficiente de correlação entre  $(Y_i, X_i)$  dado por

$$\rho_i = \frac{\tau\sigma_{x_i}^2}{\sqrt{(\sigma_{11} + \tau^2\sigma_{x_i}^2)(\sigma_{22} + \sigma_{x_i}^2)}}, \quad i = 1, \dots, n.$$

Logo, temos que

$$\begin{aligned} \Sigma_i(\theta_i) &= \begin{pmatrix} \beta^2\sigma_{x_i}^2 + \sigma_e^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_{x_i}^2 + \sigma_u^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{11} + \tau^2\sigma_{x_i}^2 & \rho_i\sqrt{(\sigma_{11} + \tau^2\sigma_{x_i}^2)(\sigma_{22} + \sigma_{x_i}^2)} \\ \rho_i\sqrt{(\sigma_{11} + \tau^2\sigma_{x_i}^2)(\sigma_{22} + \sigma_{x_i}^2)} & \sigma_{22} + \sigma_{x_i}^2 \end{pmatrix}, \end{aligned}$$

e segue que

$$J_i(\theta_i) = \begin{pmatrix} 1 & -\frac{\tau\sigma_{x_i}^2}{2(\sigma_{11} + \tau^2\sigma_{x_i}^2)\sqrt{(\sigma_{11} + \tau^2\sigma_{x_i}^2)(\sigma_{22} + \sigma_{x_i}^2)}} & 0 \\ 2\tau\sigma_{x_i}^2 & \frac{\sigma_{11}\sigma_{x_i}^2}{(\sigma_{11} + \tau^2\sigma_{x_i}^2)\sqrt{(\sigma_{11} + \tau^2\sigma_{x_i}^2)(\sigma_{22} + \sigma_{x_i}^2)}} & 0 \\ 0 & -\frac{\tau\sigma_{x_i}^2}{2(\sigma_{22} + \sigma_{x_i}^2)\sqrt{(\sigma_{11} + \tau^2\sigma_{x_i}^2)(\sigma_{22} + \sigma_{x_i}^2)}} & 1 \end{pmatrix}.$$

Então

$$\mathcal{F}(\mu_1, \mu_2) = \sum_{i=1}^n \Sigma_i^{-1}(\boldsymbol{\theta}_i) \quad \text{e} \quad \mathcal{F}(\sigma_{11}, \tau, \sigma_{22}) = \sum_{i=1}^n J_i(\boldsymbol{\theta}_i) \mathcal{F}_i(\boldsymbol{\theta}_i) J_i^T(\boldsymbol{\theta}_i),$$

de forma que

$$\mathcal{F}(\boldsymbol{\theta}^*) = \begin{pmatrix} \mathcal{F}(\mu_1, \mu_2) & \mathbf{0} \\ \mathbf{0}^T & \mathcal{F}(\sigma_{11}, \tau, \sigma_{22}) \end{pmatrix},$$

então

$$\mathcal{F}(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta}^*) \mathcal{F}(\boldsymbol{\theta}^*) \mathbf{J}^T(\boldsymbol{\theta}^*), \quad (3.48)$$

onde

$$\mathbf{J}(\boldsymbol{\theta}^*) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \mu_2 & 0 & 0 & 1 & 0 \\ \tau & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

**Observação:** Dos dados simulados na Tabela 3.2, podemos ver o comportamento dos erros padrões calculados usando a matriz de informação de Fisher associada aos parâmetros do modelo definido em (3.9) dada por (3.20) e dos erros padrões calculados usando a matriz de informação para o caso geral definido em (3.48), para este particular exemplo. Na Tabela 3.1 se apresenta o comportamento de tais erros padrões.

Olhando a Tabela 3.1, podemos ver que os valores calculados dos desvios padrões estão muito próximos entre eles, independente do tamanho da amostra, com alguma exceção no caso do desvio padrão do parâmetro  $\sigma_e^2$ . No geral poderíamos pensar que ambos métodos de estimação numericamente são equivalentes.



Tabela 3.1. Comportamento do DP-IE: Erro padrão usando a matriz de informação de Fisher dada em (3.20) e DP-IG: Erro padrão usando a matriz de informação geral dado em (3.48), quando usamos as estimativas de máxima verossimilhança na sua obtenção.

n	Método	$\alpha$	$\beta$	$\mu_x$	$\sigma_c^2$	$\sigma_u^2$
40	EMV	-1.072	0.975	-2.024	4.673	1.066
	DP-IE	(0.53)	(0.18)	(0.30)	(1.30)	(0.59)
	DP-IG	(0.54)	(0.18)	(0.30)	(1.93)	(0.56)
80	EMV	-1.040	0.986	-2.003	4.848	1.037
	DP-IE	(0.36)	(0.11)	(0.23)	(0.96)	(0.45)
	DP-IG	(0.34)	(0.10)	(0.23)	(1.39)	(0.43)
160	EMV	-1.033	0.989	-1.995	4.862	1.038
	DP-IE	(0.33)	(0.11)	(0.18)	(0.83)	(0.34)
	DP-IG	(0.30)	(0.09)	(0.18)	(1.09)	(0.33)

### 3.10 Estudo de Simulação

Nesta seção estudaremos o comportamento dos estimadores dos parâmetros de regressão do MREH com erros de medição proposto em (2.1) com as suposições em (3.1) e (3.2), assumindo a condição de identificabilidade  $\sigma_{x_i}^2, i = 1, \dots, n$ , conhecidas. Além disso estudaremos o comportamento das estatísticas do tipo Wald dadas em (2.18), (3.22) e dos erros padrões dos estimadores, obtidos usando as matrizes de informação de Fisher dadas em (3.20) e (3.48) respectivamente.

Para calcular as estimativas dos parâmetros do modelo consideramos 10,000 simulações de Monte Carlo. Para gerar os dados  $(X_i, Y_i), i = 1, \dots, n$ , usamos o seguinte conjunto de valores para os parâmetros:

$$\alpha = -1, \quad \beta = 1, \quad \mu_x = -2, \quad \sigma_c^2 = 5, \quad \text{e} \quad \sigma_u^2 = 1, \quad (3.49)$$

além disso a variância  $\sigma_{x_i}^2$  foi gerada como segue  $\sqrt{\sigma_{x_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 4), i = 1, \dots, n$  e foram mantidos fixos em todas as simulações. Os tamanhos amostrais utilizados nas simulações foram  $n = 40, 80$  e  $160$ . Em cada rodada foram calculadas as estimativas dos estimadores de momentos com seus respectivos desvios padrões, estas estimativas de momentos são usadas para iniciar o algoritmo EM proposto na Figura 3.1, permitindo obter as estimativas dos estimadores de máxima verossimilhança com seus erros padrões correspondentes. Se de fato o



procedimento de estimação é adequado, espera-se que as estimativas dos parâmetros estejam próximas dos valores dos parâmetros que foram usados na geração dos dados.

Os resultados da simulação são apresentados na Tabela 3.2, os valores que aparecem nesta tabela são médias tomadas ao longo das 10,000 amostras geradas. Observamos, que tanto o EMM como o EMV produzem resultados coerentes para os diferentes tamanhos amostrais, isto é, os valores das estimativas dos parâmetros obtidas pelo método dos momentos e de máxima verossimilhança ficam bem próximos dos valores verdadeiros utilizados na simulação. Além disso, é possível perceber que os erros padrões (em parênteses) das estimativas diminuem quando o tamanho amostral cresce, também observamos que os erros padrões das estimativas de máxima verossimilhança são menores que os erros padrões das estimativas de momentos.

Para examinar o comportamento das estatísticas do tipo Wald dadas em (2.18), (3.22), foram feitas simulações de Monte Carlo para calcular as taxas de rejeição do teste utilizando estas estatísticas ao nível nominal de 5%. A configuração para os parâmetros foi a seguinte:

$$\mu_x = -2, \quad \sigma_e^2 = 5, \quad \text{e} \quad \sigma_u^2 = 1, \quad \sqrt{\sigma_{x_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 4), \quad i = 1, \dots, n \quad (3.50)$$

A hipótese nula é  $\mathbf{H}_0 : (\alpha, \beta)^T = (0, 1)^T$  e os valores para  $(\alpha, \beta)$  utilizados para calcular as taxas de rejeição, foi a mesma grade de pontos igualmente espaçados em torno de  $(0, 1)$  utilizada na simulação do Capítulo 2, isto é:  $(-1, 0.6)$ ,  $(-1, 1)$ ,  $(-1, 1.4)$ ,  $(0, 0.6)$ ,  $(0, 1)$ ,  $(0, 1.4)$ ,  $(1, 0.6)$ ,  $(1, 1)$ ,  $(1, 1.4)$ . Esta grade de pontos nos permite monitorar não apenas o tamanho empírico, mas também o poder do teste. Os tamanhos amostrais utilizados nas simulações foram  $n = 40, 80, 160$  e para cada tripla  $(\alpha, \beta, n)$  foram realizadas 10,000 simulações de Monte Carlo, em cada rodada as estatísticas de Wald dadas em (2.18), (3.22) foram usadas para verificar as evidências contra a hipótese nula ao nível nominal de 5%. Se de fato o procedimento de estimação é adequado, espera-se que sob a hipótese nula, as taxas de rejeição sejam próximas de 5%.

A Tabela 3.3 apresenta os resultados da simulação onde observamos que as taxas de rejeição sob a hipótese nula tendem a 0.05 (5%) e as taxas de rejeição para valores de  $(\alpha, \beta)$  em torno da hipótese nula tendem a 1.00 (100%) à medida que o tamanho amostral cresce, como esperado.

### 3.11 Aplicação

Nesta seção aplicaremos os resultados deste capítulo ao conjunto de dados do Projeto WHO MONICA. Consideraremos os dos métodos de estimação aqui estudados: método dos momentos (EMM) com a matriz de covariâncias assintóticas obtida através do método delta

Tabela 3.2. Estimativas (erros-padrão) dos parâmetros de regressão do MEV estrutural heterocedástico na variável não observável, no caso  $\sqrt{\sigma_{x_i}^2} \overset{i.i.d.}{\sim} U(0.5, 4)$ ,  $i = 1, \dots, n$ . EP-IE: Erro padrão usando a matriz de informação esperada, EP-IG: Erro padrão usando a matriz de informação geral dado em (3.43) e EP-IO: Desvio padrão usando a matriz de informação observada.

n	Método	$\alpha$	$\beta$	$\mu_x$	$\sigma_e^2$	$\sigma_u^2$
40	EMM	-1.053 (5.021)	0.990 (2.477)	-2.041 (-)	4.455 (-)	0.975 (-)
	EMV	-1.072	0.975	-2.024	4.673	1.066
	EP-IE	(0.525)	(0.179)	(0.300)	(1.300)	(0.591)
	EP-IG	(0.537)	(0.183)	(0.300)	(1.934)	(0.560)
	EP-IO	(0.558)	(0.198)	(0.305)	(1.374)	(0.660)
	Verdadeiro	-1.000	1.000	-2.000	5.000	1.000
80	EMM	-1.017 (3.861)	0.998 (1.926)	-2.000 (-)	4.646 (-)	1.032 (-)
	EMV	-1.040	0.986	-2.003	4.848	1.037
	EP-IE	(0.362)	(0.119)	(0.225)	(0.961)	(0.456)
	EP-IG	(0.341)	(0.102)	(0.225)	(1.390)	(0.428)
	EP-IO	(0.377)	(0.130)	(0.224)	(0.998)	(0.497)
	Verdadeiro	-1.000	1.000	-2.000	5.000	1.000
160	EMM	-1.019 (3.258)	0.998 (1.635)	-1.997 (-)	4.721 (-)	0.997 (-)
	EMV	-1.033	0.989	-1.995	4.862	1.038
	EP-IE	(0.331)	(0.113)	(0.178)	(0.825)	(0.343)
	EP-IG	(0.298)	(0.086)	(0.178)	(1.097)	(0.329)
	EP-IO	(0.340)	(0.119)	(0.177)	(0.840)	(0.358)
	Verdadeiro	-1.000	1.000	-2.000	5.000	1.000

e método de máxima verossimilhança (EMV), considerando a estimativa dos parâmetros através do algoritmo EM, e a estimativa da matriz de covariâncias assintóticas derivada através da matriz de informação observada (EP-IO) e da matriz de informação esperada (EP-IE), respectivamente.

O MREH com erros nas variáveis sem erro na equação ( $\sigma^2 = 0$ ), estudado no capítulo 2 não é adequado para ajustar os dados do Projeto WHO MONICA, como foi mostrado por de Castro et al. (2008) através de envelopes simulados. A Tabela 3.4 apresenta as estimativas dos parâmetros estruturais  $\alpha$ ,  $\beta$ ,  $\mu_x$ ,  $\sigma_e^2$  e  $\sigma_u^2$  do modelo, com seus correspondentes erros

Tabela 3.3. Taxas de rejeição da hipótese  $\mathbf{H}_0 : (\alpha, \beta)^T = (0, 1)^T$ , usando as estatísticas de Wald dadas em (2.18) e (3.22) (ao nível nominal de 5% de significância) no MEV estrutural heterocedástico na variável não observável, para  $n = 40, 80$  e  $160$  respectivamente.

n	$\alpha$	$\mathcal{W}_{EMM}$			$\mathcal{W}_{EMV}$		
		$\beta$			$\beta$		
		0.6	1.0	1.4	0.6	1.0	1.4
40	-1	0.474	0.652	0.971	0.634	0.959	0.999
	0	0.755	<b>0.077</b>	0.632	0.953	<b>0.066</b>	0.971
	1	0.980	0.753	0.645	0.988	0.985	0.703
80	-1	0.742	0.891	1.000	0.974	0.998	1.000
	0	0.887	<b>0.065</b>	0.799	0.999	<b>0.056</b>	0.999
	1	1.000	0.790	0.591	1.000	0.996	0.987
160	-1	0.951	0.998	1.000	0.985	0.999	0.999
	0	0.993	<b>0.058</b>	0.998	0.998	<b>0.051</b>	1.000
	1	1.000	0.981	0.939	1.000	1.000	0.988

padrão (em parênteses). Podemos observar que as estimativas dos parâmetros  $\alpha$ ,  $\beta$ , e  $\sigma_e^2$  são bem diferentes se utilizarmos o EMM em vez do que EMV. No geral os erros padrão dos estimadores em ambos casos (homens e mulheres) são sempre menores se utilizarmos a estimativa da matriz de covariâncias assintóticas derivada através da matriz de informação esperada (EP-IE).

A Figura 3.2 apresenta as linhas estimadas de regressão utilizando os métodos de EMM e de EMV.

Para testar a avaliação do ajuste, nas Figuras 3.3 e 3.4 se apresentam os gráficos do QQ-plot e envelope para os dados dos homens e mulheres respectivamente. Nos gráficos se observa que o modelo ajusta melhor os dados dos homens do que das mulheres.

Os gráficos do QQ-plot e Envelope, foram construídos como em de Castro et al. (2008).

Tabela 3.4. Estimativas dos parâmetros estruturais  $\alpha$ ,  $\beta$ ,  $\mu_x$ ,  $\sigma_e^2$  e  $\sigma_u^2$ , com seus correspondentes erros padrão (em parênteses)

Sexo	Método	$\alpha$	$\beta$	$\mu_x$	$\sigma_e^2$	$\sigma_u^2$
Homens	EMM	1.197 (2.398)	3.163 (1.962)	-1.08 (-)	2.823 (-)	4.505 (-)
	EMV	-0.791	1.323	-1.138	6.746	4.208
	EP-IE	(1.840)	(1.525)	(0.350)	(2.366)	(1.065)
	EP-IO	(1.597)	(1.367)	(0.352)	(2.156)	(1.068)
Mulheres	EMM	7.007 (9.803)	4.211 (4.797)	-2.020 (-)	9.276 (-)	3.927 (-)
	EMV	2.674	2.066	-2.065	16.046	3.603
	EP-IE	(4.465)	(2.114)	(0.337)	(5.565)	(0.957)
	EP-IO	(5.577)	(2.722)	(0.337)	(6.499)	(0.980)

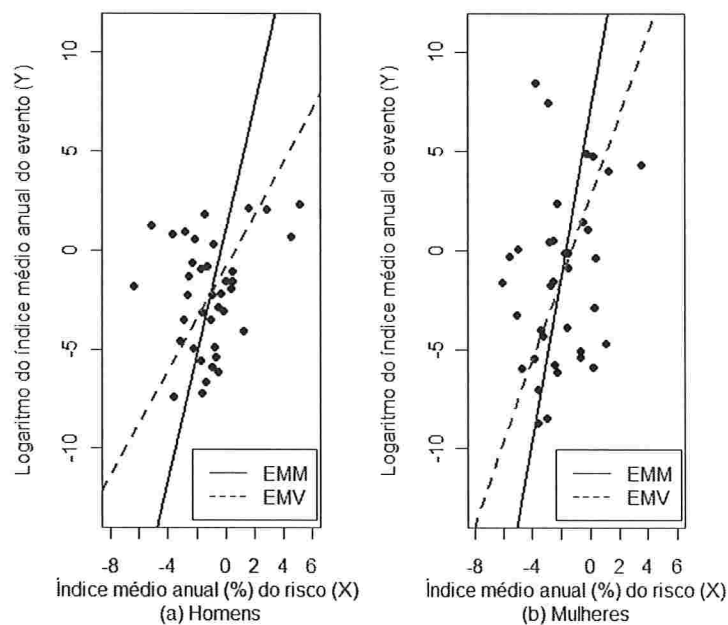


Figura 3.2. Linhas estimadas de regressão para os dados do projeto WHO MONICA

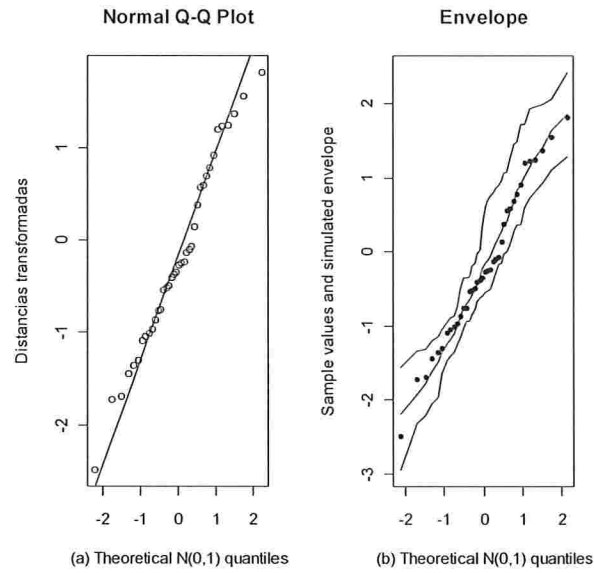


Figura 3.3. Gráfico QQ-plot e Envelope para os Homens

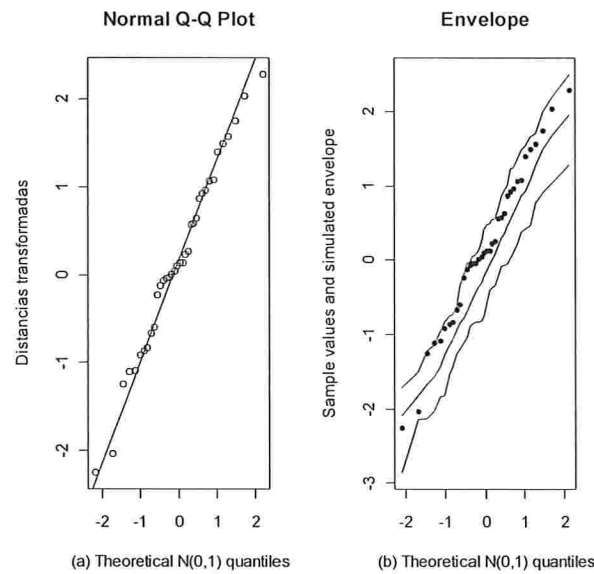


Figura 3.4. QQ-plot e Envelope das mulheres



## Capítulo 4

# Modelo Estrutural Heteroscedástico com Replicações Não-Balanceadas na Variável Exploratória

### 4.1 Introdução

Neste capítulo vamos a obter estimadores para os parâmetros do modelo definido em (2.1) com as suposições dadas em (3.1), no caso em que temos replicações apenas da variável explanatória, ou seja, para cada  $x_i$ ,  $m_i$  observações independentes,  $X_{ij}$ ,  $j = 1, \dots, m_i$ , são obtidas, isto é, amostra  $i$  é dividida em  $m_i$  subamostras, com o nível de concentração estimado por  $X_{ij}$ ,  $j = 1, \dots, m_i$ . Neste caso, não necessitamos de informações adicionais para tornar o modelo identificável. Reescrevemos o modelo considerado como

$$\begin{aligned} y_i &= \alpha + \beta x_i \\ X_{ij} &= x_i + u_{ij} \end{aligned} \quad (4.1)$$

$$Y_i = y_i + e_i \quad i = 1, \dots, n, \quad j = 1, \dots, m_i. \quad (4.2)$$

Suposições do modelo:

$$\begin{pmatrix} e_i \\ u_{ij} \\ x_i \end{pmatrix} \stackrel{\text{indep.}}{\sim} N_3 \left[ \begin{pmatrix} 0 \\ 0 \\ \mu_x \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_{x_i}^2 \end{pmatrix} \right], \quad i = 1, \dots, n, \quad j = 1, \dots, m_i. \quad (4.3)$$

De (4.3) notemos que

$$\begin{pmatrix} Y_i \\ X_{ij} \end{pmatrix} \stackrel{iid}{\sim} N_2 \left[ \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix}; \begin{pmatrix} \beta^2\sigma_{x_i}^2 + \sigma_e^2 & \beta\sigma_{x_i}^2 \\ \beta\sigma_{x_i}^2 & \sigma_{x_i}^2 + \sigma_u^2 \end{pmatrix} \right], \quad j = 1, \dots, m_i,$$

ou, equivalentemente,

$$\begin{pmatrix} Y_i \\ \mathbf{X}_i \end{pmatrix} \stackrel{indep.}{\sim} N_{m_i+1} \left[ \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \mathbf{J}_{m_i} \end{pmatrix}; \begin{pmatrix} \beta^2\sigma_{x_i}^2 + \sigma_e^2 & \beta\sigma_{x_i}^2 \mathbf{J}_{m_i}^T \\ \beta\sigma_{x_i}^2 \mathbf{J}_{m_i} & \sigma_{x_i}^2 \mathbf{J}_{m_i} \mathbf{J}_{m_i}^T + \sigma_u^2 \mathbf{I}_{m_i} \end{pmatrix} \right], \quad i = 1, \dots, n,$$

onde  $\mathbf{X}_i^T = (X_{i1}, X_{i2}, \dots, X_{im_i})$ ,  $\mathbf{J}_{m_i}$  vetor coluna de dimensão  $m_i$  com todos os elementos iguais a 1 e  $\mathbf{I}_{m_i}$  matriz identidade de ordem  $m_i$ .

