PARAMETRIZAÇÃO
ORTOGONAL E OUTRAS
CONTRIBUIÇÕES AO
MODELO COM ERRO NAS
VARIÁVEIS

João Agnaldo do Nascimento

TESE DEFENDIDA
NO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE DOUTOR
EM
ESTATÍSTICA

Área de concentração: Estatística
Orientador: Prof. Dr. Heleno Bolfarine
Co-orientador: Prof. Dr. Josemar Rodrigues

– São Paulo, Março de 1994 –

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

**PARAMETRIZAÇÃO ORTOGONAL E
OUTROS PROCEDIMENTOS NO MODELO
COM ERROS NAS VARIÁVEIS**

JOÃO AGNALDO DO NASCIMENTO

Tese de Doutorado
submetida ao Instituto de
Matemática e Estatística
da Universidade de São Paulo

Orientador: Prof. Dr. Heleno Bolfarine
Co-orientador: Prof. Dr. Josemar Rodrigues

São Paulo - 1994

PARAMETRIZAÇÃO ORTOGONAL

E OUTROS PROCEDIMENTOS NO MODELO COM ERROS NAS VARIÁVEIS

Este exemplar corresponde à redação
final da tese devidamente corrigida e defendida
por João Agnaldo do Nascimento e aprovada
pela comissão julgadora.

São Paulo, 5 de Março de 1994.

Banca examinadora:

- Prof. Dr. Heleno Bolfarine (orientador) - IME-USP
- Prof. Dr. José Galvão Leite - IME-USP
- Prof. Dr. Gauss Moutinho Cordeiro - UFPE
- Profa. Dra. Edna Miazaki - UNB
- Prof. Dr. Jorge Alberto Achcar - ICMSC-USP

*Aos meus filhos : James e Janine,
dedico este trabalho.*

Batei e a porta se abrirá. Procurai e achareis. Pedi e
vos será dado.

Agradecimentos

◇ À minha namorada Andressa pelo carinho e apoio em momentos difíceis.

◇ Aos colegas do DE/UFPB que apoiaram a minha permanência no IME-USP.

◇ Aos colegas do IME/USP, em especial à turma do futebol, à Letícia, Vitor, Pilar, Juan e Regina.

◇ À CAPES pelo apoio financeiro.

◇ Ao Prof. Dr. José Galvão Leite, por suas valiosas sugestões e discussões importantes relacionadas com vários aspectos deste trabalho.

◇ Ao Prof. Dr. Heleno Bolfarine e Prof. Dr. Josemar Rodrigues, pela competente orientação e co-orientação deste trabalho.

◇ À Dioclécia (minha mãe), Josinete, Marinho e irmãos pelo apoio recebido.

◇ E a todos que contribuíram de uma forma direta ou indireta.

Resumo

Parte deste trabalho é dedicado a utilização de transformações ortogonais no modelo com erro nas variáveis (tanto estrutural como funcional) para obtenção de inferências sobre o coeficiente angular do modelo que é geralmente o parâmetro de interesse. São utilizados procedimentos clássicos e Bayesianos. Em particular, as transformações ortogonais no modelo estrutural simplificam a implementação de estatísticas da razão de verossimilhança (com e sem correções de Bartlett) e de Wald para testar hipóteses relacionadas com o coeficiente angular do modelo. Procedimentos Bayesianos, como posteriores exatas e aproximadas para o coeficiente angular do modelo são também consideradas no modelo estrutural e funcional após uma transformação ortogonal dos parâmetros.

Procedimentos Bayesianos baseados no fator de Bayes a posteriori para a comparação de vários modelos estruturais são investigados.

Uma alternativa para a utilização das transformações ortogonais no modelo estrutural é viabilizada pelo uso do algoritmo EM.

Este algoritmo tanto pode ser utilizado para a obtenção dos estimadores de máxima verossimilhança do modelo, como também para a previsão dos valores reais e desconhecidos da variável auxiliar, que não são avaliados diretamente pelo processo de medição.

Abstract

Part of this work is dedicated to the use of orthogonal transformations in error in variables models (structural or functional) aiming at inferences about the slope parameter of the model, typically the parameter of interest. Classical and Bayesian approaches are considered. In particular, orthogonal transformations of the parameters in the model simplify the implementations of likelihood ratio statistics (with or without Bartlett corrections) and Wald statistics for testing hypotheses concerning the slope parameter. Bayesian procedures such as exact and approximate posterior distributions, are derived for the slope parameter β after an orthogonal transformation of the parameters.

Procedures based on the posterior Bayes factor for comparing several structural models are developed, leading to some new statistics.

An alternative to orthogonal transformation in structural models is available by considering the EM algorithm. This algorithm can be used to derive maximum likelihood estimators of the unk-

now parameters of the model and also to predict the unobserved real values of the auxiliary variables, which can not be observed directly by the measuring instrument.

Conteúdo

1	Introdução	1
1.1	O modelo funcional	3
1.2	O modelo estrutural	7
1.3	Considerações finais	10
1.4	Perspectiva de pesquisa futura	12
2	Parametrização ortogonal no MEV	13
2.1	Introdução	13
2.2	O MEV estrutural. Uma população	14
2.3	O MEV estrutural. Vários modelos	21
2.4	O MEV funcional	25
2.4.1	A transformação polar	25
2.4.2	Uma parametrização ortogonal	27
3	Testes da razão de verossimilhança, correções de Bartlett e testes de Wald no modelo estrutural	30
3.1	Introdução	30
3.2	Uma população com razão de variâncias ϕ conhecida	31
3.2.1	Variância σ^2 conhecida	32

3.2.2	Variância σ^2 desconhecida	33
3.2.3	A função de verossimilhança perfilada condicional	34
3.3	Uma população com coeficiente de atenuação k_x conhecido	37
3.3.1	Parâmetros μ e α conhecidos	38
3.3.2	Parâmetros μ e α desconhecidos	39
3.4	O fator de correção de Bartlett no caso k_x conhecido	41
3.4.1	Parâmetros μ e α conhecidos	43
3.4.2	Parâmetros μ e α desconhecidos	45
3.5	O teste de Wald para uma população	47
3.5.1	Variância σ^2 conhecida	47
3.5.2	Variância σ^2 desconhecida	48
3.5.3	Coeficiente de atenuação conhecido	48
3.6	Inferência para s populações com razão de variâncias $\phi_i, i = 1, \dots, s$, conhecidas	54
3.6.1	Variâncias $\sigma_i^2, i = 1, \dots, s$, conhecidas	54
3.6.2	Variâncias $\sigma_i^2, i = 1, \dots, s$, desconhecidas	59
3.7	Inferência para s populações com coeficientes de atenuação $k_{x_i}, i =$ $1, \dots, s$, conhecidos	64
3.7.1	Parâmetros μ_i e α_i conhecidos, $i = 1, \dots, s$	64
3.7.2	Parâmetros μ_i e α_i desconhecidos, $i = 1, \dots, s$	73
3.8	Alguns Modelos particulares	82
3.9	Conclusões	86
4	Inferência Bayesiana no MEV	94
4.1	Introdução	94

4.2	Variância σ^2 conhecida	95
4.3	Variância σ^2 desconhecida	100
4.4	Coeficiente de atenuação k_x conhecido	106
4.5	Inferência Bayesiana no modelo funcional	108
4.6	Inferência Bayesiana no modelo funcional com a transformação polar	112
4.6.1	Variância σ^2 conhecida com parâmetro $\alpha = 0$	113
4.6.2	Variância σ^2 desconhecida	114
4.7	O Fator de Bayes a posteriori	115
4.7.1	Introdução	115
4.7.2	A estatística da razão de verossimilhança e o FBP	117
4.8	Comparação de s modelos estruturais via fator de Bayes a posteriori	118
4.8.1	Coeficiente de atenuação k_{x_i} conhecidos	118
4.8.2	Razão de variâncias ϕ_i , $i = 1, \dots, s$ conhecidas	122
5	Previsão do valor real	128
5.1	Introdução	128
5.2	O algoritmo EM	129
5.2.1	O Modelo Estrutural com uma variável regressora	130
5.2.2	Um modelo com s variáveis regressoras	138
5.2.3	O modelo com réplicas	143
5.3	Um enfoque Bayesiano	148
5.4	Estimação de máxima verossimilhança.	151
A	Identities, prior MDIP e parametrização ortogonal	157
A.1	Identities	157
A.1.1	Identidade 1A	157

A.1.2	Identidade 2A	159
A.1.3	Identidade 3A	160
A.2	"Maximal data information priori"	160
A.2.1	Priori MDIP para o MEV estrutural e funcional	162
A.3	O método de Laplace para aproximação de Integrais	163
A.4	Parâmetros ortogonais	164
A.4.1	Construção	165
B	Programas computacionais	167
B.0.2	Programa arazão	167
B.0.3	Programa fabe	169
B.0.4	Programa fabek	172
B.0.5	Programa Sibet1	175
B.0.6	Programa Sibet2	178
B.0.7	Programa Sieswv	182
B.0.8	Programa Sies	183
B.0.9	Programa Sifuex	185
B.0.10	Programa Hbsm22	186
B.0.11	Programa Hbsm60	188

Capítulo 1

Introdução

Os modelos considerados neste trabalho são generalizações do modelo de regressão, onde duas variáveis x_i e y_i estão relacionadas através da equação

$$y_i = \alpha + \beta x_i, \quad (1.1)$$

$i = 1, \dots, n$, com α e β sendo parâmetros desconhecidos. Assumimos que os pares (x_i, y_i) são desconhecidos ou seja, nem y_i , nem x_i são observados diretamente, de modo que os valores observados X_i e Y_i obedecem às relações

$$X_i = x_i + u_i, \quad (1.2)$$

$$Y_i = y_i + e_i, \quad (1.3)$$

$i = 1, \dots, n$, onde $e_i \sim \mathcal{N}(0, \sigma_e^2)$ e $u_i \sim \mathcal{N}(0, \sigma_u^2)$. O modelo descrito por (1.1), (1.2) e (1.3) é usualmente conhecido por modelo com erro nas variáveis (MEV), e leva em conta o fato de que toda medida geralmente contém erro, quer seja este provocado pelo instrumento de medição ou pelo operador.

A estimação de β no MEV descrito acima, foi iniciada por Adcock (1877), assumindo que (u_i, e_i, x_i) são independentes com $u_i \sim \mathcal{N}(0, \sigma_u^2)$, $e_i \sim \mathcal{N}(0, \sigma_e^2)$,

$i = 1, \dots, n$, e σ_u^2 e σ_e^2 conhecidos. A técnica utilizada foi a de mínimos quadrados ortogonais (Casella e Berger, 1990), onde se minimiza a distância ortogonal (euclidiana) dos pontos observados à reta estimada com este critério. A extensão da solução de Adcock (1877) para o caso em que a razão de variâncias $\phi = \sigma_e^2/\sigma_u^2$ é suposta conhecida foi considerada em Kummel (1879). Pearson (1901) estendeu os resultados de Adcock (1877 e 1878) para o caso multivariado no sentido de que várias variáveis independentes estão presentes no modelo. Outros estimadores de β foram propostos. Dent (1935) propôs o estimador "média geométrica" para β , o qual é dado por

$$\hat{\beta} = \text{sgn}\{S_{xy}\} \left(\frac{S_{yy}}{S_{xx}} \right)^{1/2},$$

com $S_{xx} = \sum_{i=1}^n (X_i - \bar{X})^2/n$, $S_{yy} = \sum_{i=1}^n (Y_i - \bar{Y})^2/n$, $S_{xy} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})/n$ e $\text{sgn}\{.\}$ significa o sinal de $\{.\}$. Este estimador não consistente (em probabilidade) foi bastante utilizado na área de pesca (Sprenst, 1990) e ignora qualquer suposição adicional sobre σ_e^2 e σ_u^2 .

Wald (1940) introduziu o método do agrupamento, que produz estimadores consistentes para β . Este modelo foi posteriormente estendido por Bartlett (1949) e por Hooper e Theil (1958). Lindley (1947) estabelece condições para que a linearidade entre as variáveis não observáveis

$$E(y|x) = \beta_0 + \beta_1 x$$

implique na linearidade entre as variáveis observáveis, ou seja,

$$E(Y|X) = \phi_0 + \phi_1 X,$$

para algum ϕ_0 e ϕ_1 .

Logo depois da segunda guerra mundial surgiram artigos fundamentais nesta área. Podemos citar dentre outros, Lindley (1947), Geary (1949), Reiersol (1950),

Neymann e Scott (1948, 1951) e Kiefer e Wolfowitz (1956), que estudaram principalmente os problemas de falta de identificabilidade dos modelos e também a inconsistência dos estimadores usuais. Métodos Bayesianos para a estimação de β são apresentadas em Lindley (1966) e Lindley e El-Sayad (1968). Revisões importantes da literatura aparecem em Madansky (1959), Moran (1970), Nascimento (1983) e Anderson (1984). Fuller (1987) num texto recente e Kendall e Stuart (1979) tratam do problema com bastante detalhes. Vários títulos foram utilizados para o modelo definido por (1.1)-(1.3), dentre os quais podemos citar "regressão com erros em x ", "regressão com erros em ambas variáveis" e "modelos com erros nas variáveis". Neste texto vamos utilizar somente a nomenclatura: modelo funcional e modelo estrutural. Note que podemos reescrever o modelo dado por (1.1), (1.2) e (1.3) como

$$Y_i = \alpha + \beta X_i + (e_i - \beta u_i),$$

ou seja,

$$Y_i = \alpha + \beta X_i + v_i, \quad (1.4)$$

com $v_i = e_i - \beta u_i$, $i = 1, \dots, n$. Apesar de (1.4) ser similar aos modelos de regressão usuais, não é possível estimar β , de forma consistente, pelo método dos mínimos quadrados ordinários.

1.1 O modelo funcional

O modelo definido por (1.1), (1.2) e (1.3) com a suposição

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right)$$

onde x_i , é considerado como constante $i = 1, \dots, n$, é chamado de modelo funcional.

Os $n + 4$ parâmetros desse modelo, são

$$\alpha, \beta, \sigma_e^2, \sigma_u^2, x_1, \dots, x_n,$$

de modo que o número de parâmetros cresce com o tamanho da amostra. Portanto, o logaritmo da função de verossimilhança correspondente a uma amostra de tamanho n do modelo com as suposições acima é dado por

$$\ell = -\frac{n}{2} \log \sigma_e^2 - \frac{n}{2} \log \sigma_u^2 - \frac{1}{2\sigma_u^2} \sum_{i=1}^n (X_i - x_i)^2 - \frac{1}{2\sigma_e^2} \sum_{i=1}^n (Y_i - (\alpha + \beta x_i))^2. \quad (1.5)$$

Uma característica da verossimilhança em (1.5) é que ela é ilimitada. Para ver isto, basta tomar $X_i = x_i$, $i = 1, \dots, n$ e fazer $\sigma_u^2 \rightarrow 0$. O valor da função vai para ∞ , de modo que não existe solução de máxima verossimilhança. Solari (1969) mostra que se as equações que definem a primeira derivada de ℓ em (1.5) são igualadas a zero e resolvidas, o resultado é um ponto de sela e não de máximo. Zellner (1971) estuda o problema da não existência do máximo da função de verossimilhança na região do espaço paramétrico, mostrando que a solução para β não pertence à região admissível do espaço paramétrico que é dada por $\frac{|Cov(X, Y)|}{\sigma_X} \leq |\beta| \leq \frac{\sigma_Y}{|Cov(X, Y)|}$, com $Cov(X, Y) \neq 0$. Além disso, o modelo funcional não é uma decomposição ortogonal de Y , pois $Cov(v_i, X_i) = -\beta\sigma_u^2$, onde $v_i = e_i - \beta u_i$, $i = 1, \dots, n$.

Suposições adicionais são portanto necessárias para tornar factível o problema de estimação no modelo funcional. Algumas das suposições comumente adotadas são: (i) σ_e^2 e σ_u^2 conhecido (ii) σ_e^2 ou σ_u^2 conhecido e (iii) $\phi = \sigma_e^2/\sigma_u^2$ conhecido. A suposição (iii) é a mais comum e adotada por muitos autores. Veja, por exemplo, Lindley e El Sayad (1968), Zellner (1971) e Kendall e Stuart (1970). No caso em que σ_e^2 e σ_u^2 são conhecidos, derivando ℓ em (1.5) com relação a α , β e x_i , $i = 1, \dots, n$,

obtemos os seguintes estimadores de máxima verossimilhança (Barnett, 1969):

$$\begin{aligned}\hat{\alpha} &= \bar{Y} - \hat{\beta}\bar{X}, \\ \hat{x}_i &= \frac{\hat{\beta}(Y_i - \hat{\alpha}) + \phi X_i}{\hat{\beta}^2 + \phi}, \quad i = 1, \dots, n, \\ \hat{\beta} &= \frac{S_{yy} - \phi S_{xx} + \sqrt{(S_{yy} - \phi S_{xx})^2 + 4\phi S_{xy}^2}}{2S_{xy}},\end{aligned}\tag{1.6}$$

onde S_{xx} , S_{yy} e S_{xy} foram definidos anteriormente. Note que o estimador de máxima verossimilhança de β , $\hat{\beta}$, depende de σ_e^2 e σ_u^2 somente através de ϕ .

Considerando $\alpha = 0$, Sprent (1966) introduziu o método dos mínimos quadrados generalizados para estimar β , que consiste em minimizar

$$F(\mathbf{x}, \beta) = \sigma_e^{-2} \sum_{i=1}^n (Y_i - \beta x_i)^2 + \sigma_u^{-2} \sum_{i=1}^n (X_i - x_i)^2,$$

com $\mathbf{x} = (x_1, \dots, x_n)$. Após algumas manipulações algébricas, podemos escrever

$$F(\mathbf{x}, \beta) = \frac{\sum_{i=1}^n (Y_i - \beta X_i)^2}{\sigma_u^2 \beta^2 + \sigma_e^2} + (\sigma_e^{-2} \beta^2 + \sigma_u^{-2}) \sum_{i=1}^n (x_i - \hat{x}_{i,\beta})^2,\tag{1.7}$$

onde

$$\hat{x}_{i,\beta} = \frac{\beta \sigma_e^2 Y_i + \sigma_u^2 X_i}{\beta^2 \sigma_u^2 + \sigma_e^2},\tag{1.8}$$

$i = 1, \dots, n$. Segue portanto que a minimização de $F(\mathbf{x}, \beta)$ com relação à β , consiste na minimização de

$$\begin{aligned}F(\hat{\mathbf{x}}_\beta, \beta) &= \frac{\sum_{i=1}^n (Y_i - \beta X_i)^2}{\sigma_u^2 \beta^2 + \sigma_e^2} \\ &= \gamma + \frac{(S_{xx} - \gamma \sigma_u^2)}{\beta^2 \sigma_u^2 + \sigma_e^2} (\beta - \hat{\beta})^2,\end{aligned}\tag{1.9}$$

onde $\hat{\mathbf{x}}_\beta = (\hat{x}_{1,\beta}, \dots, \hat{x}_{n,\beta})$,

$$\hat{\beta} = \frac{S_{xy}}{S_{xx} - \gamma \sigma_u^2},\tag{1.10}$$

onde, como $\alpha = 0$, $S_{xx} = \sum_{i=1}^n n^{-1} X_i^2$, $S_{yy} = \sum_{i=1}^n n^{-1} Y_i^2$, $S_{xy} = \sum_{i=1}^n n^{-1} X_i Y_i$ e γ o menor autovalor de

$$\mathbf{B} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{pmatrix}, \quad (1.11)$$

ou seja,

$$\gamma = \frac{\sigma_u^2 S_{yy} + \sigma_e^2 S_{xx} - \sqrt{(\sigma_u^2 S_{yy} + \sigma_e^2 S_{xx})^2 - 4\sigma_e^2 \sigma_u^2 (S_{xx} S_{yy} - S_{xy}^2)}}{2\sigma_e^2 \sigma_u^2}.$$

Pode-se mostrar que $\hat{\beta}$ em (1.6) pode ser escrito como $\hat{\beta}$ (1.10). Portanto, de (1.8) e (1.9), temos que $F(\mathbf{x}, \beta)$ é mínimo quando x_i é substituído por $\hat{x}_{i,\beta}$ e β por $\hat{\beta}$. Além disso, de (1.9) segue também que o valor mínimo de $F(\mathbf{x}, \beta)$ é dado por

$$\gamma = \frac{\sum_{i=1}^n (Y_i - \hat{\beta} X_i)^2}{\sigma_u^2 \hat{\beta}^2 + \sigma_e^2},$$

que é uma forma alternativa para o cálculo de γ . Distribuições exatas para os estimadores acima são bastante complicadas (Anderson, 1976). Inferência assintótica é também pouco confiável, pois o número de parâmetros (parâmetros incidentais) cresce com n . Vários autores, dentre outros, Neymann e Scott (1948, 1951), Kiefer et al. (1956) e Zellner (1971) apontam que vários problemas podem ocorrer em tais situações. No Capítulo 4 apresentamos procedimentos alternativos baseados no enfoque Bayesiano, que produzem inferências confiáveis para β . No Capítulo 5, consideramos inferência para o parâmetro x_i , onde mostramos que, em grandes amostras,

$$\hat{x}_{i,\hat{\beta}} = \frac{\phi \hat{\beta} Y_i + X_i}{\hat{\beta}^2 + \phi}$$

é um estimador de Bayes com relação à determinada priori não informativa. A variância a posteriori é também determinada.

1.2 O modelo estrutural

O modelo definido por (1.1), (1.2) e (1.3), mas agora com x_i sendo uma variável aleatória, é denominado modelo estrutural. Vamos assumir no presente trabalho que

$$\begin{pmatrix} x_i \\ e_i \\ u_i \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2 & 0 & 0 \\ 0 & \sigma_e^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix} \right), \quad i = 1, \dots, n,$$

de onde segue que x_i , u_i e e_i são independentes e normalmente distribuídos, $i = 1, \dots, n$. Observamos que a nomenclatura funcional e estrutural é usada independentemente da suposição de normalidade. Temos portanto de (1.1), (1.2) e (1.3) que a distribuição conjunta de (X_i, Y_i) com as suposições acima é normal bivariada com vetor de médias e matriz de covariâncias tais que

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu \\ \alpha + \beta\mu \end{pmatrix}, \begin{pmatrix} \sigma_u^2 + \tau^2 & \beta\tau^2 \\ \beta\tau^2 & \beta^2\tau^2 + \sigma_e^2 \end{pmatrix} \right), \quad (1.12)$$

$i = 1, \dots, n$. No modelo estrutural, definido por (1.12), o estimador de máxima verossimilhança de β não pode ser obtido, pois o modelo não é identificável. A não identificação é no sentido de que dois vetores diferentes de parâmetros podem levar a mesma distribuição. Formalmente temos,

Definição 1.3.1 *Seja $Z = (X, Y)$ um vetor de variáveis aleatórias e $F_Z(a, \theta)$, $\theta \in \Theta = \{\theta = (\mu, \alpha, \beta, \tau^2, \sigma_u^2, \sigma_e^2) : -\infty < \mu, \beta < \infty; 0 < \tau^2, \sigma_u^2, \sigma_e^2 < \infty\}$ sua função de distribuição com parâmetro θ quando avaliada em $Z=a$. O parâmetro θ é identificável se para cada $\theta_1, \theta_2 \in \Theta$, com $\theta_1 \neq \theta_2$, então*

$$F_Z(a, \theta_1) \neq F_Z(a, \theta_2)$$

para algum a . Se θ é identificável diz-se que o modelo que gera F_Z é identificável.

O teorema a seguir mostra formalmente que o modelo estrutural definido acima não é identificável.

Teorema 1.2.1 *O MEV estrutural é não identificável.*

Prova: Sejam $\theta_1 = (m_1, \alpha, \beta, \tau^2, \sigma_u^2, \sigma_e^2)$ e $\theta_2 = (m_1, \alpha - \gamma m_1, \beta + \gamma, \beta \tau^2 (\beta + \gamma)^{-1}, \sigma_u^2 + \gamma \tau^2 (\beta + \gamma)^{-1}, \sigma_e^2 - \beta \gamma \tau^2)$ tal que $\gamma < |\beta|$ e $\gamma < \sigma_e^2 \beta^{-1} \tau^{-2}$. O MEV estrutural com vetor de parâmetros θ_2 é descrito por

$$X_i = x_i + u_i,$$

$$Y_i = y_i + e_i$$

e

$$y_i = (\alpha - \gamma m_1) + (\beta + \gamma)x_i,$$

com $m_j = \mu_j$, $j = 1, 2$, onde

$$u_i \sim \mathcal{N}(0, \sigma_u^2 + \gamma \tau^2 (\beta + \gamma)^{-1}), \quad e_i \sim \mathcal{N}(0, \sigma_e^2 - \beta \gamma \tau^2) \text{ e } x_i \sim \mathcal{N}(m_1, \beta \tau^2 (\beta + \gamma)^{-1}),$$

$i = 1, \dots, n$. Então, com $a = (b_1, b_2)$,

$$F_Z(a, \theta_2) = \mathbb{P}_{\theta_2}(X \leq b_1, Y \leq b_2),$$

onde

$$Z = (X, Y) \sim \mathcal{N}_2 \left[(m_1^*, m_2^*) , \begin{pmatrix} c_{xx}^* & c_{xy}^* \\ c_{xy}^* & c_{yy}^* \end{pmatrix} \right].$$

É fácil ver que

$$(m_1^*, m_2^*, c_{xx}^*, c_{xy}^*, c_{yy}^*) = (m_1, m_2, c_{xx}, c_{xy}, c_{yy}),$$

onde " * " é usado para indicar a distribuição de probabilidade sob θ_2 . Segue então que

$$F_Z(a, \theta_2) = (P_{\theta_2}(X \leq b_1, Y \leq b_2) = P_{\theta_1}(X \leq b_1, Y \leq b_2) = F_Z(a, \theta_1),$$

de modo que θ é não identificável. $\square$

Reiersol (1950) elucidou o problema da identificação do modelo estrutural mostrando que sob normalidade de e_i e u_i , a condição necessária e suficiente para que o modelo seja identificável com respeito a β é que x_i não seja normalmente distribuído. Sob o modelo estrutural, temos também que o estimador de mínimos quadrados usual,

$$\hat{\beta}_{MQO} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2},$$

não é consistente, pois sob as suposições acima,

$$S_{xx} \xrightarrow{P} \tau^2 + \sigma^2, \quad S_{yy} \xrightarrow{P} \beta^2 \tau^2 + \sigma^2, \quad S_{xy} \xrightarrow{P} \beta \tau^2, \quad (1.13)$$

de modo que,

$$\hat{\beta}_{MQO} \xrightarrow{P} k_x \beta,$$

onde $k_x = \tau^2 / (\tau^2 + \sigma_u^2)$.

As suposições comumente adotadas para tornar o modelo identificável são como no caso do modelo funcional, ou seja: (i) σ_e^2 e σ_u^2 conhecidos, (ii) σ_e^2 ou σ_u^2 conhecido e (iii) $\phi = \sigma_e^2 / \sigma_u^2$ conhecido. Pode-se também verificar que no caso do modelo estrutural, o conhecimento da razão $k_x = \tau^2 / (\tau^2 + \sigma_u^2)$ torna o modelo identificável. A razão k_x é comumente conhecida na literatura por coeficiente de atenuação ou de confiabilidade (Fuller, 1987). O coeficiente de atenuação representa o quanto da variabilidade de X_i é explicada pela variabilidade de x_i . Note

que se $\sigma_u^2 = 0$, então $k_x = 1$ e $\hat{\beta}_{MQO} \xrightarrow{P} \beta$, que corresponde ao modelo de regressão usual. Fuller (1987) apresenta uma tabela com valores de k_x associados a variáveis selecionadas em pesquisas de natureza social. O maior valor apresentado é $k_x = 0.98$ para a variável sexo e o menor é $k_x = 0,58$, para a variável índice de pobreza, o que é bastante razoável pois respostas sobre sexo não são passíveis de erros tão grandes como podem ser aquelas relativas à índices de pobreza.

Fuller (1987) apresenta procedimentos de estimação de parâmetros no modelo estrutural baseados principalmente em resultados assintóticos associados à distribuição dos estimadores obtidos. Conforme veremos nos Capítulos 3 e 4, inferências mais precisas e também de forma mais simples são obtidas para β utilizando uma reparametrização ortogonal do modelo descrito acima. No Capítulo 5, discutimos a utilização do algoritmo EM como uma forma alternativa de se estimar β . Este método também possibilita a previsão dos valores desconhecidos de x .

1.3 Considerações finais

Em grande parte deste trabalho, discutimos a possibilidade de se considerar parametrização ortogonal (ver definição no apêndice A) do modelo com erro nas variáveis como um método de simplificação de inferências sobre β . Como é bastante conhecido, algumas consequências da ortogonalidade são a simplificação do cálculo dos estimadores e também a independência assintótica do estimador de máxima verossimilhança do parâmetro de interesse com os demais parâmetros (parâmetros "nuisances" ou de perturbação). Para tal fim o Capítulo 2 é dedicado exclusivamente a um estudo da ortogonalidade nos modelos estrutural e funcional. Extensões são apresentadas para resultados que aparecem em Wong (1989), Rodrigues e Cor-

dani (1990) e Bolfarine e Cordani (1992). Consideramos também a possibilidade da utilização da transformação polar como um método de se obter ortogonalidade entre o parâmetro de interesse e o parâmetro "nuisance".

No Capítulo 3, discutimos inferências para β no modelo estrutural, usando os métodos da razão de verossimilhança e também testes de hipóteses baseados na estatística de Wald. Conforme veremos, tais procedimentos são bastante simplificados pela utilização de uma transformação ortogonal dos parâmetros no modelo. O cálculo de correções de Bartlett para a estatística da razão de verossimilhança, que é usualmente bastante complicado (Cordeiro, 1992), torna-se factível em alguns casos quando o parâmetro de interesse é ortogonal aos parâmetros "nuisances".

No Capítulo 4, inferências para β nos modelos estrutural e funcional (após uma transformação ortogonal dos parâmetros) são consideradas através do enfoque Bayesiano, obtendo-se distribuições a posteriori exatas e aproximadas com relação a priori "Maximal data information priori" (MDIP) proposta por Zellner (1977, 1990). A possibilidade da comparação de modelos estruturais via fator de Bayes a posteriori é também investigada.

No Capítulo 5 discutimos a previsão para o valor real x . No caso do modelo estrutural consideramos a possibilidade da utilização do algoritmo EM que também pode ser empregado na estimação dos parâmetros desconhecidos do modelo. Neste caso consideramos modelos mais gerais que os considerados nos capítulos anteriores. Em particular, derivamos estimadores de máxima verossimilhança em um modelo com réplicas (Chan e Mak, 1979) e também em um modelo com s variáveis regressoras. No modelo funcional a estimação de x é efetuada através dos enfoques Bayesiano e de máxima verossimilhança, este último com a transformação polar.

No apêndice são apresentadas algumas identidades utilizadas com frequência

no texto, uma descrição da priori MDIP e a definição e construção de parâmetros ortogonais. Incluímos também alguns programas computacionais utilizados nas ilustrações numéricas, que são importantes para ilustrar o comportamento numérico dos procedimentos propostos no texto.

1.4 Perspectiva de pesquisa futura

Este trabalho tem continuidade e algumas sugestões são as seguintes:

- Considerações de famílias mais gerais de distribuições para os erros do modelo, por exemplo, as elípticas.
- Desenvolvimento de testes baseados na estatística do score, nas suas versões não corrigida e corrigida (Cordeiro e Ferrari, 1991) para as hipóteses consideradas no Capítulo 3 e comparações com as estatísticas da razão de verossimilhança e Wald.
- Investigar a possibilidade do desenvolvimento de transformações ortogonais no MEV estrutural com σ_e^2 ou σ_u^2 desconhecidos.
- Investigar a possibilidade da utilização de procedimentos baseados no algoritmo EM no modelo polinomial (Fuller, 1987).
- Utilizar o algoritmo EM no modelo Ultra-estrutural (Dolby, 1977).
- Considerar a possibilidade da utilização do algoritmo EM em modelos mais gerais que aqueles considerados na Seção 5.2.2.

Capítulo 2

Parametrização ortogonal no MEV

2.1 Introdução

A maioria dos resultados deste capítulo estão relacionados com a transformação ortogonal no sentido de Cox e Reid (1987). Estas transformações são obtidas a partir das equações diferenciais derivadas em Cox e Reid e tornam o parâmetro de interesse β ortogonal aos demais parâmetros do modelo transformado.

Temos interesse na ortogonalidade de β com os demais parâmetros pela simplificação de alguns resultados a serem apresentados nos capítulos seguintes e pelas vantagens decorrentes da transformação ortogonal (vide apêndice A). Inicialmente abordaremos as transformações ortogonais destinadas ao MEV estrutural e em seguida para o MEV funcional onde inclui-se também a transformação polar como uma forma alternativa de se obter ortogonalidade entre β e os demais parâmetros do modelo.

2.2 O MEV estrutural. Uma população

Como vimos no Capítulo 1, o modelo estrutural é não identificável quando todos os parâmetros são desconhecidos. Por isso, duas situações são geralmente abordadas para contornar o problema da falta de identificação do modelo. Uma estabelece que a razão de variâncias $\phi = \sigma_e^2/\sigma_u^2$ é conhecida enquanto a outra estabelece que o coeficiente de atenuação $k_x = \tau^2/(\tau^2 + \sigma_u^2)$ é conhecido. A suposição de que ϕ seja conhecido implica que ϕ pode ser tomado, sem perda de generalidade, como sendo igual a um ($\phi = 1$) ou equivalentemente que $\sigma_e^2 = \sigma_u^2 = \sigma^2$. Isto pode ser visto sem dificuldades transformando-se a equação $y_i = \alpha + \beta x_i + e_i$ na equação $y_i^* = \alpha^* + \beta^* x_i + e_i^*$, onde $y_i^* = y_i/\phi^{1/2}$, $\alpha^* = \alpha/\phi^{1/2}$ e $e_i^* = e_i/\phi^{1/2}$, $i = 1, \dots, n$. Note que $Var(e_i^*) = Var(u_i) = \sigma_u^2$.

Admitindo ϕ conhecido, Wong (1989), propôs transformações ortogonais para os casos σ^2 conhecido e desconhecido que são descritas nos teoremas que seguem. Destes resultados, pode-se obter um interessante corolário que pode ser utilizado no modelo abordado em Lindley e El-Sayad (1968).

Teorema 2.2.1 *Seja (X_i, Y_i) , $i = 1, \dots, n$, uma amostra aleatória de (X, Y) , onde $\sigma_e^2 = \sigma_u^2 = \sigma^2$, com*

$$(X, Y) \sim \mathcal{N} \left[(\mu, \alpha + \beta\mu), \begin{pmatrix} \sigma^2 + \tau^2 & \beta\tau^2 \\ \beta\tau^2 & \beta^2\tau^2 + \sigma^2 \end{pmatrix} \right]. \quad (2.1)$$

Consideramos β como parâmetro de interesse e $\theta = (\alpha, \mu, \tau^2, \sigma^2)$ como parâmetro de perturbação. Sendo σ^2 conhecido, as equações de Cox e Reid para este modelo são dadas por

$$\begin{aligned} (\tau^2 + \sigma^2) \frac{\partial \alpha}{\partial \beta} + \beta \sigma^2 \frac{\partial \mu}{\partial \beta} &= -\mu(\tau^2 + \sigma^2), \\ \beta \frac{\partial \alpha}{\partial \beta} + (\beta^2 + 1) \frac{\partial \mu}{\partial \beta} &= -\mu\beta \end{aligned} \quad (2.2)$$

e

$$\frac{1}{2}(\beta^2 + 1) \frac{\partial \tau^2}{\partial \beta} = -\beta \tau^2,$$

com um conjunto solução dado por

$$\lambda_0 = \mu, \quad \lambda_1 = \alpha + \beta \mu, \quad \lambda_2 = \tau^2(\beta^2 + 1) + \sigma^2. \quad (2.3)$$

Prova: Pode-se mostrar que os elementos da matriz de informação de (β, θ) são dados por

$$i_{\mu\tau^2} = i_{\alpha\tau^2}, \quad i_{\mu\mu} = \sigma^2(\beta^2 + 1)D^{-1}, \quad i_{\alpha\alpha} = (\tau^2 + \sigma^2)D^{-1},$$

$$i_{\mu\beta} = \mu\beta\sigma^2D^{-1}, \quad i_{\alpha\beta} = \mu(\tau^2 + \sigma^2)D^{-1}$$

$$i_{\alpha\mu} = \beta\sigma^2D^{-1}, \quad i_{\tau^2\tau^2} = \sigma^4(\beta^2 + 1)^2(2D^2)^{-1},$$

onde $D = \tau^2\sigma^2(\beta^2 + 1) + \sigma^4$. Com os elementos da matriz de informação acima, constrói-se as equações diferenciais parciais de Cox e Reid (Apêndice A) que levam ao sistema de equações diferenciais parciais (2.2). Pode-se efetuar a prova verificando-se diretamente que (2.3) é solução de (2.2). Uma forma alternativa de concluir a prova é obtida substituindo (2.3) na verossimilhança original que segue de uma amostra aleatória de (2.1), obtendo-se com (β, λ) , $\lambda = (\lambda_0, \lambda_1, \lambda_2)$,

$$\begin{aligned} \ell(\beta, \lambda) &= -n \log(2\pi) - \frac{1}{2}n \log \sigma^2 - \frac{1}{2}n \log \lambda_2 - \left[2\lambda_2\sigma^2(\beta^2 + 1)\right]^{-1} \left\{(\beta^2\lambda_2 + \sigma^2) \right. \\ &\times \sum_{i=1}^n (X_i - \lambda_0)^2 - 2\beta(\lambda_2 - \sigma^2) \sum_{i=1}^n (X_i - \lambda_0)(Y_i - \lambda_1) + (\lambda_2 \\ &+ \beta^2\sigma^2) \sum_{i=1}^n (Y_i - \lambda_1)^2 \left. \right\}. \end{aligned} \quad (2.4)$$

Em seguida determina-se a matriz de informação de (β, λ) , que é dada por

$$I(\beta, \lambda) = \begin{bmatrix} \frac{n(\lambda_2 - \sigma^2)^2}{\lambda_2 \sigma^2 (\beta^2 + 1)^2} & 0 & 0 & 0 \\ 0 & \frac{n(\beta^2 \lambda_2 + \sigma^2)}{\lambda_2 \sigma^2 (\beta^2 + 1)} & -\frac{n\beta(\lambda_2 - \sigma^2)}{\lambda_2 \sigma^2 (\beta^2 + 1)} & 0 \\ 0 & -\frac{n\beta(\lambda_2 - \sigma^2)}{\lambda_2 \sigma^2 (\beta^2 + 1)} & \frac{n(\lambda_2 + \beta^2 \sigma^2)}{\lambda_2 \sigma^2 (\beta^2 + 1)} & 0 \\ 0 & 0 & 0 & \frac{n}{2\lambda_2^2} \end{bmatrix}. \quad (2.5)$$

Pela definição de parâmetros ortogonais (veja o Apêndice A), conclui-se a prova.

□

Teorema 2.2.2 *Sejam (X_i, Y_i) , $i = 1, \dots, n$, uma amostra aleatória que segue de (2.1), β o parâmetro de interesse e $\theta = (\mu, \alpha, \tau^2, \sigma^2)$ o vetor de parâmetros de perturbação. Se σ^2 é desconhecido então as equações de Cox e Reid para este modelo são dadas por*

$$\begin{aligned} (\tau^2 + \sigma^2) \frac{\partial \alpha}{\partial \beta} + \beta \sigma^2 \frac{\partial \mu}{\partial \beta} &= -\mu(\tau^2 + \sigma^2), \\ \beta \frac{\partial \alpha}{\partial \beta} + (\beta^2 + 1) \frac{\partial \mu}{\partial \beta} &= -\mu\beta, \\ \frac{1}{2}(\beta^2 + 1) \frac{\partial \tau^2}{\partial \beta} &= -\beta\tau^2, \end{aligned} \quad (2.6)$$

e

$$\frac{1}{2} \left\{ (\beta^2 \tau^2 + \sigma^2)^2 + 2\beta^2 \tau^4 + (\tau^2 + \sigma^2)^2 \right\} \frac{\partial \sigma^2}{\partial \beta} + \frac{1}{2} \left\{ \tau^4 (\beta^2 + 1) \right\} \frac{\partial \tau^2}{\partial \beta} = -\beta \tau^2 \sigma^4,$$

com um conjunto solução dado por

$$\lambda_0 = \mu, \quad \lambda_1 = \alpha + \beta\mu, \quad \lambda_2 = \tau^2(\beta^2 + 1) + \sigma^2 \quad e \quad \lambda_3 = \sigma^2. \quad (2.7)$$

Prova: Os elementos da matriz de informação de $(\beta, \theta) = (\alpha, \mu, \tau^2, \sigma^2)$ são os mesmos do teorema anterior acrescidos de

$$i_{\sigma^2 \mu} = i_{\sigma^2 \alpha} = 0, \quad i_{\sigma^2 \sigma^2} = \frac{1}{2D^3} \left[\tau^2(\beta^2 + 1) + 2\sigma^2 \right]^2,$$

$$i_{\sigma^2\beta} = D^{-2}\sigma^2\tau^2\beta\tau^2(\beta^2 + 1) + 2\sigma^2 - D^{-1}\beta\tau^2$$

e

$$i_{\sigma^2\sigma^2} = [\tau^2(\beta^2 + 1) + 2\sigma^2]\sigma^2(\beta^2 + 1) - (2D)^{-1}(\beta^2 + 1).$$

Com estes elementos, constrói-se o sistema (2.6). Outra maneira de provar o resultado é determinar a verossimilhança do modelo transformado por (2.7). Sendo esta igual a (2.4) com λ_3 no lugar de σ^2 , determina-se a matriz de informação de (β, λ) , $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$, de onde segue que

$$I(\beta, \lambda) = \begin{bmatrix} \frac{n(\lambda_2 - \lambda_3)^2}{\lambda_2\lambda_3(\beta^2 + 1)^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{n(\beta^2\lambda_2 + \lambda_3)}{\lambda_2\lambda_3(\beta^2 + 1)} & -\frac{n\beta(\lambda_2 - \lambda_3)}{\lambda_2\lambda_3(\beta^2 + 1)} & 0 & 0 \\ 0 & -\frac{n\beta(\lambda_2 - \lambda_3)}{\lambda_2\lambda_3(\beta^2 + 1)} & \frac{n(\lambda_2 + \beta^2\lambda_3)}{\lambda_2\lambda_3(\beta^2 + 1)} & 0 & 0 \\ 0 & 0 & 0 & \frac{n}{2\lambda_2^2} & -\frac{n}{\lambda_2\lambda_3} \\ 0 & 0 & 0 & -\frac{n}{\lambda_2\lambda_3} & \frac{n}{2\lambda_3^2} \end{bmatrix}. \quad (2.8)$$

Observando-se o elemento da matriz de informação (2.8) associado a β , que denotamos por $i_{\beta\beta}$, a prova é concluída diretamente da definição de parâmetros ortogonais.

□

Corolário 2.2.1 Se $\alpha = 0$ e σ^2 é conhecido, então β é ortogonal a $\lambda = (\lambda_0, \lambda_1, \lambda_2)$, com λ_0 e λ_2 dados em (2.3) e $\lambda_1 = \beta\mu$. Se σ^2 for desconhecido então β é ortogonal a $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$, com λ_0, λ_1 e λ_2 como acima e $\lambda_3 = \sigma^2$.

Prova: Sendo $\alpha = 0$ e σ^2 conhecido, segue do Teorema 2.2.1 que a transformação

$$\lambda_0 = \mu, \quad \lambda_1 = \beta\mu \quad e \quad \lambda_2 = \tau^2(\beta^2 + 1) + \sigma^2,$$

torna β ortogonal a $\lambda = (\lambda_0, \lambda_1, \lambda_2)$. Se σ^2 for desconhecido, segue do Teorema 2.2.1 que β é ortogonal a $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$ com $\lambda_3 = \sigma^2$ e $\lambda_0, \lambda_1, \lambda_2$ dados acima.

□

Corolário 2.2.2 Se $\alpha = \mu = 0$ e σ^2 é conhecido, então β é ortogonal a $\lambda = \lambda_2$ dado em (2.3) e se σ^2 for desconhecido, β é ortogonal a $\lambda = (\lambda_2, \lambda_3)$ com λ_2 como acima e $\lambda_3 = \sigma^2$.

Prova: Sendo $\alpha = \mu = 0$ e σ^2 conhecido, segue do Teorema 2.2.1 que a transformação $\lambda_2 = \tau^2(\beta^2 + 1) + \sigma^2$ torna β ortogonal a λ_2 . Por outro lado se σ^2 for desconhecido, segue do Teorema 2.2.1 que a transformação $\lambda_2 = \tau^2(\beta^2 + 1) + \sigma^2$, $\lambda_3 = \sigma^2$ torna β ortogonal á (λ_2, λ_3) . $\square$

Considerando o MEV com coeficiente de atenuação, k_x , conhecido Bolfarine e Cordani (1990) e (1993) apresentam transformações ortogonais com μ e α conhecidos e μ e α desconhecidos. Estes resultados estão enunciados nos dois teoremas seguintes. O primeiro trata do caso μ e α conhecidos ($\alpha = 0$, sem perda de generalidade).

Teorema 2.2.3 Seja (X_i, Y_i) , $i = 1, \dots, n$, uma amostra aleatória de (X, Y) onde

$$(X, Y) \sim \mathcal{N} \left[(\mu, \beta\mu) , \begin{pmatrix} \frac{\tau^2}{k_x} & \beta\tau^2 \\ \beta\tau^2 & \beta^2\tau^2 + \sigma_e^2 \end{pmatrix} \right], \quad (2.9)$$

com $k_x = \tau^2/(\tau^2 + \sigma_u^2)$, sendo β o parâmetro de interesse e $\theta = (\mu, \tau^2, \sigma_e^2)$ o parâmetro de perturbação. Se μ for conhecido então as equações de Cox e Reid para este modelo são dadas por

$$(1 + k^2) \frac{\partial \sigma_e^2}{\partial \beta} + k(k + 1) \beta^2 \frac{\partial \tau^2}{\partial \beta} = -2k(k + 1) \beta \tau^2, \quad (2.10)$$

$$k(k + 1) \beta^2 \tau^4 \frac{\partial \sigma_e^2}{\partial \beta} + [2k\beta^2 D + \sigma_e^4(k + 1)^2] \frac{\partial \tau^2}{\partial \beta} = -2\beta^3 k^2 \tau^6, \quad (2.11)$$

onde $k = (1 - k_x)/k_x$ e $D = k\beta^2\tau^4 + (k + 1)\tau^2\sigma_e^2$. Uma solução para o sistema de equações (2.10)-(2.11) é a transformação

$$\lambda_0 = (k + 1)\sigma_e^2 + k\lambda_1\beta^2, \quad \lambda_1 = \tau^2, \quad (2.12)$$

que torna β ortogonal à $\lambda = (\lambda_0, \lambda_1)$.

Prova: A transformação (2.12) substituída na função de log-verossimilhança correspondente a uma amostra aleatória de tamanho n de (2.9) produz a log-verossimilhança

$$\begin{aligned} \ell(\beta, \lambda) = & -n \log(2\pi) - \frac{n}{2} \log(\lambda_0 \lambda_1) - \frac{1}{\lambda_0 \lambda_1} \left\{ \frac{\lambda_0 + \lambda_1 \beta^2}{k+1} \sum_{i=1}^n (X_i - \mu)^2 \right. \\ & \left. - 2\beta \lambda_1 \sum_{i=1}^n (X_i - \mu)(Y_i - \beta \mu) + \lambda_1 (k+1) \sum_{i=1}^n (Y_i - \beta \mu)^2 \right\}. \end{aligned} \quad (2.13)$$

A matriz de informação de (λ, β) determinada a partir de (2.13) é igual a

$$\mathbf{I}(\lambda, \beta) = \begin{pmatrix} \frac{n}{2\lambda_0^2} & 0 & 0 \\ 0 & \frac{n}{2\lambda_1^2} & 0 \\ 0 & 0 & \frac{n[\lambda_1 + \mu^2(k+1)^2]}{\lambda_0} \end{pmatrix} \quad (2.14)$$

e da definição de parâmetros ortogonais percebe-se que β é ortogonal à λ . $\square$

No teorema seguinte consideramos o caso em que o intercepto α e a média μ são também desconhecidos.

Teorema 2.2.4 *Seja (X_i, Y_i) , $i = 1, \dots, n$, uma amostra aleatória de (X, Y) , onde*

$$(X, Y) \sim \mathcal{N} \left[(\mu, \alpha + \beta \mu), \begin{pmatrix} \frac{\tau^2}{k_x} & \beta \tau^2 \\ \beta \tau^2 & \beta^2 \tau^2 + \sigma_e^2 \end{pmatrix} \right] \quad (2.15)$$

com $k_x = \tau^2/(\tau^2 + \sigma_u^2)$, sendo β parâmetro de interesse e $\theta = (\alpha, \mu, \tau^2, \sigma^2)$. Então as equações de Cox e Reid para este modelo são dadas por (2.10), (2.11),

$$(\sigma_e^2 + k\beta^2\tau^2) \frac{\partial \mu}{\partial \beta} + k\beta\tau^2 \frac{\partial \alpha}{\partial \beta} = -k\mu\beta\tau^2 \quad (2.16)$$

e

$$k\beta \frac{\partial \mu}{\partial \beta} + (k+1) \frac{\partial \alpha}{\partial \beta} = -(k+1)\mu, \quad (2.17)$$

admitindo como solução

$$\lambda_0 = (k + 1)\sigma_e^2 + k\lambda_1\beta^2 \quad (2.18)$$

$$\lambda_1 = \tau^2, \quad \lambda_2 = \mu \quad e \quad \lambda_3 = \alpha + \beta\mu. \quad (2.19)$$

Prova: A transformação (2.18)-(2.19) substituída na função de verossimilhança correspondente a uma amostra aleatória de tamanho n de (2.15) produz a log-verossimilhança

$$\begin{aligned} \ell(\boldsymbol{\lambda}, \beta) = & -n \log(2\pi) - \frac{n}{2} \log(\lambda_0 \lambda_1) - \frac{1}{\lambda_0 \lambda_1} \left\{ \frac{\lambda_0 + \lambda_1 \beta^2}{k + 1} \sum_{i=1}^n (X_i - \lambda_2)^2 \right. \\ & \left. - 2\beta \lambda_1 \sum_{i=1}^n (X_i - \lambda_2)(Y_i - \lambda_3) + \lambda_1(k + 1) \sum_{i=1}^n (Y_i - \lambda_3)^2 \right\}, \quad (2.20) \end{aligned}$$

onde $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$. A matriz de informação de $(\boldsymbol{\lambda}, \beta)$ determinada a partir de (2.20) é igual a

$$\mathbf{I}(\boldsymbol{\lambda}, \beta) = \begin{pmatrix} \frac{n}{2\lambda_0^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{n}{2\lambda_1^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{n(\lambda_0 + \lambda_1\beta^2)}{(k+1)\lambda_0\lambda_1} & \frac{-n\beta}{\lambda_0} & 0 \\ 0 & 0 & \frac{-n\beta}{\lambda_0} & \frac{n(k+1)}{\lambda_0} & 0 \\ 0 & 0 & 0 & 0 & \frac{n\lambda_1}{\lambda_0} \end{pmatrix} \quad (2.21)$$

e da definição de parâmetros ortogonais percebe-se que β é ortogonal à $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$. $\square$

Rodrigues e Cordani (1990) consideram uma parametrização ortogonal alternativa ao enfoque considerado por Bolfarine e Cordani (1990). Considerando a distribuição conjunta de (X_i, Y_i) na parametrização original que é dada em (2.9) temos, que a distribuição condicional de Y_i dado X_i , é tal que

$$Y_i|X_i \sim \mathcal{N}(\beta k_x, \sigma_u^2 k_x \beta^2 + \sigma_e^2),$$

$i = 1, \dots, n$. Então β é ortogonal à $\lambda = \sigma_u^2 k_x^2 \beta^2 + \sigma_e^2$ no modelo condicional e a função de log-verossimilhança com esta reparametrização é dada por

$$\ell(\beta, \lambda) = \frac{n}{2} \log(\lambda) - \frac{1}{2\lambda} \sum_{i=1}^n (Y_i - \beta k_x X_i)^2.$$

2.3 O MEV estrutural. Vários modelos

Wong (1991) considera uma extensão do MEV estrutural definido em (2.1), considerando as seguintes equações

$$\begin{aligned} X_{ij} &= x_{ij} + u_{ij}, \\ Y_{ij} &= y_{ij} + e_{ij}, \\ y_{ij} &= \alpha_i + \beta_i x_{ij}, \end{aligned} \tag{2.22}$$

$j = 1, \dots, n_i$, $i = 1, \dots, s$. O problema de interesse primordial é a comparação dos coeficientes angulares $\beta_1, \dots, \beta_s$, com as suposições usuais de independência entre e_i , u_i e x_i , com $e_i \sim \mathcal{N}(0, \sigma_{ei}^2)$, $u_i \sim \mathcal{N}(0, \sigma_{ui}^2)$, $x_i \sim \mathcal{N}(\mu_i, \tau_i^2)$, $i = 1, \dots, s$. Supondo também que $\sigma_{ei}^2 = \sigma_{ui}^2 = \sigma_i^2$ (que corresponde a suposição $\phi_i = \sigma_{ei}^2 / \sigma_{ui}^2$ conhecido) tem-se, neste modelo que,

$$(X_{ij}, Y_{ij}) \sim \mathcal{N} \left[(\mu_i, \alpha_i + \beta_i \mu_i), \begin{pmatrix} \sigma_i^2 + \tau_i^2 & \beta_i \tau_i^2 \\ \beta_i \tau_i^2 & \beta_i^2 \tau_i^2 + \sigma_i^2 \end{pmatrix} \right], \tag{2.23}$$

$j = 1, \dots, n_i$, $i = 1, \dots, s$. Os Teoremas 2.1.1 e 2.1.2 podem ser estendidos para esta situação mais geral, ou seja, se σ_i^2 , for conhecido, $i = 1, \dots, s$, a transformação

$$\lambda_{0i} = \mu_i, \quad \lambda_{1i} = \alpha_i + \beta_i \mu_i \quad \text{e} \quad \lambda_{2i} = \tau_i^2 (\beta_i^2 + 1) + \sigma_i^2, \tag{2.24}$$

torna β_i ortogonal a $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i})$, $i = 1, \dots, s$. Neste caso, a matriz de informação de (β, λ) , onde $\beta = (\beta_1, \dots, \beta_s)$ e $\lambda = (\lambda_{0i}, \dots, \lambda_{0s}, \lambda_{1i}, \dots, \lambda_{1s}, \lambda_{2i}, \dots, \lambda_{2s})$ é diagonal

em blocos, que denotamos por

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \text{diag} [\mathbf{I}_1(\beta_1, \lambda_{01}, \lambda_{11}, \lambda_{21}), \dots, \mathbf{I}_s(\beta_s, \lambda_{0s}, \lambda_{1s}, \lambda_{2s})], \quad (2.25)$$

onde

$$\mathbf{I}_i(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \begin{bmatrix} \frac{n_i(\lambda_{2i} - \sigma_i^2)^2}{\lambda_{2i}\sigma_i^2(\beta_i^2 + 1)^2} & 0 & 0 & 0 \\ 0 & \frac{n_i(\beta_i^2\lambda_{2i} + \sigma_i^2)}{\lambda_{2i}\sigma_i^2(\beta_i^2 + 1)} & -\frac{n_i\beta_i(\lambda_{2i} - \sigma_i^2)}{\lambda_{2i}\sigma_i^2(\beta_i^2 + 1)} & 0 \\ 0 & -\frac{n_i\beta_i(\lambda_{2i} - \sigma_i^2)}{\lambda_{2i}\sigma_i^2(\beta_i^2 + 1)} & \frac{n_i(\lambda_{2i} + \beta_i^2\sigma_i^2)}{\lambda_{2i}\sigma_i^2(\beta_i^2 + 1)} & 0 \\ 0 & 0 & 0 & \frac{n_i}{2\lambda_{2i}^2} \end{bmatrix}, \quad (2.26)$$

$i = 1, \dots, s$. Se σ_i^2 , for desconhecido, $i = 1, \dots, s$, a transformação

$$\lambda_{0i} = \mu_i, \quad \lambda_{1i} = \alpha_i + \beta_i\mu_i, \quad \lambda_{2i} = \tau_i^2(\beta_i^2 + 1) + \sigma_i^2 \quad e \quad \lambda_{3i} = \sigma_i^2 \quad (2.27)$$

torna β_i ortogonal à $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i})$, $i = 1, \dots, s$. Neste caso a matriz de informação de $(\boldsymbol{\beta}, \boldsymbol{\lambda})$, onde $\boldsymbol{\lambda} = (\lambda_{0i}, \dots, \lambda_{0s}, \lambda_{1i}, \dots, \lambda_{1s}, \lambda_{2i}, \dots, \lambda_{2s}, \lambda_{3i}, \dots, \lambda_{3s})$ é dada por

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \text{diag} [\mathbf{I}_1(\beta_1, \lambda_{01}, \lambda_{11}, \lambda_{21}, \lambda_{31}), \dots, \mathbf{I}_s(\beta_s, \lambda_{0s}, \lambda_{1s}, \lambda_{2s}, \lambda_{3s})], \quad (2.28)$$

onde

$$\mathbf{I}_i = \begin{bmatrix} \frac{n_i(\lambda_{2i} - \lambda_{3i})^2}{\lambda_{2i}\lambda_{3i}(\beta_i^2 + 1)^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{n_i(\beta_i^2\lambda_{2i} + \lambda_{3i})}{\lambda_{2i}\lambda_{3i}(\beta_i^2 + 1)} & -\frac{n_i\beta_i(\lambda_{2i} - \lambda_{3i})}{\lambda_{2i}\lambda_{3i}(\beta_i^2 + 1)} & 0 & 0 \\ 0 & -\frac{n_i\beta_i(\lambda_{2i} - \lambda_{3i})}{\lambda_{2i}\lambda_{3i}(\beta_i^2 + 1)} & \frac{n_i(\lambda_{2i} + \beta_i^2\lambda_{3i})}{\lambda_{2i}\lambda_{3i}(\beta_i^2 + 1)} & 0 & 0 \\ 0 & 0 & 0 & \frac{n_i}{2\lambda_{2i}^2} & -\frac{n_i}{\lambda_{2i}\lambda_{3i}} \\ 0 & 0 & 0 & -\frac{n_i}{\lambda_{2i}\lambda_{3i}} & \frac{n_i}{2\lambda_{3i}^2} \end{bmatrix}, \quad (2.29)$$

$i = 1, \dots, s$.