## 4.2 Estimadores alternativos para os parâmetros $\sigma_u^2$ , $\bar{\sigma}_x^2$ e $\sigma_{x_i}^2$ envolvendo replicações da variável explanatória

A estimação de  $\sigma_{x_i}^2$  e  $\sigma_u^2$  (ou de  $a_i = \frac{\sigma_{x_i}^2}{\sigma_u^2}$ ) pode ser incorporada dentro do processo de estimação considerando a extensão do modelo (2.1), apresentada em (4.1) e (4.2) com as suposições dadas em (4.3).

**Definição 4.2.1.** Dada uma seqüência  $\{R_{ij}, j = 1, \dots, m_i, i = 1, \dots, n\}$ , definimos as médias  $\bar{R}_i$ ,  $\bar{R}_j$  e  $\bar{R}$  da seguinte maneira,

$$\bar{R}_i = \sum_{j=1}^{m_i} \frac{R_{ij}}{m_i}, \quad \bar{R}_j = \sum_{i=1}^n \frac{R_{ij}}{n}, \quad \bar{R} = \frac{\sum_{i=1}^n m_i \bar{R}_i}{\sum_{i=1}^n m_i}.$$

Agora de (4.1) e da definição (4.2.1) temos que:

$$\bar{X}_i = x_i + \bar{u}_i, \quad \bar{X}_{.j} = \bar{x} + \bar{u}_j, \quad \bar{X} = \bar{x} + \bar{u}, \quad \text{segue que}$$

$$X_{ij} - \bar{X}_i = u_{ij} - \bar{u}_i, \quad j = 1, \dots, m_i, \quad i = 1, \dots, n, \quad (4.4)$$



$$\bar{X}_i - \bar{X} = (x_i - \bar{x}) + (\bar{u}_i - \bar{u}), \quad i = 1, \dots, n, \quad (4.5)$$

$$X_{ij} - \bar{X}_{.j} = (x_i - \bar{x}) + (u_{ij} - \bar{u}_{.j}), \quad j = 1, \dots, m_i, \quad i = 1, \dots, n. \quad (4.6)$$

Usaremos os seguintes resultados (Veja Hocking (1996), Teorema 2.3, Corolário 2.3), para obter os estimadores para  $\sigma_u^2$ ,  $\bar{\sigma}_x^2$  e  $\sigma_{x_i}^2$  respectivamente.

**Teorema 4.1.** (Função geradora)

Seja  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , então a forma quadrática  $q = \mathbf{Y}^T \mathbf{A} \mathbf{Y}$  tem função geradora de momentos dada por:

$$m_q(t) = |I - 2t\mathbf{A}\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} \boldsymbol{\mu}^T (I - (I - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1}) \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right\},$$

para  $t < t_0$  o menor autovalor da matriz  $\mathbf{A}\boldsymbol{\Sigma}$ .

**Corolário 1.** (Momentos)

1. Os dois primeiros momentos de  $q$  são:

$$E(q) = \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} \quad (4.7)$$

$$\text{Var}(q) = 2 \text{tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}^T \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\mu} \quad (4.8)$$

2. Se  $q_1$  e  $q_2$  são formas quadráticas com matrizes  $\mathbf{A}_1$  e  $\mathbf{A}_2$ , então

$$\text{cov}(q_1, q_2) = 2 \text{tr}(\mathbf{A}_1 \boldsymbol{\Sigma} \mathbf{A}_2 \boldsymbol{\Sigma}) + 4\boldsymbol{\mu}^T \mathbf{A}_1 \boldsymbol{\Sigma} \mathbf{A}_2 \boldsymbol{\mu} \quad (4.9)$$

3. Se  $\mathbf{l} = \mathbf{B}\mathbf{Y}$  é uma forma linear, então

$$\text{cov}(\mathbf{l}, \mathbf{q}) = 2 \mathbf{B}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\mu} \quad (4.10)$$

**Proposição 4.1.** Seja

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^n (m_i - 1)}.$$

Então  $\hat{\sigma}_u^2$  é um estimador de  $\sigma_u^2$ , não viesado com variância dada por

$$\text{Var}(\hat{\sigma}_u^2) = \frac{2\sigma_u^4}{\sum_{i=1}^n (m_i - 1)}.$$

**Prova:** De (4.4) Notemos que

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^{m_i} (X_{ij} - \bar{X}_i)^2 &= \sum_{i=1}^n \sum_{j=1}^{m_i} (u_{ij} - \bar{u}_i)^2 \\ &= \mathbf{u}^T \mathbf{A} \mathbf{u} \end{aligned}$$

onde  $\mathbf{A} = \text{diag} \left\{ \left( \mathbf{I}_{m_1} - \frac{1}{m_1} \mathbf{J}_{m_1} \mathbf{J}_{m_1}^T \right), \left( \mathbf{I}_{m_2} - \frac{1}{m_2} \mathbf{J}_{m_2} \mathbf{J}_{m_2}^T \right), \dots, \left( \mathbf{I}_{m_n} - \frac{1}{m_n} \mathbf{J}_{m_n} \mathbf{J}_{m_n}^T \right) \right\}$ ,

$\mathbf{u}^T = (u_{11}, u_{12}, \dots, u_{1m_1}, u_{21}, \dots, u_{2m_2}, \dots, u_{n1}, u_{n2}, \dots, u_{nm_n}) \sim \text{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , em que  $\boldsymbol{\mu} = \mathbf{0}$  e

$\boldsymbol{\Sigma} = \sigma_u^2 \mathbf{I}_N$ , pois de (4.3) temos que  $u_{ij} \stackrel{i.i.d.}{\sim} \text{N}(0, \sigma_u^2)$ , além disso  $N = \sum_{i=1}^n m_i$ .

Após algumas manipulações algébricas obtemos

$$\mathbf{A} \boldsymbol{\Sigma} = \sigma_u^2 \text{diag} \left\{ \left( \mathbf{I}_{m_1} - \frac{1}{m_1} \mathbf{J}_{m_1} \mathbf{J}_{m_1}^T \right), \left( \mathbf{I}_{m_2} - \frac{1}{m_2} \mathbf{J}_{m_2} \mathbf{J}_{m_2}^T \right), \dots, \left( \mathbf{I}_{m_n} - \frac{1}{m_n} \mathbf{J}_{m_n} \mathbf{J}_{m_n}^T \right) \right\},$$

$$\mathbf{A} \boldsymbol{\Sigma} \mathbf{A} \boldsymbol{\Sigma} = \sigma_u^4 \text{diag} \left\{ \left( \mathbf{I}_{m_1} - \frac{1}{m_1} \mathbf{J}_{m_1} \mathbf{J}_{m_1}^T \right), \left( \mathbf{I}_{m_2} - \frac{1}{m_2} \mathbf{J}_{m_2} \mathbf{J}_{m_2}^T \right), \dots, \left( \mathbf{I}_{m_n} - \frac{1}{m_n} \mathbf{J}_{m_n} \mathbf{J}_{m_n}^T \right) \right\},$$

$$\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} = 0 \quad \text{e} \quad \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\Sigma} \mathbf{A} \boldsymbol{\mu} = 0.$$

Então, usando (4.7) temos que

$$\begin{aligned}
\mathbb{E} \left( \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^n (m_i - 1)} \right) &= \frac{1}{\sum_{i=1}^n (m_i - 1)} \mathbb{E} \left( \sum_{i=1}^n \sum_{j=1}^{m_i} (u_{ij} - \bar{u}_i)^2 \right) \\
&= \frac{\sigma_u^2 \operatorname{tr} \left\{ \operatorname{diag} \left[ \left( \mathbf{I}_{m_1} - \frac{1}{m_1} \mathbf{J}_{m_1} \mathbf{J}_{m_1}^T \right), \dots, \left( \mathbf{I}_{m_n} - \frac{1}{m_n} \mathbf{J}_{m_n} \mathbf{J}_{m_n}^T \right) \right] \right\}}{\sum_{i=1}^n (m_i - 1)} \\
&= \sigma_u^2 \frac{\sum_{i=1}^n m_i - n}{\sum_{i=1}^n (m_i - 1)} \\
&= \sigma_u^2,
\end{aligned}$$

e de (4.8) segue

$$\begin{aligned}
\operatorname{Var} \left( \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^n (m_i - 1)} \right) &= \frac{1}{(\sum_{i=1}^n (m_i - 1))^2} \operatorname{Var} \left( \sum_{i=1}^n \sum_{j=1}^{m_i} (u_{ij} - \bar{u}_i)^2 \right) \\
&= 2\sigma_u^4 \frac{\operatorname{tr} \left\{ \operatorname{diag} \left[ \left( \mathbf{I}_{m_1} - \frac{1}{m_1} \mathbf{J}_{m_1} \mathbf{J}_{m_1}^T \right), \dots, \left( \mathbf{I}_{m_n} - \frac{1}{m_n} \mathbf{J}_{m_n} \mathbf{J}_{m_n}^T \right) \right] \right\}}{(\sum_{i=1}^n (m_i - 1))^2} \\
&= 2\sigma_u^4 \frac{\sum_{i=1}^n m_i - n}{(\sum_{i=1}^n (m_i - 1))^2} \\
&= \frac{2\sigma_u^4}{\sum_{i=1}^n (m_i - 1)}.
\end{aligned}$$

□

**Proposição 4.2.** Seja

$$\hat{\sigma}_x^2 = \frac{1}{n-1} \left[ \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 - \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right) \hat{\sigma}_u^2 \right].$$

Então  $\hat{\sigma}_x^2$  é um estimador de  $\bar{\sigma}_x^2$ , não viesado com variância dada por

$$\begin{aligned}
\operatorname{Var}(\hat{\sigma}_x^2) &= \frac{2}{(n-1)^2} \left[ \frac{(n-2)}{n} \sum_{i=1}^n (\hat{\sigma}_{x_i}^2)^2 + (\hat{\sigma}_x^2)^2 + (a_1 + a_2) (\hat{\sigma}_u^2)^2 \right. \\
&\quad \left. + 2 \left( \frac{(n-2)}{n} \sum_{i=1}^n \frac{\hat{\sigma}_{x_i}^2}{m_i} + a_3 \hat{\sigma}_x^2 \right) \hat{\sigma}_u^2 \right],
\end{aligned}$$

onde

$$a_1 = \sum_{i=1}^n \left( \frac{1}{m_i} - \frac{1}{N} \right)^2 + \frac{n(n-1)}{N^2}, \quad a_2 = \frac{1}{\sum_{i=1}^n (m_i - 1)} \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right)^2, \quad e$$

$$a_3 = \frac{1}{n} \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{(n-1)}{\bar{m}} \right)$$

Prova: De (4.5) temos que

$$\begin{aligned} \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 &= \sum_{i=1}^n [(x_i - \bar{x}) + (\bar{u}_i - \bar{u})]^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (\bar{u}_i - \bar{u})^2 + 2 \sum_{i=1}^n (x_i - \bar{x})(\bar{u}_i - \bar{u}) \end{aligned}$$

segue

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}, \quad \text{onde} \quad \mathbf{A} = \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \mathbf{J}_n^T, \quad \mathbf{x}^T = (x_1, \dots, x_n) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

com  $\boldsymbol{\mu} = \mu_x \mathbf{J}_n$ , e  $\boldsymbol{\Sigma} = \text{diag}(\sigma_{x_1}^2, \sigma_{x_2}^2, \dots, \sigma_{x_n}^2)$ , pois de (4.3),  $x_i \stackrel{indep.}{\sim} N(\mu_x, \sigma_{x_i}^2)$ .

Após várias manipulações algébricas

$$\text{tr}(\mathbf{A}\boldsymbol{\Sigma}) = \left( \frac{n-1}{n} \right) \sum_{i=1}^n \sigma_{x_i}^2, \quad \text{tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}) = \sum_{i=1}^n \left[ \left( \frac{n-2}{n} \right) \sigma_{x_i}^4 + \frac{1}{n^2} \sigma_{x_i}^2 \sum_{i=1}^n \sigma_{x_i}^2 \right],$$

$$\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} = 0 \quad e \quad \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\Sigma} \mathbf{A} \boldsymbol{\mu} = 0.$$

Similarmente temos

$$\sum_{i=1}^n (\bar{u}_i - \bar{u})^2 = \mathbf{u}^T \mathbf{A} \mathbf{u},$$

onde

$$\mathbf{A} = \text{diag} \left( \frac{1}{m_1^2} \mathbf{J}_{m_1} \mathbf{J}_{m_1}^T, \dots, \frac{1}{m_n^2} \mathbf{J}_{m_n} \mathbf{J}_{m_n}^T \right) - \frac{1}{N} \left[ \frac{1}{m_1} \mathbf{J}_N \mathbf{J}_{m_1}^T \quad \frac{1}{m_2} \mathbf{J}_N \mathbf{J}_{m_2}^T \quad \dots \quad \frac{1}{m_n} \mathbf{J}_N \mathbf{J}_{m_n}^T \right].$$

$\mathbf{u}^T = (u_{11}, u_{12}, \dots, u_{1m_1}, u_{21}, \dots, u_{2m_2}, \dots, u_{n1}, u_{n2}, \dots, u_{nm_n}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , em que

$$\boldsymbol{\mu} = \mathbf{0} \quad \text{e} \quad \boldsymbol{\Sigma} = \sigma_u^2 \mathbf{I}_N, \quad N = \sum_{i=1}^n m_i.$$

Assim

$$\text{tr}(\mathbf{A}\boldsymbol{\Sigma}) = \sigma_u^2 \left[ \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{m_i^2} - \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{m_i} \right] = \sigma_u^2 \left[ \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right],$$

$$\begin{aligned} \text{tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}) &= \sigma_u^4 \left[ \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{m_i^3} - \frac{2}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{m_i^2} + \frac{1}{N^2} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{n}{m_i} \right] \\ &= \sigma_u^4 \left[ \sum_{i=1}^n \left( \frac{1}{m_i} - \frac{1}{N} \right)^2 + \frac{n(n-1)}{N^2} \right], \end{aligned}$$

$$\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} = 0 \quad \text{e} \quad \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\Sigma} \mathbf{A} \boldsymbol{\mu} = 0.$$

Então considerando os resultados obtidos na proposição (4.1) temos que

$$\begin{aligned} E(\hat{\sigma}_x^2) &= E \left( \frac{1}{n-1} \left[ \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 - \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right) \hat{\sigma}_u^2 \right] \right) \\ &= \frac{1}{n-1} \left[ E \left( \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 \right) - \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right) E(\hat{\sigma}_u^2) \right] \\ &= \frac{1}{n-1} \left[ E \left( \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 \right) - \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right) \sigma_u^2 \right], \end{aligned}$$

de (4.7) levando em conta que

$$\begin{aligned} E \left( \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 \right) &= E \left( \sum_{i=1}^n (x_i - \bar{x})^2 \right) + E \left( \sum_{i=1}^n (\bar{u}_i - \bar{u})^2 \right) \\ &= \left( \frac{n-1}{n} \right) \sum_{i=1}^n \sigma_{x_i}^2 + \sigma_u^2 \left[ \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right] \\ &= (n-1) \bar{\sigma}_x^2 + \left[ \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right] \sigma_u^2 \end{aligned}$$

e substituindo obtemos  $E(\hat{\sigma}_x^2) = \bar{\sigma}_x^2$ .

Para o cálculo da variância temos

$$\begin{aligned}
\text{Var}(\hat{\sigma}_x^2) &= \frac{1}{(n-1)^2} \text{Var} \left[ \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 - \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right) \hat{\sigma}_u^2 \right] \\
&= \frac{1}{(n-1)^2} \left[ \text{Var} \left( \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 \right) + \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right)^2 \text{Var}(\hat{\sigma}_u^2) \right] \\
&\quad - 2 \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right) \text{Cov} \left( \sum_{i=1}^n (\bar{X}_i - \bar{X})^2, \hat{\sigma}_u^2 \right),
\end{aligned}$$

e pela independência de  $x_i$  e  $u_{kj} \forall i, k, j$  e de (4.8) segue

$$\begin{aligned}
\text{Var} \left( \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 \right) &= \text{Var} \left( \sum_{i=1}^n (x_i - \bar{x})^2 \right) + \text{Var} \left( \sum_{i=1}^n (\bar{u}_i - \bar{u})^2 \right) \\
&\quad + 4 \sum_{i=1}^n \text{Var} [(x_i - \bar{x})(\bar{u}_i - \bar{u})] \\
&= 2 \sum_{i=1}^n \left[ \left( \frac{n-2}{n} \right) \sigma_{x_i}^4 + \frac{1}{n^2} \sigma_{x_i}^2 \sum_{i=1}^n \sigma_{x_i}^2 \right] + 2 \sigma_u^4 \left[ \sum_{i=1}^n \left( \frac{1}{m_i} - \frac{1}{N} \right)^2 + \frac{n(n-1)}{N^2} \right] \\
&\quad + 4 \sum_{i=1}^n \left[ \left( \frac{n-2}{n} \sigma_{x_i}^2 + \frac{1}{n} \bar{\sigma}_x^2 \right) \left( \frac{1}{m_i} - \frac{1}{N} \right) \sigma_u^2 \right] \\
&= 2 \left( \frac{n-2}{n} \sum_{i=1}^n \sigma_{x_i}^4 + (\bar{\sigma}_x^2)^2 \right) + 2 \left[ \sum_{i=1}^n \left( \frac{1}{m_i} - \frac{1}{N} \right)^2 + \frac{n(n-1)}{N^2} \right] \sigma_u^4 \\
&\quad + 4 \left[ \frac{n-2}{n} \sum_{i=1}^n \frac{\sigma_{x_i}^2}{m_i} + \frac{1}{n} \bar{\sigma}_x^2 \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{n-1}{\bar{m}} \right) \right] \sigma_u^2,
\end{aligned}$$

e

$$\text{Cov} \left( \sum_{i=1}^n (\bar{X}_i - \bar{X})^2, \hat{\sigma}_u^2 \right) = \frac{1}{\sum_{i=1}^n (m_i - 1)} \text{Cov} \left( \sum_{i=1}^n (\bar{u}_i - \bar{u})^2, \sum_{i=1}^n \sum_{j=1}^{m_i} (u_{ij} - \bar{u}_i)^2 \right),$$

segue de (4.9) que

$$\text{Cov} \left( \sum_{i=1}^n (\bar{u}_i - \bar{u})^2, \sum_{i=1}^n \sum_{j=1}^{m_i} (u_{ij} - \bar{u}_i)^2 \right) = 0.$$

Finalmente, considerando os resultados obtidos na proposição (4.1) e substituindo temos

$$\begin{aligned}
\text{Var}(\hat{\sigma}_x^2) &= \frac{2}{(n-1)^2} \left\{ \frac{(n-2)}{n} \sum_{i=1}^n (\hat{\sigma}_{x_i}^2)^2 + (\hat{\sigma}_x^2)^2 + \left[ \sum_{i=1}^n \left( \frac{1}{m_i} - \frac{1}{\bar{N}} \right)^2 + \frac{n(n-1)}{N^2} \right] (\hat{\sigma}_u^2)^2 \right. \\
&\quad \left. + 2 \left[ \frac{(n-2)}{n} \sum_{i=1}^n \frac{\hat{\sigma}_{x_i}^2}{m_i} + \frac{1}{n} \hat{\sigma}_x^2 \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{(n-1)}{\bar{m}} \right) \right] \hat{\sigma}_u^2 \right. \\
&\quad \left. + \left( \sum_{i=1}^n \frac{1}{m_i} - \frac{1}{\bar{m}} \right)^2 \frac{(\hat{\sigma}_u^2)^2}{\sum_{i=1}^n (m_i - 1)} \right\} \\
&= \frac{2}{(n-1)^2} \left[ \frac{(n-2)}{n} \sum_{i=1}^n (\hat{\sigma}_{x_i}^2)^2 + (\hat{\sigma}_x^2)^2 + (a_1 + a_2) (\hat{\sigma}_u^2)^2 \right. \\
&\quad \left. + 2 \left( \frac{(n-2)}{n} \sum_{i=1}^n \frac{\hat{\sigma}_{x_i}^2}{m_i} + a_3 \hat{\sigma}_x^2 \right) \hat{\sigma}_u^2 \right].
\end{aligned}$$

□

**Proposição 4.3.** Seja

$$\hat{\sigma}_{x_i}^2 = \frac{n}{(n-2)m_i} \sum_{j=1}^{m_i} (X_{ij} - \bar{X}_{.j})^2 - \frac{1}{n-2} \hat{\sigma}_x^2 - \frac{(n-1)}{(n-2)} \hat{\sigma}_u^2, \quad i = 1, \dots, n.$$

Então  $\hat{\sigma}_{x_i}^2$  é um estimador não viesado para  $\sigma_{x_i}^2$ ,  $i = 1, \dots, n$ .

**Prova:** Considerando os resultados obtidos nas proposições (4.1) e (4.2) temos

$$\begin{aligned}
E(\hat{\sigma}_{x_i}^2) &= \frac{n}{(n-2)m_i} \sum_{j=1}^{m_i} E[(X_{ij} - \bar{X}_{.j})^2] - \frac{1}{n-2} E(\hat{\sigma}_x^2) - \frac{(n-1)}{(n-2)} E(\hat{\sigma}_u^2) \\
&= \frac{n}{(n-2)m_i} \sum_{j=1}^{m_i} E[(x_i - \bar{x}) + (u_{ij} - \bar{u}_{.j})]^2 - \frac{1}{n-2} \bar{\sigma}_x^2 - \frac{(n-1)}{(n-2)} \sigma_u^2,
\end{aligned}$$

mas

$$\begin{aligned}
E[(x_i - \bar{x}) + (u_{ij} - \bar{u}_{.j})]^2 &= E(x_i - \bar{x})^2 + E(u_{ij} - \bar{u}_{.j})^2 \\
&= \text{Var}(x_i - \bar{x}) + \text{Var}(u_{ij} - \bar{u}_{.j}) \\
&= \left( \sigma_{x_i}^2 - \frac{2}{n} \sigma_{x_i}^2 + \frac{1}{n^2} \sum_{i=1}^n \sigma_{x_i}^2 \right) + \left( \sigma_u^2 - \frac{2}{n} \sigma_u^2 + \frac{1}{n} \sigma_u^2 \right) \\
&= \frac{1}{n} [(n-2)\sigma_{x_i}^2 + \bar{\sigma}_x^2 + (n-1)\sigma_u^2],
\end{aligned}$$



substituindo obtemos  $E(\hat{\sigma}_{x_i}^2) = \sigma_{x_i}^2, \quad i = 1, \dots, n.$  □

### 4.3 Estimação de máxima verossimilhança

Nesta seção vamos a obter a solução de máxima verossimilhança sob o modelo definido em (4.1), (4.2) e (4.3). Como veremos, é um caso sem solução explícita para as equações de verossimilhança e será resolvido através do algoritmo EM.