De forma análoga, usando a independência das s amostras tomadas nas s populações, alguns casos particulares de (2.23) geram transformações ortogonais que serão utilizadas futuramente no Capítulo 3. Representamos estes casos particulares com os seguintes modelos:

Modelo 1. Se em (2.23) $\alpha_i = \alpha$, $\mu_i = \mu$, $\beta_i = \beta$, $\tau_i^2 = \tau^2$ e $\sigma_i^2 = \sigma^2$, para $i = 1, \dots, s$, então com σ^2 conhecido, β é ortogonal à $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i})$, onde

$$\lambda_{0i} = \mu, \quad \lambda_{1i} = \alpha + \beta\mu, \quad \lambda_{2i} = \tau^2(\beta^2 + 1) + \sigma^2. \quad (2.30)$$

Modelo 2. Análogo ao Modelo 1 com σ^2 desconhecido. Então β é ortogonal $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i})$, onde

$$\lambda_{0i} = \mu, \quad \lambda_{1i} = \alpha + \beta\mu, \quad \lambda_{2i} = \tau^2(\beta^2 + 1) + \sigma^2, \quad \lambda_{3i} = \sigma^2, \quad (2.31)$$

$i = 1, \dots, s$. Os Modelos 1 e 2 são considerados em Dolby (1976), Barnett (1969) e Cox (1976).

Modelo 3. Se em (2.23), $\alpha_i = \alpha$, $\mu_i = \mu$, $\beta_i = \beta$, para $i = 1, \dots, s$, então β é ortogonal à $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i})$, onde

$$\lambda_{0i} = \mu, \quad \lambda_{1i} = \alpha + \beta\mu, \quad \lambda_{2i} = \tau_i^2(\beta^2 + 1) + \sigma_i^2, \quad (2.32)$$

com σ_i^2 , $i = 1, \dots, s$, conhecido.

Modelo 4. Análogo ao Modelo 3 com σ_i^2 , $i = 1, \dots, s$, desconhecido. Então β é ortogonal à $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i})$, onde

$$\lambda_{0i} = \mu, \quad \lambda_{1i} = \alpha + \beta\mu, \quad \lambda_{2i} = \tau_i^2(\beta^2 + 1) + \sigma_i^2, \quad \lambda_{3i} = \sigma_i^2, \quad (2.33)$$

$i = 1, \dots, s$.

Modelo 5. Se em (2.23), $\alpha_i = \alpha$, $\beta_i = \beta$, $\tau_i^2 = \tau^2$, e $\sigma_i^2 = \sigma^2$ com σ^2 conhecido, então β é ortogonal à $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i})$, onde

$$\lambda_{0i} = \mu_i, \quad \lambda_{1i} = \alpha + \beta\mu_i, \quad \lambda_{2i} = \tau^2 + (\beta^2 + 1) + \sigma^2, \quad (2.34)$$

$i = 1, \dots, s.$

Modelo 6. Análogo ao Modelo 5 com σ^2 desconhecido. Então β é ortogonal à $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i})$, onde

$$\lambda_{0i} = \mu_i, \quad \lambda_{1i} = \alpha + \beta\mu_i, \quad \lambda_{2i} = \tau^2(\beta^2 + 1) + \sigma^2, \quad e \quad \lambda_{3i} = \sigma^2, \quad (2.35)$$

$i = 1, \dots, s.$

Modelo 7. Se em (2.23), $\alpha_i = \alpha$, $\beta_i = \beta$ e $\sigma_i^2 = \sigma^2$ com σ^2 conhecido, então β é ortogonal à $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i})$, onde

$$\lambda_{0i} = \mu_i, \quad \lambda_{1i} = \alpha + \beta\mu_i, \quad \lambda_{2i} = \tau_i^2 + (\beta^2 + 1) + \sigma^2, \quad (2.36)$$

$i = 1, \dots, s.$

Modelo 8. Análogo ao Modelo 7 com σ^2 desconhecido. Então β é ortogonal à $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i})$, onde

$$\lambda_{0i} = \mu_i, \quad \lambda_{1i} = \alpha + \beta\mu_i, \quad \lambda_{2i} = \tau_i^2(\beta^2 + 1) + \sigma^2, \quad e \quad \lambda_{3i} = \sigma^2, \quad (2.37)$$

$i = 1, \dots, s.$

De modo similar, estende-se as transformações ortogonais apresentadas nos Teoremas 2.1.3 e 2.1.4 para o MEV estrutural dado em (2.22) com coeficiente de atenuação na i -ésima população, k_{xi} , $i = 1, \dots, s$, conhecido. Ou seja, neste modelo

$$(X_{ij}, Y_{ij}) \sim \mathcal{N} \left[(\mu_i, \alpha_i + \beta\mu_i), \begin{pmatrix} \frac{\tau_i^2}{k_{xi}} & \beta_i\tau_i^2 \\ \beta_i\tau_i^2 & \beta_i^2\tau_i^2 + \sigma_{ei}^2 \end{pmatrix} \right], \quad (2.38)$$

onde $k_{xi} = \tau_i^2/(\tau_i^2 + \sigma_{ui}^2)$, $i = 1, \dots, s$, $j = 1, \dots, n_i$. Se μ_i , e α_i $i = 1, \dots, s$, são conhecidos pode-se mostrar que β_i é ortogonal à $(\lambda_{0i}, \lambda_{1i})$, onde

$$\lambda_{0i} = (k_i + 1)\sigma_{ei}^2 + k_i\lambda_{1i}\beta_i^2 \quad e \quad \lambda_{1i} = \tau_i^2, \quad (2.39)$$

com $k_i = (1 - k_{x_i})/k_{x_i}$, $i = 1, \dots, s$. Se μ_i e α_i , $i = 1, \dots, s$, são desconhecidos pode-se mostrar que β é ortogonal à $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i})$, onde

$$\lambda_{0i} = (k_i + 1)\sigma_{\epsilon i}^2 + k_i \lambda_{1i} \beta_i^2 \text{ e } \lambda_{1i} = \tau_i^2, \quad \lambda_{2i} = \mu_i, \quad \lambda_{3i} = \alpha_i + \beta_i \mu_i, \quad (2.40)$$

$i = 1, \dots, s$.

2.4 O MEV funcional

No MEV funcional considerado a seguir, duas situações são consideradas para tornar o modelo identificável. Uma estabelece que apenas a razão de variâncias $\phi = \sigma_e^2/\sigma_u^2$ é conhecida enquanto que a outra estabelece que as variâncias σ_u^2 e σ_e^2 são conhecidas. Pode-se admitir, como no caso estrutural, sem perda de generalidade, que quando ϕ é conhecido, então $\sigma_e^2 = \sigma_u^2 = \sigma^2$, ou seja, $\phi = 1$. Duas transformações ortogonais são aqui apresentadas. Uma segue das equações de Cox e Reid e a outra é obtida a partir de uma transformação polar dos parâmetros do modelo.

2.4.1 A transformação polar

A transformação polar tem sido muito utilizada no MEV. Veja por exemplo, Madansky (1959), Kendall and Stewart (1961), Moran (1971) e Florens et al. (1974). Esta transformação é utilizada neste trabalho principalmente porque, quando $\phi = 1$, tem-se que o parâmetro de interesse θ é ortogonal ao parâmetro de perturbação $\omega = (\omega_1, \dots, \omega_n)$, definidos a seguir.

Considerando o MEV funcional apresentado no Capítulo 1, a transformação polar aplicada a este modelo transforma (x_i, y_i) em (ω_i, θ) de modo que

$$x_i = \omega_i \cos \theta, \quad y_i = \alpha + \omega_i \sin \theta \text{ e } \beta = \tan \theta, \quad (2.41)$$

com $-\infty < \omega_i < \infty$, $i = 1, \dots, n$, e $0 < \theta < 2\pi$. Uma discussão de algumas propriedades da transformação polar aplicada ao MEV é encontrada em Florens et al. (1974). A função de log-verossimilhança para uma amostra de tamanho n , com a transformação polar (2.41) pode ser escrita como

$$\begin{aligned} \ell(\omega, \alpha, \beta, \sigma_e^2, \sigma_u^2, \theta) &\propto -\frac{n}{2} \log(\sigma_u^2) - \frac{n}{2} \log(\sigma_e^2) - \frac{1}{2\sigma_u^2} \sum_{i=1}^n (X_i - \omega_i \cos \theta)^2 \\ &\quad - \frac{1}{2\sigma_e^2} \sum_{i=1}^n (Y_i - \alpha - \omega_i \sin \theta)^2. \end{aligned} \quad (2.42)$$

Supondo a razão de variâncias $\phi = \sigma_e^2/\sigma_u^2$ conhecida e substituindo ϕ em (2.42), obtem-se a matriz de informação $\mathbf{I}(\gamma)$ para $\gamma = (\omega, \alpha, \sigma_u^2, \theta)$, a qual é dada por

$$\mathbf{I}(\gamma) = \sigma_u^{-2} \left(\begin{array}{ccc|ccc} & & & \frac{\sin \theta}{\phi} & \frac{-\alpha \sin \theta}{\phi \sigma_u^{-2}} & i_{\omega_1 \theta}^* \\ & & & \vdots & \vdots & \vdots \\ & & & \frac{\sin \theta}{\phi} & \frac{-\alpha \sin \theta}{\phi \sigma_u^{-2}} & i_{\omega_n \theta}^* \\ \hline \frac{\sin \theta}{\phi} & \dots & \frac{\sin \theta}{\phi} & n\phi^{-1} & \frac{-n\alpha}{\phi \sigma_u^2} & \frac{\cos \theta \sum_i \omega_i}{\phi} \\ \frac{-\alpha \sin \theta}{\phi \sigma_u^2} & \dots & \frac{-\alpha \sin \theta}{\phi \sigma_u^2} & \frac{n\alpha}{\phi \sigma_u^2} & \frac{n}{\sigma_u^2} \left(1 + \frac{\alpha^2}{\phi \sigma_u^2}\right) & \frac{-\alpha \cos \theta \sum_i \omega_i}{\phi \sigma_u^2} \\ i_{\omega_n \theta}^* & \dots & i_{\omega_n \theta}^* & \frac{\cos \theta \sum_i \omega_i}{\phi} & \frac{-\alpha \cos \theta \sum_i \omega_i}{\phi \sigma_u^2} & i_{\theta \theta}^* \end{array} \right), \quad (2.43)$$

onde neste caso, $i_{\omega_i \theta}^* = (\phi^{-1} - 1)\omega_i \cos \theta \sin \theta$, $b = \cos^2(\theta) + \phi^{-1} \sin^2(\theta)$, e $i_{\theta \theta}^* = \sin^2(\theta) + \phi^{-1} \sin \theta \sum_{i=1}^n \omega_i$.

Se $\alpha = 0$ e $\phi = 1$, temos que $\mathbf{I}(\gamma)$ dada em (2.43) se reduz a

$$\mathbf{I}(\gamma) = \sigma_u^{-2} \left(\begin{array}{ccc|cc} & & & \sin \theta & 0 \\ & & & \vdots & \vdots \\ & & & \sin \theta & 0 \\ \hline \sin \theta & \dots & \sin \theta & n & \cos \theta \sum_i \omega_i \\ 0 & \dots & 0 & \cos \theta \sum_i \omega_i & i_{\theta \theta}^* \end{array} \right), \quad (2.44)$$

com $\gamma = (\omega, \sigma_u^2, \theta)$, e neste caso tem-se que o parâmetro θ é ortogonal ao vetor de parâmetros ω , resultado este equivalente ao da parametrização de Cox e Reid

(1987), pois se $\phi = 1$ tem-se em (2.43) que $i_{\omega, \theta}^* = 0$, $j = 1, \dots, n$. Contudo, mesmo para $\phi = 1$, observamos que θ não é ortogonal a σ_u^2 .

2.4.2 Uma parametrização ortogonal

Supondo que $\alpha = 0$ e que as variâncias σ_e^2 , σ_u^2 são conhecidas, a função de verossimilhança para uma amostra de tamanho n de valores observados (X_i, Y_i) , $i = 1, \dots, n$, do MEV funcional definido no Capítulo 1, pode ser escrita como

$$L(\beta, \mathbf{x}) \propto \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n \frac{(Y_i - \beta x_i)^2}{\sigma_e^2} + \sum_{i=1}^n \frac{(X_i - x_i)^2}{\sigma_u^2} \right] \right\}, \quad (2.45)$$

onde $\mathbf{x} = (x_1, \dots, x_n)$ é o vetor de valores não observáveis que chamamos de reais.

A matriz de informação de $(\beta, \mathbf{x})$, denotada por $\mathbf{I}(\beta, \mathbf{x})$ é dada por

$$\mathbf{I}(\beta, \mathbf{x}) = \begin{bmatrix} \sum_{i=1}^n \frac{x_i^2}{\sigma_e^2} & \frac{\beta x_1}{\sigma_e^2} & \dots & \frac{\beta x_n}{\sigma_e^2} \\ \frac{\beta x_1}{\sigma_e^2} & \frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\beta x_n}{\sigma_e^2} & 0 & \dots & \frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \end{bmatrix}.$$

Escrevendo as equações de Cox e Reid para (2.45), onde β é o parâmetro de interesse, obtem-se

$$i_{x_i x_i} \frac{\partial x_i}{\partial \beta} = -i_{\beta x_i},$$

$i = 1, \dots, n$. Resolvendo-se estas equações diferenciais obtem-se um conjunto de transformações dadas por

$$x_i = \frac{\lambda_i}{\left(\frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right)^{1/2}}, \quad (2.46)$$

$i = 1, \dots, n$. Substituindo (2.46) em (2.45), a nova função de verossimilhança é

dada por

$$L(\beta, \lambda) \propto \exp \left\{ -\frac{1}{2} \left[\frac{1}{\sigma_e^2} \sum_{i=1}^n \left(Y_i - \frac{\beta \lambda_i}{\left(\frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right)^{1/2}} \right)^2 + \frac{1}{\sigma_u^2} \sum_{i=1}^n \left(X_i - \frac{\lambda_i}{\left(\frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right)^{1/2}} \right)^2 \right] \right\}, \quad (2.47)$$

onde $\lambda = (\lambda_1, \dots, \lambda_n)$. A matriz de informação de (β, λ) agora é dada por

$$\mathbf{I}(\beta, \lambda) = \begin{bmatrix} \frac{\sum_{i=1}^n x_i^2}{\beta^2 \sigma_u^2 + \sigma_e^2} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \mathbf{I}_{n-1} & \\ 0 & & & \end{bmatrix}, \quad (2.48)$$

o que mostra que β é ortogonal a $\lambda = (\lambda_0, \dots, \lambda_n)$, onde $\mathbf{I}_{n-1}$ é a matriz de identidade de ordem $n - 1$. Se a razão de variâncias $\phi = \sigma_e^2/\sigma_u^2$ é conhecida, ou seja $\sigma_e^2 = \sigma_u^2 = \sigma^2$, com σ^2 desconhecido, tem-se que a função de verossimilhança (2.45) se modifica para

$$L(\beta, \sigma^2, \mathbf{x}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (Y_i - \beta x_i)^2 + \sum_{i=1}^n (X_i - x_i)^2 \right] \right\}. \quad (2.49)$$

A matriz de informação de $(\beta, \sigma^2, \mathbf{x})$, denotada por $\mathbf{I}(\beta, \sigma^2, \mathbf{x})$ é dada por

$$\mathbf{I}(\beta, \sigma^2, \mathbf{x}) = \begin{bmatrix} \frac{\sum_{i=1}^n x_i^2}{\sigma^2} & 0 & \beta x_1 & \dots & \beta x_n \\ 0 & \frac{3n-1}{2\sigma^4} & 0 & \dots & 0 \\ \beta x_1 & 0 & \frac{\beta^2+1}{\sigma^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta x_n & 0 & 0 & \dots & \frac{\beta^2+1}{\sigma^2} \end{bmatrix}.$$

Note que σ^2 é ortogonal a $(\beta, x_1, \dots, x_n)$, pois neste caso β e $\mathbf{x}$ são parâmetros de locação e σ^2 é parâmetro de escala, que na distribuição normal são ortogonais. Escrevendo as equações de Cox e Reid para (2.49), onde β é o parâmetro de interesse,

obtem-se

$$\begin{cases} i_{x_i x_i} \frac{\partial x_i}{\partial \beta} = -i_{\beta x_i}, & i = 1, \dots, n \\ i_{\sigma^2 \sigma^2} \frac{\partial \sigma^2}{\partial \beta} = -i_{\beta \sigma^2} & . \end{cases} \quad (2.50)$$

Pode-se mostrar que a transformação

$$\begin{cases} x_i = \frac{\lambda_i}{\sqrt{\beta^2 + 1}}, & i = 1, \dots, n \\ \lambda_0 = \sigma^2 \end{cases} \quad (2.51)$$

torna β ortogonal à $(\lambda_0, \mathbf{x})$, onde $\mathbf{x} = (x_1, \dots, x_n)$. Esta transformação pode ser estendida sem dificuldades para o caso de várias retas.

□□□

Capítulo 3

Testes da razão de verossimilhança, correções de Bartlett e testes de Wald no modelo estrutural

3.1 Introdução

Neste capítulo consideramos o MEV estrutural com uma e s populações onde tratamos o caso em que o coeficiente de atenuação k_{x_i} é conhecido, $i = 1, \dots, s$, e também situações onde a razão de variâncias $\phi_i = \sigma_{ei}^2 / \sigma_{ui}^2$, $i = 1, \dots, s$, é conhecida o que nos leva a considerar, sem perda de generalidade, que $\sigma_{ei}^2 = \sigma_{ui}^2 = \sigma_i^2$, com σ_i^2 conhecido ou desconhecido, $i = 1, \dots, s$.

Na inferência para o parâmetro β , as transformações ortogonais consideradas no Capítulo 2 para o caso de uma população, fornecem simplificações quando

utilizamos o teste da razão de verossimilhança e o teste de Wald. O mesmo ocorre para o caso de s populações.

Principais contribuições deste capítulo são o desenvolvimento de inferências para β usando a verossimilhança perfilada condicional (Cox e Reid, 1987), a introdução do teste de Wald para testar hipóteses relativas à β usando transformações ortogonais e a correção de Bartlett para a estatística da razão de verossimilhança com uma e s populações quando o coeficiente de atenuação k_{x_i} é conhecido, $i = 1, \dots, s$. Neste último caso consideramos situações onde μ_i e α_i são conhecidos e situações onde μ_i e α_i são desconhecidos, $i = 1, \dots, s$. Consideramos também a possibilidade da utilização da estatística da razão de verossimilhança condicional estendida para testar hipóteses do tipo $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, com β_0 especificado, nos casos em que a razão de variâncias ou o coeficiente de atenuação são conhecidos.

3.2 Uma população com razão de variâncias ϕ conhecida

Conforme visto no Capítulo 2, no caso estrutural, a função de verossimilhança na parametrização original com $\phi = 1$ é obtida a partir da distribuição conjunta de (X_i, Y_i) , que escrevemos como

$$(X_i, Y_i) \sim \mathcal{N} \left[(\mu, \alpha + \beta\mu), \begin{pmatrix} \sigma^2 + \tau^2 & \beta\tau^2 \\ \beta\tau^2 & \beta^2\tau^2 + \sigma^2 \end{pmatrix} \right], \quad (3.1)$$

$i = 1, \dots, n$.

Se σ^2 for conhecido, a transformação ortogonal (2.3) nos leva à log-verossimilhança

(2.4). Se σ^2 for desconhecido, a transformação ortogonal (2.7) nos leva à log-verossimilhança (2.4) com λ_3 no lugar de σ^2 . Inicialmente abordaremos o caso em que a variância σ^2 é conhecida.

3.2.1 Variância σ^2 conhecida

Derivando (2.4), onde $\sigma^2 = \lambda_3$ em relação aos parâmetros desconhecidos λ_0 , λ_1 , λ_2 e λ_3 obtem-se os estimadores de máxima verossimilhança $\hat{\lambda}_0$, $\hat{\lambda}_1$ e $\hat{\lambda}_2$ que são dados por

$$\begin{aligned}\hat{\lambda}_0 &= \bar{X}, \\ \hat{\lambda}_1 &= \bar{Y}, \\ \hat{\lambda}_2 &= \frac{S_{xx} + 2\hat{\beta}S_{xy} + \hat{\beta}^2S_{yy}}{\hat{\beta}^2 + 1},\end{aligned}\tag{3.2}$$

sendo que $\hat{\beta}$ é a raiz da equação quadrática

$$\hat{\beta}^2 S_{xy} + \hat{\beta}(S_{xx} - S_{yy}) - S_{xy} = 0,\tag{3.3}$$

com o mesmo sinal de S_{xy} , ou seja,

$$\hat{\beta} = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}}{2S_{xy}}.$$

Neste caso o elemento $i_{\beta\beta}$ da matriz de informação (2.5) associado a β é

$$i_{\beta\beta} = \frac{n(\lambda_2 - \sigma^2)^2}{\lambda_2 \sigma^2 (\beta^2 + 1)^2},$$

que pode ser usado para construir um intervalo de confiança aproximado para o parâmetro de interesse β , ou seja, para n grande, temos que

$$\left(\hat{\beta} \mp z_{\alpha/2} \hat{i}_{\beta\beta}^{-1/2}\right)\tag{3.4}$$

é um intervalo de confiança aproximado para β com coeficiente de confiança $1 - \alpha$ onde $z_{\alpha/2}$ é o percentil $(1 - \frac{\alpha}{2})100\%$ da distribuição normal padrão e $\hat{i}_{\beta\beta}$ é o estimador de máxima verossimilhança de $i_{\beta\beta}$.

Substituindo os estimadores de máxima verossimilhança de λ_0 , λ_1 , λ_2 e β dados em (3.2) e (3.3) em (2.4), obtem-se $\ell(\hat{\beta}, \hat{\lambda})$, onde $\hat{\lambda} = (\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2)$ e sob a hipótese nula $H_0 : \beta = \beta_0$, obtem-se $\ell(\beta_0, \hat{\lambda}_0)$, onde $\hat{\lambda}_0$ é o estimador de máxima verossimilhança de λ quando $\beta = \beta_0$. A estatística da razão de verossimilhança para testar $H_0 : \beta = \beta_0$ é então dada por

$$\begin{aligned} W &= 2 [\ell(\hat{\beta}, \hat{\lambda}) - \ell(\beta_0, \hat{\lambda}_0)] \\ &= n \left\{ \log \left[\frac{S_{xx} + 2\beta_0 S_{xy} + \beta_0^2 S_{yy}}{\beta_0^2 + 1} \right] + \frac{\beta_0^2 S_{xx} - 2\beta_0 S_{xy} + S_{yy}}{\sigma^2(\beta_0^2 + 1)} \right. \\ &\quad \left. - \log \left[\frac{S_{xx} + 2\hat{\beta} S_{xy} + \hat{\beta}^2 S_{yy}}{\hat{\beta}^2 + 1} \right] - \frac{\hat{\beta}^2 S_{xx} - 2\hat{\beta} S_{xy} + S_{yy}}{\sigma^2(\hat{\beta}^2 + 1)} \right\}, \end{aligned} \quad (3.5)$$

que tem distribuição assintótica qui-quadrado com um grau de liberdade (que denotamos por χ_1^2). Wong (1989) usando resultados gerais para correções da estatística da razão de verossimilhança obtidos por Lawley (1956) (ver Seção 3.4) mostra que

$$E(W) = 1 - \frac{2\sigma^2}{n(\lambda_2 - \sigma^2)} + \frac{2\sigma^4}{n(\lambda_2 - \sigma^2)^2} + O(n^{-2}), \quad (3.6)$$

de modo que a distribuição da estatística

$$W^* = \frac{W}{1 - \frac{2\sigma^2}{n(\hat{\lambda}_2 - \sigma^2)} + \frac{2\sigma^4}{n(\hat{\lambda}_2 - \sigma^2)^2}}$$

é mais próxima da distribuição χ_1^2 do que a distribuição da estatística dada em (3.5). Em (3.6), $O(n^{-2})$ indica termos de ordem n^{-2} .

3.2.2 Variância σ^2 desconhecida

Derivando a função de log-verossimilhança (2.4) em relação a λ_0 , λ_1 , λ_2 , λ_3 e β obtem-se os estimadores (3.2), (3.3) e

$$\hat{\lambda}_3 = \frac{\hat{\beta}^2 S_{xx} - 2\hat{\beta} S_{xy} + S_{yy}}{\hat{\beta}^2 + 1}. \quad (3.7)$$

De forma análoga à anterior, um intervalo de confiança para β com coeficiente de confiança $1 - \alpha$ tem a forma (3.4) mas agora com

$$i_{\beta\beta} = \frac{n(\lambda_2 - \lambda_3)^2}{\lambda_2 \lambda_3 (\beta^2 + 1)^2}. \quad (3.8)$$

Substituindo os estimadores de máxima verossimilhança dados em (3.2), (3.3) e (3.7) em (2.4) com λ_3 no lugar de σ^2 , de forma análoga à seção anterior, pode-se escrever a estatística da razão de verossimilhança para testar $H_0 : \beta = \beta_0$, após uma série de manipulações algébricas, como

$$W = n \left\{ \log \left[\frac{(\beta_0^2 S_{xx} - 2\beta_0 S_{xy} + S_{yy})(S_{xx} + 2\beta_0 S_{xy} + \beta_0^2 S_{yy})}{(\beta_0^2 + 1)^2} \right] - \log \left[\frac{(\hat{\beta}^2 S_{xx} - 2\hat{\beta} S_{xy} + S_{yy})(S_{xx} + 2\hat{\beta} S_{xy} + \hat{\beta}^2 S_{yy})}{(\hat{\beta}^2 + 1)^2} \right] \right\}, \quad (3.9)$$

que tem distribuição assintótica qui-quadrado com um grau de liberdade. Wong (1989) mostra, neste caso, que

$$E(W) = 1 + \frac{2,5}{n} + O(n^{-2}), \quad (3.10)$$

de modo que a estatística

$$W^* = \frac{W}{(1 + \frac{2,5}{n})} \quad (3.11)$$

tem distribuição mais próxima da distribuição χ_1^2 que a distribuição de W dada em (3.9).

3.2.3 A função de verossimilhança perfilada condicional

Nesta seção consideramos a verossimilhança perfilada condicional (ou corrigida) para o modelo estrutural com ϕ conhecido. Esta verossimilhança é penalizada com o fator de correção dado em Cox e Reid (1987) e é obtida para o parâmetro

de interesse β , que é ortogonal aos parâmetros de perturbação $\lambda_0, \lambda_1, \lambda_2$ e λ_3 . Segundo Cox e Reid (1987), a log-verossimilhança perfilada condicional é dada por

$$\ell_{cp}(\beta) = \ell_p(\beta) - C(\beta), \quad (3.12)$$

onde $\hat{\lambda}_\beta$ é o estimador de máxima verossimilhança de $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$ dado β , $\ell_p(\beta) = \ell(\beta, \hat{\lambda}_\beta)$ é a log-verossimilhança perfilada obtida a partir da log-verossimilhança ℓ definida em (2.4) quando σ^2 é conhecida e por (2.4) com λ_3 no lugar de σ^2 , quando este for desconhecido, e

$$C(\beta) = \frac{1}{2} \log \left[\det \left(\mathbf{J}(\beta, \hat{\lambda}_\beta) \right) \right], \quad (3.13)$$

onde $\mathbf{J}(\beta, \hat{\lambda}_\beta)$ é a matriz de informação observada de λ dado β . Uma das vantagens da verossimilhança perfilada condicional é que ela produz em geral estimadores mais eficientes que os correspondentes estimadores obtidos a partir da verossimilhança usual (Cox e Reid, 1987).

Variância σ^2 conhecida

No caso em que σ^2 é conhecido, pode-se mostrar que

$$C(\beta) = \frac{1}{2} \log \left[\frac{n^3(\beta^2 + 1)^3}{2\sigma^2(S_{xx} + 2\beta S_{xy} + \beta^2 S_{yy})^3} \right], \quad (3.14)$$

de modo que a estatística da razão baseada na $\ell_{cp}(\beta)$ para testar $H_0 : \beta = \beta_0$, com β_0 conhecido é dada por

$$W_c = 2 \left[\ell_{cp}(\tilde{\beta}) - \ell_{cp}(\beta_0) \right], \quad (3.15)$$

onde $\ell_{cp}(\tilde{\beta})$ é definida em (3.12) com $C(\beta)$ dado em (3.14),

$$\ell_p(\beta) = n \left\{ \log \left(\frac{S_{xx} + 2\beta S_{xy} + \beta^2 S_{yy}}{\beta^2 + 1} \right) + \frac{\beta^2 S_{xx} - 2\beta S_{xy} + S_{yy}}{\sigma^2(\beta^2 + 1)} \right\}$$

$-\frac{n}{2}$

e

$$\tilde{\beta} = \sup_{\beta} \{ \ell_{cp}(\beta) \}.$$

Portanto, de (3.15) temos que

$$W_c = W - C(\tilde{\beta}) + C(\beta_0), \quad (3.16)$$

com W dado em (3.5), que tem distribuição assintótica qui-quadrado com um grau de liberdade (Cox e Reid, 1987). Neste caso $\tilde{\beta}$ é obtido por métodos numéricos, de modo que não temos uma representação explícita para W_c .

Variância σ^2 desconhecida

No caso em que σ^2 é desconhecido, temos que

$$C(\beta) = \frac{1}{2} \log \left[\frac{n^4}{4\lambda_2^3(\beta)\lambda_3^3(\beta)} \right],$$

com

$$\lambda_2(\beta) = \frac{S_{xx} + 2\beta S_{xy} + \beta^2 S_{yy}}{\beta^2 + 1} \quad (3.17)$$

e

$$\lambda_3(\beta) = \frac{\beta^2 S_{xx} - 2\beta S_{xy} + S_{yy}}{\beta^2 + 1}. \quad (3.18)$$

Como, neste caso,

$$\ell_p(\beta) = \underbrace{n}_{-n/2} \{ \log [\lambda_2(\beta)\lambda_3(\beta)] \},$$

temos que a estatística da razão de verossimilhança obtida a partir da verossimilhança perfilada condicional (3.12) é dada por

$$\begin{aligned} W_c &= 2 \left[\ell_{cp}(\tilde{\beta}) - \ell_{cp}(\beta_0) \right], \\ &= W - C(\tilde{\beta}) + C(\beta_0), \end{aligned} \quad (3.19)$$

onde W é dado em (3.9), Neste caso, segue que (Wong, 1989)

$$W_c = \frac{n-3}{n}W,$$

que tem distribuição assintótica χ_1^2 , com W dada por (3.9). De (3.10) temos que

$$\begin{aligned} E(W_c) &= \frac{n-3}{n}E(W) \\ &= 1 - \frac{0,5}{n} + O(n^{-2}), \end{aligned}$$

o que prova que a distribuição de W_c é mais próxima da distribuição χ_1^2 do que a distribuição de W . Neste caso, considerando

$$W_c^* = \frac{W_c}{\left(1 - \frac{0,5}{n}\right)},$$

pode-se mostrar que $W_c^* = W^* + O_p(n^{-2})$, que tem distribuição assintótica qui-quadrado com 1 grau de liberdade.

3.3 Uma população com coeficiente de atenuação k_x conhecido

Nesta seção consideramos o MEV estrutural com o coeficiente de atenuação, k_x , conhecido. Para este caso duas transformações ortogonais foram apresentadas no Capítulo 2. A primeira dada em (2.12), para quando μ e α são conhecidos ($\alpha = 0$, sem perda de generalidade) e a segunda dada por (2.18)-(2.19) para o caso em que μ e α são desconhecidos. Com a verossimilhança obtida a partir desta transformação ortogonal, consideramos a estimação de máxima verossimilhança dos parâmetros do modelo e intervalos de confiança para o parâmetro β . Testes para $H_0 : \beta = \beta_0$ são também derivados utilizando-se a verossimilhança usual e a verossimilhança perfilada condicional.

3.3.1 Parâmetros μ e α conhecidos

Sem perda de generalidade consideramos $\alpha = 0$. A transformação ortogonal (2.12) substituída na função de log-verossimilhança obtida a partir de uma amostra aleatória de tamanho n de (2.15) nos fornece a log-verossimilhança (2.13). Derivando (2.13) com relação aos parâmetros do modelo, obtem-se os seguintes estimadores de máxima verossimilhança:

$$\hat{\beta} = \frac{\sum_i^n (X_i - \mu)Y_i + (k+1)\mu \sum_i^n Y_i}{S_{x\mu}(k+1)^{-1} + 2\mu \sum_{i=1}^n (X_i - \mu) + n[(k+1)\mu]^2}, \quad (3.20)$$

$$\hat{\lambda}_0 = \frac{k_x \hat{\beta}^2 S_{x\mu} - 2\hat{\beta} \sum_i^n (X_i - \mu)(Y_i - \mu\hat{\beta}) + (k+1) \sum_i^n (Y_i - \mu\hat{\beta})^2}{n} \quad (3.21)$$

e

$$\hat{\lambda}_1 = k_x S_{x\mu} n^{-1}, \quad (3.22)$$

onde $S_{x\mu} = \sum_{i=1}^n (X_i - \mu)^2$. Note que se $\mu = 0$, temos de (3.20) que

$$\hat{\beta} = \frac{(k+1) \sum_i^n X_i Y_i}{\sum_i^n X_i^2}, \quad (3.23)$$

que é igual ao estimador de mínimos quadrados quando $k = 0$, ou seja, quando não há erro de medida em X_i . De (2.14) segue que

$$i_{\beta\beta}^{-1} = \frac{\lambda_0}{n[\lambda_1 + (k+1)^2 \mu^2]},$$

é a variância assintótica do estimador de máxima verossimilhança. Portanto, como em (3.4), um intervalo de confiança aproximado para β é dado por $(\hat{\beta} \mp z_{\alpha/2} i_{\beta\beta}^{-1/2})$, com coeficiente de confiança $1 - \alpha$.

A verossimilhança perfilada para o modelo acima pode ser escrita como

$$\ell_p(\beta) = -n \log(2\pi) - \frac{n}{2} \log(\lambda_0(\beta)), \quad (3.24)$$

onde

$$\lambda_0(\beta) = \left\{ k_x \beta^2 S_{x\mu} - 2\beta \sum_i (X_i - \mu)(Y_i - \mu\beta) + (k+1) \sum_i (Y_i - \mu\beta)^2 \right\} / n \quad (3.25)$$

e $S_{x\mu} = \sum_{i=1}^n (X_i - \mu)^2$. Após algumas manipulações algébricas, temos que a estatística da razão de verossimilhança que segue de (3.24) é igual a

$$\begin{aligned} W &= 2 \left\{ \ell_p(\hat{\beta}) - \ell_p(\beta_0) \right\} \\ &= n \log \left[\frac{k_x \beta_0^2 S_{x\mu} - 2\beta_0 \sum_i (X_i - \mu)(Y_i - \mu\beta_0) + (k+1) \sum_i (Y_i - \mu\beta_0)^2}{k_x \hat{\beta}^2 S_{x\mu} - 2\hat{\beta} \sum_i (X_i - \mu)(Y_i - \mu\hat{\beta}) + (k+1) \sum_i (Y_i - \mu\hat{\beta})^2} \right], \end{aligned} \quad (3.26)$$

que tem distribuição assintótica qui-quadrado com um grau de liberdade.

Pode-se mostrar sem dificuldade que a verossimilhança perfilada condicional para este modelo é igual a

$$\ell_{cp}(\beta) = \ell_p(\beta) = -n \log(2\pi) - \frac{(n-2)}{2} \log(\lambda_0(\beta)), \quad (3.27)$$

onde, neste caso,

$$\lambda_0(\beta) = k_x \beta^2 S_{x\mu} - 2\beta S_{xy} + (k+1) S_{yy}. \quad (3.28)$$

A estatística da razão de verossimilhança usando (3.27) é dada por

$$\begin{aligned} W_c &= 2 \left\{ \ell_{cp}(\hat{\beta}) - \ell_{cp}(\beta_0) \right\} \\ &= \frac{(n-2)}{n} W, \end{aligned} \quad (3.29)$$

que tem distribuição assintótica qui-quadrado com um grau de liberdade.

3.3.2 Parâmetros μ e α desconhecidos

A transformação ortogonal (2.18)-(2.19) substituída na função log-verossimilhança obtida a partir de uma amostra aleatória de tamanho n de (2.15) nos fornece a log-verossimilhança (2.20). Derivando (2.20) em relação aos parâmetros (β, λ) ,

onde $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$ e igualando as derivadas a zero, tem-se os seguintes estimadores de máxima verossimilhança:

$$\hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{k_x \sum_{i=1}^n (X_i - \bar{X})^2}, \quad (3.30)$$

$$\hat{\lambda}_0 = k_x \hat{\beta}^2 S_{xx} - 2\hat{\beta} S_{xy} + (k+1)S_{yy}, \quad (3.31)$$

$$\hat{\lambda}_1 = k_x S_{xx}, \quad (3.32)$$

$$\hat{\lambda}_2 = \bar{X}, \quad \hat{\lambda}_3 = \bar{Y}, \quad (3.33)$$

onde, neste caso,

$$S_{xx} = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S_{xy} = \sum_{i=1}^n n^{-1} (X_i - \bar{X})(Y_i - \bar{Y}) \quad e \quad S_{yy} = \sum_{i=1}^n n^{-1} (Y_i - \bar{Y})^2. \quad (3.34)$$

De (2.21), segue que a variância assintótica de $\hat{\beta}$ é dada por

$$i_{\beta\beta}^{-1} = \frac{\lambda_0}{n\lambda_1}. \quad (3.35)$$

O estimador (3.30) foi originalmente proposto por Fuller (1987), usando argumentos de não viciosidade. Utilizando a transformação ortogonal (2.18)-(2.19) segue então que $\hat{\beta}$ é na verdade um estimador de máxima verossimilhança para β .

Como em (3.4), para n grande,

$$\left(\hat{\beta} \mp z_{\alpha/2} \left(\frac{\hat{\lambda}_0}{n\hat{\lambda}_1} \right)^{1/2} \right),$$

é um intervalo de confiança aproximado para β com coeficiente de confiança $1 - \alpha$.

A estatística da razão de verossimilhança para testar $H_0 : \beta = \beta_0$ que segue de (3.24) e (3.25) é igual a

$$\begin{aligned} W &= 2 \{ \ell_p(\hat{\beta}) - \ell_p(\beta_0) \} \\ &= n \log \left[\frac{k_x \beta_0^2 S_{xx} - 2\beta_0 S_{xy} + (k+1)S_{yy}}{k_x \hat{\beta}^2 S_{xx} - 2\hat{\beta} S_{xy} + (k+1)S_{yy}} \right], \end{aligned} \quad (3.36)$$

onde S_{xx} , S_{yy} e S_{xy} são respectivamente as variâncias e covariância definidas em (3.34). A distribuição assintótica de W é qui-quadrado com um grau de liberdade. Neste caso, temos que a verossimilhança perfilada condicional é dada por

$$\ell_{cp}(\beta) = \ell_p(\beta) = -n \log(2\pi) - \frac{(n-3)}{2} \log(\lambda_0(\beta)), \quad (3.37)$$

onde

$$\lambda_0(\beta) = k_x \beta^2 S_{xx} - 2\beta S_{xy} + (k+1)S_{yy},$$

com S_{xx} , S_{xy} e S_{yy} definidos em (3.34). A estatística da razão de verossimilhança que segue de (3.37) é dada por

$$\begin{aligned} W_c &= 2 \{ \ell_p(\hat{\beta}) - \ell_p(\beta_0) \} \\ &= (n-3) \log \left[\frac{k_x \beta_0^2 S_{xx} - 2\beta_0 S_{xy} + (k+1)S_{yy}}{k_x \hat{\beta}^2 S_{xx} - 2\hat{\beta} S_{xy} + (k+1)S_{yy}} \right] \\ &= \frac{(n-3)}{n} W, \end{aligned} \quad (3.38)$$

onde W é dado em (3.36). A distribuição assintótica de W_c também é qui-quadrado com um grau de liberdade.

3.4 O fator de correção de Bartlett no caso k_x conhecido

Seja $\ell(\theta)$ o log da função de verossimilhança que depende de um parâmetro $\theta = (\theta_1, \dots, \theta_{p+q})$. Usando a notação considerada por Lawley (1956), seja

$$\begin{aligned} \ell_r &= \frac{\partial \ell}{\partial \theta_r}, & \ell_{rs} &= \frac{\partial^2 \ell}{\partial \theta_r \partial \theta_s}, & \ell_{rst} &= \frac{\partial^3 \ell}{\partial \theta_r \partial \theta_s \partial \theta_t}, \quad \dots \\ v_{rs} &= E(\ell_{rs}), & v_{rst} &= E(\ell_{rst}), & \dots \\ (v_{rs})_t &= \frac{\partial (v_{rs})}{\partial \theta_t}, & (v_{rst})_u &= \frac{\partial (v_{rst})}{\partial \theta_u}, & (v_{rs})_{tu} &= \frac{\partial^2 (v_{rs})}{\partial \theta_t \partial \theta_u}, \quad \dots \end{aligned} \quad (3.39)$$

e assim por diante. Portanto, temos $p + q$ parâmetros no modelo e suponhamos que a hipótese nula H_0 especifica que $\theta_{p+1} = \dots = \theta_{p+q} = 0$ de modo que sob H_0 , temos p parâmetros no modelo. Sendo $\hat{\ell}_p$ a função de verossimilhança maximizada sob H_0 e $\hat{\ell}_{p+q}$ a função de verossimilhança calculada no estimador de máxima verossimilhança, $\hat{\theta}$, Lawley (1956) mostra, no caso globalmente ortogonal que

$$\mathbb{E}(W) = 2\mathbb{E}\{\hat{\ell}_{p+q} - \hat{\ell}_p\} = q + \epsilon_p + O(n^{-2}),$$

onde

$$\epsilon_p = \epsilon_{p+q} - \epsilon_q,$$

com

$$\epsilon_k = A_k + B_k, \quad k = p, p + q, \quad (3.40)$$

sendo

$$A_k = \sum_{r,s=1}^k \left\{ \frac{1}{4} v_{rrss} - (v_{rrs})_s + (v_{rs})_{rs} \right\} / [v_{rr} v_{ss}]$$

e

$$B_k = \sum_{r,s,t=1}^k \frac{\frac{1}{6} v_{rst}^2 + \frac{1}{4} v_{rss} v_{rtt} - v_{rst} (v_{rs})_t - v_{rss} (v_{rt})_t + (v_{rs})_t (v_{rt})_s + (v_{rs})_s (v_{rt})_t}{[-v_{rr} v_{ss} v_{tt}]}.$$

Conforme estabelecido por Lawley (1956), a distribuição da estatística

$$W^* = \frac{2(\hat{\ell}_{p+q} - \hat{\ell}_p)}{1 + \frac{\epsilon_{p+q} - \epsilon_p}{q}}$$

é mais próxima da distribuição χ_q^2 que a distribuição de $W = 2(\hat{\ell}_{p+q} - \hat{\ell}_p)$. O fator $1 + \frac{\epsilon_{p+q} - \epsilon_p}{q}$ é conhecido como fator de correção de Bartlett. Expandindo este fator em Taylor, podemos também representar a estatística corrigida como

$$W^* = 2(\hat{\ell}_{p+q} - \hat{\ell}_p) \left[1 - \frac{(\epsilon_{p+q} - \epsilon_p)}{q} \right].$$

Com esta representação o fator de correção é dado por

$$1 - \frac{(\epsilon_{p+q} - \epsilon_p)}{q}.$$

No caso em que ϵ_{p+q} e ϵ_p envolvem parâmetro desconhecido, estes são substituídos pelos seus estimadores de máxima verossimilhança. A parametrização ortogonal global permite utilizar (3.40), onde para maior clareza, denota-se $\theta = (\theta^{(1)}, \theta^{(2)})$, onde θ é um vetor de parâmetros de dimensão $(p+q) \times 1$, $\theta^{(1)}$ é de dimensão $p \times 1$ e $\theta^{(2)}$ é de dimensão $q \times 1$ com $\theta^{(1)}$ ortogonal à $\theta^{(2)}$. O problema de interesse é então testar a hipótese $H_0 : \theta^{(1)} = \theta_0$ contra $H_1 : \theta^{(1)} \neq \theta_0$. Sob H_1 , a função de verossimilhança tem $p+q$ parâmetros e sob H_0 o número de parâmetros é reduzido para p . Lembramos que para a maioria dos modelos, principalmente nos modelos com erros nas variáveis, é bastante árdua a obtenção dos fatores de correção de Bartlett. Em algumas situações especiais, é possível a obtenção de expressões bastante gerais para os fatores de correção (ver por exemplo, Cordeiro, 1983, onde expressões gerais são obtidas para fatores de correção nos modelos lineares generalizados). Uma expressão para o fator de correção no caso geral é também dada em Lawley (1956).

3.4.1 Parâmetros μ e α conhecidos

Note que podemos escrever a verossimilhança (2.13) na forma

$$\ell(\lambda, \beta) = -n \log(2\pi) - \frac{n}{2} \log(\lambda_0 \lambda_1) - \sum_{i=1}^n \frac{a_i^2}{\lambda_1(k+1)} - \frac{1}{\lambda_0} P(\beta), \quad (3.41)$$

onde

$$a_i = (X_i - \mu) \text{ e } P(\beta) = c_0 + c_1\beta + c_2\beta^2, \quad (3.42)$$

com

$$c_0 = \frac{k+1}{2} \sum_{i=1}^n X_i^2, \quad c_1 = - \sum_{i=1}^n a_i X_i - \mu(k+1) \sum_{i=1}^n X_i \quad (3.43)$$

e

$$c_2 = \frac{\sum_{i=1}^n a_i^2}{2(k+1)} + \mu \sum_{i=1}^n a_i + \frac{n\mu^2(k+1)}{2}. \quad (3.44)$$

Para calcularmos as expressões que aparecem em (3.39) é necessário calcular as esperanças de c_0 , c_1 e c_2 que aparecem em (3.43) e (3.44), ou seja,

$$\mathbb{E}(c_0) = \frac{n}{2}[\lambda_0 + \lambda_1\beta^2 + (k+1)\beta^2\mu^2], \quad (3.45)$$

$$\mathbb{E}(c_1) = -n\beta[\lambda_1 + \mu^2(k+1)], \quad \mathbb{E}(c_2) = \frac{n}{2}[\lambda_1 + \mu^2(k+1)], \quad (3.46)$$

$$\mathbb{E}(P(\beta)) = \frac{n\lambda_0}{2}. \quad (3.47)$$

As equações (3.45), (3.46) e (3.47) permitem um cálculo mais simples para as expressões em (3.39), e conseqüentemente, facilitam o cálculo de ϵ_k , $k = p, p+q$ dados em (3.40). Após uma longa série de manipulações algébricas mostra-se para a hipótese $H_0 : \beta = \beta_0$, que o fator de Bartlett é dado por

$$\epsilon_{p+q} - \epsilon_p = \epsilon_3 - \epsilon_2 = \frac{1,5}{n}.$$

Considerando o termo de ordem n^{-1} , segue portanto que

$$\begin{aligned} \mathbb{E}(W) &= q + (\epsilon_{p+q} - \epsilon_p) \\ &= 1 + \frac{1,5}{n}. \end{aligned} \quad (3.48)$$

Então a estatística W corrigida com (3.48), que denotamos por W^* , é igual a

$$W^* = \frac{W}{\left(1 + \frac{1,5}{n}\right)}, \quad (3.49)$$

cuja distribuição está mais próxima da distribuição qui-quadrado com um grau de liberdade do que a distribuição de W .

De (3.29) e (3.48) temos, até a ordem n^{-1} , que

$$\mathbb{E}(W_c) = 1 - \frac{0,5}{n}.$$

Note que W_c tem distribuição com média mais próxima da distribuição χ_1^2 que a média da distribuição de W . Podemos também corrigir W_c definindo

$$W_c^* = \frac{W_c}{\left(1 - \frac{0,5}{n}\right)}, \quad (3.50)$$

cujas distribuição está próxima da distribuição qui-quadrado com um grau de liberdade, para valores grandes de n . Note também que $W_c^* = W^* + O_p(n^{-2})$. Estudos de simulação apresentados na Ilustração numérica 3.3 mostram que a estatística corrigida (3.49) é bastante eficiente, mesmo para n pequeno.

3.4.2 Parâmetros μ e α desconhecidos

Podemos escrever a verossimilhança (2.20) na forma

$$\ell(\lambda, \beta) = -n \log(2\pi) - \frac{n}{2} \log(\lambda_0 \lambda_1) - \sum_{i=1}^n \frac{(X_i - \lambda_2)^2}{2\lambda_1(k+1)} - \frac{1}{\lambda_0} P(\beta), \quad (3.51)$$

onde

$$P(\beta) = d_0 + d_1\beta + d_2\beta^2, \quad (3.52)$$

com

$$d_0 = \frac{k+1}{2} \sum_{i=1}^n (Y_i - \lambda_3)^2, \quad d_1 = - \sum_{i=1}^n (X_i - \lambda_2)(Y_i - \lambda_3), \quad e \quad d_2 = \frac{1}{2(k+1)} \sum_{i=1}^n (X_i - \lambda_2)^2. \quad (3.53)$$

No cálculo das expressões (3.40) é necessário obtermos as esperanças de (3.52) e (3.53), que são dadas por:

$$\mathbb{E}(d_0) = \frac{n}{2}(\lambda_0 + \lambda_1\beta^2), \quad (3.54)$$

$$\mathbb{E}(d_1) = -n\lambda_1\beta, \quad \mathbb{E}(d_2) = \frac{n\lambda_1}{2} \quad (3.55)$$

e

$$\mathbb{E}(P(\beta)) = \frac{n\lambda_0}{2}. \quad (3.56)$$

As equações (3.54), (3.55) e (3.56) simplificam as expressões que aparecem em (3.39), facilitando o cálculo de ϵ_k , $k = p, p + q$ dados em (3.40). Para a hipótese $H_0 : \beta = \beta_0$,

$$\epsilon_{p+q} - \epsilon_p = A_{p+q} + B_{p+q} - (A_p + B_p), \quad (3.57)$$

onde $p + q = 5$ e $p = 4$. Nesta situação, o cálculo de (3.40) requer a análise de 230 termos e portanto, bem mais trabalhoso do que situação em que μ e α são conhecidos. Após uma série de manipulações algébricas, tem-se que

$$\epsilon_5 = A_5 + B_5 = \frac{37}{6n}, \quad (3.58)$$

$$\epsilon_4 = A_4 + B_4 = \frac{25}{6n} - \frac{\lambda_1 \beta_0^2}{n(\lambda_0 + \lambda_1 \beta_0^2)} + \frac{\lambda_0^2 + \lambda_1^2 \beta_0^4}{n(\lambda_0 + \lambda_1 \beta_0^2)^2}, \quad (3.59)$$

de modo que

$$\epsilon_5 - \epsilon_4 = \frac{2}{n} + \frac{\lambda_1 \beta_0^2}{n(\lambda_0 + \lambda_1 \beta_0^2)} - \frac{\lambda_0^2 + \lambda_1^2 \beta_0^4}{n(\lambda_0 + \lambda_1 \beta_0^2)^2}. \quad (3.60)$$

Considerando o termo de ordem n^{-1} , como em (3.48), segue que

$$\begin{aligned} \mathbb{E}(W) &= q + (\epsilon_{p+q} - \epsilon_p) \\ &= 1 + (\epsilon_5 - \epsilon_4). \end{aligned} \quad (3.61)$$

A estatística W dada em (3.36) corrigida com (3.61), denotada por W^* , ou seja,

$$W^* = \frac{W}{1 + \epsilon_5 - \epsilon_4}, \quad (3.62)$$

tem distribuição mais próxima da distribuição qui-quadrado assintótica com um grau de liberdade do que a distribuição de W . Um estudo do comportamento de W^* é apresentado na Ilustração numérica 3.2. Neste estudo são também apresentados resultados numéricos que ilustram a eficiência da correção na aproximação de $\mathbb{E}(W)$.

3.5 O teste de Wald para uma população

Conforme veremos a seguir o teste de Wald tem em muitos casos a vantagem de ser mais simples do ponto de vista computacional do que o teste da razão de verossimilhança, quer seja este obtido a partir da verossimilhança corrigida ou não corrigida. A estatística da razão de verossimilhança utiliza os estimadores de máxima verossimilhança dos parâmetros sob a hipótese nula, que em geral são difíceis de serem obtidos do ponto de vista numérico. No caso do modelo estrutural, os estimadores de máxima verossimilhança usados na estatística de Wald possuem expressões explícitas e de fácil cálculo. A parametrização ortogonal contribui bastante para esta simplificação numérica. A estatística de Wald para testar a hipótese $H_0 : \mathbf{r}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \mathbf{0}$, onde $\mathbf{r}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ é um vetor de funções e $\mathbf{R}$ uma matriz associada à hipótese nula, tem a forma (Buse, 1982),

$$W_a = [\mathbf{r}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})]' [\mathbf{R}\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})\mathbf{R}']^{-1} [\mathbf{r}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})], \quad (3.63)$$

onde $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\lambda}}$ são, respectivamente, os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ e $\boldsymbol{\lambda}$, e o número de graus de liberdade é igual ao número de funções do vetor $\mathbf{r}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ impostas pela hipótese nula. Nesta seção denotamos por W_a a estatística de Wald para a hipótese $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, considerando σ^2 conhecido, σ^2 desconhecido e k_x conhecido.

3.5.1 Variância σ^2 conhecida

Se $\phi = 1$ e a variância σ^2 é conhecida, para testar a hipótese $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ temos que $\mathbf{r}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \boldsymbol{\beta} - \boldsymbol{\beta}_0$, $\mathbf{R} = (1, 0, 0, 0)$ e $\mathbf{R}\mathbf{I}^{-1}(\boldsymbol{\beta}, \boldsymbol{\lambda})\mathbf{R}' = i_{\beta\beta}^{-1}$ é obtido de (2.5). Então, de (3.63) podemos escrever

$$W_a = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' i_{\beta\beta}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$

$$= \frac{n(\hat{\lambda}_2 - \sigma^2)^2(\hat{\beta} - \beta_0)^2}{\hat{\lambda}_2 \sigma^2 (\hat{\beta}^2 + 1)^2}, \quad (3.64)$$

onde $\hat{\beta}$ segue de (3.3). A distribuição assintótica de W_a é qui-quadrado com um grau de liberdade, conforme mencionado acima. Note que o teste de Wald é mais simples de ser implementado numericamente. É importante notar também que, a estatística do teste de Wald, coincide com o quadrado da estatística do teste obtida a partir da distribuição assintótica do estimador de máxima verossimilhança $\hat{\beta}$, considerado na Seção 3.2.1.

3.5.2 Variância σ^2 desconhecida

Se $\phi = 1$ e a variância σ^2 é desconhecida, para a hipótese $H_0 : \beta = \beta_0$ tem-se que $\mathbf{r}(\beta, \boldsymbol{\lambda}) = \mathbf{r}(\beta, \lambda_0, \lambda_1, \lambda_2, \lambda_3) = \beta - \beta_0$, $\mathbf{R} = (1, 0, 0, 0, 0)$ e $\mathbf{R}\mathbf{I}^{-1}(\beta, \boldsymbol{\lambda})\mathbf{R}' = i_{\beta\beta}^{-1}$ dado em (2.8). Então, de (3.63),

$$\begin{aligned} W_a &= (\hat{\beta} - \beta_0)^2 i_{\beta\beta} \\ &= \frac{n(\hat{\lambda}_2 - \hat{\lambda}_3)^2 (\hat{\beta} - \beta_0)^2}{\hat{\lambda}_2 \hat{\lambda}_3 (\hat{\beta}^2 + 1)^2}, \end{aligned} \quad (3.65)$$

onde $\hat{\beta}$ segue de (3.3) e a distribuição assintótica de W_a é qui-quadrado com um grau de liberdade (Buse, 1982). Neste caso, a estatística do teste de Wald também coincide com o quadrado da estatística do teste que segue da distribuição assintótica de $\hat{\beta}$ para testar $H_0 : \beta = \beta_0$.