#### 4.3.1 Máxima verossimilhança

O vetor de dados observados  $\mathbf{Z}_i = (Y_i, \mathbf{X}_i^T)^T$ , onde  $\mathbf{X}_i = (X_{i1}, \dots, X_{im_i})^T$ , sob a suposição em (4.3) tem distribuição conjunta dada por  $\mathbf{Z}_i \stackrel{indep}{\sim} N_{m_i+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$ , em que:

$$\boldsymbol{\mu} = \mathbf{a} + \mu_x \mathbf{b} = \begin{pmatrix} \alpha + \beta \mu_x \\ \mu_x \mathbf{J}_{m_i} \end{pmatrix}, \quad e \quad (4.11)$$

$$\boldsymbol{\Sigma}_i = \sigma_{x_i}^2 \mathbf{b} \mathbf{b}^T + \boldsymbol{\Psi} = \begin{pmatrix} \beta^2 \sigma_{x_i}^2 + \sigma_e^2 & \beta \sigma_{x_i}^2 \mathbf{J}_{m_i}^T \\ \beta \sigma_{x_i}^2 \mathbf{J}_{m_i} & \sigma_{x_i}^2 \mathbf{J}_{m_i} \mathbf{J}_{m_i}^T + \sigma_u^2 \mathbf{I}_{m_i} \end{pmatrix}, \quad (4.12)$$

onde

$$\mathbf{a} = \begin{pmatrix} \alpha \\ \mathbf{0}_{m_i} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \beta \\ \mathbf{J}_{m_i} \end{pmatrix}, \quad e \quad \boldsymbol{\Psi} = \begin{pmatrix} \sigma_e^2 & \mathbf{0}^T \\ \mathbf{0} & \sigma_u^2 \mathbf{I}_{m_i} \end{pmatrix}, \quad i = 1, \dots, n. \quad (4.13)$$

Visto que já temos na Proposição (4.3) um estimador disponível para  $\sigma_{x_i}^2, i = 1, \dots, n$  (Gleser (1992)), nosso parâmetro de interesse será  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2)^T$ . Assim a função de log-verossimilhança é dada por

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n l_i(\boldsymbol{\theta}), \quad \text{onde}$$

$$l_i(\boldsymbol{\theta}) = \text{constante} - \frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} (\mathbf{Z}_i - \mathbf{a} - \mu_x \mathbf{b})^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{Z}_i - \mathbf{a} - \mu_x \mathbf{b}), \quad (4.14)$$

onde o determinante e inversa da matriz de covariâncias  $\Sigma_i$ , são dados por

$$|\Sigma_i| = |\Psi|(1 + \sigma_{x_i}^2 \mathbf{b}^T \Psi^{-1} \mathbf{b}) = \sigma_e^2 \sigma_u^{2m_i} \sigma_{x_i}^2 k_i \quad \text{e} \quad \Sigma_i^{-1} = \Psi^{-1} - k_i^{-1} \Psi^{-1} \mathbf{b} \mathbf{b}^T \Psi^{-1}, \quad (4.15)$$

com

$$k_i = \sigma_{x_i}^{-2} (1 + \sigma_{x_i}^2 \mathbf{b}^T \Psi^{-1} \mathbf{b}) = \sigma_{x_i}^{-2} + \beta^2 \sigma_e^{-2} + m_i \sigma_u^{-2}, \quad i = 1, \dots, n.$$

Substituindo (4.15) em (4.14) e após várias manipulações algébricas resulta

$$l_i(\boldsymbol{\theta}) \propto -\frac{1}{2} \log |\Sigma_i| - \frac{1}{2} [w_i - k_i^{-1} f_i^2], \quad (4.16)$$

onde

$$w_i = \sigma_e^{-2} (Y_i - \alpha - \beta \mu_x)^2 + \sigma_u^{-2} \sum_{j=1}^{m_i} (X_{ij} - \mu_x)^2$$

$$f_i = \beta \sigma_e^{-2} (Y_i - \alpha - \beta \mu_x) + \sigma_u^{-2} \sum_{j=1}^{m_i} (X_{ij} - \mu_x), \quad i = 1, \dots, n.$$

### 4.3.2 Equações de verossimilhança

Usando (3.16) para derivar (4.16) em relação aos parâmetros do modelo obtemos

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha} = 0 \implies \sigma_e^{-2} \sum_{i=1}^n (Y_i - \alpha - \beta \mu_x) - \beta \sigma_e^{-2} \sum_{i=1}^n k_i^{-1} f_i = 0, \quad (4.17)$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \beta} = 0 \implies \sigma_e^{-2} \sum_{i=1}^n [\mu_x (Y_i - \alpha - \beta \mu_x) + k_i^{-1} f_i (Y_i - \alpha - 2\beta \mu_x)]$$

$$- \beta \sigma_e^{-2} \sum_{i=1}^n (k_i^{-1} + f_i^2 k_i^{-2}) = 0, \quad (4.18)$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \mu_x} = 0 \implies \sum_{i=1}^n \sigma_{x_i}^{-2} k_i^{-1} f_i = 0, \quad (4.19)$$

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma_e^2} = 0 \implies & -\sum_{i=1}^n (\sigma_e^{-2} - \sigma_e^{-4} \beta^2 k_i^{-1}) + \sigma_e^{-4} \sum_{i=1}^n (Y_i - \alpha - \beta \mu_x)^2 \\ & - 2\beta \sigma_e^{-4} \sum_{i=1}^n k_i^{-1} f_i(Y_i - \alpha - \beta \mu_x) + \beta^2 \sigma_e^{-4} \sum_{i=1}^n k_i^{-2} f_i^2 = 0, \end{aligned} \quad (4.20)$$

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma_u^2} = 0 \implies & -\sum_{i=1}^n m_i (\sigma_u^{-2} - \sigma_u^{-4} k_i^{-1}) + \sigma_u^{-4} \sum_{i=1}^n \sum_{j=1}^{m_i} (X_{ij} - \mu_x)^2 \\ & - 2\sigma_u^{-4} \sum_{i=1}^n \sum_{j=1}^{m_i} k_i^{-1} f_i(X_{ij} - \mu_x) + \sigma_u^{-4} \sum_{i=1}^n m_i k_i^{-2} f_i^2 = 0. \end{aligned} \quad (4.21)$$

Com a presença de  $k_i$  nos denominadores dos termos dos somatórios sobre  $i$  nas equações de verossimilhança, conclui-se, que não pode haver soluções explícitas, neste caso a utilização do algoritmo EM pode se apresentar como uma alternativa a ser considerada, devido a que a maximização da log-verossimilhança completa (que inclui os dados não observados) é em geral mais simples que a log-verossimilhança observada.

### 4.3.3 Algoritmo EM

Sob a suposição feita em (4.3) o vetor de dados completos  $\mathbf{W}_i = (Y_i, \mathbf{X}_i^T, x_i)^T$ , tem distribuição conjunta dada por

$$\mathbf{W}_i \stackrel{indep}{\sim} N_{m_i+1}(\mathbf{v}, \boldsymbol{\Omega}_i), \quad i = 1, \dots, n, \quad (4.22)$$

onde

$$\mathbf{v} = \begin{pmatrix} \alpha + \beta \mu_x \\ \mu_x \mathbf{J}_{m_i} \\ \mu_x \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Omega}_i = \begin{pmatrix} \beta^2 \sigma_{x_i}^2 + \sigma_e^2 & \beta \sigma_{x_i}^2 \mathbf{J}_{m_i}^T & \beta \sigma_{x_i}^2 \\ \beta \sigma_{x_i}^2 \mathbf{J}_{m_i} & \sigma_{x_i}^2 \mathbf{J}_{m_i} \mathbf{J}_{m_i}^T + \sigma_u^2 \mathbf{I}_{m_i} & \sigma_{x_i}^2 \mathbf{J}_{m_i} \\ \beta \sigma_{x_i}^2 & \sigma_{x_i}^2 \mathbf{J}_{m_i}^T & \sigma_{x_i}^2 \end{pmatrix}. \quad (4.23)$$

A função de log-verossimilhança associada ao modelo em (4.16) é dada por

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{i=1}^n l_i(\boldsymbol{\theta}), \quad \text{onde} \\ l_i(\boldsymbol{\theta}) &= \text{constante} - \frac{1}{2} \log |\boldsymbol{\Omega}_i| - \frac{1}{2} (\mathbf{w}_i - \mathbf{v})^T \boldsymbol{\Omega}_i^{-1} (\mathbf{w}_i - \mathbf{v}). \end{aligned} \quad (4.24)$$

As expressões para o determinante e inversa da matriz de covariâncias  $\Omega_i$ , neste caso são

$$|\Omega_i| = \sigma_{x_i}^2 \sigma_e^2 \sigma_u^{2m_i} \quad e \quad \Omega_i^{-1} = \begin{pmatrix} \sigma_e^{-2} & \mathbf{0}^T & -\beta \sigma_e^{-2} \\ \mathbf{0} & \sigma_u^{-2} \mathbf{I}_{m_i} & -\sigma_u^{-2} \mathbf{J}_{m_i} \\ -\beta \sigma_e^{-2} & -\sigma_u^{-2} \mathbf{J}_{m_i}^T & \sigma_{x_i}^{-2} + \beta^2 \sigma_e^{-2} + m_i \sigma_u^{-2} \end{pmatrix}.$$

Substituindo, as expressões de acima em (4.24), operações algébricas conduzem a

$$\begin{aligned} l_{c_i}(\boldsymbol{\theta}) &\propto -\frac{1}{2} \ln(\sigma_e^2) - \frac{m_i}{2} \ln(\sigma_u^2) - \frac{1}{2} \ln(\sigma_{x_i}^2) - \frac{1}{2} \sigma_e^{-2} (Y_i - \alpha - \beta \mu_x)^2 \\ &\quad + \beta \sigma_e^{-2} (Y_i - \alpha - \beta \mu_x) x_i - \beta \sigma_e^{-2} \mu_x (Y_i - \alpha - \beta \mu_x) - \frac{1}{2} \sigma_u^{-2} \sum_{j=1}^{m_i} (X_{ij} - \mu_x)^2 \\ &\quad + \sigma_u^{-2} \sum_{j=1}^{m_i} (X_{ij} - \mu_x) x_i - \sigma_u^{-2} \mu_x \sum_{j=1}^{m_i} (X_{ij} - \mu_x) - \frac{1}{2} (\sigma_{x_i}^{-2} + \beta^2 \sigma_e^{-2} + m_i \sigma_u^{-2}) x_i^2 \\ &\quad + (\sigma_{x_i}^{-2} + \beta^2 \sigma_e^{-2} + m_i \sigma_u^{-2}) \mu_x x_i - \frac{1}{2} (\sigma_{x_i}^{-2} + \beta^2 \sigma_e^{-2} + m_i \sigma_u^{-2}) \mu_x^2. \end{aligned}$$

Agora na etapa E do algoritmo EM temos que

$$Q(\boldsymbol{\theta}) = \sum_{i=1}^n Q_i(\boldsymbol{\theta}), \quad \text{onde}$$

$$\begin{aligned} Q_i(\boldsymbol{\theta}) &= E(l_{c_i}(\boldsymbol{\theta} | \mathbf{W}_i) | Y_i, \mathbf{X}_i) = E(l_{c_i}(\boldsymbol{\theta} | x_i) | Y_i, \mathbf{X}_i) \\ &\propto -\frac{1}{2} \ln(\sigma_e^2) - \frac{m_i}{2} \ln(\sigma_u^2) - \frac{1}{2} \ln(\sigma_{x_i}^2) - \frac{1}{2} \sigma_e^{-2} (Y_i - \alpha - \beta \mu_x)^2 \\ &\quad + \beta \sigma_e^{-2} (Y_i - \alpha - \beta \mu_x) \hat{x}_i - \beta \sigma_e^{-2} \mu_x (Y_i - \alpha - \beta \mu_x) - \frac{1}{2} \sigma_u^{-2} \sum_{j=1}^{m_i} (X_{ij} - \mu_x)^2 \\ &\quad + \sigma_u^{-2} \sum_{j=1}^{m_i} (X_{ij} - \mu_x) \hat{x}_i - \sigma_u^{-2} \mu_x \sum_{j=1}^{m_i} (X_{ij} - \mu_x) - \frac{1}{2} (\sigma_{x_i}^{-2} + \beta^2 \sigma_e^{-2} + m_i \sigma_u^{-2}) \hat{x}_i^2 \\ &\quad + (\sigma_{x_i}^{-2} + \beta^2 \sigma_e^{-2} + m_i \sigma_u^{-2}) \mu_x \hat{x}_i - \frac{1}{2} (\sigma_{x_i}^{-2} + \beta^2 \sigma_e^{-2} + m_i \sigma_u^{-2}) \mu_x^2, \end{aligned}$$

onde

$$\hat{x}_i = E(x_i) \quad e \quad \hat{x}_i^2 = v(x_i) + (\hat{x}_i)^2, \quad i = 1, \dots, n,$$

com

$$v(x_i) = \text{Var}(x_i|Y_i, X_i) = \frac{\sigma_{x_i}^2 \sigma_c^2 \sigma_u^2}{(\beta^2 \sigma_{x_i}^2 + \sigma_c^2) \sigma_u^2 + m_i \sigma_c^2 \sigma_{x_i}^2}$$
$$E(x_i) = E(x_i|Y_i, X_i) = \mu_x + V(x_i) \left[ \beta \sigma_c^{-2} (Y_i - \alpha - \beta \mu_x) + \sigma_u^{-2} \sum_{j=1}^{m_i} (X_{ij} - \mu_x) \right].$$

Na etapa M do algoritmo, devemos determinar  $\hat{\theta}$  de modo que  $Q(\hat{\theta})$  seja máxima. Então resolvendo o sistema de equações

$$\frac{\partial Q(\theta)}{\partial \psi} = 0, \quad \text{para } \psi = \alpha, \beta, \mu_x, \sigma_c^2, \sigma_u^2,$$

obtemos expressões iterativas para os parâmetros.

Na figura (4.1) se apresenta um esboço do esquema iterativo do algoritmo EM.

início

- Fornecer valores iniciais  $\theta^{(0)} = (\alpha^{(0)}, \beta^{(0)}, \mu_x^{(0)}, \sigma_{x_i}^2, \sigma_e^2, \sigma_u^2)$ ,
- passo E: Avaliar

$$V(x_i)^{(p)} = \frac{\sigma_{x_i}^{2(p)} \sigma_e^{2(p)} \sigma_u^{2(p)}}{(\beta^{2(p)} \sigma_{x_i}^{2(p)} + \sigma_e^{2(p)}) \sigma_u^{2(p)} + m_i \sigma_{x_i}^{2(p)} \sigma_e^{2(p)}}$$

$$E(x_i)^{(p)} = \mu_x^{(p)} + V(x_i)^{(p)} \left[ \beta^{(p)} \sigma_e^{-2(p)} (Y_i - \alpha^{(p)} - \beta^{(p)} \mu_x^{(p)}) + \sigma_u^{-2(p)} \sum_{j=1}^{m_i} (X_{ij} - \mu_x^{(p)}) \right]$$

$$\hat{x}_i^{(p)} = E(x_i)^{(p)}, \quad \hat{\sigma}_{x_i}^{2(p)} = V(x_i)^{(p)} + \left( \hat{x}_i^{(p)} \right)^2, \quad i = 1, \dots, n,$$

- passo M: Avaliar

$$\mu_x^{(p+1)} = \frac{\sum_{i=1}^n \sigma_{x_i}^{-2(p)} \hat{x}_i^{(p)}}{\sum_{i=1}^n \sigma_{x_i}^{-2(p)}}, \quad \beta^{(p+1)} = \frac{\sum_{i=1}^n Y_i \hat{x}_i^{(p)} - n \bar{Y} \bar{\hat{x}}^{(p)}}{\sum_{i=1}^n \hat{x}_i^{2(p)} - n (\bar{\hat{x}}^{(p)})^2},$$

$$\alpha^{(p+1)} = \bar{Y} - \bar{\hat{x}}^{(p)} \beta^{(p+1)},$$

$$\sigma_e^{2(p+1)} = \frac{1}{n} \sum_{i=1}^n \left[ (Y_i - \alpha^{(p+1)}) - \beta^{(p+1)} \hat{x}_i^{(p)} \right]^2,$$

$$\sigma_u^{2(p+1)} = \frac{1}{\sum_{i=1}^n m_i} \left[ \sum_{i=1}^n \sum_{j=1}^{m_i} X_{ij}^2 - 2 \sum_{i=1}^n \hat{x}_i^{(p)} \left( \sum_{j=1}^{m_i} X_{ij} \right) + \sum_{i=1}^n m_i \hat{x}_i^{2(p)} \right],$$

- repetir passos E e M até que  $|\theta^{(p)} - \theta^{(p+1)}| < \epsilon$ , para algum  $\epsilon > 0$ ,

fim.

Figura 4.1. Algoritmo EM

## 4.4 Matriz de covariância assintótica dos EMV

### 4.4.1 Informação observada

O logaritmo da função de verossimilhança para este modelo segue de (4.16). Denotaremos a matriz de informação observada por

$$\mathbf{H}(\hat{\theta}) = - \sum_{i=1}^n \mathbf{H}_i(\hat{\theta}), \text{ onde } \mathbf{H}_i(\hat{\theta}) = \frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta = \hat{\theta}}, \quad (4.25)$$

sendo  $\hat{\theta}$  o EMV de  $\theta$  obtido na seção anterior. Usando a restrição de identificabilidade  $\sigma_{x_i}^2$ ,  $i = 1, \dots, n$ ,  $\mathbf{H}_i(\theta)$ , a matriz de derivadas de segunda ordem de  $l_i(\theta)$  em relação ao parâmetro  $\theta = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2)^T$ , é representada por:

$$\mathbf{H}_i(\hat{\theta}) = \begin{pmatrix} H_{i\alpha\alpha} & H_{i\alpha\beta} & H_{i\alpha\mu_x} & H_{i\alpha\sigma_e^2} & H_{i\alpha\sigma_u^2} \\ H_{i\beta\alpha} & H_{i\beta\beta} & H_{i\beta\mu_x} & H_{i\beta\sigma_e^2} & H_{i\beta\sigma_u^2} \\ H_{i\mu_x\alpha} & H_{i\mu_x\beta} & H_{i\mu_x\mu_x} & H_{i\mu_x\sigma_e^2} & H_{i\mu_x\sigma_u^2} \\ H_{i\sigma_e^2\alpha} & H_{i\sigma_e^2\beta} & H_{i\sigma_e^2\mu_x} & H_{i\sigma_e^2\sigma_e^2} & H_{i\sigma_e^2\sigma_u^2} \\ H_{i\sigma_u^2\alpha} & H_{i\sigma_u^2\beta} & H_{i\sigma_u^2\mu_x} & H_{i\sigma_u^2\sigma_e^2} & H_{i\sigma_u^2\sigma_u^2} \end{pmatrix} \Big|_{\theta = \hat{\theta}}, \quad i = 1, \dots, n. \quad (4.26)$$