3.5.3 Coeficiente de atenuação conhecido

Se a média μ for conhecida e $\alpha = 0$, seguindo a notação dada em (3.63), tem-se para $H_0 : \beta = \beta_0$ que o vetor de funções é dado por $\mathbf{r}(\beta, \boldsymbol{\lambda}) = \beta - \beta_0$ e o

vetor $\mathbf{R} = (1, 0, 0)$. Então,

$$[\mathbf{R}\mathbf{I}^{-1}(\boldsymbol{\beta}, \boldsymbol{\lambda})\mathbf{R}'] = i_{\beta\beta}^{-1},$$

onde $i_{\beta\beta}$ é o elemento da matriz de informação associado a β , dado em (2.14).

Portanto, segue de (3.63) que

$$\begin{aligned} W_a &= (\hat{\beta} - \beta_0)^2 \hat{i}_{\beta\beta} \\ &= n(\hat{\beta} - \beta_0)^2 \left[\frac{\hat{\lambda}_1 + \mu^2(k+1)}{\hat{\lambda}_0} \right], \end{aligned} \quad (3.66)$$

onde $\hat{i}_{\beta\beta}$ é o estimador de máxima verossimilhança de $i_{\beta\beta}$ e $\hat{\beta}$ segue de (3.20). W_a tem distribuição assintótica qui-quadrado com um grau de liberdade. Também, neste caso, o teste de Wald é bem mais simples de ser implementado do que o teste da razão de verossimilhança, e pode ser obtido a partir da distribuição assintótica de $\hat{\beta}$.

Se a média μ e o intercepto α forem desconhecidos tem-se, para a hipótese $H_0 : \beta = \beta_0$, o vetor de funções $\mathbf{r}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \beta - \beta_0$ e a matriz $\mathbf{R} = (1, 0, 0, 0, 0)$. Então, ocorre mais uma vez que

$$[\mathbf{R}\mathbf{I}^{-1}(\boldsymbol{\beta}, \boldsymbol{\lambda})\mathbf{R}'] = i_{\beta\beta},$$

onde $i_{\beta\beta}$ é o elemento da matriz de informação associado à β , dado em (2.21).

Seguindo a equação (3.63), após um cálculo simples, tem-se que a estatística de Wald para este teste é dada por

$$\begin{aligned} W_a &= (\hat{\beta} - \beta_0)^2 \hat{i}_{\beta\beta} \\ &= n(\hat{\beta} - \beta_0)^2 \frac{\hat{\lambda}_1}{\hat{\lambda}_0}, \end{aligned} \quad (3.67)$$

com $\hat{\beta}$ dado em (3.30). A distribuição assintótica de W_a é qui-quadrado com um grau de liberdade.

Tabela 3.1: Aceitação de H_0 em 1000 amostras, $\alpha = 0,05$

n	N_1	N_2	N_3	N_4
05	839	913	977	774
10	913	942	957	891
20	944	955	961	942
30	927	935	937	927
50	940	947	948	943
100	952	954	955	940

Ilustração numérica 3.1

Foram geradas 1000 amostras aleatórias de tamanho n de acordo com o MEV estrutural com razão de variância $\phi = 1$, $\tau^2 = 1,0$, $\beta = \beta_0 = 1,5$, $\mu = 30,0$, $\alpha = 0,5$ e $\sigma^2 = 1,0$. Para cada amostra gerada, calculou-se a estatística da razão de verossimilhança W dada em (3.9), a estatística da razão de verossimilhança corrigida com o fator de Bartlett W^* dada em (3.11), a estatística da razão de verossimilhança corrigida W_c dada em (3.19) e a estatística de Wald W_a dada em (3.65). Determinou-se o número de vezes em que aceita-se a hipótese nula $H_0 : \beta = \beta_0$, com $\beta_0 = 1,5$ e σ^2 desconhecido, aos níveis 0,05 e 0,10. Denotamos este número por N_1 para W , N_2 para W^* , N_3 para W_c e N_4 para W_a . Os resultados desta ilustração numérica estão apresentados nas Tabelas 3.1 e 3.2. Note que para n pequeno, os níveis de significância observados com as estatísticas W^* e W_c estão mais próximos dos níveis teóricos do que os níveis observados com as estatísticas W e W_a . Para n grande o comportamento das quatro estatísticas é bem próximo.

□

Tabela 3.2: Aceitação de H_0 em 1000 amostras, $\alpha = 0,10$

n	N_1	N_2	N_3	N_4
05	743	858	946	730
10	853	896	913	849
20	892	911	924	896
30	886	895	900	891
50	890	896	900	887
100	894	900	901	886

Ilustração numérica 3.2

Foram geradas 1000 amostras aleatórias de tamanho n de acordo com o MEV estrutural com $k_x = 2/3$, $\mu = 30,0$, $\alpha = 0,5$, $\beta = 1,5$ e $\sigma_u^2 = 1,0$. Para cada amostra gerada, calculou-se a estatística da razão de verossimilhança W dada em (3.36), a estatística da razão de verossimilhança corrigida com o fator de Bartlett W^* , dada em (3.62), a estatística da razão de verossimilhança corrigida W_c , dada em (3.38) e a estatística de Wald W_a dada em (3.67). Determinou-se o número de vezes em que se aceita a hipótese nula $H_0 : \beta = \beta_0$, com $\beta_0 = 1,5$ aos níveis 0,05 e 0,10 e μ e α desconhecidos. Denotamos este número por N_1 para W , N_2 para W^* , N_3 para W_c e N_4 para W_a , respectivamente. Para ilustrar a precisão da correção de Bartlett neste caso, calculou-se também estimadores de $\mathbb{E}(W)$, $\mathbb{E}(W^*)$ e $\mathbb{E}(W_c)$ dados, respectivamente, por $\hat{\mathbb{E}}(W) = \sum_{i=1}^{1000} W_i/1000$, $\hat{\mathbb{E}}(W^*) = \sum_{i=1}^{1000} W_i^*/1000$ e $\hat{\mathbb{E}}(W_c) = \sum_{i=1}^{1000} (W_c)_i/1000$. Os resultados desta ilustração numérica estão apresentados nas Tabelas 3.3 e 3.4. Note que para n pequeno, os níveis de significância observados com a estatística corrigida W^* , são bem mais

Tabela 3.3: Aceitação de H_0 em 1000 amostras, $\alpha = 0,05$

n	N_1	N_2	N_3	N_4	$\hat{IE}(W)$	$\hat{IE}(W^*)$	$\hat{IE}(W_c)$
05	847	955	979	770	1,86804	0,98801	0,74625
10	907	946	957	878	1,36038	1,05210	0,95214
20	940	957	965	930	1,07135	0,95171	0,90888
30	944	948	951	935	1,07146	0,99192	0,96394
50	943	949	949	934	1,08159	1,03353	1,01567

Tabela 3.4: Aceitação de H_0 em 1000 amostras, $\alpha = 0,10$

n	N_1	N_2	N_3	N_4
05	763	911	948	721
10	853	895	907	815
20	877	902	910	865
30	897	899	910	892
50	890	901	902	882

próximos dos níveis teóricos do que os níveis observados com a estatística da razão de verossimilhança sem correção. O comportamento de W_c também é bastante satisfatório. Note também que, para n pequeno, $\hat{IE}(W^*)$ e $\hat{IE}(W_c)$ estão bem mais próximos de 1 do que $\hat{IE}(W)$. Para n grande, o comportamento das estatísticas é bastante próximo.

Tabela 3.5: Aceitação de H_0 em 1000 amostras, $\alpha = 0,10$

n	N_1	N_2	N_3	N_4
05	835	886	935	794
10	892	907	913	885
20	899	908	911	893
30	885	897	903	868
50	903	905	908	898

Ilustração numérica 3.3

Neste exemplo foram geradas 1000 amostras aleatórias de tamanho n satisfazendo ao modelo estrutural com coeficiente de atenuação $k_x = 0,75$, $\sigma_u^2 = 1,0$, $\beta = 4,0$, $\mu = 0,0$ e $\alpha = 0,0$. Como na ilustração numérica anterior, para cada amostra gerada calculou-se W dada em (3.26), W^* dada em (3.49), W_c dada em (3.30) e W_a dada em (3.66). Foi determinado o número de vezes em que a hipótese nula $H_0 : \beta = \beta_0$, $\beta_0 = 4,0$, com μ e α conhecidos é aceita aos níveis $\alpha = 0,10$ e $\alpha = 0,05$. O número de vezes em que a hipótese nula acima é aceita nas 1000 amostras geradas é denotado, respectivamente, por N_1 para W , N_2 para W^* , N_3 para W_c e N_4 para W_a . Os valores N_j , $j = 1, \dots, 4$, estão apresentados nas Tabelas 3.5 e 3.6 (que se encontram no final do capítulo, disposição esta escolhida pelo editor)).

A correção de Bartlett para W melhora sensivelmente a precisão do teste, aproximando-se mesmo para n pequeno, dos valores nominais. Nota-se também que este estudo de simulação parece indicar que o teste da razão de verossimilhança é melhor que o teste de Wald para valores pequenos de n . A estatística de Wald

W_a apesar de computacionalmente mais simples, apresenta um comportamento não muito satisfatório para valores pequenos de n . A medida que n aumenta, ambas estatísticas aproximam-se rapidamente do nível nominal do teste, o que é esperado. O comportamento da estatística da razão de verossimilhança corrigida também é bastante satisfatório.

3.6 Inferência para s populações com razão de variâncias ϕ_i , $i = 1, \dots, s$, conhecidas

Nesta Seção, que é uma extensão da Seção 3.2 (uma população), consideramos o modelo (2.22), analisado sob duas situações diferentes: σ_i^2 conhecido e σ_i^2 desconhecido, $i = 1, \dots, s$.

3.6.1 Variâncias σ_i^2 , $i = 1, \dots, s$, conhecidas

Considere uma amostra aleatória de tamanho n_i de (X_{ij}, Y_{ij}) como em (2.23), onde σ_i^2 , $i = 1, \dots, s$, são conhecidos e distintos, de tamanhos n_i , $i = 1, \dots, s$, onde $n' = \sum_{i=1}^s n_i$. A transformação ortogonal (2.24) torna β_i ortogonal à $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i})$, $i = 1, \dots, n$. Substituindo (2.24) na função de log-verossimilhança original, obtem-se com $\beta = (\beta_1, \dots, \beta_s)$ e $\lambda = (\lambda_{0i}, \dots, \lambda_{0s}, \lambda_{11}, \dots, \lambda_{1s}, \lambda_{21}, \dots, \lambda_{2s})$,

$$\ell(\beta, \lambda) = \sum_{i=1}^s \ell_i(\beta_i, \lambda_{0i}, \lambda_{1i}, \lambda_{2i}), \quad (3.68)$$

onde

$$\begin{aligned} \ell_i(\beta_i, \lambda_{0i}, \lambda_{1i}, \lambda_{2i}) &= -n_i \log(2\pi) - \frac{n_i}{2} \log(\sigma_i^2) - \frac{n_i}{2} \log \lambda_{2i} - \left[2\sigma_i^2 \lambda_{2i} (\beta^2 + 1) \right]^{-1} \\ &\times \left\{ (\beta^2 \lambda_{2i} + \sigma_i^2) \sum_{i,k} (X_{ik} - \lambda_{0i})^2 - 2\beta(\lambda_{2i} - \sigma_i^2) \sum_{i,k} (X_{ik} \right. \end{aligned}$$

$$- \lambda_{0i})(Y_{ik} - \lambda_{1i}) + (\lambda_{2i} + \beta^2 \sigma_i^2) \sum_{i,k} (Y_{ik} - \lambda_{1i})^2 \Big\}. \quad (3.69)$$

Derivando (3.68) em relação aos parâmetros β_i , λ_{0i} , λ_{1i} , e λ_{2i} e igualando estas derivadas à zero, obtem-se os seguintes estimadores de máxima verossimilhança:

$$\begin{aligned} \hat{\lambda}_{0i} &= \bar{X}_i = \sum_{j=1}^{n_i} n_i^{-1} X_{ij}, \\ \hat{\lambda}_{1i} &= \bar{Y}_i = \sum_{j=1}^{n_i} n_i^{-1} Y_{ij}, \\ \hat{\lambda}_{2i} &= \frac{S_{xx}^{(i)} + 2\hat{\beta}_i S_{xy}^{(i)} + \hat{\beta}_i^2 S_{yy}^{(i)}}{\hat{\beta}_i^2 + 1}, \end{aligned} \quad (3.70)$$

com $\hat{\beta}_i$ dado pela solução da equação

$$\hat{\beta}_i^2 S_{xy}^{(i)} + \hat{\beta}_i (S_{xx}^{(i)} - S_{yy}^{(i)}) - S_{xy}^{(i)} = 0, \quad (3.71)$$

onde

$$S_{xx}^{(i)} = \sum_{j=1}^{n_i} n_i^{-1} (X_{ij} - \bar{X}_i)^2, \quad S_{xy}^{(i)} = \sum_{j=1}^{n_i} n_i^{-1} (X_{ij} - \bar{X}_i)(Y_{ij} - \bar{Y}_i) \quad e \quad S_{yy}^{(i)} = \sum_{j=1}^{n_i} n_i^{-1} (Y_{ij} - \bar{Y}_i)^2, \quad (3.72)$$

$i = 1, \dots, s$.

Um intervalo de confiança aproximado para β_i pode ser obtido usando o elemento $i_{\beta_i \beta_i}$ da matriz de informação (2.25), ou seja,

$$i_{\beta_i \beta_i} = \frac{n_i (\lambda_{2i} - \sigma_i^2)^2}{\lambda_{2i} \sigma_i^2 (\beta_i^2 + 1)^2}, \quad (3.73)$$

$i = 1, \dots, s$, de modo que, como nas seções anteriores,

$$\left(\hat{\beta}_i \mp z_{\alpha/2} \hat{i}_{\beta_i \beta_i}^{-1/2} \right),$$

é um intervalo de confiança aproximado para β_i com coeficiente de confiança $1 - \alpha$, $i = 1, \dots, s$. Sob a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$, os estimadores de máxima verossimilhança de λ_{0i} , λ_{1i} e λ_{2i} são como em (3.70), mas com $\tilde{\beta}$, o estimador de

máxima verossimilhança de β , satisfazendo à equação

$$\sum_{i=1}^s n_i \frac{[S_{xx}^{(i)} + 2\tilde{\beta}S_{xy}^{(i)} + \tilde{\beta}^2S_{yy}^{(i)} - \sigma_i^2(\tilde{\beta} + 1)] [\tilde{\beta}S_{xx}^{(i)} + (\tilde{\beta}^2 - 1)S_{xy}^{(i)} - \tilde{\beta}S_{yy}^{(i)}]}{\sigma_i^2 [S_{xx}^{(i)} + 2\tilde{\beta}S_{xy}^{(i)} + \tilde{\beta}^2S_{yy}^{(i)}]} = 0. \quad (3.74)$$

No caso particular em que $\mu_i = \mu$, $\sigma_i^2 = \sigma^2$, e $\tau_i^2 = \tau^2$, $i = 1, \dots, s$, $\tilde{\beta}$ é obtido a partir da solução da equação

$$A\tilde{\beta}^2 + B\tilde{\beta} - A = 0, \quad (3.75)$$

onde

$$A = \sum_{i=1}^s \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})(X_{ij} - \bar{X})n^{-1} \quad e \quad B = \sum_{i=1}^s \sum_{j=1}^{n_i} [(X_{ij} - \bar{X})^2 - (Y_{ij} - \bar{Y})^2] n^{-1},$$

ou seja, $\tilde{\beta}$ é da forma (3.3), mas agora com o mesmo sinal de A.

O teste da razão de verossimilhança e o fator de correção de Bartlett

Considerando testes da razão de verossimilhança para a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, onde β_0 é especificado, tem-se que a estatística da razão de verossimilhança para esta hipótese é dada por

$$W = 2 \sum_{i=1}^s [\ell_i(\beta_0, \hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i}) - \ell_i(\hat{\beta}_i, \hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i})], \quad (3.76)$$

onde ℓ_i é definida por (3.69). Após algumas manipulações algébricas, tem-se que a estatística da razão de verossimilhança é dada por (Wong, 1991)

$$\sum_{i=1}^s n_i \left\{ \log [f_i(\beta_0)] + \frac{g_i(\beta_0)}{\sigma_i^2} - \log [f_i(\hat{\beta}_i)] - \frac{g_i(\hat{\beta}_i)}{\sigma_i^2} \right\}, \quad (3.77)$$

onde

$$f_i(\beta) = \frac{S_{xx}^{(i)} + 2\beta S_{xy}^{(i)} + \beta^2 S_{yy}^{(i)}}{\beta^2 + 1} \quad (3.78)$$

e

$$g_i(\beta) = \frac{\beta^2 S_{xx}^{(i)} - 2\beta S_{xy}^{(i)} + S_{yy}^{(i)}}{\beta^2 + 1}, \quad (3.79)$$

$i = 1, \dots, s$. A distribuição assintótica de W é qui-quadrado com s graus de liberdade.

Sob a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$, onde β não é especificado, a estatística da razão de verossimilhança é dada por

$$W = \sum_{i=1}^s n_i \left\{ \log [f_i(\tilde{\beta})] + \frac{g_i(\tilde{\beta})}{\sigma_i^2} - \log [f_i(\hat{\beta}_i)] - \frac{g_i(\hat{\beta}_i)}{\sigma_i^2} \right\}, \quad (3.80)$$

onde $\tilde{\beta}$ satisfaz a (3.74) e f_i, g_i dados por (3.78) e (3.79). Neste caso, a distribuição assintótica de W é qui-quadrado com $s - 1$ graus de liberdade. A correção de Bartlett para W é considerada em Wong (1991), onde mostra-se que

$$E(W) = s - 1 - D + O(n_*^{-2}), \quad (3.81)$$

com $n_* = \min \{n_1, \dots, n_s\}$ e

$$D = \sum_{i=1}^s \left\{ \frac{2\sigma_i^2}{n_i(\lambda_{2i} - \sigma_i^2)} + \frac{\sigma_i^4}{2n_i(\lambda_{2i} - \sigma_i^2)^2} \right\} + \frac{1 + \sum_{i=1}^s [(\lambda_{2i} - 3\sigma_i^2)/\lambda_{2i}]}{\sum_{i=1}^s [n_i(\lambda_{2i} - \sigma_i^2)^2/(\lambda_{2i}\sigma_i^2)]} + \frac{\sum_{i=1}^s [n_i(\lambda_{2i} - \sigma_i^2)^2/\lambda_{2i}^2]}{2 \left[\sum_{i=1}^s \{n_i(\lambda_{2i} - \sigma_i^2)^2/(\lambda_{2i}\sigma_i^2)\} \right]^2}.$$

Portanto a estatística

$$W^* = \frac{W}{1 - \frac{D}{s-1}}, \quad (3.82)$$

tem distribuição mais próxima da distribuição qui-quadrado com $s - 1$ graus de liberdade do que a distribuição de W . Em (3.82), os parâmetros desconhecidos são substituídos por seus estimadores de máxima verossimilhança.

O Teste de Wald

Conforme mencionado anteriormente, a estatística de Wald para testar as hipóteses $H_0 : \beta_1 = \dots = \beta_s = \beta_0$ e $H_0 : \beta_1 = \dots = \beta_s = \beta$, com β não especificado,

apresenta a vantagem numérica de utilizar apenas os estimadores dos parâmetros sob o modelo geral, evitando-se assim resolver a equação (3.74).

Para a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, onde β_0 é não especificado, seguindo a notação da Seção 3.4 temos que $[r(\beta, \lambda)]' = (\beta_1 - \beta_0, \dots, \beta_s - \beta_0)$, onde $\mathbf{R}$ é uma matriz de dimensão $s \times 4s$ associada à hipótese nula. Após algumas simplificações, a estatística deste teste é dada por

$$W_a = \sum_{i=1}^s \hat{i}_{\beta_i \beta_i} (\hat{\beta}_i - \beta_0)^2, \quad (3.83)$$

com distribuição aproximada qui-quadrado com s graus de liberdade, onde $\hat{i}_{\beta_i \beta_i}$ é dado por (3.73) e $\hat{i}_{\beta_i \beta_i}$ é seu estimador de máxima verossimilhança calculado sob o modelo geral. Portanto o teste de Wald neste caso é bem mais simples de ser implementado que o teste da razão de máxima verossimilhança. Note também que a estatística (3.83), com $s = 2$, coincide com o quadrado da estatística do teste obtido através da distribuição assintótica de $\hat{\beta}_1$ e $\hat{\beta}_2$ para testar $H_0 : \beta_1 = \beta_2 = \beta_0$.

Para a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$ tem-se que o vetor $[r(\beta, \lambda)]' = (\beta_1 - \beta_2, \beta_2 - \beta_3, \dots, \beta_{s-1} - \beta_s)$ e $\mathbf{R}$ é uma matriz $(s-1) \times 4s$, de modo que $\mathbf{R}\mathbf{I}^{-1}(\beta, \lambda)\mathbf{R}' = \mathbf{A}$, onde $\mathbf{A}$ tem dimensão $(s-1) \times (s-1)$ dada por

$$\mathbf{A} = \begin{bmatrix} \hat{i}_{\beta_1 \beta_1}^{-1} + \hat{i}_{\beta_2 \beta_2}^{-1} & -\hat{i}_{\beta_2 \beta_2}^{-1} & 0 & 0 & \dots & 0 & 0 \\ -\hat{i}_{\beta_2 \beta_2}^{-1} & \hat{i}_{\beta_2 \beta_2}^{-1} + \hat{i}_{\beta_3 \beta_3}^{-1} & -\hat{i}_{\beta_3 \beta_3}^{-1} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \hat{i}_{\beta_{s-2} \beta_{s-2}}^{-1} + \hat{i}_{\beta_{s-1} \beta_{s-1}}^{-1} & -\hat{i}_{\beta_{s-1} \beta_{s-1}}^{-1} \\ 0 & 0 & 0 & 0 & \dots & -\hat{i}_{\beta_{s-1} \beta_{s-1}}^{-1} & \hat{i}_{\beta_{s-1} \beta_{s-1}}^{-1} + \hat{i}_{\beta_s \beta_s}^{-1} \end{bmatrix}. \quad (3.84)$$

De (3.63), a estatística de Wald é dada por

$$W_a = (\hat{\beta}_1 - \hat{\beta}_2, \dots, \hat{\beta}_{s-1} - \hat{\beta}_s)' \hat{\mathbf{A}}^{-1} (\hat{\beta}_1 - \hat{\beta}_2, \dots, \hat{\beta}_{s-1} - \hat{\beta}_s), \quad (3.85)$$

onde $\hat{\mathbf{A}}$ é formada com os estimadores de máxima verossimilhança dos elementos da matriz $\mathbf{A}$. A distribuição assintótica de W_a é qui-quadrado com $s - 1$ graus de liberdade. Se $s = 2$ há uma expressão explícita para (3.85), ou seja,

$$W_a = \frac{(\hat{\beta}_1 - \hat{\beta}_2)^2}{\hat{i}_{\beta_1\beta_1}^{-1} + \hat{i}_{\beta_2\beta_2}^{-1}}, \quad (3.86)$$

com distribuição assintótica qui-quadrado com um grau de liberdade, sendo $i_{\beta_i\beta_i}$ dado por (3.73) e $\hat{\beta}_i$ por (3.71), $i = 1, \dots, n$.

3.6.2 Variâncias σ_i^2 , $i = 1, \dots, s$, desconhecidas

Considere uma amostra aleatória de tamanho n_i de (X_{ij}, Y_{ij}) , $i = 1, \dots, s$, $j = 1, \dots, n_i$, como em (2.23), onde σ_i^2 , $i = 1, \dots, s$, são desconhecidos e distintos e $n' = \sum_{i=1}^s n_i$. A parametrização ortogonal (2.27) torna β_i ortogonal á $(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i})$ para todo i . Substituindo (2.27) na função log-verossimilhança original, obtem-se com $\beta = (\beta_1, \dots, \beta_s)$ e $\lambda = (\lambda_{01}, \dots, \lambda_{0s}, \lambda_{11}, \dots, \lambda_{1s}, \lambda_{21}, \dots, \lambda_{2s}, \lambda_{31}, \dots, \lambda_{3s})$,

$$\ell(\beta, \lambda) = \sum_{i=1}^s \ell_i(\beta_i, \lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i}), \quad (3.87)$$

onde

$$\begin{aligned} \ell_i(\beta_i, \lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i}) &= -n_i \log(2\pi) - \frac{n_i}{2} \log(\lambda_{2i}\lambda_{3i}) - \left[2\lambda_{2i}\lambda_{3i}(\beta_i^2 + 1)\right]^{-1} \\ &\times \left\{ (\beta_i^2\lambda_{2i} + \lambda_{3i}) \sum_{i,k} (X_{ik} - \lambda_{0i})^2 - 2\beta_i(\lambda_{2i} - \lambda_{3i}) \sum_{i,k} (X_{ik} \right. \\ &\left. - \lambda_{0i})(Y_{ik} - \lambda_{1i}) + (\lambda_{2i} + \beta_i^2\lambda_{3i}) \sum_{i,k} (Y_{ik} - \lambda_{1i})^2 \right\}, \quad (3.88) \end{aligned}$$

$i = 1, \dots, s$.

Derivando (3.87) em relação aos parâmetros β_i , λ_{0i} , λ_{1i} , λ_{2i} , e λ_{3i} e igualando estas derivadas à zero, obtem-se os estimadores de máxima verossimilhança $\hat{\lambda}_{0i}$, $\hat{\lambda}_{1i}$, $\hat{\lambda}_{3i}$ e $\hat{\beta}_i$ que são dados, respectivamente, por (3.70) e (3.71), ou seja são os

mesmos do caso σ_i^2 conhecido e o estimador de máxima verossimilhança de λ_{3i} é dado por

$$\hat{\lambda}_{3i} = \frac{\hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + S_{yy}^{(i)}}{\hat{\beta}_i^2 + 1}, \quad (3.89)$$

$i = 1, \dots, s$. Um intervalo de confiança aproximado para β_i pode ser obtido usando o elemento $i_{\beta_i\beta_i}$ da matriz de informação (2.28), onde

$$i_{\beta_i\beta_i} = \frac{n_i(\lambda_{2i} - \lambda_{3i})^2}{\lambda_{2i}\lambda_{3i}(\beta_i^2 + 1)^2}, \quad (3.90)$$

$i = 1, \dots, s$. Então, um intervalo de confiança para β_i tem a mesma expressão que o intervalo (3.4) com $\hat{i}_{\beta_i\beta_i}$ dado em (3.90).

Sob a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$, os estimadores de máxima verossimilhança de λ_{0i} , λ_{1i} , λ_{2i} , e λ_{3i} são como dados em (3.70), (3.71), (3.89) e $\tilde{\beta}$, o estimador de máxima verossimilhança de β sob H_0 , satisfaz à equação

$$\sum_{i=1}^s n_i \frac{[(\tilde{\beta}^2 + 1)S_{xx}^{(i)} - 4\tilde{\beta}S_{xy}^{(i)} - (\tilde{\beta}^2 - 1)S_{yy}^{(i)}] [\tilde{\beta}S_{xx} + (\tilde{\beta}^2 - 1)S_{xy}^{(i)} - \tilde{\beta}S_{yy}^{(i)}]}{f_i(\tilde{\beta})g_i(\tilde{\beta})} = 0, \quad (3.91)$$

onde as funções f_i , g_i , $i = 1, \dots, s$, estão definidas em (3.78) e (3.79). Se $\mu_i = \mu$, $\sigma_i^2 = \sigma^2$ e $\tau_i^2 = \tau^2$, $i = 1, \dots, s$, $\tilde{\beta}$ é calculado a partir de (3.75).

O teste da razão de verossimilhança e da razão de verossimilhança condicional estendida

Considerando a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, onde β_0 é especificado tem-se que a estatística da razão de verossimilhança para esta hipótese é dada por

$$W = 2 \sum_{i=1}^s \left[\ell_i(\beta_0, \hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i}, \hat{\lambda}_{3i}) - \ell_i(\hat{\beta}_i, \hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i}, \hat{\lambda}_{3i}) \right], \quad (3.92)$$

onde ℓ_i é definida por (3.88). Após algumas manipulações algébricas, tem-se que a estatística da razão de verossimilhança (Wong, 1991) é dada por

$$W = \sum_{i=1}^s n_i \left\{ \log [f_i(\beta_0)g_i(\beta_0)] - \log [f_i(\hat{\beta}_i)g_i(\hat{\beta}_i)] \right\}, \quad (3.93)$$

onde f_i, g_i , $i = 1, \dots, s$, são dados em (3.78) e (3.79) e $\hat{\beta}_i$, $i = 1, \dots, s$, por (3.71). A distribuição assintótica de W é qui-quadrado com s graus de liberdade. Wong (1991), mostra que

$$\mathbb{E}(W) = s + \sum_{i=1}^s \frac{2,5}{n_i} + O(n_*^{-2}),$$

onde $n_* = \min \{n_1, \dots, n_s\}$, de modo que corrigindo W dada em (3.93) tem-se que

$$W^* = \frac{W}{1 + \sum_{i=1}^s \frac{2,5}{sn_i}}. \quad (3.94)$$

Portanto, W^* tem distribuição mais próxima da distribuição qui-quadrado com s graus de liberdade que a distribuição de W .

A seguir vamos introduzir uma nova estatística para testar $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, que denominamos estatística da razão de verossimilhança condicional estendida. Ela é motivada pela estatística da razão de verossimilhança condicional, W_c , dada na Seção 3.3.1 e é definida por

$$W_c = \sum_{i=1}^s \left(\frac{n_i - 3}{n_i} \right) W^{(i)}, \quad (3.95)$$

onde

$$W^{(i)} = n_i \left\{ \log [f_i(\beta_0)g_i(\beta_0)] - \log [f_i(\hat{\beta}_i)g_i(\hat{\beta}_i)] \right\},$$

com f_i, g_i , dados em (3.78) e (3.79), $i = 1, \dots, s$. Note que podemos calcular, até termos de ordem n_*^{-2} , o valor esperado de W_c pois cada $W^{(i)}$ é equivalente a W dado na equação (3.9). De (3.10), temos então que

$$\mathbb{E}(W_c) = \sum_{i=1}^s \left(\frac{n_i - 3}{n_i} \right) \mathbb{E}(W^{(i)})$$

$$\begin{aligned}
&= \sum_{i=1}^s \left(\frac{n_i - 3}{n_i} \right) \left(1 + \frac{2,5}{n_i} \right) + O(n_*^{-2}) \\
&= \sum_{i=1}^s \left(1 - \frac{0,5}{n_i} \right) + O(n_*^{-2}) \\
&= s - \sum_{i=1}^s \frac{0,5}{n_i} + O(n_*^{-2}).
\end{aligned} \tag{3.96}$$

Isto mostra que a média da distribuição de W_c é mais próxima da média da χ_s^2 do que a média da distribuição de W dada em (3.92). A correção de Bartlett para (3.95) é dada por

$$W_c^* = \frac{W_c}{\left(1 - \sum_{i=1}^s \frac{0,5}{sn_i} \right)}, \tag{3.97}$$

com distribuição assintótica χ_s^2 .

Sob a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$, onde β é não especificado, a estatística da razão de verossimilhança para esta hipótese é dada por

$$W = \sum_{i=1}^s n_i \left\{ \log [f_i(\tilde{\beta})g_i(\tilde{\beta})] - \log [f_i(\hat{\beta}_i)g_i(\hat{\beta}_i)] \right\}, \tag{3.98}$$

com distribuição assintótica χ_{s-1}^2 . O fator de Bartlett para a correção de (3.98) (Wong, 1991) é dado por

$$\begin{aligned}
D = \frac{1}{s-1} & \left(\sum_{i=1}^s \frac{5}{n_i} + \frac{4 - \sum_{i=1}^s [(3\lambda_{2i}^2 - 2\lambda_{2i}\lambda_{3i} + 3\lambda_{3i}^2)/(\lambda_{2i}\lambda_{3i})]}{\sum_{i=1}^s [n_i(\lambda_{2i} - \lambda_{3i})^2/(\lambda_{2i}\lambda_{3i})]} \right. \\
& \left. + \frac{\sum_{i=1}^s [n_i(\lambda_{2i} - \lambda_{3i})^4/(\lambda_{2i}\lambda_{3i})]}{2 \left[\sum_{i=1}^s [n_i(\lambda_{2i} - \lambda_{3i})^2/(\lambda_{2i}\lambda_{3i})] \right]^2} \right).
\end{aligned}$$

O Teste de Wald

Para a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, onde β_0 é especificado, seguindo a notação da seção 3.5, tem-se que $[r(\beta, \lambda)]' = (\beta_1 - \beta_0, \dots, \beta_s - \beta_0)$ e a matriz $\mathbf{R}$ associada à hipótese nula é de dimensão $s \times 5s$. Seguindo a equação (3.63), após

algum cálculo, concluí-se que a estatística deste teste é dada por

$$\begin{aligned} W_a &= \sum_{i=1}^s \hat{i}_{\beta_i \beta_i} (\hat{\beta}_i - \beta_0)^2, \\ &= \sum_{i=1}^n \frac{(\hat{\lambda}_{2i} - \hat{\lambda}_{3i})^2}{\hat{\lambda}_{2i} \hat{\lambda}_{3i} (\hat{\beta}_i^2 + 1)^2} (\hat{\beta}_i - \beta_0)^2. \end{aligned} \quad (3.99)$$

Para a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$ tem-se que o vetor $[\mathbf{r}(\boldsymbol{\beta}, \boldsymbol{\lambda})]' = (\beta_1 - \beta_2, \beta_2 - \beta_3, \dots, \beta_{s-1} - \beta_s)$, $\mathbf{R}$ é uma matriz de dimensão $(s-1) \times 5s$ e $\mathbf{R}\mathbf{I}^{-1}(\boldsymbol{\beta}, \boldsymbol{\lambda})\mathbf{R}' = \mathbf{A}_*$, onde $\mathbf{A}_*$ tem os mesmos elementos que $\mathbf{A}$ dada em (3.84), mas com $i_{\beta_i \beta_i}$ dada por (3.90). De (3.63) segue-se que a estatística deste teste é dada por

$$W_a = (\hat{\beta}_1 - \hat{\beta}_2, \dots, \hat{\beta}_{s-1} - \hat{\beta}_s)' \hat{\mathbf{A}}_*^{-1} (\hat{\beta}_1 - \hat{\beta}_2, \dots, \hat{\beta}_{s-1} - \hat{\beta}_s), \quad (3.100)$$

sendo a matriz $\hat{\mathbf{A}}_*$ formada com os estimadores de máxima verossimilhança dos elementos da matriz $\mathbf{A}_*$ e $\hat{\beta}_i$, $i = 1, \dots, s$, são dados por (3.71). A distribuição assintótica de W_a é qui-quadrado com $s-1$ graus de liberdade.

Se $s = 2$, a estatística de Wald para a hipótese acima tem a forma explícita

$$W_a = \frac{(\hat{\beta}_1 - \hat{\beta}_2)^2}{\hat{i}_{\beta_1 \beta_1}^{-1} + \hat{i}_{\beta_2 \beta_2}^{-1}}, \quad (3.101)$$

com distribuição assintótica qui-quadrado com um grau de liberdade.

Ilustração numérica 3.4

Foram geradas 1000 amostras aleatórias de tamanho $n_1 = n_2 = n$ de acordo com o MEV estrutural com razão de variância $\phi_i = 1$, para o caso em que o parâmetro σ_i^2 é desconhecido e $s = 2$, com $\tau_i^2 = 1, 0$, $\beta_i = 1, 5$, $i = 1, 2$, $\mu_1 = 30, 0$, $\mu_2 = 40, 0$, $\alpha_1 = 0, 5$, $\alpha_2 = 3, 0$, $\sigma_1^2 = \sigma_2^2 = 1, 0$. Para cada amostra, calculou-se a estatística da razão de verossimilhança W dada em (3.93), a estatística da razão de verossimilhança corrigida com o fator de Bartlett, W^* dada em (3.94), a estatística da razão de verossimilhança condicional estendida W_c dada em (3.95) e

a estatística de Wald, W_a dada em (3.99). Determinou-se o número de vezes em que aceita-se a hipótese nula $H_0 : \beta_1 = \beta_2 = \beta_0$, com $\beta_0 = 1, 5$, nas 1000 amostras geradas, aos níveis 0,05 e 0,10. Este número é denotado, respectivamente, por N_1 para W , N_2 para W^* , N_3 para W_c e N_4 para W_a . Os resultados desta ilustração numérica estão apresentados nas Tabelas 3.7 e 3.8. Note que para n pequeno, os níveis de significância observados com as estatísticas W^* e W_c estão mais próximos dos níveis nominais do que os níveis observados com as estatísticas W e W_a . O comportamento de W_a não é muito satisfatório para n pequeno.

3.7 Inferência para s populações com coeficientes de atenuação k_{x_i} , $i = 1, \dots, s$, conhecidos

O modelo com erro nas variáveis com coeficientes de atenuação conhecidos é analisado aqui em dois casos, como anteriormente. O primeiro supõe $\alpha_i = 0$ e a média μ_i , $i = 1, \dots, s$, conhecida, enquanto que o segundo, mais geral, considera α_i e μ_i desconhecidas, $i = 1, \dots, s$.

3.7.1 Parâmetros μ_i e α_i conhecidos, $i = 1, \dots, s$.

Sem perda de generalidade, tomamos $\alpha_i = 0$. Sob a suposição de que o coeficiente de atenuação $k_{x_i} = \tau_i^2 / (\tau_i^2 + \sigma_{ui}^2)$, correspondente ao i -ésimo modelo é conhecido, $i = 1, \dots, s$, a distribuição conjunta de (X_{ij}, Y_{ij}) é normal bivariada com vetor de médias $(\mu_i, \beta_i \mu_i)$ e matriz de covariâncias

$$\begin{pmatrix} \tau_i^2(k_i + 1) & \beta_i \tau_i^2 \\ \beta_i \tau_i^2 & \beta_i^2 \tau_i^2 + \sigma_{ei}^2 \end{pmatrix},$$

onde $k_i = (1 - k_{x_i})/k_{x_i}$, $i = 1, \dots, s$. Assumindo que μ_i é conhecido e $\alpha_i = 0$, $i = 1, \dots, s$, a transformação ortogonal considerada em (2.12) estende-se sem dificuldades para esta situação mais geral, obtendo-se a transformação ortogonal dada em (2.39). Tomando uma amostra aleatória de tamanho n_i de (X_{ij}, Y_{ij}) , $i = 1, \dots, s$, $j = 1, \dots, n_i$, com $n' = \sum_{i=1}^s n_i$, obtem-se a função de verossimilhança original (não transformada). Substituindo (2.39) nesta verossimilhança, tem-se com $\beta = (\beta_1, \dots, \beta_s)$ e $\lambda = (\lambda_{01}, \dots, \lambda_{0s}, \lambda_{11}, \dots, \lambda_{1s})$,

$$\ell(\lambda, \beta) = \sum_{i=1}^s \ell_i(\lambda_{0i}, \lambda_{1i}, \beta_i), \quad (3.102)$$

onde

$$\begin{aligned} \ell_i(\lambda_{0i}, \lambda_{1i}, \beta_i) = & -n_i \log(2\pi) - \frac{n_i}{2} \log(\lambda_{0i} \lambda_{1i}) - \frac{1}{\lambda_{0i} \lambda_{1i}} \left\{ \frac{\lambda_{0i} + \lambda_{1i} \beta_i^2}{k_i + 1} \right. \\ & \times \sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2 - 2\beta_i \lambda_{1i} \sum_{j=1}^{n_i} (X_{ij} - \mu_i)(Y_{ij} - \beta_i \mu_i) \\ & \left. + \lambda_{1i} (k_i + 1) \sum_{j=1}^{n_i} (Y_{ij} - \beta_i \mu_i)^2 \right\}. \end{aligned} \quad (3.103)$$

De (3.102), a matriz de informação de (λ, β) , denotada por $\mathbf{I}(\lambda, \beta)$ é dada por

$$\mathbf{I}(\lambda, \beta) = \text{diag}(\mathbf{I}_1(\lambda_{01}, \lambda_{11}, \beta_1), \dots, \mathbf{I}_s(\lambda_{0s}, \lambda_{1s}, \beta_s)), \quad (3.104)$$

onde

$$\mathbf{I}_i(\lambda_{0i}, \lambda_{1i}, \beta_i) = \begin{pmatrix} \frac{n_i}{2\lambda_{0i}^2} & 0 & 0 \\ 0 & \frac{n_i}{2\lambda_{1i}^2} & 0 \\ 0 & 0 & \frac{n_i [\lambda_{1i} + \mu_i^2 (k_i + 1)^2]}{\lambda_{0i}} \end{pmatrix} \quad (3.105)$$

$i = 1, \dots, s$.

Derivando (3.102) em relação à β_i , λ_{0i} e λ_{1i} , $i = 1, \dots, s$, temos que os estimadores de máxima verossimilhança são dados, respectivamente, por

$$\hat{\beta}_i = \frac{\sum_{j=1}^{n_i} Y_{ij} (X_{ij} - \mu_i) - (k_i + 1) \mu_i \sum_{j=1}^{n_i} Y_{ij}}{[S_{xx}^{(i)} / (k_i + 1)] + 2\mu_i \sum_{j=1}^{n_i} (X_{ij} - \mu_i) + n_i \mu_i^2}, \quad (3.106)$$

$$\hat{\lambda}_{1i} = \frac{S_{xx}^{(i)}}{k_i + 1}, \quad (3.107)$$

$$\hat{\lambda}_{0i} = (k_i + 1)^{-1} \hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (1 + k_i) S_{yy}^{(i)}, \quad (3.108)$$

onde, neste caso,

$$\begin{aligned} S_{xx}^{(i)} &= n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2, \\ S_{yy}^{(i)} &= n_i^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\beta}_i \mu_i)^2, \\ S_{xy}^{(i)} &= n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \mu_i)(Y_{ij} - \hat{\beta}_i \mu_i), \end{aligned} \quad (3.109)$$

$i = 1, \dots, s$. Usando-se (3.104) pode-se determinar um intervalo de confiança aproximado para $\hat{\beta}_i$, ou seja, de (3.104) tem-se que a variância assintótica de $\hat{\beta}_i$ é

$$i_{\beta_i \beta_i}^{-1} = \frac{\lambda_{0i}}{n_i [\lambda_{1i} + \mu_i (k_i + 1)^2]} \quad (3.110)$$

e sendo $\hat{i}_{\beta_i \beta_i}^{-1}$ o seu estimador de máxima verossimilhança, tem-se de forma análoga às anteriores que um intervalo de confiança para β_i é dado por

$$\left(\hat{\beta}_i \mp z_{\alpha/2} [\hat{i}_{\beta_i \beta_i}]^{-1/2} \right),$$

$i = 1, \dots, s$. De forma análoga, no caso em que $\beta_1 = \dots = \beta_s = \beta$, um intervalo de confiança aproximado pode ser dado para β calculando $i_{\beta\beta}$, o elemento da matriz de informação associado à β . Este elemento é calculado a partir de (3.102) tendo β no lugar de β_i , $i = 1, \dots, s$, e é igual a

$$i_{\beta\beta} = \sum_{i=1}^s i_{\beta_i \beta_i}, \quad (3.111)$$

onde $i_{\beta_i \beta_i}$ é dado em (3.110). Temos então de (3.111) que um intervalo de confiança aproximado para β com coeficiente de confiança $1 - \gamma$ é dado por

$$\left(\hat{\beta} \mp z_{\alpha/2} [\hat{i}_{\beta\beta}]^{-1/2} \right).$$

A estatística da razão de verossimilhança

Sob a hipótese nula $H_0 : \beta_1 = \dots = \beta_s = \beta$, os estimadores de máxima verossimilhança de β , λ_{0i} e λ_{1i} são tais que

$$\hat{\beta} = \frac{\sum_{i=1}^s \sum_{j=1}^{n_i} (X_{ij} - \mu_i)(\hat{\lambda}_{0i}\hat{\lambda}_{1i})^{-1}Y_{ij} + (k_i + 1) \sum_{i=1}^s n_i \mu_i (\hat{\lambda}_{0i}\hat{\lambda}_{1i})^{-1}\bar{Y}_{.i}}{\sum_{i=1}^s \frac{S_{xx}^{(i)}}{(k_i + 1)\hat{\lambda}_{0i}\hat{\lambda}_{1i}} + 2 \sum_{i=1}^s \frac{(X_{ij} - \mu_i)\mu_i}{\hat{\lambda}_{0i}\hat{\lambda}_{1i}} + \sum_{i=1}^s \frac{(k_i + 1)n_i \mu_i^2}{\hat{\lambda}_{0i}\hat{\lambda}_{1i}}}, \quad (3.112)$$

$$\hat{\lambda}_{0i} = (k_i + 1)^{-1} \hat{\beta}^2 S_{xx0}^{(i)} - 2\hat{\beta} S_{xy0}^{(i)} + (k_i + 1) S_{yy0}^{(i)} \quad (3.113)$$

e

$$\hat{\lambda}_{1i} = \frac{S_{xx}^{(i)}}{k_i + 1}, \quad (3.114)$$

onde $S_{xx0} = n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \hat{\beta}\mu_i)^2$ e $S_{xy0}^{(i)}$ e $S_{yy0}^{(i)}$ são dados em (3.109) com $\hat{\beta}$ no lugar de $\hat{\beta}_i$. Usando a verossimilhança perfilada denotada por $\ell(\hat{\lambda}, \hat{\beta})$, onde ℓ é definida em (3.102), tem-se que a estatística da razão de verossimilhança é igual a

$$W = 2 \sum_{i=1}^s \left[\ell_i(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\beta}) - \ell_i(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\beta}_i) \right]. \quad (3.115)$$

Usando a identidade (A.1) do apêndice A, podemos escrever que

$$\begin{aligned} \ell_i(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\beta}_i) &= -n_i \log(2\pi) - \frac{n_i}{2} \log \left[(k_i + 1)^{-1} \hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)} \right] \\ &\quad - \frac{n_i}{2} \log \left[(k_i + 1)^{-1} S_{xx}^{(i)} \right] - n_i, \end{aligned} \quad (3.116)$$

$i = 1, \dots, s$. De modo análogo, obtem-se $\ell_i(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\beta})$ que substituído juntamente com (3.116) em (3.115) levam á estatística

$$\begin{aligned} W &= \sum_{i=1}^s n_i \left\{ \log \left[(k_i + 1)^{-1} \hat{\beta}^2 S_{xx}^{(i)} - 2\hat{\beta} S_{xy0}^{(i)} + (k_i + 1) S_{yy0}^{(i)} \right] \right. \\ &\quad \left. - \log \left[(k_i + 1)^{-1} \hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)} \right] \right\}, \end{aligned} \quad (3.117)$$

com distribuição assintótica qui-quadrado com $s - 1$ graus de liberdade. O estimador $\hat{\beta}$ em (3.112) pode ser colocado numa forma mais simples utilizando (3.107) e

(3.108), ou seja, podemos escrever

$$\sum_{i=1}^s \frac{(\beta B_i - A_i)}{k_{x_i} \beta^2 - 2\beta S_{xy}^{(i)} + k_{x_i}^{-1} S_{yy}^{(i)}} = 0, \quad (3.118)$$

onde

$$A_i = \sum_{j=1}^{n_i} (X_{ij} - \mu_i) Y_{ij} + (k_i + 1) \mu_i \sum_{j=1}^{n_i} Y_{ij}$$

e

$$B_i = (k_i + 1) S_{xx}^{(i)} + 2 \sum_{j=1}^{n_i} (X_{ij} - \mu_i) \mu_i + n_i \mu_i^2 (k_i + 1),$$

$i = 1, \dots, s$, cuja solução deve ser obtida por métodos numéricos.

No caso particular em que $\mu_i = \mu$, $\tau_i^2 = \tau^2$, $k_{x_i} = k_x$, $i = 1, \dots, s$, segue que o estimador de β sob a hipótese nula $H_0 : \beta_1 = \dots = \beta_s = \beta$, é dado por

$$\hat{\beta} = \frac{\sum_{i=1}^s \sum_{j=1}^{n_i} X_{ij} Y_{ij}}{k_x \sum_{i=1}^s \sum_{j=1}^{n_i} X_{ij}^2}. \quad (3.119)$$

Vamos considerar agora a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, onde β_0 é um valor especificado de β . De maneira análoga ao caso anterior, obtem-se a estatística da razão de verossimilhança, que é dada por

$$\begin{aligned} W = & \sum_{i=1}^s n_i \left\{ \log \left[(k_i + 1)^{-1} \beta_0^2 S_{xx}^{(i)} - 2\beta_0 S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)} \right] \right. \\ & \left. - \log \left[(k_i + 1)^{-1} \hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)} \right] \right\}, \end{aligned} \quad (3.120)$$

com $\hat{\beta}_i$ dado pela solução de (3.106). A distribuição assintótica de W dada em (3.120) é qui-quadrado com s graus de liberdade.

O fator de correção de Bartlett

Para facilitar o cálculo da correção de Bartlett, para o modelo desta seção, escrevemos ℓ_i dado em (3.103) em uma forma equivalente, ou seja,

$$\ell_i(\lambda_{0i}, \lambda_{1i}, \beta_i) = -n_i \log(2\pi) - \frac{n_i}{2} \log(\lambda_{0i} \lambda_{1i})$$

$$- \sum_{j=1}^{n_i} \frac{a_{ij}^2}{2\lambda_{1i}(k_i + 1)} - \frac{P_i(\beta_i)}{\lambda_{0i}}, \quad (3.121)$$

onde $a_{ij} = X_{ij} - \mu_i$,

$$P_i(\beta_i) = c_{0i} + c_{1i}\beta_i + c_{2i}\beta_i^2, \quad (3.122)$$

$$c_{0i} = \frac{k_i + 1}{2} \sum_{j=1}^{n_i} Y_{ij}^2, \quad c_{1i} = - \sum_{j=1}^{n_i} a_{ij} Y_{ij} - \mu_i(k_i + 1) \sum_{j=1}^{n_i} Y_{ij}, \quad (3.123)$$

$$c_{2i} = \frac{\sum_{j=1}^{n_i} a_{ij}^2}{2(k_i + 1) + \mu_i \sum_{j=1}^{n_i} a_{ij}} + \frac{n_i \mu_i^2 (k_i + 1)}{2}. \quad (3.124)$$

Para o cálculo das expressões (3.39) precisamos das esperanças de (3.122), (3.123) e (3.124) que são dadas por

$$E[c_{0i}] = \frac{n_i}{2} [\lambda_{0i} + \lambda_{1i}\beta_i^2 + (k_i + 1)\beta_i^2 \mu_i^2], \quad (3.125)$$

$$E[c_{1i}] = -n_i \beta_i [\lambda_{1i} + \mu_i^2 (k_i + 1)] \quad (3.126)$$

e

$$E[c_{2i}] = -\frac{1}{2\beta_i} E[c_{1i}], \quad (3.127)$$

$i = 1, \dots, s$. As equações (3.125), (3.126) e (3.127) permitem um cálculo mais simples para as expressões dadas em (3.39) e conseqüentemente facilitam o cálculo de ϵ_k , $k = p, p + q$ dados em (3.40). Após longas manipulações algébricas, mostra-se, para a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, com $p + q = 3s$ representando o número de parâmetros do modelo sob a hipótese nula, que

$$A_{p+q} = A_{3s} = \sum_{i=1}^s \frac{8}{n_i}, \quad (3.128)$$

$$B_{p+q} = B_{3s} = - \sum_{i=1}^s \frac{19}{6n_i}, \quad (3.129)$$

$$A_p = A_{2s} = \sum_{i=1}^s \frac{6}{n_i},$$

$$B_p = B_{2s} = - \sum_{i=1}^s \frac{8}{3n_i}.$$

Portanto, de (3.40) temos que

$$\epsilon_{p+q} - \epsilon_p = \epsilon_{3s} - \epsilon_{2s} = \sum_{i=1}^s \frac{1,5}{n_i}. \quad (3.130)$$

Considerando os termos de ordem $O(n_*^{-1})$, com $n_* = \min \{n_1, \dots, n_s\}$ segue que

$$\begin{aligned} \mathbb{E}(W) &= s + (A_{p+q} + B_{p+q} - A_p - B_p) + O(n_*^{-2}) \\ &= s + \sum_{i=1}^s \frac{1,5}{n_i} + O(n_*^{-2}). \end{aligned} \quad (3.131)$$

Portanto a estatística (3.120) corrigida com (3.131) fica

$$W^* = \frac{W}{\left(1 + \sum_{i=1}^s \frac{1,5}{sn_i}\right)}, \quad (3.132)$$

que, conforme mostrado em Lawley (1957), tem distribuição assintótica mais próxima da distribuição qui-quadrado com s graus de liberdade do que a distribuição de (3.120). Para corrigir W em (3.115), que é usado para testar a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$ onde β é não especificado, temos um trabalho algébrico bem mais extenso. Neste caso, A_{p+q} e B_{p+q} são dados em (3.128) e (3.129), e com $p = 2s + 1$ tem-se que

$$\begin{aligned} A_p &= A_{2s+1} = s \sum_{i=1}^s \left(\frac{3}{n_i} + \frac{b_i}{a} \right), \\ B_p &= B_{2s+1} = \sum_{i=1}^s \left(\frac{n_i \lambda_{0i}^2 b_i}{3a} - \frac{5n_i b_i^2}{6a} - \frac{8}{n_i} \right), \end{aligned}$$

onde

$$a = n\beta, \quad b_i = \frac{\lambda_{1i} + \mu_i^2(k_i + 1)}{\lambda_{0i}},$$

$i = 1, \dots, s$. Considerando termos até a ordem n_*^{-1} , onde $n_* = \min \{n_1, \dots, n_s\}$, tem-se que

$$\begin{aligned} \mathbb{E}(W) &= s + (A_{3s} + B_{3s} - A_{2s+1} - B_{2s+1}) + O(n_*^{-2}) \\ &= s + D + O(n_*^{-2}), \end{aligned} \quad (3.133)$$

com

$$D = \sum_{i=1}^s \frac{1.5}{n_i} - \frac{2}{a} \sum_{i=1}^s b_i - \frac{1}{3a^2} \sum_{i=1}^s n_i b_i^2 \lambda_{0i}^2 + \frac{5}{6a^2} \sum_{i=1}^s n_i b_i^2.$$

Portanto a estatística (3.115) corrigida com (3.133) é

$$W^* = \frac{W}{\left(1 + \frac{D}{s}\right)}, \quad (3.134)$$

onde W é dada por (3.115).

A estatística da razão de verossimilhança perfilada condicional

A seguir vamos introduzir uma nova estatística para testar $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, como na Seção 3.6.2, que denominamos estatística da razão de verossimilhança condicional estendida. Ela é motivada pela estatística do teste da razão de verossimilhança condicional perfilada dada em (3.29) para testar $H_0 : \beta = \beta_0$ e é definida por

$$W_c = \sum_{i=1}^s \frac{n_i - 2}{n_i} R^{(i)}, \quad (3.135)$$

onde

$$R^{(i)} = n_i \log \left[\frac{k_{x_i} \beta_0^2 S_{xx}^{(i)} - 2\beta_0 \sum_{j=1}^{n_i} (X_{ij} - \mu_i)(Y_{ij} - \mu_i \beta_0) + (k_i + 1) \sum_{j=1}^{n_i} (Y_{ij} - \mu_i \beta_0)^2}{k_{x_i} \hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i \sum_{j=1}^{n_i} (X_{ij} - \mu_i)(Y_{ij} - \mu_i \hat{\beta}_i) + (k_i + 1) \sum_{j=1}^{n_i} (Y_{ij} - \mu_i \hat{\beta}_i)^2} \right], \quad (3.136)$$

onde $S_{xx}^{(i)}$ é dado por (3.109). Note que W_c é uma extensão da estatística da razão de verossimilhança obtida a partir da verossimilhança condicional dada em (3.29), para o caso de s populações. Verificamos a seguir que a distribuição da estatística W_c tem a média mais próxima da média da distribuição χ_s^2 que a média da distribuição da estatística da razão de verossimilhança, W . Como cada $R^{(i)}$ em (3.135) é equivalente a (3.26), temos de (3.48), até termos de ordem $n_*^{-1} =$

$\min \{n_1, \dots, n_s\}$, que

$$\mathbb{E}(R^{(i)}) = 1 + \frac{1,5}{n_i},$$

de modo que,

$$\begin{aligned}\mathbb{E}(W_c) &= \sum_{i=1}^s \left(1 - \frac{0,5}{n_i}\right) + O(n_*^{-2}) \\ &= s - \sum_{i=1}^s \frac{0,5}{n_i} + O(n_*^{-2}),\end{aligned}$$

que é mais próxima de s que a média de W .

A correção de Bartlett para (3.135) é igual a

$$W_c^* = \frac{W}{1 - \sum_{i=1}^s \left(\frac{0,5}{sn_i}\right)}, \quad (3.137)$$

que tem distribuição próxima da distribuição qui-quadrado com s graus de liberdade.