Fazendo  $c_i = k_i$ ,  $t_i = w_i$ , e  $h_i = f_i$  em (3.16) e (3.17), após várias manipulações algébricas, obtemos as seguintes expressões para os elementos da matriz  $H_i$  (veja Apêndice A):

$$\begin{aligned} H_{i\alpha\alpha} &= -\sigma_e^{-2} + k_i^{-1} \beta^2 \sigma_e^{-4}, \\ H_{i\beta\beta} &= -\sigma_e^{-2} k_i^{-1} - \mu_x^2 \sigma_e^{-2} + k_i^{-1} [\sigma_e^{-4} (Y_i - \alpha - 2\beta\mu_x)^2 - 2f_i \mu_x \sigma_e^{-2}] \\ &\quad - 4\beta f_i \sigma_e^{-4} k_i^{-2} (Y_i - \alpha - 2\beta\mu_x) + f_i^2 k_i^{-2} (4\beta^2 \sigma_e^{-4} k_i^{-1} - \sigma_e^{-2}) + 2\beta^2 k_i^{-2} \sigma_e^{-4}, \\ H_{i\mu_x\mu_x} &= -(\beta^2 \sigma_e^{-2} + m_i \sigma_u^{-2}) + k_i^{-1} (\beta^2 \sigma_e^{-2} + m_i \sigma_u^{-2})^2, \end{aligned}$$



$$\begin{aligned}
H_{i_{\sigma_e^2 \sigma_e^2}} &= -\frac{1}{2}(-\sigma_e^{-4} + 2\beta^2 \sigma_e^{-6} k_i^{-1} - \beta^4 \sigma_e^{-8} k_i^{-2}) - \sigma_e^{-6} (Y_i - \alpha - \beta \mu_x)^2 \\
&\quad + k_i^{-1} [\beta^2 \sigma_e^{-8} (Y_i - \alpha - \beta \mu_x)^2 + 2\beta f_i \sigma_e^{-6} (Y_i - \alpha - \beta \mu_x)] \\
&\quad - 2f_i \beta^3 \sigma_e^{-8} k_i^{-2} (Y_i - \alpha - \beta \mu_x) + f_i^2 k_i^{-2} (k_i^{-1} \beta^4 \sigma_e^{-8} - \beta^2 \sigma_e^{-6}),
\end{aligned}$$

$$\begin{aligned}
H_{i_{\sigma_u^2 \sigma_u^2}} &= \frac{1}{2} m_i (\sigma_u^{-4} - 2\sigma_u^{-6} k_i^{-1} + m_i \sigma_u^{-8} k_i^{-2}) - \sigma_u^{-6} \sum_{j=1}^{m_i} (X_{ij} - \mu_x)^2 \\
&\quad + k_i^{-1} \left[ \sigma_u^{-8} \left( \sum_{j=1}^{m_i} (X_{ij} - \mu_x) \right)^2 + 2f_i \sigma_u^{-6} \sum_{j=1}^{m_i} (X_{ij} - \mu_x) \right] \\
&\quad - 2m_i f_i \sigma_u^{-8} k_i^{-2} \sum_{j=1}^{m_i} (X_{ij} - \mu_x) + m_i f_i^2 \sigma_u^{-6} (m_i \sigma_u^{-8} k_i^{-1} - \sigma_u^{-6}),
\end{aligned}$$

$$\begin{aligned}
H_{i_{\alpha\beta}} &= H_{i_{\beta\alpha}} \\
&= -\mu_x \sigma_e^{-2} - \beta \sigma_e^{-4} k_i^{-1} (Y_i - \alpha - 2\beta \mu_x) - f_i \sigma_e^{-2} k_i^{-1} \\
&\quad + 2\beta^2 f_i \sigma_e^{-4} k_i^{-2},
\end{aligned}$$

$$\begin{aligned}
H_{i_{\alpha\mu_x}} &= H_{i_{\mu_x\alpha}} \\
&= -\beta \sigma_e^{-2} + k_i^{-1} (\beta^3 \sigma_e^{-4} + m_i \beta \sigma_u^{-2} \sigma_e^{-2}),
\end{aligned}$$

$$\begin{aligned}
H_{i_{\alpha\sigma_e^2}} &= H_{i_{\sigma_e^2\alpha}} \\
&= -\sigma_e^{-4} (Y_i - \alpha - \beta \mu_x) \\
&\quad + k_i^{-1} [\beta^2 \sigma_e^{-6} (Y_i - \alpha - \beta \mu_x) + \beta f_i \sigma_e^{-4}] - \beta^3 f_i \sigma_e^{-6} k_i^{-2},
\end{aligned}$$

$$\begin{aligned}
H_{i_{\alpha\sigma_u^2}} &= H_{i_{\sigma_u^2\alpha}} \\
&= k_i^{-1} \beta \sigma_e^{-2} \sigma_u^{-4} \sum_{j=1}^{m_i} (X_{ij} - \mu_x) - m_i \beta f_i \sigma_e^{-2} \sigma_u^{-4} k_i^{-2},
\end{aligned}$$

$$\begin{aligned}
H_{i_{\beta\sigma_e^2}} &= H_{i_{\sigma_e^2\beta}} \\
&= \beta \sigma_e^{-4} k_i^{-1} - \beta^3 \sigma_e^{-6} k_i^{-2} - \mu_x \sigma_e^{-4} (Y_i - \alpha - \beta \mu_x) \\
&\quad - k_i^{-1} [\beta \sigma_e^{-6} (Y_i - \alpha - \beta \mu_x) (Y_i - \alpha - 2\beta \mu_x) + f_i \sigma_e^{-4} (Y_i - \alpha - 2\beta \mu_x)] \\
&\quad + f_i k_i^{-2} [\beta^2 \sigma_e^{-6} (Y_i - \alpha - 2\beta \mu_x) + 2\beta^2 \sigma_e^{-6} (Y_i - \alpha - \beta \mu_x)] \\
&\quad - f_i^2 k_i^{-2} (2k_i^{-1} \beta^3 \sigma_e^{-6} - \beta \sigma_e^{-4}),
\end{aligned}$$

$$\begin{aligned}
H_{i\beta\mu_x} &= H_{i\mu_x\beta} \\
&= -k_i^{-1} [\sigma_e^{-2}(\beta^2\sigma_e^{-2} + m_i\sigma_u^{-2})(Y_i - \alpha - 2\beta\mu_x) + 2\beta f_i\sigma_e^{-2}] \\
&\quad + \sigma_e^{-2}(Y_i - \alpha - 2\beta\mu_x) + 2\beta f_i\sigma_e^{-2}k_i^{-2}(\beta^2\sigma_e^{-2} + m_i\sigma_u^{-2}),
\end{aligned}$$

$$\begin{aligned}
H_{i\delta\sigma_u^2} &= H_{i\sigma_u^2\delta} \\
&= -m_i\beta\sigma_e^{-2}\sigma_u^{-4}k_i^{-2} - \sigma_e^{-2}\sigma_u^{-4}k_i^{-1}(Y_i - \alpha - 2\beta\mu_x) \sum_{j=1}^{m_i} (X_{ij} - \mu_x) \\
&\quad + m_i f_i \sigma_u^{-4} k_i^{-2} \left[ 2\beta\sigma_e^{-2} \sum_{j=1}^{m_i} (X_{ij} + \mu_x) + \sigma_e^{-2}(Y_i - \alpha - 2\beta\mu_x) \right] \\
&\quad - 2m_i\beta f_i^2 \sigma_e^{-2} \sigma_u^{-4} k_i^{-3},
\end{aligned}$$

$$\begin{aligned}
H_{i\mu_x\sigma_e^2} &= H_{i\sigma_e^2\mu_x} \\
&= -\beta\sigma_e^{-4}(Y_i - \alpha - \beta\mu_x) + k_i^{-1} [\beta\sigma_e^{-4}(Y_i - \alpha - \beta\mu_x)(\beta^2\sigma_e^{-2} + m_i\sigma_u^{-2}) \\
&\quad + \beta^2 f_i \sigma_e^{-4}] - \beta^2 f_i \sigma_e^{-4} k_i^{-2} (\beta^2\sigma_e^{-2} + m_i\sigma_u^{-2}),
\end{aligned}$$

$$\begin{aligned}
H_{i\mu_x\sigma_u^2} &= H_{i\sigma_u^2\mu_x} \\
&= -\sigma_u^{-4} \sum_{j=1}^{m_i} (X_{ij} - \mu_x) + k^{-1} \left[ \sigma_u^{-4} (\beta^2\sigma_e^{-2} + m_i\sigma_u^{-2}) \sum_{j=1}^{m_i} (X_{ij} - \mu_x) + m_i f_i \sigma_u^{-4} \right] \\
&\quad - m_i f_i k_i^{-2} \sigma_u^{-4} (\beta^2\sigma_e^{-2} + \sigma_u^{-2}),
\end{aligned}$$

$$\begin{aligned}
H_{i\sigma_e^2\sigma_u^2} &= H_{i\sigma_u^2\sigma_e^2} \\
&= \frac{1}{2} m_i \beta^2 \sigma_e^{-4} \sigma_u^{-4} k_i^{-2} + \beta \sigma_e^{-4} \sigma_u^{-4} k_i^{-1} (Y_i - \alpha - \beta\mu_x) \sum_{j=1}^{m_i} (X_{ij} - \mu_x) \\
&\quad - f_i k_i^{-2} \left[ \beta^2 \sigma_e^{-4} \sigma_u^{-4} \sum_{j=1}^{m_i} (X_{ij} - \mu_x) + m_i \beta \sigma_e^{-4} \sigma_u^{-4} (Y_i - \alpha - \beta\mu_x) \right] \\
&\quad + m_i f_i^2 k_i^{-3} \beta^2 \sigma_e^{-4} \sigma_u^{-4}.
\end{aligned}$$

---

## Capítulo 5

# Enfoque Bayesiano

### 5.1 Introdução

Inicialmente, sob enfoque da estatística clássica, nós desenvolvemos uma teoria que nos permitiu obter os estimadores (com seus respectivos desvios-padrão) dos parâmetros de um MREH com erros nas variáveis (no caso em que as variâncias dos erros de medição e a variância da variável não observada  $x_i$  são heterocedásticas). A seguir estudaremos o MREH com erros nas variáveis na formulação Bayesiana.

No restante do capítulo concentramo-nos na análise bayesiana do MREH com erros nas variáveis, apenas no caso em que a variância da variável não observada  $x_i$  é considerada heterocedástica. Estudamos tanto o modelo com restrição de identificabilidade como o modelo sem restrição de identificabilidade, isto é, pelo fato de que a metodologia Bayesiana não está limitada apenas a situações de identificabilidade.

Como na maioria dos casos, em nosso estudo, a distribuição *a posteriori* e as distribuições marginais e condicionais *a posteriori* não têm uma forma fechada (apenas seus núcleos têm forma conhecida), procedimentos numéricos computacionais tipo MCMC (veja Apêndice C.1), devem ser usados para amostrar a partir destas distribuições. Neste caso, a implementação dos métodos MCMC foi feita com o auxílio do software computacional WinBUGS (veja Apêndice C.2) desenvolvido no projeto BUGS (Spiegelhalter et al. (2007)).

Finalmente baseando-se em uma amostra de dados simulados, são analisados os modelos com e sem restrição de identificabilidade. Usando WinBUGS e conforme descrito em Ntzoufras (2009), obtemos os diagnósticos de convergência e as estimativas de várias

medidas de resumo da distribuição *a posteriori*, tais como, média, mediana, desvio padrão, etc..

## 5.2 O modelo estrutural sob a ótica bayesiana

Consideremos o modelo em (2.1) e as suposições feitas em (3.2) e (3.3), isto é,

$$\begin{aligned} y_i &= \alpha + \beta x_i \\ X_i &= x_i + u_i \\ Y_i &= y_i + e_i \quad i = 1, \dots, n, \end{aligned} \quad (5.1)$$

com as suposições

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{indep.}}{\sim} N_3 \left[ \begin{pmatrix} 0 \\ 0 \\ \mu_x \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_{x_i}^2 \end{pmatrix} \right], \quad i = 1, \dots, n, \quad (5.2)$$

que levam ao modelo estrutural heterocedástico

$$\begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix} \stackrel{\text{ind}}{\sim} N_3 \left[ \begin{pmatrix} \alpha + \beta \mu_x \\ \mu_x \\ \mu_x \end{pmatrix}; \begin{pmatrix} \beta^2 \sigma_{x_i}^2 + \sigma_e^2 & \beta \sigma_{x_i}^2 & \beta \sigma_{x_i}^2 \\ \beta \sigma_{x_i}^2 & \sigma_{x_i}^2 + \sigma_u^2 & \sigma_{x_i}^2 \\ \beta \sigma_{x_i}^2 & \sigma_{x_i}^2 & \sigma_{x_i}^2 \end{pmatrix} \right] \quad i = 1, \dots, n. \quad (5.3)$$

Chamaremos

$$\mathbf{Z}_{\text{obs}} = \mathbf{Z}, \text{ com } \mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n), \text{ onde } \mathbf{Z}_i = \{(Y_i, X_i)^T, i = 1, \dots, n\} \text{ e}$$

$$\mathbf{W}_c = \mathbf{W}, \text{ com } \mathbf{W} = (\mathbf{W}_1, \dots, \mathbf{W}_n), \text{ onde } \mathbf{W}_i = \{(Y_i, X_i, x_i)^T, i = 1, \dots, n\},$$

ao conjunto dos dados observados e completos, respectivamente. Então, sob modelo em (3.9), temos que a função de verossimilhança baseada nos dados observados é dada por

$$\begin{aligned} L(\boldsymbol{\theta} | \mathbf{Z}_{\text{obs}}) &\propto (\sigma_e^2)^{-\frac{n}{2}} (\sigma_u^2)^{-\frac{n}{2}} \prod_{i=1}^n (\sigma_{x_i}^2 c_i)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma_e^2} \sum_{i=1}^n (Y_i - \alpha - \beta \mu_x)^2 \right. \right. \\ &\quad \left. \left. + \frac{1}{\sigma_u^2} \sum_{i=1}^n (X_i - \mu_x)^2 - \sum_{i=1}^n c_i^{-1} \left[ \frac{\beta}{\sigma_e^2} (Y_i - \alpha - \beta \mu_x) + \frac{1}{\sigma_u^2} (X_i - \mu_x) \right]^2 \right] \right\} \quad (5.4) \end{aligned}$$

onde

$$c_i = \frac{1}{\sigma_{x_i}^2} + \frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} = \frac{\sigma_e^2 \sigma_u^2 + \beta^2 \sigma_{x_i}^2 \sigma_u^2 + \sigma_{x_i}^2 \sigma_e^2}{\sigma_{x_i} \sigma_e^2 \sigma_u^2}, \quad i = 1, \dots, n, \quad (5.5)$$

e sob modelo em (5.3), a função de verossimilhança baseada nos dados completos é dada por:

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{W}_c) \propto & (\sigma_e^2)^{-\frac{n}{2}} (\sigma_u^2)^{-\frac{n}{2}} \prod_{i=1}^n (\sigma_{x_i}^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma_e^2} \sum_{i=1}^n (Y_i - \alpha - \beta \mu_x)^2 \right. \right. \\ & \left. \left. + \frac{1}{\sigma_u^2} \sum_{i=1}^n (X_i - x_i)^2 + \sum_{i=1}^n \frac{1}{\sigma_{x_i}^2} (x_i - \mu_x)^2 \right] \right\}, \end{aligned} \quad (5.6)$$

em que  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_u^2, \sigma_e^2, \sigma_{x_1}^2, \dots, \sigma_{x_n}^2)^\top$  é o vetor de parâmetros desconhecidos. Além disso acrescentamos a suposição de que o vetor de parâmetros  $\boldsymbol{\theta}$  tem distribuição *a priori*  $\pi(\boldsymbol{\theta})$ . O espaço paramétrico do modelo é o conjunto

$$\Theta = \{ \boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_u^2, \sigma_e^2, \sigma_{x_1}^2, \dots, \sigma_{x_n}^2)^\top : \alpha, \beta, \mu_x \in \mathbb{R} \text{ e } \sigma_e^2, \sigma_u^2, \sigma_{x_1}^2, \dots, \sigma_{x_n}^2 \in \mathbb{R}^+ \}.$$

### 5.2.1 Distribuição *a posteriori*

A distribuição *a posteriori* de  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_u^2, \sigma_e^2, \sigma_{x_1}^2, \dots, \sigma_{x_n}^2)^\top$  baseada nos dados observados é dada por

$$p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}}) \propto L(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}})\pi(\boldsymbol{\theta}),$$

onde  $L(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}})$  é a verossimilhança dada em (5.4). Portanto,

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}}) \propto & (\sigma_e^2)^{-\frac{n}{2}} (\sigma_u^2)^{-\frac{n}{2}} \prod_{i=1}^n (\sigma_{x_i}^2 c_i)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma_e^2} \sum_{i=1}^n (Y_i - \alpha - \beta \mu_x)^2 \right. \right. \\ & \left. \left. + \frac{1}{\sigma_u^2} \sum_{i=1}^n (X_i - \mu_x)^2 - \sum_{i=1}^n c_i^{-1} \left[ \frac{\beta}{\sigma_e^2} (Y_i - \alpha - \beta \mu_x) + \frac{1}{\sigma_u^2} (X_i - \mu_x) \right]^2 \right] \right\} \pi(\boldsymbol{\theta}) \end{aligned} \quad (5.7)$$

Para a distribuição *a posteriori* de  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_u^2, \sigma_e^2, \sigma_{x_1}^2, \dots, \sigma_{x_n}^2)^\top$  baseada nos dados completos é importante ressaltar que a distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}})$  também é uma distribuição marginal de  $p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{Z}_{\text{obs}})$ , onde  $\mathbf{x} = (x_1, \dots, x_n)$  é o vetor de variáveis latentes (não observáveis), ou seja

$$p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}}) = \int_{\mathcal{X} = \mathbb{R}^n} p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{Z}_{\text{obs}}) d\mathbf{x},$$

e levando em conta a regra de Bayes temos

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}}) &\propto \int_{\mathbf{x} = \mathbb{R}^n} p(\mathbf{Z}_{\text{obs}}|\boldsymbol{\theta}, \mathbf{x})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\mathbf{x} \\ &\propto \int_{\mathbf{x} = \mathbb{R}^n} L(\boldsymbol{\theta}|\mathbf{W}_c)\pi(\boldsymbol{\theta})d\mathbf{x}, \end{aligned}$$

onde  $L(\boldsymbol{\theta}|\mathbf{W}_c)$  é a verossimilhança dada em (5.6). Assim, a distribuição *a posteriori* de  $\boldsymbol{\theta}$  baseada nos dados observados, também pode ser escrita como

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}}) &\propto \int_{\mathbf{x} = \mathbb{R}^n} (\sigma_c^2 \sigma_u^2)^{-\frac{n}{2}} \prod_{i=1}^n (\sigma_{x_i}^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma_c^2} \sum_{i=1}^n (Y_i - \alpha - \beta \mu_x)^2 \right. \right. \\ &\quad \left. \left. + \frac{1}{\sigma_u^2} \sum_{i=1}^n (X_i - x_i)^2 + \sum_{i=1}^n \frac{1}{\sigma_{x_i}^2} (x_i - \mu_x)^2 \right] \right\} \pi(\boldsymbol{\theta}) d\mathbf{x}. \end{aligned} \quad (5.8)$$

### 5.2.2 Distribuição preditiva

Seja  $\mathbf{Z}_s$  uma observação futura do vetor  $(Y, X)^T$ , isto é,  $\mathbf{Z}_s = (Y_s, X_s)^T$ , onde

$$[\mathbf{Z}_s|\boldsymbol{\theta}] \sim N_2 \left[ \begin{pmatrix} \alpha + \beta \mu_x \\ \mu_x \end{pmatrix}; \begin{pmatrix} \beta^2 \sigma_{x_s}^2 + \sigma_c^2 & \beta \sigma_{x_s}^2 \\ \beta \sigma_{x_s}^2 & \sigma_{x_s}^2 + \sigma_u^2 \end{pmatrix} \right].$$

Então a distribuição preditiva *a posteriori* nesse modelo é dada por

$$\begin{aligned} p(\mathbf{Z}_s|\mathbf{Z}_{\text{obs}}) &= \int_{\Theta} p(\mathbf{Z}_s, \boldsymbol{\theta}|\mathbf{Z}_{\text{obs}})d\boldsymbol{\theta} = \int_{\Theta} p(\mathbf{Z}_s|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}})d\boldsymbol{\theta} \\ &= \int_{\Theta} p(Y_s|X_s, \boldsymbol{\theta})p(X_s|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}})d\boldsymbol{\theta}, \end{aligned} \quad (5.9)$$

onde  $p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}})$  é a distribuição *a posteriori* do parâmetro  $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \sigma_u^2, \sigma_c^2, \sigma_{x_1}^2, \dots, \sigma_{x_n}^2)^T$ , além disso  $p(Y_s|X_s, \boldsymbol{\theta})$  e  $p(X_s|\boldsymbol{\theta})$  representam respectivamente as distribuições condicionais de



$[Y_s|X_s, \boldsymbol{\theta}]$  e  $[X_s|\boldsymbol{\theta}]$ , em que

$$[Y_s|X_s, \boldsymbol{\theta}] \sim N\left(\alpha + \beta\mu_x + \frac{\beta\sigma_{x_s}^2}{\sigma_{x_s}^2 + \sigma_u^2}(X_s - \mu_x), \beta^2\sigma_{x_s}^2 + \sigma_e^2 - \frac{\beta^2(\sigma_{x_s}^2)^2}{\sigma_{x_s}^2 + \sigma_u^2}\right) \quad (5.10)$$

$$[X_s|\boldsymbol{\theta}] \sim N(\mu_x, \sigma_{x_s}^2 + \sigma_u^2). \quad (5.11)$$

A integração em (5.9) não tem solução analítica, entretanto aproximações podem ser obtidas usando métodos MCMC ("Markov-Chain-Monte Carlo"). Ver Gamerman and Lopes (2006) para maiores detalhes.

Assim, para gerar uma observação  $\mathbf{Z}_s$  futura, geramos primeiro um vetor  $\boldsymbol{\theta}$  da distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}})$ , a seguir, com  $\boldsymbol{\theta}$  fixado, geramos um valor  $X_s$  da distribuição condicional  $p(X_s|\boldsymbol{\theta})$  em (5.11) e finalmente geramos um valor para  $Y_s$  da distribuição condicional  $p(Y_s|X_s, \boldsymbol{\theta})$  em (5.10).

Agora se estamos interessados em gerar um valor  $Y_s$  futuro dado que o valor futuro  $X_s = X_0$  é fixo, isto é, se quisermos gerar um valor  $Y_s$  da distribuição preditiva  $p(Y_s|X_s = X_0, \mathbf{Z}_{\text{obs}})$ , basta gerar  $\boldsymbol{\theta}$  da distribuição  $p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}})$  e logo gerar  $Y_s$  da distribuição  $p(Y_s|X_s = X_0, \mathbf{Z}_{\text{obs}})$ , levando em conta que

$$p(Y_s|X_s = X_0, \mathbf{Z}_{\text{obs}}) = \int_{\Theta} p(Y_s|X_0, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}})d\boldsymbol{\theta}.$$

Semelhantemente, podemos gerar uma observação futura  $X_s$  da distribuição preditiva  $p(X_s|Y_s = Y_0, \mathbf{Z}_{\text{obs}})$ , com

$$[X_s|Y_s = Y_0, \boldsymbol{\theta}] \sim N\left(\mu_x + \frac{\beta\sigma_{x_s}^2}{\beta^2\sigma_{x_s}^2 + \sigma_e^2}(Y_0 - \alpha - \beta\mu_x), \sigma_{x_s}^2 + \sigma_u^2 - \frac{\beta^2(\sigma_{x_s}^2)^2}{\beta^2\sigma_{x_s}^2 + \sigma_e^2}\right).$$

**Obs.:** O modelo estrutural bayesiano com erros nas variáveis, como sabemos, não é identificável (a função de verossimilhança não é identificável). Em princípio, teoricamente, isso não é problema. Contudo, nas aplicações, o algoritmo de Gibbs para amostrar da distribuição *a posteriori* de modelos não identificáveis pode apresentar problemas de convergência. Ver



Gelfand and Sahu (1999) para maiores detalhes. Além disso, se a distribuição *a priori* é não informativa, o problema pode piorar bastante e a recomendação de Gelfand and Sahu, nesse caso é que se escolha uma distribuição *a priori* "adequada", que não seja nem muito pouco informativa, mas também que não seja muito precisa para não limitar o aprendizado proveniente dos dados. Segundo Gelfand, o problema é achar essa distribuição *a priori* "adequada" o que geralmente demanda um trabalho considerável de simulação. Uma maneira de evitar o problema computacional decorrente da falta de identificabilidade é simplesmente acrescentar uma restrição de identificabilidade ao modelo.