O teste de Wald

A estatística de Wald para testar $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, com a notação usada em (3.63) tem vetor de funções $[\mathbf{r}(\beta, \lambda)]' = (\beta_1 - \beta_0, \dots, \beta_s - \beta_0)$, sendo $\mathbf{R}$ uma matriz $s \times 3s$ associada à hipótese nula. Seguindo a equação (3.63) obtem-se a estatística de Wald para este teste que é dada por

$$W_a = \sum_{i=1}^s (\hat{\beta}_i - \beta_0)^2 \hat{i}_{\beta_i \beta_i}. \quad (3.138)$$

A estatística W_a tem distribuição assintótica qui-quadrado com s graus de liberdade onde $\hat{i}_{\beta_i \beta_i}$ é o estimador de máxima verossimilhança do elemento da matriz de informação (3.104) associado à β . Neste caso a estatística acima também coincide com o quadrado da estatística do teste obtido a partir da distribuição assintótica do estimador de máxima verossimilhança.

No caso da hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$, onde β é não especificado, a estatística de Wald, de forma análoga à anterior, é dada por

$$W_a = [\mathbf{r}(\hat{\beta}, \hat{\lambda})]' \hat{\mathbf{A}} [\mathbf{r}(\hat{\beta}, \hat{\lambda})], \quad (3.139)$$

onde $\mathbf{r}(\hat{\beta}, \hat{\lambda})$ é o estimador de máxima verossimilhança de $\mathbf{r}(\beta, \lambda)$ e $\hat{\mathbf{A}}$ a matriz dada por (3.84) com o estimador de máxima verossimilhança $\hat{i}_{\beta_j \beta_j}$ no lugar de

$$i_{\beta_j \beta_j} = \frac{n_j [\lambda_{1j} + \mu_j^2 (k_j^2 + 1)^2]}{\lambda_{0j}},$$

$j = 1, \dots, s$. A distribuição assintótica de W_a é qui-quadrado com $s - 1$ graus de liberdade. Se $s = 2$, (3.139) se reduz a

$$W_a = \frac{(\hat{\beta}_1 - \hat{\beta}_2)^2}{\hat{i}_{\beta_1 \beta_1}^{-1} + \hat{i}_{\beta_2 \beta_2}^{-1}}. \quad (3.140)$$

3.7.2 Parâmetros μ_i e α_i desconhecidos, $i = 1, \dots, s$

Sob a suposição de que o coeficiente de atenuação $k_{x_i} = \tau_i^2 / (\tau_i^2 + \sigma_{u_i}^2)$ correspondente ao i -ésimo modelo é conhecido, $i = 1, \dots, s$, a distribuição conjunta de (X_{ij}, Y_{ij}) é normal bivariada com vetor de médias $(\mu_i, \alpha_i + \beta_i \mu_i)$ e matriz de covariâncias

$$\begin{pmatrix} \tau_i^2 (k_i + 1) & \beta_i \tau_i^2 \\ \beta_i \tau_i^2 & \beta_i^2 \tau_i^2 + \sigma_{ei}^2 \end{pmatrix},$$

onde $k_i = (1 - k_{x_i}) / k_{x_i}$, $i = 1, \dots, s$. A parametrização ortogonal considerada em (2.18)-(2.19) estende-se sem dificuldades para esta situação mais geral, obtendo-se a parametrização ortogonal dada em (2.40). Tomando-se uma amostra aleatória de tamanho n_i de (X_{ij}, Y_{ij}) , com $n' = \sum_{i=1}^s n_i$ satisfazendo ao modelo acima descrito com $i = 1, \dots, s$, e $j = 1, \dots, n_i$, obtém-se a função de verossimilhança original (sem

a transformação ortogonal nos parâmetros). Substituindo (2.40) nesta verossimilhança, tem-se com $\beta = (\beta_1, \dots, \beta_s)$ e

$$\lambda = (\lambda_{01}, \dots, \lambda_{0s}, \lambda_{11}, \dots, \lambda_{1s}, \lambda_{21}, \dots, \lambda_{2s}, \lambda_{31}, \dots, \lambda_{3s}),$$

que

$$\ell(\lambda, \beta) = \sum_{i=1}^s \ell_i(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i}, \beta_i), \quad (3.141)$$

onde

$$\begin{aligned} \ell_i(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i}, \beta_i) = & \\ & -n_i \log(2\pi) - \frac{n_i}{2} \log(\lambda_{0i} \lambda_{1i}) - \frac{1}{\lambda_{0i} \lambda_{1i}} \left\{ \frac{\lambda_{0i} + \lambda_{1i} \beta_i^2}{k_i + 1} \sum_{j=1}^{n_i} (X_{ij} - \lambda_{2i})^2 \right. \\ & \left. - 2\beta_i \lambda_{1i} \sum_{j=1}^{n_i} (X_{ij} - \lambda_{2i})(Y_{ij} - \lambda_{3i}) + \lambda_{1i}(k_i + 1) \sum_{j=1}^{n_i} (Y_{ij} - \lambda_{3i})^2 \right\}, \end{aligned} \quad (3.142)$$

$i = 1, \dots, s$. De (3.141), a matriz de informação de (λ, β) é igual a

$$\mathbf{I}(\lambda, \beta) = \text{diag}(\mathbf{I}_1(\lambda_{01}, \lambda_{11}, \lambda_{21}, \lambda_{31}, \beta_1), \dots, \mathbf{I}_s(\lambda_{0s}, \lambda_{1s}, \lambda_{2s}, \lambda_{3s}, \beta_s)), \quad (3.143)$$

onde

$$\mathbf{I}_i(\lambda_{0i}, \lambda_{1i}, \beta_i) = \begin{pmatrix} \frac{n_i}{2\lambda_{0i}^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{n_i}{2\lambda_{1i}^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{n_i(\lambda_{0i} + \lambda_{1i}\beta_i^2)}{(k_i + 1)\lambda_{0i}\lambda_{1i}} & -\frac{n_i\beta_i}{\lambda_{0i}} & 0 \\ 0 & 0 & -\frac{n_i\beta_i}{\lambda_{0i}} & \frac{n_i(k_i + 1)}{\lambda_{0i}} & 0 \\ 0 & 0 & 0 & 0 & \frac{n_i\lambda_{1i}}{\lambda_{0i}} \end{pmatrix}, \quad (3.144)$$

$i = 1, \dots, s$.

Derivando (3.141) em relação aos parâmetros $\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i}, \beta_i$ e igualando estas derivadas a zero, obtem-se os seguintes estimadores de máxima verossimilhança:

$$\hat{\lambda}_{2i} = \bar{X}_i, \quad \hat{\lambda}_{3i} = \bar{Y}_i, \quad (3.145)$$

$$\hat{\beta}_i = (k_i + 1) \frac{S_{xy}^{(i)}}{S_{xx}^{(i)}}, \quad (3.146)$$

$$\hat{\lambda}_{0i} = \frac{\hat{\beta}_i^2}{(k_i + 1)} S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)}, \quad (3.147)$$

$$\hat{\lambda}_{1i} = \frac{S_{xx}^{(i)}}{(k_i + 1)}, \quad (3.148)$$

onde $S_{xx}^{(i)}$, $S_{xy}^{(i)}$ e $S_{yy}^{(i)}$ são dados por (3.72). Um intervalo de confiança aproximado para β_i pode ser obtido usando o elemento $i_{\beta_i\beta_i}$ da matriz de informação (3.143), onde

$$i_{\beta_i\beta_i} = \frac{n_i \lambda_{1i}}{\lambda_{0i}}, \quad (3.149)$$

ou seja,

$$\left[\hat{\beta}_i \mp z_{\alpha/2} \left(\frac{n_i \hat{\lambda}_{1i}}{\hat{\lambda}_{0i}} \right)^{-1/2} \right],$$

$i = 1, \dots, s$. Como em (3.111) no caso em que $\beta_1 = \dots = \beta_s = \beta$, um intervalo de confiança aproximado para β é da forma

$$\left(\hat{\beta} \mp z_{\alpha/2} \left\{ \sum_{i=1}^s \frac{n_i \hat{\lambda}_{1i}}{\hat{\lambda}_{0i}} \right\}^{-1/2} \right).$$

O Teste da razão de verossimilhança

Sob a hipótese nula, $H_0 : \beta_1 = \dots = \beta_s = \beta$, os estimadores de máxima verossimilhança são tais que: $\hat{\beta}$ satisfaz à equação

$$\sum_{i=1}^s \frac{\beta k_{x_i} S_{xx}^{(i)} - (k_i + 1) S_{xy}^{(i)}}{\beta^2 S_{xx}^{(i)} k_{x_i} - 2\beta S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)}} = 0 \quad (3.150)$$

e os demais estimadores sob H_0 são como em (3.145)-(3.148) com $\hat{\beta}$ no lugar de $\hat{\beta}_i$, $i = 1, \dots, s$. Se $\mu_i = \mu$, $\tau_i^2 = \tau^2$, $k_{x_i} = k_x$ e $\alpha_i = \alpha$, temos neste caso que

$$\hat{\beta} = \frac{\sum_{i=1}^s \sum_{j=1}^{n_i} (X_{ij} - \bar{X})(Y_{ij} - \bar{Y})}{k_x \sum_{i=1}^s \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2}.$$

Usando a verossimilhança perfilada denotada por $\ell(\hat{\lambda}, \hat{\beta})$, onde ℓ é definida em (3.141), tem-se que a estatística da razão de verossimilhança é igual a

$$W = 2 \sum_{i=1}^s \left[\ell_i(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i}, \hat{\lambda}_{3i}, \hat{\beta}) - \ell_i(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i}, \hat{\lambda}_{3i}, \hat{\beta}_i) \right]. \quad (3.151)$$

Após algumas manipulações algébricas tem-se que

$$\begin{aligned} \ell_i(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i}, \hat{\lambda}_{3i}, \hat{\beta}) = \\ -n_i \log(2\pi) - \frac{n_i}{2} \log \left[(k_i + 1)^{-1} \hat{\beta}^2 S_{xx}^{(i)} - 2\hat{\beta} S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)} \right] \\ - \frac{n_i}{2} \log \left[(k_i + 1)^{-1} S_{xx}^{(i)} \right] - n_i. \end{aligned} \quad (3.152)$$

De (3.152) obtem-se $\ell_i(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i}, \hat{\lambda}_{3i}, \hat{\beta}_i)$ e $\ell_i(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i}, \hat{\lambda}_{3i}, \hat{\beta})$ que substituídos em (3.151) nos fornece a estatística da razão de verossimilhança W na forma

$$\begin{aligned} W = \sum_{i=1}^s n_i \left\{ \log \left[(k_i + 1)^{-1} \hat{\beta}^2 S_{xx}^{(i)} - 2\hat{\beta} S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)} \right] \right. \\ \left. - \log \left[(k_i + 1)^{-1} \hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)} \right] \right\}, \end{aligned} \quad (3.153)$$

que tem distribuição assintótica qui-quadrado com $s - 1$ graus de liberdade.

Sob a hipótese nula $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, onde β_0 é especificado, de maneira análoga ao caso anterior, obtem-se a estatística da razão de verossimilhança que é dada por

$$\begin{aligned} W = \sum_{i=1}^s n_i \left\{ \log \left[(k_i + 1)^{-1} \hat{\beta}_0^2 S_{xx}^{(i)} - 2\hat{\beta}_0 S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)} \right] \right. \\ \left. - \log \left[(k_i + 1)^{-1} \hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (k_i + 1) S_{yy}^{(i)} \right] \right\}, \end{aligned} \quad (3.154)$$

e que tem distribuição assintótica qui-quadrado com s graus de liberdade, sendo $S_{xx}^{(i)}$, $S_{xy}^{(i)}$ e $S_{yy}^{(i)}$ definidos em (3.72).

O fator de correção de Bartlett

Consideramos em primeiro lugar o caso da hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta_0$.

O cálculo da correção de Bartlett fica facilitado escrevendo-se ℓ_i dada em (3.121)

em outra forma equivalente, ou seja:

$$\begin{aligned}\ell_i(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i}, \beta_i) &= -n_i \log(2\pi) - \frac{n_i}{2} \log(\lambda_{0i} \lambda_{1i}) \\ &\quad - \sum_{j=1}^{n_i} \frac{(X_{ij} - \lambda_{2i}^2)}{2\lambda_{1i}(k_i + 1)} - \frac{P_i(\beta_i)}{\lambda_{0i}},\end{aligned}\quad (3.155)$$

onde

$$c_{0i} = \frac{k_i + 1}{2} \sum_{j=1}^{n_i} (Y_{ij} - \lambda_{3i}), \quad c_{1i} = - \sum_{j=1}^{n_i} (X_{ij} - \lambda_{2i}(Y_{ij} - \lambda_{3i})), \quad (3.156)$$

$$c_{2i} = \frac{1}{2(k_i + 1)} \sum_{j=1}^{n_i} (X_{ij} - \lambda_{2i})^2, \quad (3.157)$$

e

$$P_i(\beta_i) = c_{0i} + c_{1i}\beta_i + c_{2i}\beta_i^2, \quad (3.158)$$

$i = 1, \dots, s$. Para o cálculo da expressão (3.39) precisamos das esperanças de (3.156), (3.157) e (3.158) que são dadas por

$$\mathbb{E}(c_{0i}) = \frac{n_i}{2}(\lambda_{0i} + \lambda_{1i}\beta_i^2), \quad (3.159)$$

$$\mathbb{E}(c_{1i}) = -n_i\lambda_{1i}\beta_i, \quad (3.160)$$

$$\mathbb{E}(c_{2i}) = \frac{n_i\lambda_{0i}}{2} \quad (3.161)$$

e

$$\mathbb{E}(P_i(\beta_i)) = \frac{n_i\lambda_{0i}}{2}, \quad (3.162)$$

$i = 1, \dots, s$. As esperanças (3.159)-(3.162) permitem um cálculo mais simples para as expressões dadas em (3.40) e conseqüentemente facilitam o cálculo de ϵ_k , $k = p, p+q$ em (3.40). Após uma longa série de manipulações algébricas mostra-se para a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, com $p + q = 5s$ representando o número de parâmetros sob a hipótese alternativa e $p = 4s$ o número de parâmetros do modelo

sob a hipótese nula, que

$$\begin{aligned}\epsilon_{p+q} = A_{p+q} + B_{p+q} &= A_{5s} + B_{5s} \\ &= \sum_{i=1}^s \frac{37}{6n_i},\end{aligned}$$

$$\begin{aligned}\epsilon_p = A_p + B_p &= A_{4s} + B_{4s} \\ &= \sum_{i=1}^s \frac{25}{6n_i} - \sum_{i=1}^s \frac{1}{n_i(\lambda_{0i} + \lambda_{1i}\beta_0^2)} \left[\lambda_{1i}\beta_0^2 - \frac{\lambda_{0i}^2 + \lambda_{1i}^2\beta_0^4}{n_i(\lambda_{0i} + \lambda_{1i}\beta_0^2)} \right],\end{aligned}$$

de modo que

$$\begin{aligned}\epsilon_{p+q} - \epsilon_p &= \epsilon_{5s} - \epsilon_{4s} \\ &= \sum_{i=1}^s \frac{2}{n_i} + \sum_{i=1}^s \frac{\lambda_{1i}\beta_0^2}{n_i(\lambda_{0i} + \lambda_{1i}\beta_0^2)} - \sum_{i=1}^s \frac{\lambda_{0i}^2 + \lambda_{1i}^2\beta_0^4}{n_i(\lambda_{0i} + \lambda_{1i}\beta_0^2)^2}, \quad (3.163)\end{aligned}$$

que é uma extensão natural de (3.60). Considerando termos de ordem n_*^{-1} , segue que

$$E(W) = s + (\epsilon_{5s} - \epsilon_{4s}). \quad (3.164)$$

A estatística W dada em (3.154) corrigida com (3.164) é dada por

$$W^* = \frac{W}{1 + \frac{(\epsilon_{5s} - \epsilon_{4s})}{s}}, \quad (3.165)$$

que de acordo com Lawley (1957), tem distribuição mais próxima da distribuição qui-quadrado com s graus de liberdade do que a distribuição de W dada em (3.154).

Como na seção 3.7.1, motivada pela estatística da razão de verossimilhança condicional W_c , dada em (3.38), podemos definir a estatística da razão de verossimilhança condicional estendida, que neste caso é dada por

$$W_c = \sum_{i=1}^s \frac{n_i - 3}{n_i} R^{(i)}, \quad (3.166)$$

onde

$$R^{(i)} = n_i \log \left[\frac{k_{x_i}\beta_0^2 S_{xx}^{(i)} - 2\beta_0 S_{xy}^{(i)} + (k_i + 1)S_{yy}^{(i)}}{k_{x_i}\hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (k_i + 1)S_{yy}^{(i)}} \right]. \quad (3.167)$$

A estatística W_c tem distribuição assintótica qui-quadrado com s graus de liberdade.

Consideramos agora a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$, onde β é desconhecido. Neste caso, $p + q = 5s$ e $p = 4s + 1$, de modo que após uma longa série de manipulações algébricas, temos que

$$\begin{aligned}\epsilon_{5s} &= A_{5s} + B_{5s} \\ &= \sum_{i=1}^s \frac{37}{6n_i},\end{aligned}$$

e

$$\begin{aligned}\epsilon_{4s+1} &= A_{4s+1} + B_{4s+1} \\ &= \sum_{i=1}^s \frac{25}{6n_i} - \frac{3}{a} \sum_{i=1}^s \frac{\lambda_{1i}}{\lambda_{0i}} + \beta^2 \sum_{i=1}^s \frac{\lambda_{1i}^2}{n_i(\lambda_{0i} + \lambda_{1i}\beta^2)} \\ &\quad + \sum_{i=1}^s \frac{\lambda_{0i}^2 + \lambda_{1i}^2\beta^4}{2n_i(\lambda_{0i} + \lambda_{1i}\beta^2)^2} - \frac{1}{a} \sum_{i=1}^s \frac{\lambda_{1i}^2\beta^2}{\lambda_{0i}(\lambda_{0i} + \lambda_{1i}\beta^2)} \\ &\quad + \frac{1}{a} \sum_{i=1}^s \frac{\lambda_{1i}\beta^2}{(\lambda_{0i} + \lambda_{1i}\beta^2)^2} + \frac{1}{2a^2} \sum_{i=1}^s \frac{n_i\lambda_{1i}^2}{\lambda_{0i}^2},\end{aligned}$$

onde $a = \sum_{i=1}^s n_i\lambda_{1i}/\lambda_{0i}$. Temos então que

$$\mathbb{E}(W) = s - 1 + (\epsilon_{5s} - \epsilon_{4s+1}) + O(n_*^{-2}). \quad (3.168)$$

A estatística W dada em (3.153) corrigida com o fator dado em (3.168) é dada por

$$W^* = \frac{W}{1 + \frac{(\epsilon_{5s} - \epsilon_{4s+1})}{s-1}}. \quad (3.169)$$

W^* tem distribuição mais próxima da distribuição qui-quadrado com $s-1$ graus de liberdade do que a distribuição de W dada em (3.153).

O teste de Wald

A estatística de Wald para testar $H_0 : \beta_1 = \dots = \beta_s = \beta_0$ com a notação usada em (3.36) tem vetor de funções $[\mathbf{r}(\beta, \boldsymbol{\lambda})]' = (\beta_1 - \beta_0, \dots, \beta_s - \beta_0)$ sendo $\mathbf{R}$ uma matriz

$s \times 5s$ associada à hipótese nula. Seguindo a equação (3.63) temos que a estatística de Wald para a hipótese acima é dada em (3.138) com $i_{\beta_i\beta_i}$ dado em (3.149), sendo $\hat{i}_{\beta_i\beta_i}$ seu estimador de máxima verossimilhança e $\hat{\beta}_i$ dado em (3.146). W_a neste caso tem distribuição assintótica qui-quadrado com s graus de liberdade.

Para a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$ com β não especificado, a estatística de Wald tem a forma (3.139), onde a matriz $\mathbf{A}$ é dada em (3.84) com $i_{\beta_i\beta_i}$ dado em (3.149). Neste caso, W_a tem distribuição assintótica qui-quadrado com $s - 1$ graus de liberdade. Se $s = 2$, W_a tem a forma explícita dada por (3.140), ou seja

$$W_a = \frac{(\hat{\beta}_1 - \hat{\beta}_2)^2}{\hat{i}_{\beta_1\beta_1}^{-1} + \hat{i}_{\beta_2\beta_2}^{-1}}. \quad (3.170)$$

A estatística W_a tem distribuição assintótica qui-quadrado com 1 grau de liberdade.

Ilustração numérica 3.5

Foram geradas 1000 amostras aleatórias de tamanhos $n_1 = n_2 = n$ satisfazendo ao MEV estrutural gerado com $s = 2$ para o caso em que os parâmetros μ_i e α_i são conhecidos ($\alpha_i = 0,0$, $i = 1,2$), $\beta_1 = \beta_2 = 4,0$, $k_{x_1} = 0,60$, $k_{x_2} = 0,75$, $\sigma_{u1}^2 = \sigma_{u2}^2 = 1,0$ e $\mu_1 = \mu_2 = 0,0$. Estamos interessados em testar $H_0 : \beta_1 = \beta_2 = \beta_0 = 4,0$. Para cada amostra gerada, calculou-se o valor das estatísticas W dada em (3.120), W^* em (3.132), W_c em (3.135), W_c^* em (3.137) e W_a em (3.138). Apresenta-se nas Tabelas 3.9 e 3.10, o número de vezes N_j em que aceita-se a hipótese nula nas 1000 amostras para os níveis $\alpha = 0,10$ e $0,05$, sendo que N_1 corresponde a W , N_2 corresponde a W^* , N_3 corresponde a W_c , N_4 corresponde a W_c^* e N_5 corresponde a W_a . A estatística da razão de verossimilhança com a correção de Bartlett apresenta o nível observado mais próximo do nível descritivo especialmente para valores pequenos de n , o que é desejável. Neste caso, W se comporta bem melhor que W_a . Note também que W_c se compor-

ta melhor que W , conforme verificado analiticamente no texto. Para a hipótese $H_0 : \beta_1 = \beta_2 = \beta$, onde β é não especificado, gerou-se 1000 amostras aleatórias de tamanhos $n_1 = n_2 = n$ com $s = 2$, $k_{x_1} = k_{x_2} = 0,95$, $\beta_1 = \beta_2 = 4,0$, $\sigma_u^2 = 1,0$, $\alpha_1 = \alpha_2 = 0,0$ e $\mu_1 = \mu_2 = 0,0$. Em cada amostra gerada, calculou-se o valor das estatísticas W , em (3.117), W^* , em (3.134) e W_a em (3.140). Apresenta-se nas Tabelas 3.11 e 3.12 o número de vezes N_j em que aceita-se a hipótese nula nas 1000 amostras simuladas para os níveis $\alpha = 0,10$ e $0,05$, sendo que N_1 corresponde a W , N_2 corresponde a W^* e N_3 corresponde a W_a . A correção de Bartlett melhora a aproximação de W para a distribuição qui-quadrado de mesmo para valores pequenos de n , o que é esperado. A estatística de Wald, W_a , se aproxima mais rapidamente do nível descritivo, mesmo para valores moderados de n , apresentando, especialmente neste modelo, muita facilidade para o seu cálculo. Isto mostra que a estatística da razão de verossimilhança é mais afetada pelo fato de β ser desconhecido do que o a estatística de Wald.

Ilustração numérica 3.6

Foram geradas 1000 amostras aleatórias de tamanho $n_1 = n_2 = n$ satisfazendo ao MEV estrutural para o caso em que os parâmetros μ_i e α_i são desconhecidos, com $s = 2$, $\beta_1 = \beta_2 = 1,5$, $k_{x_1} = k_{x_2} = 2/3$, $\sigma_{u1}^2 = \sigma_{u2}^2 = 1,0$ e $\mu_1 = 30,0$, $\mu_2 = 40,0$. Neste caso a hipótese de interesse é $H_o : \beta_1 = \beta_2 = \beta_0 = 1,5$. Para cada amostra gerada, calculou-se o valor das estatísticas W dada em (3.154), W^* em (3.165), W_c em (3.166) e W_a em (3.138). De forma análoga à ilustração numérica 3.2 calculamos também estimadores para $IE(W)$, $IE(W^*)$ e $IE(W_c)$. Apresenta-se nas Tabelas 3.13 e 3.14, o número de vezes N_j em que aceita-se a hipótese nula nas 1000 amostras para os níveis $\alpha = 0,05$ e $0,10$, sendo que N_1 corresponde a W , N_2 corresponde a W^* , N_3 corresponde a W_c e N_4 corresponde a W_a . A estatística

da razão de verossimilhança com a correção de Bartlett apresenta o nível mais próximo do nível descritivo mesmo quando n tem valores pequenos. Neste caso, W se comporta bem melhor que W_a . Note também que W_c se comporta melhor que W , conforme verificado no texto.

Ilustração numérica 3.7

Foram geradas 1000 amostras aleatórias de tamanhos $n_1 = n_2 = n$ satisfazendo ao MEV estrutural com $s = 2$, $\beta_1 = 1,5$, $\beta_2 = 1,5$, $k_{x_1} = k_{x_2} = 2/3$, $\sigma_{u_1}^2 = \sigma_{u_2}^2 = 1,0$, $\mu_1 = 30,0$, $\mu_2 = 40,0$, $\alpha_1 = 0,5$, $\alpha_2 = 3,0$. A hipótese de interesse é $H_o : \beta_1 = \beta_2 = \beta$, onde α_i e μ_i são desconhecidos, $i = 1, 2$. Para cada amostra gerada, calculou-se o valor das estatísticas W dada em (3.153), W^* em (3.169) e W_a em (3.170). Apresenta-se nas Tabelas 3.15 e 3.16, o número de vezes N_j em que aceita-se a hipótese nula nas 1000 amostras para os níveis $\alpha = 0,05$ e $0,10$, sendo que N_1 corresponde a W , N_2 corresponde a W^* , N_3 corresponde a W_c e N_4 corresponde a W_a . A estatística da razão de verossimilhança com a correção de Bartlett apresenta o nível observado mais próximo do nível descritivo mesmo quando n tem valores pequenos.

3.8 Alguns Modelos particulares

No capítulo 2 foram apresentadas algumas transformações ortogonais para casos particulares do modelo considerado em (2.22) e que foram denotados por Modelo 1,..., Modelo 9. Estes modelos são analisados por Barnett (1969), Dolby (1976) e outros autores, sem a utilização da parametrização ortogonal. Aqui apresentamos intervalos de confiança e o teste da razão de verossimilhança utilizando parametrizações ortogonais para os modelos 1 e 2. Para os outros modelos, as

derivações são similares.