### 5.3 O modelo estrutural identificável com particular *priori* de componentes independentes

Apresentamos o modelo estrutural normal heteroscedástico com erros nas variáveis no caso em que a distribuição *a priori* é própria e tem componentes independentes normais ou gama invertida, assumindo a restrição de identificabilidade que considera  $\sigma_{x_i}^2, i = 1, \dots, n$ , conhecida.

Supondo  $\sigma_{x_i}^2, i = 1, \dots, n$  conhecida o modelo em (3.9) torna-se identificável e o vetor de parâmetros desconhecidos é dado por  $\theta = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2)^T$ . Acrescentamos a esse modelo a suposição *a priori* de que  $\alpha, \beta, \mu_x, \sigma_e^2$  e  $\sigma_u^2$  são variáveis aleatórias independentes, com distribuição *a priori*  $\pi(\alpha), \pi(\beta), \pi(\mu_x), \pi(\sigma_e^2)$  e  $\pi(\sigma_u^2)$  respectivamente, com  $\alpha, \beta, \mu_x$  tendo distribuição normal e  $\sigma_e^2, \sigma_u^2$  tendo distribuição gama invertida, como segue:

$$\alpha \sim N(a, \sigma_\alpha^2), \quad \beta \sim N(b, \sigma_\beta^2), \quad \mu_x \sim N(l, \sigma_{\mu_x}^2) \quad (5.12)$$

$$\sigma_e^2 \sim IG(p, q), \quad \sigma_u^2 \sim IG(f, g), \quad (5.13)$$

onde  $a, b, l, f, g, p, q, \sigma_\alpha^2, \sigma_\beta^2$  e  $\sigma_{\mu_x}^2$  são os hiperparâmetros conhecidos.

**Obs.:**  $z \sim IG(c, d)$  indica que a variável aleatória  $z$  tem distribuição Gama Invertida com parâmetros  $c$  e  $d$ . Isto é denotado por

$$z \sim IG(c, d) \Leftrightarrow p(z|c, d) = \frac{d^c}{\Gamma(c)} z^{-(c+1)} \exp\{-d/z\}, \quad z > 0.$$

Então, a distribuição *a priori* do parâmetro  $\theta$  é dada por

$$\begin{aligned}\pi(\theta) &= \pi(\alpha)\pi(\beta)\pi(\mu_x)\pi(\sigma_e^2)\pi(\sigma_u^2) \\ &\propto (\sigma_e^2)^{-(p+1)}(\sigma_u^2)^{-(f+1)} \exp \left\{ \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} + \frac{(\mu_x - l)^2}{\sigma_{\mu_x}^2} + \frac{2q}{\sigma_e^2} + \frac{2g}{\sigma_u^2} \right\}\end{aligned}\quad (5.14)$$

#### 5.4 Distribuição *a posteriori*

A distribuição *a posteriori* de  $\theta = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2)^T$  baseada nos dados observados ( $\mathbf{Z}_{\text{obs}}$ ) é dada por

$$p(\theta|\mathbf{Z}_{\text{obs}}) \propto L(\theta|\mathbf{Z}_{\text{obs}})\pi(\theta),$$

onde  $L(\theta|\mathbf{Z}_{\text{obs}})$  é a função de verossimilhança baseada nos dados observados sob a restrição de identificabilidade e  $\pi(\theta)$  é a distribuição *a priori* dada em (5.14).

Então aplicando a restrição de identificabilidade em (5.7) e substituindo (5.14), segue que

$$\begin{aligned}p(\theta|\mathbf{Z}_{\text{obs}}) &\propto (\sigma_e^2)^{-\left(\frac{n}{2}+p+1\right)}(\sigma_u^2)^{-\left(\frac{n}{2}+f+1\right)} \prod_{i=1}^n c_i^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma_e^2} \sum_{i=1}^n (Y_i - \alpha - \beta\mu_x)^2 \right. \right. \\ &\quad \left. \left. + \frac{1}{\sigma_u^2} \sum_{i=1}^n (X_i - \mu_x)^2 - \sum_{i=1}^n c_i^{-1} \left[ \frac{\beta}{\sigma_e^2} (Y_i - \alpha - \beta\mu_x) + \frac{1}{\sigma_u^2} (X_i - \mu_x) \right]^2 \right. \right. \\ &\quad \left. \left. + \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} + \frac{(\mu_x - l)^2}{\sigma_{\mu_x}^2} + \frac{2q}{\sigma_e^2} + \frac{2g}{\sigma_u^2} \right] \right\}.\end{aligned}\quad (5.15)$$

Fazendo  $c_i = \frac{|\Sigma_i|}{\sigma_e^2 \sigma_u^2 \sigma_{x_i}^2}$ ,  $i = 1, \dots, n$  e substituindo em (5.15) temos que

$$\begin{aligned}p(\theta|\mathbf{Z}_{\text{obs}}) &\propto (\sigma_e^2)^{-(p+1)}(\sigma_u^2)^{-(f+1)} \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n \frac{(\sigma_u^2 + \sigma_{x_i}^2)}{|\Sigma_i|} (Y_i - \alpha - \beta\mu_x)^2 \right. \right. \\ &\quad \left. \left. + \sum_{i=1}^n \frac{(\sigma_e^2 + \beta^2 \sigma_{x_i}^2)}{|\Sigma_i|} (X_i - \mu_x)^2 - 2\beta \sum_{i=1}^n \frac{\sigma_{x_i}^2}{|\Sigma_i|} (Y_i - \alpha - \beta\mu_x)(X_i - \mu_x) \right. \right. \\ &\quad \left. \left. + \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} + \frac{(\mu_x - l)^2}{\sigma_{\mu_x}^2} + \frac{2q}{\sigma_e^2} + \frac{2g}{\sigma_u^2} \right] \right\}.\end{aligned}\quad (5.16)$$

Semelhantemente a distribuição *a posteriori* baseada nos dados completos ( $\mathbf{W}_c$ ) sob esse

modelo é dada por

$$p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}}) = \int_{\mathbf{x} = \mathbb{R}^n} p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{Z}_{\text{obs}})d\mathbf{x},$$

onde

$$p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{Z}_{\text{obs}}) \propto L(\boldsymbol{\theta}|\mathbf{W}_c)\pi(\boldsymbol{\theta}).$$

Então aplicando a restrição de identificabilidade em (5.6) e substituindo (5.14), temos que

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{Z}_{\text{obs}}) \propto & (\sigma_e^2)^{-\left(\frac{n}{2}+p+1\right)}(\sigma_u^2)^{-\left(\frac{n}{2}+f+1\right)} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma_e^2} \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 \right. \right. \\ & + \frac{1}{\sigma_u^2} \sum_{i=1}^n (X_i - x_i)^2 + \sum_{i=1}^n \frac{1}{\sigma_{x_i}^2} (x_i - \mu_x)^2 + \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} \\ & \left. \left. + \frac{(\mu_x - l)^2}{\sigma_{\mu_x}^2} + \frac{2q}{\sigma_e^2} + \frac{2g}{\sigma_u^2} \right] \right\}. \end{aligned} \quad (5.17)$$

Notemos que tanto a distribuição *a posteriori* baseada nos dados observados como a distribuição *a posteriori* baseada nos dados completos, dadas em (5.16) e (5.17) respectivamente, não tem uma forma fechada, apenas o núcleo delas têm forma conhecida. Procedimentos numéricos computacionais, devem ser usados para tratar desse tipo de problema.

#### 5.4.1 Distribuições condicionais completas baseados nos dados observados ( $\mathbf{Z}_{\text{obs}}$ )

Baseando-se na distribuição *a posteriori* dada em (5.16), obtemos a seguir as distribuições condicionais completas univariadas dos parâmetros  $\alpha$ ,  $\beta$ ,  $\mu_x$ ,  $\sigma_e^2$  e  $\sigma_u^2$ .

**Distribuição condicional completa de  $\alpha$  dado  $\beta$ ,  $\mu_x$ ,  $\sigma_e^2$ ,  $\sigma_u^2$ ,  $\sigma_x^2$  e  $\mathbf{Z}_{\text{obs}}$**

$$\begin{aligned} \pi(\alpha|\beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto & \exp \left\{ -\frac{1}{2} \left[ \left( \sum_{i=1}^n \frac{(\sigma_u^2 + \sigma_{x_i}^2)}{|\Sigma_i|} + \frac{1}{\sigma_\alpha^2} \right) \alpha^2 \right. \right. \\ & \left. \left. - 2 \left( \sum_{i=1}^n \frac{[\sigma_u^2(Y_i - \beta\mu_x) + \sigma_{x_i}^2(Y_i - \beta X_i)]}{|\Sigma_i|} + \frac{a}{\sigma_\alpha^2} \right) \alpha \right] \right\}. \end{aligned} \quad (5.18)$$

Então

$$\pi(\alpha|\beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto \exp\left\{-\frac{A}{2}[A\alpha^2 - 2B\alpha]\right\} \propto \exp\left\{-\frac{1}{2A^{-1}}\left[\alpha - \frac{B}{A}\right]^2\right\},$$

onde  $A$  e  $B$  são dadas por

$$A = \sum_{i=1}^n \frac{(\sigma_u^2 + \sigma_{x_i}^2)}{|\Sigma_i|} + \frac{1}{\sigma_\alpha^2},$$

$$B = \sum_{i=1}^n \frac{[\sigma_u^2(Y_i - \beta\mu_x) + \sigma_{x_i}^2(Y_i - \beta X_i)]}{|\Sigma_i|} + \frac{a}{\sigma_\alpha^2}.$$

Portanto

$$[\alpha|\beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}] \sim N\left(\frac{B}{A}, \frac{1}{A}\right), \quad A > 0. \quad (5.19)$$

Distribuição condicional completa de  $\mu_x$  dado  $\alpha, \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2$  e  $\mathbf{Z}_{\text{obs}}$

$$\pi(\mu_x|\alpha, \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^n \frac{1}{|\Sigma_i|}(\beta^2\sigma_u^2 + \sigma_e^2) + \frac{1}{\sigma_{\mu_x}^2}\right)\mu_x^2 - 2\left(\beta\sigma_u^2 \sum_{i=1}^n \frac{(Y_i - \alpha)}{|\Sigma_i|} + \sigma_e^2 \sum_{i=1}^n \frac{X_i}{|\Sigma_i|} + \frac{l}{\sigma_{\mu_x}^2}\right)\mu_x\right]\right\}. \quad (5.20)$$

Chamando

$$C = \sum_{i=1}^n \frac{1}{|\Sigma_i|}(\beta^2\sigma_u^2 + \sigma_e^2) + \frac{1}{\sigma_{\mu_x}^2},$$

$$D = \beta\sigma_u^2 \sum_{i=1}^n \frac{(Y_i - \alpha)}{|\Sigma_i|} + \sigma_e^2 \sum_{i=1}^n \frac{X_i}{|\Sigma_i|} + \frac{l}{\sigma_{\mu_x}^2},$$

temos que

$$\begin{aligned}\pi(\mu_x|\alpha, \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) &\propto \exp\left\{-\frac{1}{2}(C\mu_x^2 - 2D\mu_x)\right\} \\ &\propto \exp\left\{-\frac{1}{2C^{-1}}\left(\mu_x - \frac{D}{C}\right)^2\right\}, \quad C > 0,\end{aligned}$$

ou seja

$$[\mu_x|\alpha, \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}] \sim N\left(\frac{D}{C}, \frac{1}{C}\right), \quad C > 0. \quad (5.21)$$

Distribuição condicional completa de  $\beta$  dado  $\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2$  e  $\mathbf{Z}_{\text{obs}}$

$$\begin{aligned}\pi(\beta|\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) &\propto \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^n \left[\frac{(\sigma_u^2\mu_x^2 + \sigma_{x_i}^2 X_i^2)}{|\Sigma_i|}\right]\right)\beta^2\right. \\ &\quad \left.- 2\sum_{i=1}^n \left[\frac{(\sigma_u^2\mu_x + \sigma_{x_i}^2 X_i)(Y_i - \alpha)}{|\Sigma_i|}\right]\beta + \frac{(\beta - b)^2}{\sigma_\beta^2}\right\}.\end{aligned}$$

Fazendo

$$\begin{aligned}\varphi_1(\beta|\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) &= \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^n \left[\frac{(\sigma_u^2\mu_x^2 + \sigma_{x_i}^2 X_i^2)}{|\Sigma_i|}\right]\right)\beta^2\right. \\ &\quad \left.- 2\sum_{i=1}^n \left[\frac{(\sigma_u^2\mu_x + \sigma_{x_i}^2 X_i)(Y_i - \alpha)}{|\Sigma_i|}\right]\beta\right\},\end{aligned}$$

temos que

$$\pi(\beta|\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto \varphi_1(\beta|\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \exp\left\{-\frac{1}{2\sigma_\beta^2}(\beta - b)^2\right\}, \quad (5.22)$$

onde  $\varphi_1(\beta|\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}})$  é uma função uniformemente limitada em  $\beta$  e o segundo fator é o núcleo de uma  $N(b, \sigma_\beta^2)$ , com  $b$  e  $\sigma_\beta^2$  conhecidos.

Distribuição condicional completa de  $\sigma_e^2$  dado  $\alpha, \beta, \mu_x, \sigma_u^2, \sigma_x^2$  e  $\mathbf{Z}_{\text{obs}}$

$$\begin{aligned} \pi(\sigma_e^2 | \beta, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) &\propto (\sigma_e^2)^{-(p+1)} \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sigma_e^2 \sum_{i=1}^n \frac{(X_i - \mu_x)^2}{|\Sigma_i|} + \frac{2q}{\sigma_e^2} \right] \right\} \\ &\propto \varphi_2(\sigma_e^2 | \beta, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) (\sigma_e^2)^{-(p+1)} \exp \left\{ -\frac{q}{\sigma_e^2} \right\}, \end{aligned} \quad (5.23)$$

onde

$$\varphi_2(\sigma_e^2 | \beta, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) = \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sigma_e^2 \sum_{i=1}^n \frac{(X_i - \mu_x)^2}{|\Sigma_i|} \right] \right\},$$

é uma função uniformemente limitada em  $\sigma_e^2$  e o segundo fator é o núcleo de uma distribuição  $IG(p, q)$ .

Distribuição condicional completa de  $\sigma_u^2$  dado  $\alpha, \beta, \mu_x, \sigma_e^2, \sigma_x^2$  e  $\mathbf{Z}_{\text{obs}}$

$$\begin{aligned} \pi(\sigma_u^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) &\propto (\sigma_u^2)^{-(f+1)} \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sigma_u^2 \sum_{i=1}^n \frac{(Y_i - \alpha - \beta \mu_x)^2}{|\Sigma_i|} + \frac{2g}{\sigma_u^2} \right] \right\} \\ &\propto \varphi_3(\sigma_u^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) (\sigma_u^2)^{-(f+1)} \exp \left\{ -\frac{g}{\sigma_u^2} \right\}, \end{aligned} \quad (5.24)$$

onde

$$\varphi_3(\sigma_u^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) = \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sigma_u^2 \sum_{i=1}^n \frac{(Y_i - \alpha - \beta \mu_x)^2}{|\Sigma_i|} \right] \right\},$$

é uma função uniformemente limitada em  $\sigma_u^2$  e o segundo fator é o núcleo de uma distribuição  $IG(f, g)$ .

**Observação:** Notemos que as distribuições condicionais completas univariadas de  $\alpha$  e  $\mu_x$  dadas em (5.19) e (5.21) respectivamente, são Normais, entretanto que das distribuições condicionais completas univariadas de  $\beta, \sigma_e^2$  e  $\sigma_u^2$ , dadas em (5.22), (5.23) e (5.24) respectivamente, conhecemos apenas o núcleo.



## 5.5 O modelo estrutural não identificável com particular *priori* de componentes independentes

No modelo estrutural normal dado em (2.1) com as suposições feitas em (3.2) e (3.3), sem restrição de identificabilidade e com parâmetro desconhecido  $\theta = (\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_{x_1}^2, \dots, \sigma_{x_n}^2)^\top$ , acrescentamos a suposição *a priori* que  $\alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2$  e  $\sigma_{x_i}^2, i = 1, \dots, n$ , são independentes, com distribuições *a priori*  $\pi(\alpha), \pi(\beta), \pi(\mu_x), \pi(\sigma_u^2), \pi(\sigma_e^2)$  e  $\pi(\sigma_{x_i}^2) i = 1, \dots, n$  respectivamente, em que  $\alpha \sim N(a, \sigma_\alpha^2), \beta \sim N(b, \sigma_\beta^2), \mu_x \sim N(l, \sigma_{\mu_x}^2), \sigma_u^2 \sim IG(f, g), \sigma_e^2 \sim IG(p, q)$  e  $\sigma_{x_i}^2 \sim IG(c, d), i = 1, \dots, n$ , onde  $a, \sigma_\alpha^2, b, \sigma_\beta^2, l, \sigma_{\mu_x}^2, c, d, f, g, p$  e  $q$  são hiperparâmetros conhecidos. Assim, a distribuição *a priori* do parâmetro  $\theta$  é dada por

$$\begin{aligned} \pi(\theta) &= \pi(\alpha)\pi(\beta)\pi(\mu_x)\pi(\sigma_u^2)\pi(\sigma_e^2)\pi(\sigma_x^2), \quad \text{onde} \quad \pi(\sigma_x^2) = \prod_{i=1}^n \pi(\sigma_{x_i}^2) \\ &= (\sigma_e^2)^{-(p+1)}(\sigma_u^2)^{-(f+1)} \prod_{i=1}^n (\sigma_{x_i}^2)^{-(c+1)} \exp \left\{ \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} + \frac{(\mu_x - l)^2}{\sigma_{\mu_x}^2} \right. \\ &\quad \left. + \frac{2q}{\sigma_e^2} + \frac{2g}{\sigma_u^2} + 2d \sum_{i=1}^n \frac{1}{\sigma_{x_i}^2} \right\}. \end{aligned} \quad (5.25)$$

## 5.6 Distribuição *a posteriori*

A distribuição *a posteriori* baseada nos dados observados ( $\mathbf{Z}_{obs}$ ) sob esse modelo é dada em (5.7). Então substituindo  $\pi(\theta)$  por (5.25) a distribuição *a posteriori* fica como segue

$$\begin{aligned} p(\theta | \mathbf{Z}_{obs}) &\propto (\sigma_e^2)^{-\left(\frac{n}{2}+p+1\right)}(\sigma_u^2)^{-\left(\frac{n}{2}+f+1\right)} \prod_{i=1}^n \left[ c_i^{-\frac{1}{2}} (\sigma_{x_i}^2)^{-(c+\frac{3}{2})} \right] \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma_e^2} \sum_{i=1}^n (Y_i - \alpha - \beta \mu_x)^2 \right. \right. \\ &\quad \left. \left. + \frac{1}{\sigma_u^2} \sum_{i=1}^n (X_i - \mu_x)^2 - \sum_{i=1}^n c_i^{-1} \left[ \frac{\beta}{\sigma_e^2} (Y_i - \alpha - \beta \mu_x) + \frac{1}{\sigma_u^2} (X_i - \mu_x) \right]^2 \right. \right. \\ &\quad \left. \left. + \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} + \frac{(\mu_x - l)^2}{\sigma_{\mu_x}^2} + \frac{2q}{\sigma_e^2} + \frac{2g}{\sigma_u^2} + 2d \sum_{i=1}^n \frac{1}{\sigma_{x_i}^2} \right] \right\}. \end{aligned} \quad (5.26)$$



A distribuição *a posteriori* baseada nos dados completos ( $\mathbf{W}_c$ ) sob esse modelo é dada por

$$p(\boldsymbol{\theta}|\mathbf{Z}_{\text{obs}}) = \int_{\mathbf{x} = \mathbb{R}^n} p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{Z}_{\text{obs}})d\mathbf{x},$$

onde

$$p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{Z}_{\text{obs}}) \propto L(\boldsymbol{\theta}|\mathbf{W}_c)\pi(\boldsymbol{\theta})$$

com  $L(\boldsymbol{\theta}|\mathbf{W}_c)$  e  $\pi(\boldsymbol{\theta})$  dados em (5.6) e (5.25) respectivamente, ou seja

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{Z}_{\text{obs}}) \propto & (\sigma_\epsilon^2)^{-\left(\frac{n}{2}+p+1\right)}(\sigma_u^2)^{-\left(\frac{n}{2}+f+1\right)} \prod_{i=1}^n (\sigma_{x_i}^2)^{-\left(c+\frac{3}{2}\right)} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma_\epsilon^2} \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 \right. \right. \\ & + \frac{1}{\sigma_u^2} \sum_{i=1}^n (X_i - x_i)^2 + \sum_{i=1}^n \frac{1}{\sigma_{x_i}^2} (x_i - \mu_x)^2 + \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} \\ & \left. \left. + \frac{(\mu_x - l)^2}{\sigma_{\mu_x}^2} + \frac{2q}{\sigma_c^2} + \frac{2g}{\sigma_u^2} + 2d \sum_{i=1}^n \frac{1}{\sigma_{x_i}^2} \right] \right\}. \end{aligned} \quad (5.27)$$

Com base em (5.26) e (5.27), podemos ver que a distribuição *a posteriori* não tem uma forma fechada, apenas o núcleo da distribuição *a posteriori* baseada nos dados observados, dada em (5.26) e o núcleo da distribuição *a posteriori* baseada nos dados completos, dada em (5.27) têm forma conhecida. Além disso, as distribuições marginais e condicionais *a posteriori*, também não tem forma fechada. Procedimentos numéricos computacionais (tipo MCMC), devem ser usados para tratar desse tipo de problema.

### 5.6.1 Distribuições condicionais completas baseados nos dados observados ( $\mathbf{Z}_{\text{obs}}$ )

Baseando-se na distribuição *a posteriori* dada em (5.26), obtemos a seguir as distribuições condicionais completas univariadas dos parâmetros  $\alpha$ ,  $\beta$ ,  $\mu_x$ ,  $\sigma_\epsilon^2$ ,  $\sigma_u^2$  e  $\sigma_{x_i}^2$ ,  $i = 1, \dots, n$ .

Distribuição condicional completa de  $\alpha$  dado  $\beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2$  e  $\mathbf{Z}_{\text{obs}}$

De (5.26), operações algébricas conduzem a

$$\begin{aligned} \pi(\alpha|\beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto & \exp \left\{ -\frac{1}{2} \left[ \left( \frac{n}{\sigma_e^2} - \frac{\beta^2}{(\sigma_e^2)^2} \sum_{i=1}^n c_i^{-1} + \frac{1}{\sigma_\alpha^2} \right) \alpha^2 - 2 \left( \frac{n}{\sigma_e^2} (\bar{Y} - \beta\mu_x) \right. \right. \right. \\ & - \frac{\beta^2}{(\sigma_e^2)^2} \sum_{i=1}^n c_i^{-1} (Y_i - \beta\mu_x) - \frac{\beta}{\sigma_e^2 \sigma_u^2} \sum_{i=1}^n c_i^{-1} X_i \\ & \left. \left. \left. + \frac{\beta\mu_x}{\sigma_e^2 \sigma_u^2} \sum_{i=1}^n c_i^{-1} + \frac{a}{\sigma_\alpha^2} \right) \alpha \right] \right\}. \end{aligned} \quad (5.28)$$

Então

$$\pi(\alpha|\beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto \exp \left\{ -\frac{A}{2} \left[ \alpha^2 - 2\frac{B}{A}\alpha \right] \right\} \propto \exp \left\{ -\frac{1}{2A^{-1}} \left[ \alpha - \frac{B}{A} \right]^2 \right\},$$

onde  $A$  e  $B$  são dadas por

$$\begin{aligned} A &= \frac{n}{\sigma_e^2} - \frac{B^2}{(\sigma_e^2)^2} \sum_{i=1}^n c_i^{-1} + \frac{a}{\sigma_\alpha^2}, \\ B &= \frac{n}{\sigma_e^2} (\bar{Y} - \beta\mu_x) - \frac{\beta^2}{(\sigma_e^2)^2} \sum_{i=1}^n c_i^{-1} (Y_i - \beta\mu_x) - \frac{\beta}{\sigma_e^2 \sigma_u^2} \sum_{i=1}^n c_i^{-1} X_i + \frac{\beta\mu_x}{\sigma_e^2 \sigma_u^2} \sum_{i=1}^n c_i^{-1} + \frac{a}{\sigma_\alpha^2}. \end{aligned}$$

Portanto

$$[\alpha|\beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}] \sim N \left( \frac{B}{A}, \frac{1}{A} \right), \quad A > 0. \quad (5.29)$$

Distribuição condicional completa de  $\mu_x$  dado  $\alpha, \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2$  e  $\mathbf{Z}_{\text{obs}}$

Fazendo operações algébricas em (5.26), obtemos

$$\begin{aligned} \pi(\mu_x | \alpha, \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) &\propto \exp \left\{ -\frac{1}{2} \left[ \left( \frac{n\beta^2}{\sigma_e^2} + \frac{n}{\sigma_u^2} - \left[ \frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right]^2 \sum_{i=1}^n c_i^{-1} + \frac{1}{\sigma_{\mu_x}^2} \right) \mu_x^2 \right. \right. \\ &\quad - 2 \left( \frac{n\beta(\bar{Y} - \alpha)}{\sigma_e^2} + \frac{n}{\sigma_u^2} \bar{X} - \frac{\beta^3}{(\sigma_e^2)^2} \sum_{i=1}^n c_i^{-1} (Y_i - \alpha) \right. \\ &\quad - \frac{\beta^2}{\sigma_e^2 \sigma_u^2} \sum_{i=1}^n c_i^{-1} X_i - \frac{\beta}{\sigma_e^2 \sigma_u^2} \sum_{i=1}^n c_i^{-1} (Y_i - \alpha) \\ &\quad \left. \left. - \frac{1}{(\sigma_u^2)^2} \sum_{i=1}^n c_i^{-1} X_i + \frac{l}{\sigma_{\mu_x}^2} \right) \mu_x \right] \Big\}, \end{aligned} \quad (5.30)$$

e temos que

$$\begin{aligned} \pi(\mu_x | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) &\propto \exp \left\{ -\frac{1}{2} (C\mu_x^2 - 2D\mu_x) \right\} \\ &\propto \exp \left\{ -\frac{1}{2C^{-1}} \left( \mu_x - \frac{D}{C} \right)^2 \right\}, \quad C > 0, \end{aligned}$$

onde  $D$  e  $C$  são dadas por

$$\begin{aligned} C &= \frac{n\beta^2}{\sigma_e^2} + \frac{n}{\sigma_u^2} - \left[ \frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right]^2 \sum_{i=1}^n c_i^{-1} + \frac{1}{\sigma_{\mu_x}^2}, \\ D &= \frac{n\beta(\bar{Y} - \alpha)}{\sigma_e^2} + \frac{n}{\sigma_u^2} \bar{X} - \frac{\beta^3}{(\sigma_e^2)^2} \sum_{i=1}^n c_i^{-1} (Y_i - \alpha) - \frac{\beta^2}{\sigma_e^2 \sigma_u^2} \sum_{i=1}^n c_i^{-1} X_i \\ &\quad - \frac{\beta}{\sigma_e^2 \sigma_u^2} \sum_{i=1}^n c_i^{-1} (Y_i - \alpha) - \frac{1}{(\sigma_u^2)^2} \sum_{i=1}^n c_i^{-1} X_i + \frac{l}{\sigma_{\mu_x}^2}, \end{aligned}$$

ou seja

$$[\mu_x | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}] \sim N \left( \frac{D}{C}, \frac{1}{C} \right), \quad C > 0. \quad (5.31)$$

Distribuição condicional completa de  $\beta$  dado  $\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2$  e  $\mathbf{Z}_{\text{obs}}$

Considerando (5.26) e omitindo várias passagens, operações algébricas conduzem a

$$\begin{aligned} \pi(\beta|\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) &\propto \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \left( \mu_x^2 \sum_{i=1}^n \frac{(\sigma_u^2 + \sigma_{x_i}^2)}{|\Sigma_i|} \right. \right. \right. \\ &\quad \left. \left. - 2\mu_x \sum_{i=1}^n \frac{\sigma_{x_i}^2 (X_i - \mu_x)}{|\Sigma_i|} \right) \beta^2 - 2 \left( \mu_x \sum_{i=1}^n \frac{(\sigma_u^2 + \sigma_{x_i}^2)(Y_i - \alpha)}{|\Sigma_i|} \right. \right. \\ &\quad \left. \left. + \sum_{i=1}^n \frac{\sigma_{x_i}^2 (Y_i - \alpha)(X_i - \mu_x)}{|\Sigma_i|} \right) \beta + \frac{(\beta - b)^2}{\sigma_\beta^2} \right] \right\}. \end{aligned}$$

Fazendo

$$\begin{aligned} E &= \sum_{i=1}^n \frac{1}{|\Sigma_i|} \left[ \mu_x^2 (\sigma_u^2 + \sigma_{x_i}^2) - 2\mu_x \sigma_{x_i}^2 (X_i - \mu_x) \right], \\ F &= \sum_{i=1}^n \frac{1}{|\Sigma_i|} \left[ \mu_x (\sigma_u^2 + \sigma_{x_i}^2) (Y_i - \alpha) + \sigma_{x_i}^2 (Y_i - \alpha) (X_i - \mu_x) \right], \end{aligned}$$

temos que

$$\pi(\beta|\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto \varphi_1(\beta|\alpha, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \exp \left\{ -\frac{1}{2\sigma_\beta^2} (\beta - b)^2 \right\}, \quad (5.32)$$

onde

$$\varphi_1(\beta|\alpha, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) = \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (E\beta^2 - 2F\beta) \right\},$$

é uma função uniformemente limitada em  $\beta$  e o segundo fator é o núcleo de uma  $N(b, \sigma_\beta^2)$ , com  $b$  e  $\sigma_\beta^2$  conhecidos.

Distribuição condicional completa de  $\sigma_e^2$  dado  $\alpha, \beta, \mu_x, \sigma_u^2, \sigma_x^2$  e  $\mathbf{Z}_{\text{obs}}$

Partindo de (5.26) e omitindo várias passagens obtemos a seguinte expressão para a condicional completa de  $\sigma_e^2$

$$\pi(\sigma_e^2|\alpha, \beta, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto (\sigma_e^2)^{-(p+1)} \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n \frac{1}{|\Sigma_i|} \sigma_e^2 (X_i - \mu_x)^2 + \frac{2q}{\sigma_\epsilon^2} \right] \right\},$$

além disso

$$\pi(\sigma_e^2 | \alpha, \beta, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto \varphi_2(\sigma_e^2 | \beta, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) (\sigma_e^2)^{-(p+1)} \exp \left\{ -\frac{1}{2} \left[ \frac{2q}{\sigma_e^2} \right] \right\} \quad (5.33)$$

onde

$$\varphi_2(\sigma_e^2 | \beta, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) = \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n \frac{1}{|\Sigma_i|} \sigma_e^2 (X_i - \mu_x)^2 \right] \right\},$$

é uma função uniformemente limitada em  $\sigma_e^2$  e o segundo fator é o núcleo de uma distribuição  $IG(p, q)$ .

**Distribuição condicional completa de  $\sigma_u^2$  dado  $\alpha, \beta, \mu_x, \sigma_e^2, \sigma_x^2$  e  $\mathbf{Z}_{\text{obs}}$**

De (5.26), temos que

$$\pi(\sigma_u^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto (\sigma_u^2)^{-(f+1)} \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n \frac{1}{|\Sigma_i|} \sigma_u^2 (Y_i - \alpha - \beta \mu_x)^2 + \frac{2g}{\sigma_u^2} \right] \right\},$$

ou

$$\pi(\sigma_u^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) \propto \varphi_3(\sigma_u^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) (\sigma_u^2)^{-(f+1)} \exp \left\{ -\frac{2g}{\sigma_u^2} \right\} \quad (5.34)$$

onde

$$\varphi_3(\sigma_u^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_x^2, \mathbf{Z}_{\text{obs}}) = \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n \frac{1}{|\Sigma_i|} \sigma_u^2 (Y_i - \alpha - \beta \mu_x)^2 \right] \right\},$$

é uma função uniformemente limitada em  $\sigma_u^2$  e o segundo fator é o núcleo de uma distribuição  $IG(f, g)$ .

**Distribuição condicional completa de  $\sigma_{x_i}^2$  dado  $\alpha, \beta, \mu_x, \sigma_e^2$  e  $\mathbf{Z}_{\text{obs}}$ ,  $i = 1, \dots, n$**

Finalmente, após de algumas manipulações algébricas em (5.26), obtemos

$$\pi(\sigma_{x_i}^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2, \mathbf{Z}_{\text{obs}}) \propto (\sigma_{x_i}^2)^{-(c+1)} |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{|\Sigma_i|} \sigma_{x_i}^2 (Y_i - \alpha - \beta X_i)^2 + \frac{2d}{\sigma_{x_i}^2} \right] \right\},$$

$i = 1, \dots, n$ . Além disso

$$\pi(\sigma_{x_i}^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2, \mathbf{Z}_{\text{obs}}) \propto \varphi_4(\sigma_u^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2, \mathbf{Z}_{\text{obs}}) (\sigma_{x_i}^2)^{-(c+1)} \exp \left\{ -\frac{2d}{\sigma_{x_i}^2} \right\} \quad (5.35)$$

$i = 1, \dots, n$ , em que

$$\varphi_4(\sigma_{x_i}^2 | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2, \mathbf{Z}_{\text{obs}}) = |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \frac{1}{|\Sigma_i|} \sigma_{x_i}^2 (Y_i - \alpha - \beta X_i)^2 \right] \right\},$$

é uma função uniformemente limitada em  $\sigma_{x_i}^2$  e o segundo fator é o núcleo de uma distribuição  $IG(c, d)$ .

**Observação:** Pelo descrito acima, podemos dizer que as distribuições condicionais completas univariadas de  $\alpha$  e  $\mu_x$  dadas em (5.29) e (5.31) respectivamente, são Normais e das distribuições condicionais completas univariadas de  $\beta$ ,  $\sigma_e^2$ ,  $\sigma_u^2$  e  $\sigma_{x_i}^2$ ,  $i = 1, \dots, n$  dadas em (5.32), (5.33), (5.34) e (5.35) respectivamente, conhecemos apenas o núcleo.

Semelhantemente podemos obter as distribuições condicionais completas *a posteriori* baseadas nos dados completos.

### 5.6.2 Distribuições condicionais completas *a posteriori* baseadas nos dados completos ( $\mathbf{W}_c$ )

Usando a distribuição *a posteriori* em (5.27), obtemos as seguintes distribuições condicionais completas para os parâmetros do modelo.

A distribuição condicional de  $\alpha$  dado  $\beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2$  e  $\mathbf{W}_c$  é dada por

$$\pi(\alpha | \beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{W}_c) \propto \exp \left\{ -\frac{1}{2} \left( \frac{n\sigma_\alpha^2 + \sigma_e^2}{\sigma_e^2 \sigma_\alpha^2} \right) \left[ \alpha - \left( \frac{n\sigma_\alpha^2 (\bar{Y} - \beta \bar{x}) + a\sigma_e^2}{n\sigma_\alpha^2 + \sigma_e^2} \right) \right]^2 \right\}.$$

Portanto

$$[\alpha | \beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{W}_c] \sim \text{N} \left( \frac{n\sigma_\alpha^2 (\bar{Y} - \beta \bar{x}) + a\sigma_e^2}{n\sigma_\alpha^2 + \sigma_e^2}; \frac{\sigma_e^2 \sigma_\alpha^2}{n\sigma_\alpha^2 + \sigma_e^2} \right). \quad (5.36)$$

A distribuição condicional de  $\beta$  dado  $\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2$  e  $\mathbf{W}_c$  é dada por

$$\pi(\beta|\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{W}_c) \propto \exp \left\{ -\frac{1}{2} \left( \frac{\sigma_\beta^2 \sum_{i=1}^n x_i^2 + \sigma_e^2}{\sigma_e^2 \sigma_\beta^2} \right) \left[ \beta^2 - 2 \left( \frac{\sigma_\beta^2 (\sum_{i=1}^n Y_i x_i - \alpha n \bar{x}) + b \sigma_e^2}{\sigma_\beta^2 \sum_{i=1}^n x_i^2 + \sigma_e^2} \right) \beta \right] \right\},$$

assim

$$[\beta|\alpha, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{W}_c] \sim N \left( \frac{\sigma_\beta^2 (\sum_{i=1}^n Y_i x_i - \alpha n \bar{x}) + b \sigma_e^2}{\sigma_\beta^2 \sum_{i=1}^n x_i^2 + \sigma_e^2}; \frac{\sigma_e^2 \sigma_\beta^2}{\sigma_\beta^2 \sum_{i=1}^n x_i^2 + \sigma_e^2} \right). \quad (5.37)$$

A distribuição condicional de  $\mu_x$  dado  $\alpha, \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2$  e  $\mathbf{W}_c$  é dada por

$$\begin{aligned} \pi(\mu_x|\alpha, \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{W}_c) &\propto \exp \left\{ -\frac{1}{2} (A \mu_x^2 - 2B \mu_x) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} A \left( \mu_x^2 - 2 \frac{B}{A} \mu_x \right) \right\}, \end{aligned}$$

onde

$$A = \sum_{i=1}^n \left( \frac{n \sigma_{\mu_x}^2 + \sigma_{x_i}^2}{n \sigma_{x_i}^2 \sigma_{\mu_x}^2} \right), \quad B = \sum_{i=1}^n \left( \frac{n \sigma_{\mu_x}^2 x_i + l \sigma_{x_i}^2}{n \sigma_{x_i}^2 \sigma_{\mu_x}^2} \right)$$

portanto

$$[\mu_x|\alpha, \beta, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mathbf{W}_c] \sim N \left( \frac{B}{A}; \frac{1}{A} \right), \quad A > 0. \quad (5.38)$$

A distribuição condicional de  $\sigma_e^2$  dado  $\alpha, \beta, \mu_x, \sigma_u^2, \sigma_x^2$  e  $\mathbf{W}_c$  é dada por

$$\pi(\sigma_e^2|\alpha, \beta, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{W}_c) \propto (\sigma_e^2)^{-\left(\frac{n}{2} + p + 1\right)} \exp \left\{ -\frac{1}{\sigma_e^2} \left[ \frac{1}{2} \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 + q \right] \right\},$$



assim

$$[\sigma_c^2 | \alpha, \beta, \mu_x, \sigma_u^2, \sigma_x^2, \mathbf{W}_c] \sim IG \left( \frac{n}{2} + p; \frac{1}{2} \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 + q \right). \quad (5.39)$$

A distribuição de  $\sigma_u^2$  dado  $\alpha, \beta, \mu_x, \sigma_c^2, \sigma_x^2$  e  $\mathbf{W}_c$  é dada por

$$\pi(\sigma_u^2 | \alpha, \beta, \mu_x, \sigma_c^2, \sigma_x^2, \mathbf{W}_c) \propto (\sigma_u^2)^{-\left(\frac{n}{2} + f + 1\right)} \exp \left\{ -\frac{1}{\sigma_u^2} \left[ \frac{1}{2} \sum_{i=1}^n (X_i - x_i)^2 + g \right] \right\},$$

portanto

$$[\sigma_u^2 | \alpha, \beta, \mu_x, \sigma_c^2, \sigma_x^2, \mathbf{W}_c] \sim IG \left( \frac{n}{2} + f; \frac{1}{2} \sum_{i=1}^n (X_i - x_i)^2 + g \right) \quad (5.40)$$

A distribuição condicional de  $\sigma_{x_i}^2$  dado  $\alpha, \beta, \mu_x, \sigma_u^2, \sigma_c^2$  e  $\mathbf{W}_c$ ,  $i = 1, \dots, n$  é dada por

$$\pi(\sigma_{x_i}^2 | \alpha, \beta, \mu_x, \sigma_u^2, \sigma_c^2, \mathbf{W}_c) \propto (\sigma_{x_i}^2)^{-\left(c + \frac{1}{2} + 1\right)} \exp \left\{ -\frac{1}{\sigma_{x_i}^2} \left[ \frac{(x_i - \mu_x)^2}{2} + d \right] \right\}, \quad i = 1, \dots, n,$$

assim

$$[\sigma_{x_i}^2 | \alpha, \beta, \mu_x, \sigma_u^2, \sigma_c^2, \mathbf{W}_c] \sim IG \left( c + \frac{1}{2}; \frac{(x_i - \mu_x)^2}{2} + d \right), \quad i = 1, \dots, n. \quad (5.41)$$

A distribuição condicional de  $x_i$  dado  $\alpha, \beta, \mu_x, \sigma_c^2, \sigma_u^2, \sigma_x^2$  e  $\mathbf{W}_c$ ,  $i = 1, \dots, n$  é dada por

$$\begin{aligned} \pi(x_i | \alpha, \beta, \mu_x, \sigma_c^2, \sigma_u^2, \sigma_x^2, \mathbf{W}_c) &\propto \exp \left\{ -\frac{1}{2} (C_i x_i^2 - 2D_i x_i) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} C_i \left( x_i^2 - 2 \frac{D_i}{C_i} x_i \right) \right\}, \quad i = 1, \dots, n, \end{aligned}$$

onde

$$C_i = \frac{\beta^2 \sigma_u^2 \sigma_{x_i}^2 + \sigma_e^2 \sigma_{x_i}^2 + \sigma_e^2 \sigma_u^2}{\sigma_e^2 \sigma_u^2 \sigma_{x_i}^2}$$

$$D_i = \frac{\beta \sigma_u^2 \sigma_{x_i}^2 (Y_i - \alpha) + \sigma_e^2 \sigma_{x_i}^2 X_i + \mu_x \sigma_e^2 \sigma_u^2}{\sigma_e^2 \sigma_u^2 \sigma_{x_i}^2}, \quad i = 1, \dots, n,$$

portanto

$$[x_i | \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_{x_i}^2, \mathbf{Z}_{obs}] \sim N\left(\frac{D_i}{C_i}; \frac{1}{C_i}\right), \quad C_i > 0, \quad i = 1, \dots, n. \quad (5.42)$$

**Observação:** A introdução das variáveis latentes (não observadas)  $x_1, \dots, x_n$  faz com que as distribuições condicionais completas univariadas de  $\alpha$ ,  $\beta$  e  $\mu_x$  dadas em (5.36), (5.37) e (5.38) respectivamente, sejam Normais e as distribuições condicionais completas univariadas de  $\sigma_e^2$ ,  $\sigma_u^2$  e  $\sigma_{x_i}^2$ ,  $i = 1, \dots, n$  dadas em (5.39), (5.40) e (5.41) respectivamente, sejam Gammas Invertidas.

## 5.7 Simulação

Nesta seção utilizamos uma amostra de tamanho  $n = 30$  de dados simulados a partir do modelo definido em (5.1) e (5.2) com a seguinte configuração para os parâmetros

$$\alpha = -1, \quad \beta = 1, \quad \mu_x = -2, \quad \sigma_e^2 = 5, \quad \text{e} \quad \sigma_u^2 = 1,$$

além disso a variância  $\sigma_{x_i}^2$  foi gerada como segue  $\sqrt{\sigma_{x_i}^2} \stackrel{i.i.d.}{\sim} U(0.5, 4)$ ,  $i = 1, \dots, n$  e foram mantidos fixos em todas as simulações.

A simulação foi feita utilizando o software estatístico WinBUGS. A configuração de nosso modelo não está na biblioteca de WinBUGS, portanto utilizamos diretamente a função de verossimilhança para gerar as cadeias. O modelo utilizado é dado na Figura 5.1. Usamos *prioris* não informativas para os parâmetros com a finalidade de usar toda a informação contida na verossimilhança. Distribuições *a priori* normais para os parâmetros  $\alpha$ ,  $\beta$ ,  $\mu_x$  e distribuições *a priori* uniformes para os parâmetros  $\sigma_e$  e  $\sigma_u$ . Utilizamos as estimativas dos momentos como valores iniciais para as cadeias. Para o processo do diagnóstico de convergência

---

das seqüências de Gibbs, usamos uma única seqüência de Gibbs de comprimento  $T = 30,000$  e para o processo de estimação dos parâmetros, foi usada uma única seqüência de Gibbs de comprimento  $T = 10,000$ . Os resultados da simulação foram os seguintes:

a) **Diagnóstico de convergência das seqüências de Gibbs**

A seguir são apresentados alguns dispositivos gráficos para o diagnóstico de possíveis falhas na convergência do algoritmo de Gibbs. Das Figuras 5.2 - 5.5, podemos observar que os dispositivos apresentados não sugerem que exista problema na convergência do algoritmo de Gibbs.

b) **Estimação**

A Tabela 5.1 apresenta as estimativas de várias medidas de resumo da distribuição *a posteriori*. Estas estimativas foram obtidas utilizando-se uma única seqüência de Gibbs de comprimento  $T = 10,000$ . Como utilizamos *prioris* não informativas, se fato o procedimento de estimação é adequado, espera-se que as estimativas *a posteriori* dos parâmetros estejam próximas dos valores dos parâmetros que foram usados na geração dos dados. Além disso, essas estimativas deveriam também estar próximas das estimativas de máxima verossimilhança.

Na Tabela 5.1 observa-se que as estimativas do valor esperado da distribuição *a posteriori* dos parâmetros, obtidas pela média aritmética e pela mediana amostral, estimam cada um dos parâmetros com um "pequeno" viés e somente o parâmetro  $\mu_x$  parece estar superestimado, as outras estimativas estão próximas dos valores dos parâmetros. Além disso, mais importante ainda, os valores verdadeiros dos parâmetros (os usados para gerar a amostra) pertencem aos intervalos de credibilidade apresentados. A Figura 5.6 apresenta os gráficos das estimativas das densidades *posteriores* marginais dos parâmetros.

```

Model:
  model
  {
    for(i in 1:n){
      zeros[i] <- 0
      phi[i] <- log(varu*vare +pow(beta,
2)*varu*varx[i]+vare*varx[i])/2+((varx[i]+varu)*pow((y[i]-alfa-
beta*mux), 2)+(pow(beta, 2)*varx[i]+vare)*pow((x[i]-mux), 2)-2*beta*varx[i]*(y[i]-
alfa-beta*mux)*(x[i]-mux)))/(2*(varu*vare+pow(beta, 2)*varu*varx[i]+vare*varx[i]))+C
      zeros[i] ~ dpois(phi[i])
    }
    alfa ~ dunif(-20,20)
    beta ~ dunif(-20,20)
    mux ~ dunif(-20,20)
    uvare ~ dunif(0,20)
    vare<- uvare*uvare
    uvaru ~ dunif(0,20)
    varu<- uvaru*uvaru
  }
}

Data:
list(y=c(-3.644,-0.540,-2.837, 0.620,-4.464,-5.838,-4.753,-4.783,-3.973,
3.268,-0.195,-4.205,-4.062,-0.303,-4.024,-6.996,-8.757,-0.991,
-4.065,-10.100, 1.410, 1.587,-2.665,-0.321,-2.276,-1.756,-7.861,
-5.222,-0.922,-0.773), x=c(-4.870, 0.733,-2.336, 3.435,-3.597,-2.774,-3.909,-5.596,-0.888,-0.802,
0.698,-4.138,-1.178, 0.461,-1.175,-6.333,-2.146,-0.514,-4.821,-6.932,-1.293, 2.298,-0.302,-0.783,
3.186,4.112,-3.013,-1.964, 0.280, 2.827), varx=c(3.956, 10.975, 2.053, 6.472, 6.678, 1.635, 2.531,
15.199, 8.357, 7.896 ,3.758, 4.489, 13.721, 13.839, 7.854, 15.889, 7.520, 4.185, 1.988, 15.926,
2.346, 8.333, 0.788, 1.370, 12.221, 15.059, 13.034, 12.248, 9.568, 11.715),C=1000, n=30)

Initial:
list(alfa=-1.033, beta=0.989, mux=-1.995, uvare=2.205, uvaru=1.019)
list(alfa=-0.833, beta=1.089, mux=-2.095, uvare=2, uvaru=1)
list(alfa=-1.563, beta=1.789, mux=-1.095, uvare=0.92, uvaru=0.5)

```

Figura 5.1. Programa WinBUGS para o modelo definido em (5.1) e (5.2)

Tabela 5.1. Estimativas do valor esperado, erro padrão, quantis e intervalos de credibilidade para os parâmetros  $\alpha$ ,  $\beta$ ,  $\mu_x$ ,  $\sigma_e^2$  e  $\sigma_u^2$ , baseados em uma única seqüência de Gibbs de comprimento  $T=10,000$ .

Medidas de Resumo	$\alpha$	$\beta$	$\mu_x$	$\sigma_e^2$	$\sigma_u^2$
Média aritmética	-1.731	0.908	-1.471	4.835	1.861
Erro padrão	0.595	0.251	0.491	2.284	1.737
Erro MC	0.0113	0.0056	0.0069	0.0421	0.0375
Mediana	-1.757	0.880	-1.487	4.785	1.352
Intervalo de credibilidade					
Limite inferior (2.5%)	-2.830	0.486	-2.384	0.256	0.009
Limite superior (97.5%)	-0.512	1.466	-0.437	9.560	6.240

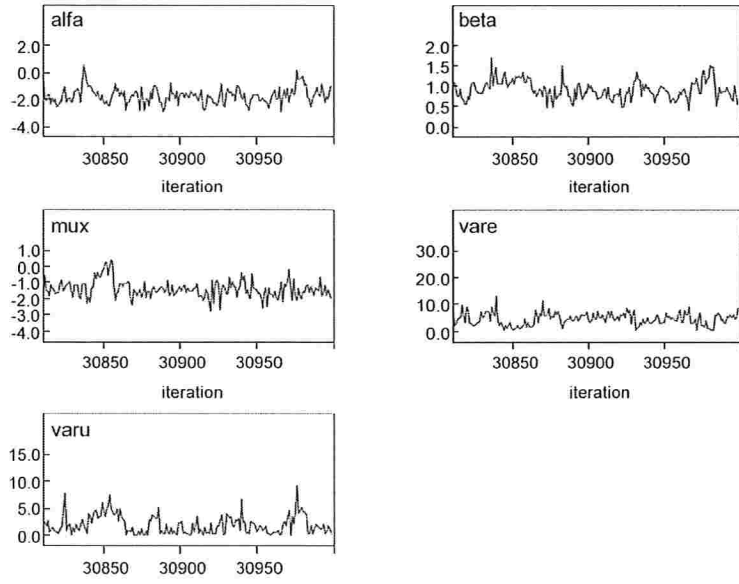


Figura 5.2. "Trace" para os parâmetros, baseados em uma única sequência de Gibbs de comprimento  $T=30,000$ .

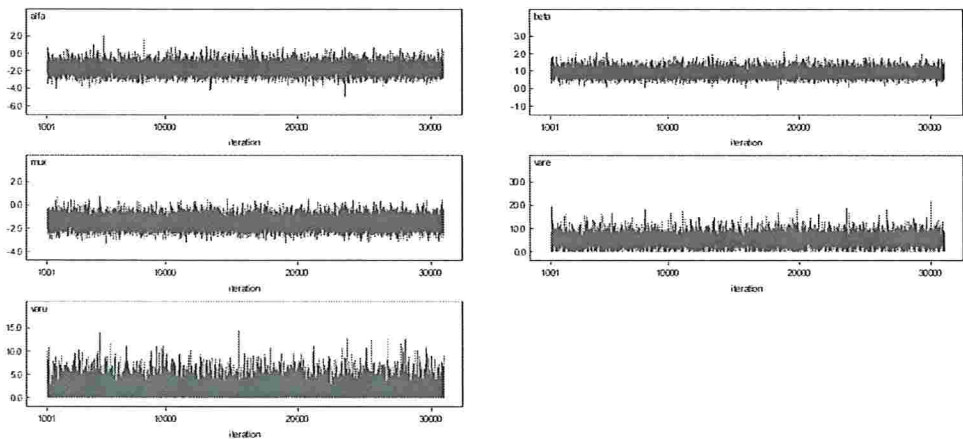


Figura 5.3. Comportamento histórico das cadeias para os parâmetros, baseados em uma única sequência de Gibbs de comprimento  $T=30,000$ .

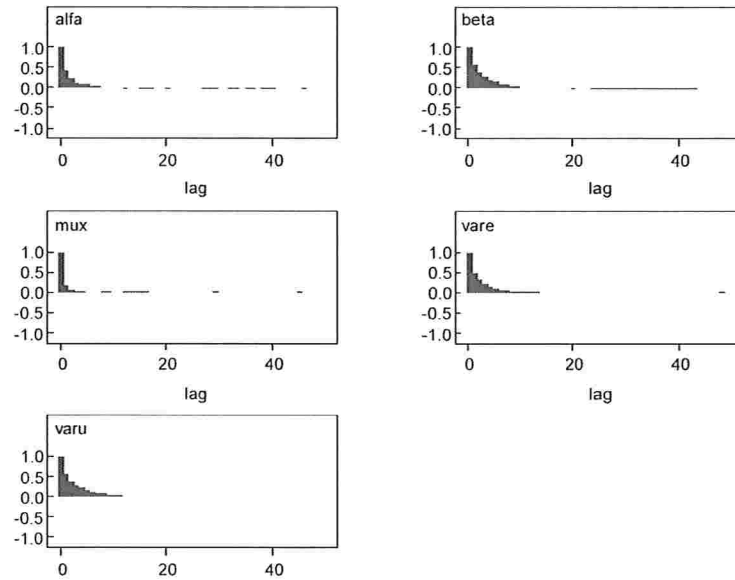


Figura 5.4. Gráficos das funções de auto-correlação dos parâmetros, baseados em uma única seqüência de Gibbs de comprimento  $T=30,000$ .

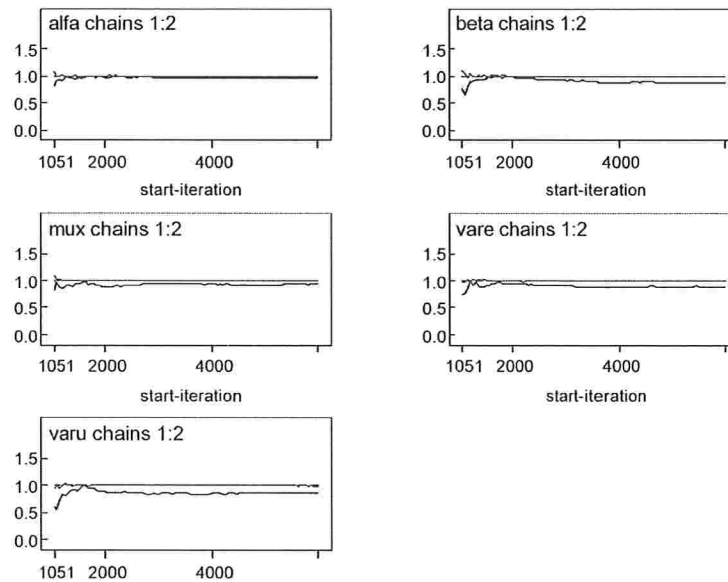


Figura 5.5. Cadeias paralelas para os parâmetros, baseados em uma única seqüência de Gibbs de comprimento  $T=10,000$ .

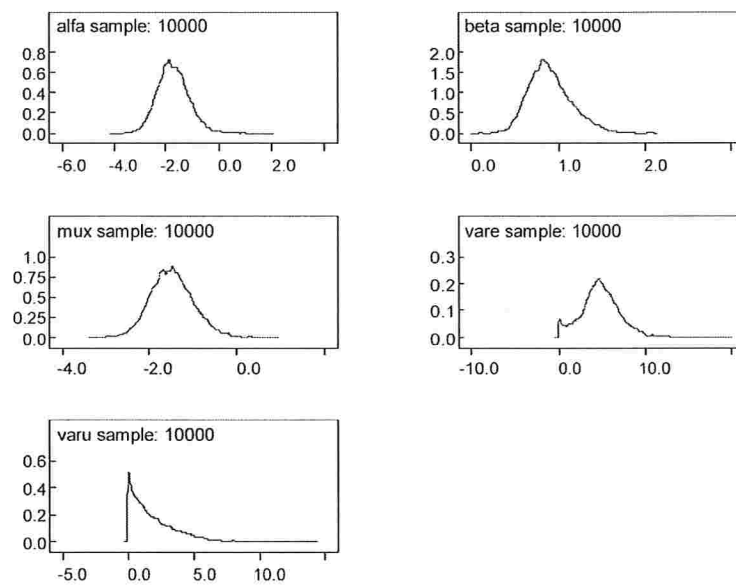


Figura 5.6. Gráficos das estimativas das densidades marginais *a posteriori* para os parâmetros, baseados em uma única seqüência de Gibbs de comprimento  $T=10,000$ .



---

## Capítulo 6

# Contribuições e Estudos futuros

### 6.1 Principais contribuições

As principais contribuições deste trabalho são as seguintes:

- No Capítulo 2 relaxamos a suposição de normalidade do vetor  $\mathbf{Z}_i = (Y_i, X_i)^T$ ,  $i = 1, \dots, n$ , imposta por Kulathinal et al. (2002) na obtenção dos EMM do modelo linear heterocedástico com erros nas variáveis e na equação. Desenvolvemos um método para a obtenção dos EMM baseado só na existência do primeiro e segundo momento do vetor aleatório  $\mathbf{Z}_i$ . Mesmo que este método tenha sido utilizado num modelo sem erro na equação, este resultado continua sendo válido no caso com erro na equação. Também desenvolvemos a matriz de covariâncias assintótica para os estimadores de momentos obtidos pelo método proposto. Além disso, determinamos a matriz de informação de Fisher por diferenciação direta da função de log-verossimilhança (Kulathinal et al. (2002) apresenta outro método alternativo), obtendo assim a distribuição assintótica dos estimadores de máxima verossimilhança. Finalmente conduzimos várias simulações a fim de avaliar o comportamento das estatísticas do tipo Wald e dos estimadores dos parâmetros de regressão.
- No Capítulo 3 desenvolvemos uma teoria assintótica para determinar os EMM e os EMV (com seus respectivos desvios padrões) dos parâmetros estruturais de um MREH com erros nas variáveis, no caso em que a variável não observada  $x_i$  é heterocedástica. Além disso propomos um método geral para obter a matriz de informação de Fisher associada aos parâmetros do modelo definido pelas equações (1.5), (1.6) e (1.7) com suas respectivas restrições de identificabilidade (casos particulares). É bom lembrar que este modelo com a configuração elaborada nesta tese, ainda não é abordado na literatura.

- No Capítulo 4 tentamos desenvolver um método que nos permita incorporar a estimação dos parâmetros  $\sigma_{x_i}^2$  e  $\sigma_u^2$  (ou  $a_i = \frac{\sigma_{x_i}}{\sigma_u}$ ,  $i = 1, \dots, n$ ) dentro do processo de estimação. O conhecimento de  $a_i$ ,  $i = 1, \dots, n$  permite o conhecimento da razão da confiabilidade (ou coeficiente de atenuação)  $K_{x_i} = \frac{\sigma_{x_i}^2}{\sigma_{x_i}^2 + \sigma_u^2}$  e, assim, resolver o problema de falta de identificação do modelo.

## 6.2 Comentários finais

Nesta tese, desenvolvemos uma metodologia para obter estimadores consistentes (com seus correspondentes desvios padrões) dos parâmetros estruturais, primeiro num modelo de regressão estrutural com erros de medição quando as variâncias das medidas dos erros variam entre as observações (Capítulo 2) e logo no caso em que a variância da covariável não observada é considerada heterocedástica (Capítulo 3).

Os modelos introduzidos como já é bem sabido apresentam problemas de falta de identificabilidade, deste modo suposições adicionais devem ser consideradas. No caso do modelo do Capítulo 2 a restrição de identificabilidade foi supor que as variâncias dos erros de medição eram conhecidas. No caso do modelo do Capítulo 3 a restrição foi considerar  $\sigma_{x_i}^2$  conhecida ou  $a_i = \sigma_{x_i}^2 / \sigma_u^2$ ,  $i = 1, \dots, n$ , conhecida, esta última restrição equivale a conhecer o coeficiente de atenuação.

Uma maneira alternativa de lidar com a informação adicional da variância é mediante o uso de réplicas. Então introduzimos uma extensão do modelo definido no Capítulo 3, com replicações apenas na variável explanatória (Capítulo 4), com a idéia de desenvolver algum método que nos permitisse estimar as restrições de identificabilidade do modelo do Capítulo 3. Finalmente para salvar definitivamente o problema de identificabilidade tentamos fazer uma inferência Bayesiana simples (Capítulo 5), para o modelo definido no Capítulo 3.

Como um comentário final, podemos dizer que as contribuições mais importantes desta tese estão nos resultados do Capítulo 3.

---

### 6.3 Sugestões para pesquisas futuras

- Estender o modelo apresentado no Capítulo 3 para o caso em que a variável não observada  $x_i$  tenha uma distribuição elíptica, que garanta uma inferência estatística mais robusta que a normal.
- Estender o modelo apresentado no Capítulo 3 para o caso em que a variável não observada  $x_i$  tenha uma distribuição assimétrica.
- Fazer estudos de influência local nos modelos propostos nos Capítulos 3 e 4.
- Programar o método MCMC (sem WinBUGS) para a teoria da abordagem Bayesiana proposta no Capítulo 5. Conduzir estudos de simulação e aplicações com o modelo.

## Apêndice A

### Informação Observada

Os elementos da matriz de informação observada, denotada por  $\mathbf{K}_i$ , no caso de replicações não-balanceadas na variável exploratória (Capítulo 4 Subseção 4.4.1), tem a seguinte expressão geral

$$\begin{aligned} K_{i\gamma\psi} = & -\frac{1}{2} \frac{\partial^2}{\partial\gamma\partial\psi} \ln |\Sigma_i| - \frac{1}{2} \frac{\partial^2 t_i}{\partial\gamma\partial\psi} + c_i^{-1} \left[ \frac{\partial h_i}{\partial\gamma} \frac{\partial h_i}{\partial\psi} + h_i \frac{\partial^2 h_i}{\partial\gamma\partial\psi} \right] \\ & - h_i c_i^{-2} \left[ \frac{\partial c_i}{\partial\gamma} \frac{\partial h_i}{\partial\psi} + \frac{\partial h_i}{\partial\gamma} \frac{\partial c_i}{\partial\psi} \right] + h_i^2 c_i^{-2} \left[ c_i^{-1} \frac{\partial c_i}{\partial\gamma} \frac{\partial c_i}{\partial\psi} - \frac{1}{2} \frac{\partial^2 c_i}{\partial\gamma\partial\psi} \right], \end{aligned} \quad (\text{A.1})$$

onde  $\gamma, \psi = \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2$ .

Derivando (4.16) em relação a  $\psi$  (e fazendo  $k_i = c_i$ ,  $w_i = t_i$ , e  $f_i = h_i$ ), obtemos as derivadas de primeira ordem

$$\frac{\partial l_i(\boldsymbol{\theta})}{\partial\psi} = -\frac{1}{2} \frac{\partial}{\partial\psi} \ln |\Sigma_i| - \frac{1}{2} \frac{\partial t_i}{\partial\psi} + c_i^{-1} h_i \frac{\partial h_i}{\partial\psi} - \frac{1}{2} h_i^2 c_i^{-2} \frac{\partial c_i}{\partial\psi}, \quad (\text{A.2})$$

onde  $\psi = \alpha, \beta, \mu_x, \sigma_e^2, \sigma_u^2$ , com

$$\begin{aligned} \frac{\partial \ln |\Sigma_i|}{\partial\alpha} &= \frac{\partial c_i}{\partial\alpha} = 0, & \frac{\partial t_i}{\partial\alpha} &= -2\sigma_e^{-2}(Y_i - \alpha - \beta\mu_x), & \frac{\partial h_i}{\partial\alpha} &= -\beta\sigma_e^{-2}, \\ \frac{\partial \ln |\Sigma_i|}{\partial\beta} &= 2\beta\sigma_e^{-2}c_i^{-1}, & \frac{\partial t_i}{\partial\beta} &= -2\mu_x\sigma_e^{-2}(Y_i - \alpha - \beta\mu_x), & \frac{\partial h_i}{\partial\beta} &= \sigma_e^{-2}(Y_i - \alpha - 2\beta\mu_x), \end{aligned}$$

$$\begin{aligned}
\frac{\partial c_i}{\partial \beta} &= 2\beta\sigma_\epsilon^{-2}, & \frac{\partial t_i}{\partial \mu_x} &= -2\sigma_\epsilon^{-2}\beta(Y_i - \alpha - \beta\mu_x) - 2\sigma_u^{-2}\sum_{j=1}^{m_i}(X_{ij} - \mu_x), \\
\frac{\partial \ln |\Sigma_i|}{\partial \mu_x} &= \frac{\partial c_i}{\partial \mu_x} = 0 & \frac{\partial h_i}{\partial \mu_x} &= -\beta^2\sigma_\epsilon^{-2} - m_i\sigma_u^{-2}, \\
\frac{\partial t_i}{\partial \sigma_\epsilon^2} &= -\sigma_\epsilon^{-4}(Y_i - \alpha - \beta\mu_x), & \frac{\partial h_i}{\partial \sigma_\epsilon^2} &= -\beta\sigma_\epsilon^{-4}(Y_i - \alpha - \beta\mu_x), & \frac{\partial c_i}{\partial \sigma_\epsilon^2} &= -\beta^2\sigma_\epsilon^{-4}, \\
\frac{\partial \ln |\Sigma_i|}{\partial \sigma_\epsilon^2} &= \sigma_\epsilon^{-2} - \beta^2c_i^{-1}\sigma_\epsilon^{-4}, & \frac{\partial c_i}{\partial \sigma_u^2} &= -m_i\sigma_u^{-4}, & \frac{\partial h_i}{\partial \sigma_u^2} &= -\sigma_u^{-4}\sum_{j=1}^{m_i}(X_{ij} - \mu_x) \\
\frac{\partial \ln |\Sigma_i|}{\partial \sigma_u^2} &= m_i\sigma_u^{-2} - m_i\sigma_u^{-4}c_i^{-1}, & \text{e } \frac{\partial t_i}{\partial \sigma_u^2} &= -\sigma_\epsilon^{-4}\sum_{j=1}^{m_i}(X_{ij} - \mu_x)^2, & i &= 1, \dots, n.
\end{aligned}$$

Derivando mais uma vez a expressão em (A.2) em relação a  $\gamma = \alpha, \beta, \mu_x, \sigma_\epsilon^2, \sigma_u^2$ , obtemos a expressão em (A.1), onde

$$\begin{aligned}
\frac{\partial^2 \ln |\Sigma_i|}{\partial \alpha^2} &= \frac{\partial^2 c_i}{\partial \alpha^2} = \frac{\partial^2 h_i}{\partial \alpha^2} = \frac{\partial^2 \ln |\Sigma_i|}{\partial \beta \partial \alpha} = \frac{\partial^2 c_i}{\partial \beta \partial \alpha} = 0, & \frac{\partial^2 t_i}{\partial \alpha^2} &= 2\sigma_\epsilon^{-2}, & \frac{\partial^2 h_i}{\partial \beta \partial \alpha} &= -\sigma_\epsilon^{-2}, \\
\frac{\partial^2 \ln |\Sigma_i|}{\partial \mu_x \partial \alpha} &= \frac{\partial^2 c_i}{\partial \mu_x \partial \alpha} = \frac{\partial^2 h_i}{\partial \mu_x \partial \alpha} = 0, & \frac{\partial^2 t_i}{\partial \beta \partial \alpha} &= 2\sigma_\epsilon^{-2}\mu_x, & \frac{\partial^2 t_i}{\partial \mu_x \partial \alpha} &= 2\sigma_\epsilon^{-2}\beta, \\
\frac{\partial^2 \ln |\Sigma_i|}{\partial \sigma_\epsilon^2 \partial \alpha} &= \frac{\partial^2 \ln |\Sigma_i|}{\partial \sigma_u^2 \partial \alpha} = \frac{\partial^2 c_i}{\partial \sigma_\epsilon^2 \partial \alpha} = \frac{\partial^2 c_i}{\partial \sigma_u^2 \partial \alpha} = \frac{\partial^2 t_i}{\partial \sigma_u^2 \partial \alpha} = \frac{\partial^2 h_i}{\partial \sigma_u^2 \partial \alpha} = 0, & \frac{\partial^2 h_i}{\partial \sigma_\epsilon^2 \partial \alpha} &= \beta\sigma_\epsilon^{-4}, \\
\frac{\partial^2 t_i}{\partial \sigma_\epsilon^2 \partial \alpha} &= 2\sigma_\epsilon^{-4}(Y_i - \alpha - \beta\mu_x), & \frac{\partial^2 \ln |\Sigma_i|}{\partial \beta^2} &= 2\sigma_\epsilon^{-2}c_i^{-1} - 4\beta^2c_i^{-2}\sigma_\epsilon^{-4}, & \frac{\partial^2 t_i}{\partial \beta^2} &= 2\mu_x^2\sigma_\epsilon^{-2}, \\
\frac{\partial^2 h_i}{\partial \beta^2} &= -2\mu_x\sigma_\epsilon^{-2}, & \frac{\partial^2 c_i}{\partial \beta^2} &= 2\sigma_\epsilon^{-2}, & \frac{\partial^2 \ln |\Sigma_i|}{\partial \mu_x \partial \beta} &= \frac{\partial^2 c_i}{\partial \mu_x \partial \beta} = 0, & \frac{\partial^2 h_i}{\partial \mu_x \partial \beta} &= -2\beta\sigma_\epsilon^{-2} \\
\frac{\partial^2 t_i}{\partial \mu_x \partial \alpha} &= -2\sigma_\epsilon^{-2}(Y_i - \alpha - 2\beta\mu_x), & \frac{\partial^2 \ln |\Sigma_i|}{\partial \sigma_u^2 \partial \beta} &= 2m_i\beta c_i^{-2}\sigma_\epsilon^{-2}\sigma_u^{-4}, & \frac{\partial^2 c_i}{\partial \sigma_\epsilon^2 \partial \beta} &= -2\beta\sigma_\epsilon^{-4}, \\
\frac{\partial^2 c_i}{\partial \sigma_u^2 \partial \beta} &= \frac{\partial^2 t_i}{\partial \sigma_u^2 \partial \beta} = \frac{\partial^2 h_i}{\partial \sigma_u^2 \partial \beta} = 0, & \frac{\partial^2 \ln |\Sigma_i|}{\partial \sigma_\epsilon^2 \partial \beta} &= 2\beta c_i^{-1}\sigma_\epsilon^{-4}(\beta^2 c_i^{-1}\sigma_\epsilon^{-2} - 1), \\
\frac{\partial^2 t_i}{\partial \sigma_\epsilon^2 \partial \beta} &= 2\mu_x\sigma_\epsilon^{-4}(Y_i - \alpha - \beta\mu_x), & \frac{\partial^2 h_i}{\partial \sigma_\epsilon^2 \partial \beta} &= -\sigma_\epsilon^{-4}(Y_i - \alpha - \beta\mu_x),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 t_i}{\partial \mu_x^2} &= 2\beta^2 \sigma_c^{-2} + 2m_i \sigma_u^{-2}, & \frac{\partial^2 \ln |\Sigma_i|}{\partial \mu_x^2} &= \frac{\partial^2 h_i}{\partial \mu_x^2} = \frac{\partial^2 c_i}{\partial \mu_x^2} = 0, & \frac{\partial^2 h_i}{\partial \sigma_u^2 \partial \mu_x} &= m_i \sigma_u^{-4}, \\
\frac{\partial^2 t_i}{\partial \sigma_u^2 \partial \mu_x} &= 2\sigma_u^{-4} \sum_{j=1}^{m_i} (X_{ij} - \mu_x), & \frac{\partial^2 \ln |\Sigma_i|}{\partial \sigma_u^2 \partial \mu_x} &= \frac{\partial^2 c_i}{\partial \sigma_u^2 \partial \mu_x} = 0, & \frac{\partial^2 h_i}{\partial \sigma_e^2 \partial \mu_x} &= \beta^2 \sigma_e^{-4}, \\
\frac{\partial^2 t_i}{\partial \sigma_e^2 \partial \mu_x} &= 2\beta \sigma_e^{-4} (Y_i - \alpha - \beta \mu_x), & \frac{\partial^2 \ln |\Sigma_i|}{\partial \sigma_e^2 \partial \mu_x} &= \frac{\partial^2 c_i}{\partial \sigma_e^2 \partial \mu_x} = 0, & \frac{\partial^2 c_i}{\partial (\sigma_u^2)^2} &= 2m_i \sigma_u^{-6}, \\
\frac{\partial^2 \ln |\Sigma_i|}{\partial (\sigma_u^2)^2} &= -m_i \sigma_u^{-4} (1 - 2\sigma_u^{-2} c_i^{-1} + m_i \sigma_u^{-4} c_i^{-2}), & \frac{\partial^2 t_i}{\partial (\sigma_u^2)^2} &= 2\sigma_u^{-6} \sum_{j=1}^{m_i} (X_{ij} - \mu_x)^2, \\
\frac{\partial^2 h_i}{\partial (\sigma_u^2)^2} &= 2\sigma_u^{-6} \sum_{j=1}^{m_i} (X_{ij} - \mu_x), & \frac{\partial^2 \ln |\Sigma_i|}{\partial \sigma_e^2 \partial \sigma_u^2} &= -m_i \beta^2 c_i^{-2} \sigma_e^{-4} \sigma_u^{-4}, \\
\frac{\partial^2 h_i}{\partial \sigma_e^2 \partial \sigma_u^2} &= \frac{\partial^2 c_i}{\partial \sigma_e^2 \partial \sigma_u^2} = \frac{\partial^2 t_i}{\partial \sigma_e^2 \partial \sigma_u^2} = 0, & \frac{\partial^2 t_i}{\partial (\sigma_e^2)^2} &= 2\sigma_e^{-6} (Y_i - \alpha - \beta \mu_x), \\
\frac{\partial^2 \ln |\Sigma_i|}{\partial (\sigma_e^2)^2} &= -\sigma_e^{-4} + 2\beta^2 c_i^{-1} \sigma_e^{-6} - \beta^4 c_i^{-2} \sigma_e^{-8}, & \frac{\partial^2 c_i}{\partial (\sigma_u^2)^2} &= 2\beta^2 \sigma_e^{-6}, \quad e \\
\frac{\partial^2 h_i}{\partial (\sigma_e^2)^2} &= 2\beta \sigma_e^{-6} (Y_i - \alpha - \beta \mu_x), & i &= 1, \dots, n.
\end{aligned}$$



## Apêndice B

# Informação de Fisher e Reparametrizações

### B.1 Reparametrizações

Seja  $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$  um modelo paramétrico, onde o objeto  $\theta$ , o conjunto  $\Theta$  e a função que liga  $P_\theta$  a  $\theta$  se denominam parâmetro, espaço paramétrico e parametrização respectivamente.

O método de máxima verossimilhança é invariante sob reparametrizações, isto é.

Sejam  $(P_\theta^1, \theta \in \Theta)$  e  $(P_\lambda^2, \lambda \in \Lambda)$ , com  $P_{g(\theta)}^2 = P_\theta^1$ , dos parametrizações identificáveis do modelo paramétrico  $\mathcal{P}$ . Seja  $g$  uma transformação bijetiva de  $\Theta$  em  $\Lambda$ , de maneira que  $\lambda = g(\theta)$  define uma reparametrização. Além disso, seja  $\hat{\mathbf{P}}$  o estimador de máxima verossimilhança de  $\mathbf{P} \in \mathcal{P}$ . Então os estimadores de máxima verossimilhança ficam determinados por  $P_{\hat{\theta}}^1 = \hat{\mathbf{P}}$  e  $P_{\hat{\lambda}}^2 = \hat{\mathbf{P}}$ , implicando que  $P_{g(\hat{\theta})}^2 = \hat{\mathbf{P}}$ . A identificabilidade da parametrização implica então que  $\hat{\lambda} = g(\hat{\theta})$ .

### B.2 Informação de Fisher

A informação de Fisher não é invariante sob reparametrizações, por isso é necessário ser cuidadoso na sua interpretação.

Seja  $\lambda = g(\theta)$  com  $g$  inversível e continuamente diferenciável. Usaremos  $L_1$  e  $L_2$  para distinguir entre as funções de verossimilhança escritas em termos de  $\theta$  e  $\lambda$  respectivamente. Consideremos a igualdade

$$L_2(\mathbf{Z}, g(\theta)) = L_1(\mathbf{Z}, \theta),$$



tomando logaritmo e pela regra da cadeia temos:

$$\begin{aligned}\frac{\partial \log L_1}{\partial \boldsymbol{\theta}} &= \frac{\partial \log L_2}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} \\ &= \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} \frac{\partial \log L_2}{\partial \boldsymbol{\lambda}}.\end{aligned}$$

Em consequência tomando variância segue que

$$\text{Var} \left( \frac{\partial \log L_1}{\partial \boldsymbol{\theta}} \right) = \left( \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} \right) \text{Var} \left( \frac{\partial \log L_2}{\partial \boldsymbol{\lambda}} \right) \left( \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} \right)^T.$$

portanto

$$\mathcal{F}^{(1)}(\boldsymbol{\theta}) = \left( \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} \right) \mathcal{F}^{(2)}(\boldsymbol{\lambda}) \left( \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} \right)^T,$$

onde  $\mathcal{F}^{(1)}$  e  $\mathcal{F}^{(2)}$  é a informação de Fisher correspondente ao vetor de parâmetros  $\boldsymbol{\theta}$  e  $\boldsymbol{\lambda}$  respectivamente.

Assim, a informação de Fisher para o vetor de parâmetros  $\boldsymbol{\lambda}$  fica como segue

$$\mathcal{F}^{(2)}(\boldsymbol{\lambda}) = \left( \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\lambda}} \right) \mathcal{F}^{(1)}(\boldsymbol{\theta}) \left( \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\lambda}} \right)^T,$$

onde  $J = \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\lambda}}$  é conhecido como o Jacobiano da transformação.

---

## Apêndice C

# Modelagem Bayesiana

### C.1 Métodos MCMC

Em geral, em modelos ligeiramente complexos, é comum o núcleo da densidade a posteriori não apresentar forma conhecida e, portanto, torna-se necessário a utilização de métodos numéricos para aproximar a distribuição a posteriori de interesse. Nas duas últimas décadas a inferência bayesiana vem experimentando um grande avanço devido à introdução de técnicas de simulação estocástica que permitem, de forma relativamente simples, obter amostras da distribuição objetivo.

Suponha que temos uma distribuição  $\pi(\boldsymbol{\theta})$ ,  $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^p$  (em geral a posteriori) que é difícil de ser amostrada, já seja porque o número de parâmetros  $p$  é muito grande, ou porque a própria distribuição é intratável, então uma solução para o problema seria obtida se pudéssemos construir uma cadeia de Markov homogênea e aperiódica com espaço de estados  $\Omega$  e distribuição estacionária igual a  $\pi(\boldsymbol{\theta})$ , e daí então amostrar dessa cadeia. Isto é porque se uma cadeia de Markov é irreduzível e aperiódica a distribuição estacionária (ou distribuição invariante) é única e coincide com a distribuição limite (ou distribuição de equilíbrio), portanto, para  $n$  suficientemente grande a observação gerada na  $n$ -ésima iteração (da cadeia) essencialmente é uma observação da distribuição de  $\pi(\boldsymbol{\theta})$  (assim como todas as observações geradas a partir da  $(n+1)$ -ésima iteração).

Para implementar a situação acima temos na literatura os algoritmos denominados métodos de Monte Carlo via Cadeias de Markov (MCMC). Um MCMC, para simular de uma densidade  $\pi(\cdot)$ , é qualquer método que produza uma cadeia de Markov homogênea, ergódica e irreduzível, cuja distribuição estacionária seja a distribuição de interesse ( $\pi(\cdot)$ ). Em geral,

nossa distribuição objetivo será a distribuição a posteriori de  $\theta$ .

Existe na literatura uma quantidade muito grande de algoritmos do tipo MCMC (Robert and Casella (1999) e Chen et al. (2000)). Os algoritmos mais conhecidos são o amostrador de Gibbs proposto por Geman and Geman (1984) e popularizado na comunidade estatística por (Gelfand and Smith (1990)) e o algoritmo de Metropolis-Hastings inicialmente proposto por Metropolis et al. (1953) e estendido por Hastings (1970). Muitos autores na literatura, têm estabelecido condições para a convergência de vários desses algoritmos MCMC, Roberts and Smith (1993) por exemplo estabelecem condições suficientes relativamente simples sobre  $\pi(\theta)$  de modo que as cadeias induzidas pelos algoritmos de Gibbs e Metropolis-Hastings (com distribuição estacionária  $\pi(\theta)$ ) sejam irredutíveis e aperiódicas e portanto converjam em distribuição para uma variável com distribuição  $\pi(\theta)$ . Roberts and Tweedie (1996) mostram a ergodicidade geométrica de uma classe bastante ampla de algoritmos tipo Metropolis-Hastings. Para mais detalhes sobre métodos MCMC, uma boa referência é Gamerman and Lopes (2006) e Congdon (2006).

## C.2 Programa WinBUGS

O programa computacional WinBUGS (Bayesian inference Using Gibbs Sampler for Windows) é direcionado para análise bayesiana de modelos estatísticos complexos, utilizando técnicas de MCMC. Consiste de um conjunto de funções que permite a especificação do modelo e das distribuições de probabilidade para todos os seus componentes aleatórios (observações e parâmetros).

O WinBUGS é dotado da capacidade de reconhecer formas conjugadas, log-concavidade, distribuições com amplitudes restritas, etc. Com base nesta classificação, o algoritmo de amostragem mais eficiente é selecionado para a simulação. Se nenhuma destas propriedades é reconhecida, uma mensagem avisa a incapacidade de escolher o método de atualização. Por exemplo, quando formas conjugadas condicionais são reconhecidas a amostra é eficientemente gerada das distribuições condicionais completas. Se não existir conjugação, as amostras são geradas por métodos de rejeição, de rejeição adaptivo ou pelo algoritmo Metropolis-Hastings, conforme descrito em Gamerman and Lopes (2006). O método para amostragem é escolhido seguindo uma hierarquia. Em cada caso um método é usado se nenhum outro anterior for apropriado.

---

A linguagem do sistema e a entrada e saída de dados obedecem a mesma sintaxe da linguagem de programação S-Plus ou R. O WinBUGS permite, também, que o modelo seja especificado a partir de uma representação gráfica (Doodle). Se desejado, a linguagem gráfica pode ser transformada para linguagem texto, mais flexível do que a primeira.

Como grandes grandes atrativos para o uso do WinBUGS podemos citar: a possibilidade da especificação de uma variedade de distribuições a priori e a geração de amostras das densidades condicionais completas por diferentes métodos, a versão 1.4 já permite a especificação de distribuições multivariadas; a amostragem por bloco e a especificação gráfica do modelo; o monitoramento de qualquer função dos parâmetros do modelo; a obtenção dos valores amostrados de cada parâmetro monitorado, a cada  $M$  iterações a partir de uma determinada iteração  $L$  (os valores  $M$  e  $L$  são especificados pelo usuário); o fornecimento automaticamente de resumos decorrentes da amostra obtida (média, mediana, desvio padrão e intervalo de credibilidade); resultados para diagnósticos de convergência; e outras facilidades como a construção imediata de gráficos para a análise de cadeia.

É importante ressaltar que o WinBUGS é um programa de livre acesso e pode ser rodado, também, na linguagem de programação R, através do pacote R2WinBUGS (Sturtz et al. (2005)).

---

## Referências Bibliográficas

- Arellano-Valle, R. B. and Bolfarine, H. (1998). Weak differential measurement error models. *Statistics and Probability Letters*, **40**:279–287. 2
- Arellano-Valle, R. B., Bolfarine, H., and Gasco, L. (2002). Measurement Error Models with Nonconstant Covariance Matrices. *Journal of Multivariate Analysis*, **82**:395–415. 14, 37
- Boyles, R. A. (1983). On the convergence of EM algorithm. *Journal of the Royal Statistical Society, B*, **45**:47–50. 18
- Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models*. Chapman & Hall/CRC, Boca Raton, second edition. 2
- Chen, M. H., Shao, Q. M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag. 122
- Cheng, C. and Van Ness, J. (1999). *Statistical Regression with Measurement Error*. Arnold Publishers, London. 2, 9, 13
- Cheng, C.-L. and Riu, J. (2006). On Estimating Linear Relationships When Both Variables Are Subject to Heteroscedastic Measurement Errors. *Technometrics*, **48**(4):1089–1101. 3
- Congdon, P. (2006). *Bayesian Statistical Modelling*. John Wiley & Sons, England, second edition. 122
- de Castro, M., Galea, M., and Bolfarine, H. (2008). Hypothesis testing in an errors-in-variables model with heteroscedastic measurement errors. *Statistics in Medicine*, **27**:5217–5234. 3, 29, 65, 66
- Dempster, A. P., Laird, N. M., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, B*, **39**:1–38. 17, 18
- Dolby, G. R. (1976). The ultrastructural relation: A synthesis of the functional and structural relations. *Biometrika*, **63**:39–50. 50
- Fuller, W. A. (1987). *Measurement Error Models*. John Wiley, New York. 2, 3, 9
- Galea-Rojas, M., de Castilho, M. V., H. Bolfarine, H., and de Castro, M. (2003). Detection of analytical bias. *Analyst*, **128**:1070–1081. 3



- Gamerman, D. and Lopes, H. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman, second edition. 91, 122
- Gelfand, A. E. and Sahu, S. K. (1999). Identifiability improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, **94**:247–253. 92
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**:398–409. 122
- Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**:721–741. 122
- Gleser, L. J. (1992). The Importance of Assessing Measurement Reliability in Multivariate Regression. *Journal of the American Statistical Association*, **87**:696–707. 78
- Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, **57**:97–109. 122
- Hocking, R. R. (1996). *Methods and Applications of Linear Models*. Wiley-Interscience. 71
- Johnson, N., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, volume 1. John Wiley & Sons, New York, second edition. 53
- Kulathinal, S. B., Kuulasmaa, K., and Gasbarra, D. (2002). Estimation of an errors-in-variables regression model when the variances of the measurement errors vary between the observations. *Statistics in Medicine*, **21**:1089–1101. 3, 5, 6, 9, 13, 28, 50, 113
- Lange, K. L., Little, R. J., and Taylor, J. (1989). Robust statistical modelling using the t-distribution. *Journal of the American Statistical Association*, **84**:881–896. 53
- Magnus, J. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester, third edition. 14, 56, 57, 58
- Mardia, K. V., Kent, J. T., and Bibby, J. M. (1989). *Multivariate Analysis*. Academic Press Limited, seventh edition. 20
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. W., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machine. *Journal of Chemical Physics*, **21**:1087–1091. 122
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons. 19, 44, 52
- Ntzoufras, I. (2009). *Bayesian Modelling Using WinBUGS*. John Wiley & Sons, New Jersey. 87
- Patriota, A. G. (2010). *Modelos heterocedastico com erros nas variaveis*. PhD thesis, IME-USP. 3

- Patriota, A. G. and Bolfarine, H. (2008). A Heteroscedastic Polynomial Regression with Measurement Error in Both Axes. *The Indian Journal of Statistics*, **70-B**:267–282. 3
- Patriota, A. G., Bolfarine, H., and de Castro, M. (2009). A heteroscedastic structural errors-in-variables model with equation error. *Statistical Methodology*, **6**:408–423. 3
- R. Development Core Team (2011). *R: A language and environment for statistical computing*. R. Foundation for Statistical Computing, Vienna, Austria. 1
- Ripley, B. D. and Thompson, M. (1987). Regression techniques for the detection of analytical bias. *Analyst*, **112**:377–383. 3
- Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag. 122
- Roberts, G. O. and Smith, A. F. M. (1993). Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithm. *Stochastic Processes and their Applications*, **49**:207–216. 122
- Roberts, G. O. and Tweedie, R. I. (1996). Geometric Convergence and Central Limit Theorems for Multidimensional Hastings and Metropolis Algorithms. *Biometrika*, **83**:95–110. 122
- Sen, P. K. and Singer, J. M. (1993). *Large sample methods in statistics: an introduction with applications*. Chapman & Hall. 16, 27, 38, 41
- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2007). Winbugs user manual: Version 1.4.3. Technical report, Cambridge: Medical Research Council Biostatistics. 87
- Sprent, P. (1990). Some history of functional and structural relationships. *American Mathematical Society*, **112**:3–15. 2
- Sturtz, S., Ligges, U., and Gelman, A. (2005). R2winbugs: A Package for Running winbugs from R. *Journal of Statistical Software*, **12**:1–16. 123
- Wilson, E. B. and Hilferty, M. M. (1931). The Distribution of Chi-Square. *Proceedings of the National Academy of Science*, **17**:684–688. 53
- Wu, C. F. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, **11**:95–103. 18