Modelo 1. Neste modelo, os valores observados (X_{ij}, Y_{ij}) têm distribuição normal bivariada

$$(X_{ij}, Y_{ij}) \sim \mathcal{N}_2 \left[(\mu, \alpha + \beta\mu), \begin{pmatrix} \sigma^2 + \tau^2 & \beta\tau^2 \\ \beta\tau^2 & \beta^2\tau^2 + \sigma^2 \end{pmatrix} \right], \quad (3.171)$$

$i = 1, \dots, s$ e $j = 1, \dots, n_i$. Com uma amostra aleatória de tamanho n_i , $i = 1, \dots, s$, onde $n = \sum_{i=1}^s n_i$ de (3.171) temos a função de verossimilhança desta amostra, com $\sigma_{ei}^2 = \sigma_{ui}^2 = \sigma^2$, $i = 1, \dots, s$, conhecido. Na parametrização ortogonal, a função de verossimilhança com $\lambda = (\lambda_0, \lambda_1, \lambda_2)$ pode ser escrita como

$$\begin{aligned} \ell(\beta, \lambda) &= -2n \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{n}{2} \log \lambda_2 - \left[2\sigma^2 \lambda_2 (\beta^2 + 1) \right]^{-1} \\ &\times \left\{ (\beta^2 \lambda_2 + \sigma^2) \sum_{i,j} (X_{ij} - \lambda_0)^2 - 2\beta(\lambda_2 - \sigma^2) \sum_{i,j} (X_{ij} - \lambda_0)(Y_{ij} - \lambda_1)^2 \right. \\ &\left. + (\lambda_2 + \beta^2 \sigma^2) \sum_{i,j} (Y_{ij} - \lambda_1)^2 \right\}. \end{aligned} \quad (3.172)$$

Derivando (3.172) em relação a β , λ_0 , λ_1 , λ_2 e igualando estas quatro derivadas à zero, após um longo cálculo obtem-se os estimadores de máxima verossimilhança de λ_0 , λ_1 e λ_2 ,

$$\hat{\lambda}_0 = n^{-1} \sum_{i,j} X_{ij} = \bar{X}, \quad \hat{\lambda}_1 = n^{-1} \sum_{i,j} Y_{ij} = \bar{Y}, \quad (3.173)$$

$$\hat{\lambda}_2 = \frac{S_{xx} + 2\hat{\beta}S_{xy} + \hat{\beta}^2 S_{yy}}{\hat{\beta}^2 + 1}, \quad (3.174)$$

onde $\hat{\beta}$ corresponde à maior raiz da equação quadrática

$$\beta^2 S_{xx} - \beta(S_{xx} - S_{yy}) - S_{xy} = 0, \quad (3.175)$$

com

$$S_{xx} = n^{-1} \sum_{i=1}^s \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2, \quad S_{xy} = n^{-1} \sum_{i=1}^s \sum_{j=1}^{n_i} (X_{ij} - \bar{X})(Y_{ij} - \bar{Y}), \quad (3.176)$$

e

$$S_{yy} = \sum_{i=1}^s \sum_{j=1}^{n_i} n^{-1} (Y_{ij} - \bar{Y})^2. \quad (3.177)$$

A matriz de informação de (β, λ) é a mesma de (2.5) com $n = \sum_{i=1}^s n_i$ e a variância assintótica de $\hat{\beta}$ é dada por

$$i_{\beta\beta}^{-1} = \frac{\lambda_2 \sigma^2 (\beta^2 + 1)}{n(\lambda_2 - \sigma^2)^2}. \quad (3.178)$$

Um intervalo de confiança aproximado para β , usando (3.178) pode ser dado por

$$P_r \left[\hat{\beta} - z_{\frac{\alpha}{2}} i_{\beta\beta}^{-1} \leq \beta \leq \hat{\beta} + z_{\frac{\alpha}{2}} i_{\beta\beta}^{-1} \right] \simeq 1 - \alpha. \quad (3.179)$$

O teste da hipótese $H_0 : \beta = \beta_0$ pode ser efetuado usando-se a estatística da razão de verossimilhança, ou seja, W

$$W = 2 \left[\ell(\hat{\beta}, \hat{\lambda}) - \ell(\beta_0, \hat{\lambda}_0) \right], \quad (3.180)$$

onde $\hat{\lambda}_0$ é o estimador de máxima verossimilhança de λ sob H_0 . A distribuição assintótica de W é qui-quadrado com um grau de liberdade, $\chi_{(1)}^2$, e pode ser escrita na forma

$$W = n \log(\lambda_2(\beta_0)) + \frac{\lambda_3(\beta_0)}{\sigma^2} - \log(\lambda_2(\hat{\beta})) - \frac{\lambda_3(\hat{\beta})}{\sigma^2}, \quad (3.181)$$

onde

$$\lambda_2(z) = \frac{S_{xx} + 2zS_{xy} + z^2S_{yy}}{z^2 + 1}, \quad (3.182)$$

e

$$\lambda_3(z) = \frac{z^2S_{xx} - 2zS_{xy} + S_{yy}}{z^2 + 1}. \quad (3.183)$$

Utilizando as correções de Wong (1989), a estatística (3.181) corrigida com o fator de Bartlett fica,

$$W^* = \frac{W_1}{1 - \frac{2\sigma^2}{n(\hat{\lambda}_2 - \sigma^2)} + \frac{\sigma^4}{n(\hat{\lambda}_2 - \sigma^2)^2}}, \quad (3.184)$$

com

$$n = \sum_{i=1}^s n_i$$

e onde $\hat{\lambda}_2$ é dado em (3.174) e $\hat{\beta}$ em (3.175).

Modelo 2. Neste modelo, os valores observados (X_{ij}, Y_{ij}) $i = 1, \dots, s, j = 1, \dots, n_i$ têm a mesma distribuição dada em (3.171) com $\sigma_{ei}^2 = \sigma_{u1}^2 = \sigma^2$ e σ^2 desconhecido. Com uma amostra aleatória de tamanho $n_i, i = 1, \dots, s$, onde $n = \sum_{i=1}^s n_i$ de (3.171) temos a função de verossimilhança desta amostra. Na transformação ortogonal, a função de verossimilhança com $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$ pode ser escrita como em (3.172), com λ_3 no lugar de σ^2 . Derivando esta função de verossimilhança em relação a $\beta, \lambda_1, \lambda_2$ e λ_3 e igualando estas derivadas a zero, obtemos os estimadores de máxima verossimilhança $\hat{\lambda}_0$ e $\hat{\lambda}_1$ dados em (3.173), $\hat{\lambda}_2$ dado em (3.174), $\hat{\beta}$ dado em (3.175) e

$$\hat{\lambda}_3 = \frac{\hat{\beta}^2 S_{xx} - 2\hat{\beta} S_{xy} + S_{yy}}{\hat{\beta}^2 + 1},$$

onde S_{xx}, S_{xy} e S_{yy} dados respectivamente em (3.176)-(3.177).

O elemento da matriz de informação associado à β é dado por

$$i_{\beta\beta} = \frac{n(\lambda_2 - \lambda_3)^2}{\lambda_2 \lambda_3 (\beta^2 + 1)^2},$$

que pode ser usado para construir um intervalo de confiança aproximado para β , ou seja,

$$(\hat{\beta} \mp z_{\alpha/2} \hat{i}_{\beta\beta}^{-1/2}).$$

Para testar $H_0 : \beta = \beta_0$, com β_0 conhecido, podemos usar a estatística da razão de verossimilhança,

$$W = \left[\log(\lambda_2(\beta_0) \lambda_3(\beta_0)) - \log(\lambda_2(\hat{\beta}) \lambda_3(\hat{\beta})) \right], \quad (3.185)$$

que tem distribuição assintótica qui-quadrado com um grau de liberdade, com $\lambda_2(\cdot)$ e $\lambda_3(\cdot)$ dados respectivamente por (3.182) e (3.183). De (3.10), temos que

$$E(W) = 1 + \frac{2,5}{n} + O(n_*^{-2})$$

$$n = \sum_{i=1}^s n_i,$$

de modo que

$$W^* = \frac{W}{\left(1 + \frac{2,5}{n}\right)},$$

com W dado em (3.185). Portanto, W^* tem distribuição nula mais próxima da distribuição qui-quadrado assintótica com um grau de liberdade do que a distribuição de W . Estimação e testes de hipóteses nos outros modelos podem ser estudados de maneira similar.

3.9 Conclusões

Os estudos de simulação desenvolvidos neste capítulo para o caso em que k_x é conhecido, parecem indicar para o caso de uma única população com α e μ conhecidos, que a estatística da razão de verossimilhança corrigida, W^* tem um melhor comportamento com relação a proximidade do nível descritivo. Para amostras de tamanhos moderados, a estatística W_c também tem um comportamento bastante satisfatório. O mesmo parece ocorrer no caso em que μ e α são desconhecidos.

Para o caso em que temos mais de uma população, a correção de Bartlett com k_x conhecido não parece ser tão eficiente para n pequeno, como no caso de uma única população, embora a estatística da razão de verossimilhança com a correção de Bartlett apresente o comportamento melhor que o da estatística da razão de verossimilhança sem a referida correção. Tal fato também parece ter sido notado

por Wong (1990), quando ϕ_i é conhecido. Para tais situações, a estatística de Wald parece se comportar de modo bastante satisfatório, principalmente quando n é grande apesar de que para $s \geq 3$ seu cálculo pode não ser tão simples. Nos casos em que W_c foi definida, ela também apresenta um comportamento bastante satisfatório, apesar de não envolver a correção de Bartlett.

Os resultados dos estudos de simulação mostram também que quando a razão de variâncias ϕ é conhecida, a estatística de Wald tem um comportamento satisfatório para valores grandes de n e no caso de até duas populações é computacionalmente simples. Para valores pequenos e moderados de n , a estatística de Wald parece não ser recomendável.

□□□

Tabela 3.6: Aceitação de H_0 em 1000 amostras, $\alpha = 0,05$

n	N_1	N_2	N_3	N_4
05	905	951	969	845
10	932	949	953	903
20	945	953	956	935
30	942	949	953	937
50	943	943	946	939

Tabela 3.7: Aceitação de H_0 em 1000 amostras, $\alpha = 0,05$

n	N_1	N_2	N_3	N_4
05	778	912	976	679
10	892	940	965	825
20	941	956	962	909
30	929	946	948	902
50	936	942	943	925
100	942	948	949	933

Tabela 3.8: Aceitação de H_0 em 1000 amostras, $\alpha = 0,10$

n	N_1	N_2	N_3	N_4
05	696	837	953	623
10	830	885	916	777
20	867	900	911	867
30	876	896	899	857
50	877	887	891	882
100	900	902	903	897.

Tabela 3.9: Número de aceitação de H_0 em 1000 amostras, $\alpha = 0,10$

n	N_1	N_2	N_3	N_4	N_5
05	811	893	931	780	766
10	884	905	923	865	809
20	885	905	912	878	854
30	885	897	903	883	868
50	897	907	910	897	867

Tabela 3.10: Número de aceitação de H_0 em 1000 amostras, $\alpha = 0,05$

n	N_1	N_2	N_3	N_4	N_5
05	893	934	977	863	814
10	931	955	960	919	865
20	939	954	957	936	900
30	943	952	955	943	909
50	946	950	952	944	911.

Tabela 3.11: Aceitação de H_0 em 1000 amostras, $\alpha = 0,10$

n	N_1	N_2	N_3
05	799	838	805
10	858	875	859
20	902	913	904
30	882	886	885
50	888	890	888 .

Tabela 3.12: Aceitação de H_0 em 1000 amostras, $\alpha = 0,05$

n	N_1	N_2	N_3
05	870	905	874
10	916	930	912
20	949	957	952
30	927	932	928
50	937	946	937

Tabela 3.13: Aceitação de H_0 em 1000 amostras, $\alpha = 0,05$

n	N_1	N_2	N_3	N_4	$\hat{IE}(W)$	$\hat{IE}(W^*)$	$\hat{IE}(W_c)$
05	789	927	975	661	3,98681	2,22734	1,59472
10	893	927	962	842	2,65283	2,06746	1,85698
20	935	959	963	915	2,21045	1,97600	1,88789
30	939	948	952	923	2,17206	2,01230	1,954858
50	942	949	952	936	2,15252	2,05752	2,023336

Tabela 3.14: Aceitação de H_0 em 1000 amostras, $\alpha = 0,10$

n	N_1	N_2	N_3	N_4	$\hat{IE}(W)$	$\hat{IE}(W^*)$	$\hat{IE}(W_c)$
05	694	883	940	568	3,98681	2,22734	1,59472
10	820	888	918	779	2,65283	2,06746	1,85698
20	866	902	916	837	2,21045	1,97600	1,88789
30	882	893	905	871	2,17206	2,01230	1,954858
50	880	892	895	871	2,15252	2,05752	2,023336

Tabela 3.15: Aceitação de H_0 em 1000 amostras, $\alpha = 0,05$

n	N_1	N_2	N_3
05	815	888	829
10	905	934	902
20	935	949	939
30	939	948	942
50	950	954	950

Tabela 3.16: Aceitação de H_0 em 1000 amostras, $\alpha = 0,10$

n	N_1	N_2	N_3
05	734	825	753
10	829	881	836
20	891	905	893
30	874	893	878
50	891	901	892

Capítulo 4

Inferência Bayesiana no MEV

4.1 Introdução

Neste capítulo abordamos problemas de inferência sobre β utilizando o enfoque Bayesiano. As transformações ortogonais consideradas no Capítulo 2 facilitam as integrações a serem efetuadas e possibilitam a obtenção de posteriores exatas nos modelos funcional e estrutural. Um estudo pioneiro feito por Lindley e El-Sayad (1968), para o caso estrutural, obtem posteriores aproximadas para β no caso particular em que $\mu = \alpha = 0$ e ϕ é conhecido. Para valores grandes de n obtemos aproximações normais para as posteriores de β utilizando um enfoque alternativo ao proposto por Lindley e El-Sayad (1968). Obtemos a mesma expressão para a variância assintótica que aquela obtida em Lindley e El Sayad (1968). Portanto, parte deste Capítulo pode ser visto como uma rediscussão de Lindley e El-Sayad (1968) e Lindley (1966). Usando a transformação ortogonal considerada no Capítulo 2, verificamos também que no caso em que k_x é conhecido, a distribuição a posteriori de β é uma distribuição $t - Student$, o que justifica alguns resultados clássicos apresentados em Fuller (1987). Para a comparação de vários modelos

estruturais, consideramos a possibilidade da utilização do fator de Bayes a posteriori proposto por Aitkin (1990). Este enfoque produz procedimentos alternativos aos do Capítulo 3, bastante simples de serem implementados. Um dos principais resultados é o Teorema 4.8.1, onde são estudadas algumas propriedades do fator de Bayes a posteriori utilizado para testar hipóteses do tipo $H_0 : \beta_1 = \dots = \beta_s = \beta$, com β especificado ou não, no modelo estrutural.

4.2 Variância σ^2 conhecida

Nesta Seção assumimos que $\phi = 1$, ou seja, que $\sigma_e^2 = \sigma_u^2 = \sigma^2$, com σ^2 conhecido. Considere a identidade

$$\begin{aligned} a_1 \sum_{i=1}^n (X_i - \lambda_0)^2 - 2a_2 \sum_{i=1}^n (X_i - \lambda_0)(Y_i - \lambda_1) + a_3 \sum_{i=1}^n (Y_i - \lambda_1)^2 &= a_1 S_{xx} - 2a_2 S_{xy} + \\ &+ a_3 S_{yy} + na_1 \left[\lambda_0 - \left\{ \bar{X} - \frac{a_2}{a_1}(\bar{Y} - \lambda_1) \right\} \right]^2 + n \left[\left(a_3 - \frac{a_2^2}{a_1} \right) (\bar{Y} - \lambda_1)^2 \right], \end{aligned} \quad (4.1)$$

onde

$$a_1 = \beta^2 \lambda_2 + \sigma^2, \quad a_2 = \beta(\lambda_2 - \sigma^2) \quad e \quad a_3 = \lambda_2 + \beta^2 \sigma^2. \quad (4.2)$$

Utilizando (4.1) na log-verossimilhança (2.4) e denotando

$$L(\beta, \boldsymbol{\lambda} | \mathbf{X}, \mathbf{Y}) = \exp \ell(\beta, \boldsymbol{\lambda}),$$

onde $\mathbf{X} = (X_1, \dots, X_n)$ e $\mathbf{Y} = (Y_1, \dots, Y_n)$, podemos escrever

$$\begin{aligned} L(\beta, \boldsymbol{\lambda} | \mathbf{X}, \mathbf{Y}) &\propto (\lambda_2 \sigma^2)^{-n/2} \exp \left\{ -\frac{n}{2D} (a_1 S_{xx} - 2a_2 S_{xy} + a_3 S_{yy}) \right\} \\ &\times \exp \left\{ -\frac{n}{2D} a_1 \left[\lambda_0 - \left(\bar{X} - \frac{a_2}{a_1}(\bar{Y} - \lambda_1) \right) \right]^2 \right\} \\ &\times \exp \left\{ -\frac{n}{2D} \left(\frac{a_2^2}{a_1} - a_3 \right) (\bar{Y} - \lambda_1)^2 \right\}, \end{aligned} \quad (4.3)$$

onde $D = \lambda_2 \sigma^2 (\beta^2 + 1)$. Conforme verificado no Apêndice A, temos que a priori MDIP para este caso, é tal que

$$p(\beta, \lambda_0, \lambda_1, \lambda_2) \propto \lambda_2^{-1/2}. \quad (4.4)$$

Combinando a priori (4.4) com a verossimilhança (4.3) e integrando em λ_0 e λ_1 , temos que a posteriori conjunta de β e λ_2 é dada por

$$\pi(\beta, \lambda_2 | \mathbf{X}, \mathbf{Y}) \propto (\lambda_2)^{-(n+1)/2} d_1 d_2 \exp \left\{ -\frac{n}{2D} (a_1 S_{xx} - 2a_2 S_{xy} + a_3 S_{yy}) \right\}, \quad (4.5)$$

onde

$$d_1 = \left(\frac{D}{n a_1} \right)^{1/2} \text{ e } d_2 = \left(\frac{D}{n [a_3 - (a_2^2/a_1)]} \right)^{1/2},$$

onde S_{xx} , S_{xy} e S_{yy} são como em (3.34). Integrando (4.5) em λ_2 , obtem-se a posteriori marginal de β , que é dada por

$$\pi(\beta | \mathbf{X}, \mathbf{Y}) \propto \left(\frac{\beta^2 + 1}{S_{xx} + 2\beta S_{xy} + \beta^2 S_{yy}} \right)^{\frac{n-2}{2}} \exp \left\{ -\frac{n}{2\sigma^2} \frac{(\beta^2 S_{xx} - 2\beta S_{xy} + S_{yy})}{\beta^2 + 1} \right\}. \quad (4.6)$$

A moda desta posteriori pode ser obtida por métodos numéricos resolvendo-se uma equação polinomial de grau 5 em β , ou seja,

$$b_5 \beta^5 + b_4 \beta^4 + b_3 \beta^3 + b_2 \beta^2 + b_1 \beta + b_0 = 0, \quad (4.7)$$

onde

$$\begin{aligned} b_5 &= -2cS_{yy} - nS_{yy}S_{xx}, \quad c = (n-2)\sigma^2, \\ b_4 &= cS_{yy} - 2cS_{xy} - S_{yy}S_{xy} - 2nS_{xy}S_{xx}, \\ b_3 &= 2cS_{xy} - 3cS_{yy} - 2nS_{xy}^2 + nS_{yy}^2 - nS_{xx}^2, \\ b_2 &= cS_{xx} - 2cS_{xy} - nS_{xy}S_{xx} + 3S_{xy}S_{xx}, \\ b_1 &= cS_{xx} - 2cS_{yy} + 2nS_{xy}^2 + nS_{xx}S_{yy}, \\ b_0 &= -2cS_{xy} + nS_{xx}S_{yy}. \end{aligned} \quad (4.8)$$

Equações polinomiais similares aparecem em Barnett (1969) e Dolby et al. (1987). Como se vê, determinar a moda exata é trabalhoso. Quando n é grande, o resultado seguinte apresenta uma aproximação normal para (4.6), o que torna mais simples a obtenção de inferências sobre β .

Teorema 4.2.1 *A distribuição assintótica a posteriori de β é dada por*

$$\beta|\mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\hat{\beta}, \hat{\sigma}_\beta^2), \quad (4.9)$$

onde $\hat{\beta}$ é dado em (1.6) e

$$\hat{\sigma}_\beta^2 = \frac{\sigma^2(S_{xx} + S_{yy} + \sqrt{\Delta})\hat{\beta}(\hat{\beta}^2 + 1)}{nS_{xy}(S_{xx} + S_{yy} + \sqrt{\Delta} - 2\sigma^2)},$$

com

$$\Delta = (S_{yy} - S_{xx})^2 + 4S_{xy}^2.$$

Prova: De (4.6) temos que

$$\log \pi(\beta|\mathbf{X}, \mathbf{Y}) \propto -\frac{n}{2} \frac{Q_2}{\sigma^2(\beta^2 + 1)} - \frac{n-2}{2} \log \left(\frac{Q_1}{\beta^2 + 1} \right),$$

onde $\mathbf{X} = (X_1, \dots, X_n)$, $\mathbf{Y} = (Y_1, \dots, Y_n)$,

$$Q_2 = \beta^2 S_{xx} - 2\beta S_{xy} + S_{yy} \text{ e } Q_1 = S_{xx} + 2\beta S_{xy} + \beta^2 S_{yy}. \quad (4.10)$$

Para n suficientemente grande, $n-2 \approx n$ e como $Q_1 + Q_2 = (\beta^2 + 1)(S_{xx} + S_{yy})$, temos que,

$$\log \pi(\beta|\mathbf{X}, \mathbf{Y}) \propto -\frac{n}{2} \left\{ \frac{Q_2}{\sigma^2(\beta^2 + 1)} + \log \left[S_{xx} + S_{yy} - \frac{Q_2}{(\beta^2 + 1)} \right] \right\}. \quad (4.11)$$

De (1.9), podemos escrever que

$$\frac{Q_2}{\beta^2 + 1} = \gamma + \frac{(S_{xx} - \gamma)}{\beta^2 + 1}(\beta - \hat{\beta})^2, \quad (4.12)$$

com γ dado em (1.10) e com $\sigma_e^2 = \sigma_u^2 = 1$. De (4.11) e (4.12), temos, para n grande, que

$$\begin{aligned} \log \pi(\beta|\mathbf{X}, \mathbf{Y}) &\propto -\frac{n}{2} \left\{ \frac{1}{\sigma^2} \left[\gamma + \frac{(S_{xx} - \gamma)}{\beta^2 + 1} (\beta - \hat{\beta})^2 \right] \right. \\ &\quad \left. + \log \left(1 - \frac{(S_{xx} - \gamma)}{(S_{xx} + S_{yy} - \gamma)} \frac{(\beta - \hat{\beta})^2}{\beta^2 + 1} \right) \right\}. \end{aligned}$$

De (1.13) e como

$$\gamma \xrightarrow{P} \sigma^2, \quad (4.13)$$

quando $n \rightarrow \infty$, segue que

$$\hat{\beta} \xrightarrow{P} \beta. \quad (4.14)$$

Então, de (4.14) e usando o fato de que $\log(1 + x) \approx x$ para x próximo de zero, podemos escrever, para n grande,

$$\log \pi(\beta|\mathbf{X}, \mathbf{Y}) \propto -\frac{n}{2} \left[\frac{\gamma}{\sigma^2} + d \frac{(\beta - \hat{\beta})^2}{\beta^2 + 1} \right], \quad (4.15)$$

onde

$$d = \frac{(S_{xx} - \gamma)(S_{xx} + S_{yy} - \gamma - \sigma^2)}{\sigma^2(S_{xx} + S_{yy} - \gamma)}.$$

Portanto, maximizar (4.15) com respeito a β , corresponde a minimizar $nd(\beta - \hat{\beta})^2/(\beta^2 + 1)$, de modo que $\hat{\beta}$ maximiza $\log \pi(\beta|\mathbf{X}, \mathbf{Y})$. A variância assintótica correspondente a distribuição a posteriori de β dado $\mathbf{X}$ e $\mathbf{Y}$, denotada por $\hat{\sigma}_\beta^2$ pode ser obtida a partir da segunda derivada de (4.15), ou seja,

$$\begin{aligned} \hat{\sigma}_\beta^2 &= \left[\frac{-\partial^2 \log \pi(\beta|\mathbf{X}, \mathbf{Y})}{\partial \beta^2} \Big|_{\beta=\hat{\beta}} \right]^{-1} \\ &= \frac{\sigma^2(\hat{\beta}^2 + 1)(S_{xx} + S_{yy} - \gamma)}{n(S_{xx} - \gamma)(S_{xx} + S_{yy} - \gamma - \sigma^2)} \\ &= \frac{\sigma^2(S_{xx} + S_{yy} + \sqrt{\Delta})(\hat{\beta}^2 + 1)\hat{\beta}}{nS_{xy}(S_{xx} + S_{yy} + \sqrt{\Delta} - 2\sigma^2)}, \end{aligned} \quad (4.16)$$

concluindo-se assim a prova. $\square$

De (1.13), (4.13) e (4.14), tem-se que, para $n \rightarrow \infty$,

$$\hat{\sigma}_\beta^2 \xrightarrow{P} \frac{\lambda_2 \sigma^2 (\beta^2 + 1)^2}{(\lambda_2 - \sigma^2)^2},$$

onde o lado direito coincide com a variância assintótica do estimador de máxima verossimilhança $\hat{\beta}$, que segue de (2.5). Este resultado estabelece uma relação entre a variância da distribuição a posteriori de β e a variância do estimador de máxima verossimilhança $\hat{\beta}$ em grandes amostras.

Desprezando o primeiro fator no lado direito de (4.6), ou seja, considerando apenas

$$\pi(\beta|\mathbf{X}, \mathbf{Y}) \propto \exp \left\{ -\frac{n}{2\sigma^2} \frac{(\beta^2 S_{xx} - 2\beta S_{xy} + S_{yy})}{\beta^2 + 1} \right\}, \quad (4.17)$$

não é difícil ver, usando o enfoque do Teorema 4.2.1 que a aproximação normal para a posteriori (4.17) é dada por

$$\beta|\mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\hat{\beta}, \hat{\sigma}_\beta^2), \quad (4.18)$$

com

$$\hat{\sigma}_\beta^2 = \frac{\sigma^2(\hat{\beta}^2 + 1)}{n(S_{xx} - \gamma)}. \quad (4.19)$$

Note de (4.16) que a variância assintótica dada pelo Teorema 4.2.1 é maior que a variância assintótica dada em (4.19). Uma explicação para este fato é que em (4.17) se despreza um fator importante na posteriori $\pi(\beta|\mathbf{X}, \mathbf{Y})$, fator este que também contém informação sobre a variabilidade de β .

Um intervalo aproximado de $100(1 - \frac{\alpha}{2})\%$ de credibilidade para β é dado por

$$(\hat{\beta} \mp z_{\alpha/2} \hat{\sigma}_\beta), \quad (4.20)$$

sendo $\hat{\sigma}_\beta$ dado no Teorema 4.2.1 e $z_{\alpha/2}$ representando o percentil $100(1 - \frac{\alpha}{2})\%$ da distribuição normal padrão.

Exemplo 4.1. Os dados da Tabela 4.1 são apresentados em Madansky (1959), onde x representa o diâmetro da calota observada durante o ensaio de dureza Brinell e y o valor da dureza Brinell. Neste caso, $n = 50$ e σ_e^2 e σ_u^2 são conhecidos, com $\sigma_e^2 = 7500$ e $\sigma_u^2 = 50$. Transformando-se os dados originais de modo que $\sigma_e^2 = \sigma_u^2 = 50$, a matriz $\mathbf{B}$ dada em (1.11) se reduz a

$$\mathbf{B} = \begin{pmatrix} 715,19 & 191,74 \\ 191,74 & 92,87 \end{pmatrix}.$$

Então usando o Teorema 4.2.1 temos que a posteriori de β é aproximadamente normal com $\hat{\beta} = 0,28376$ e $\hat{\sigma}_\beta^2 = 0,00161$. Segue portanto de (4.20) que um intervalo aproximado de 95% de credibilidade para β é igual a $(0,20506 ; 0,36246)$. Por outro lado, a variância assintótica obtida a partir de (4.19) é igual a 0,00151, de modo que um intervalo aproximado de 95% de credibilidade para β neste caso é dado por $(0,20754 ; 0,35999)$. Portanto, parece claro que o fator desprezado na posteriori (4.17) contém informação importante sobre o parâmetro β . Note também que com $\phi = 150$, $\hat{\beta} = \sqrt{150}(0,28376) = 3,47533$, que é o valor originalmente obtido por Madansky (1959). Também, neste caso, $\sigma_\beta^2 = (180)(0,00161) = 0,24500$.

□

4.3 Variância σ^2 desconhecida

Neste caso, assumimos que $\phi = 1$, ou seja, $\sigma_e^2 = \sigma_u^2 = \sigma^2$, com σ^2 desconhecido, de modo que a função de verossimilhança é como em (2.4), com λ_3 no lugar de σ^2 . Considerando a priori MDIP para (β, λ) (ver Apêndice A), temos que

$$p(\beta, \lambda) \propto (\lambda_2 \lambda_3)^{-1/2}.$$

Tabela 4.1: Dados de um ensaio de Dureza Brinell

Baixa Temperatura		Alta Temperatura	
x	y	x	y
229	845	277	900
230	810	285	800
235	750	285	815
235	750	285	815
235	755	285	925
235	755	285	965
235	765	285	970
235	795	285	970
235	795	285	975
235	930	285	975
239	905	285	975
241	760	285	1010
241	760	285	1030
241	760	285	1045
241	795	285	1150
241	800	285	1160
241	805	293	940
241	815	293	1005
241	825	293	1005
241	825	293	1015
241	835	293	1040
241	870	302	935
241	875	302	1075
241	960	302	1095
241	1050	321	1140

Usando a identidade (4.1), segue que a posteriori conjunta de (β, λ) , $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$, é dada por

$$\begin{aligned} \pi(\beta, \lambda | \mathbf{X}, \mathbf{Y}) &\propto (\lambda_2 \lambda_3)^{-(n-1)/2} \exp \left\{ -\frac{n}{2D} (a_1 S_{xx} - 2a_2 S_{xy} + a_3 S_{yy}) \right\} \\ &\times \exp \left\{ -\frac{n}{2D} a_1 \left[\lambda_0 - \left(\bar{X} - \frac{a_2}{a_1} (\bar{Y} - \lambda_1) \right) \right]^2 \right\} \\ &\times \exp \left\{ -\frac{n}{2D} \frac{a_2^2}{a_1} (\bar{Y} - \lambda_1)^2 + n a_3 (\bar{Y} - \lambda_1)^2 \right\}, \end{aligned} \quad (4.21)$$

onde

$$D = \lambda_2 \lambda_3 (\beta^2 + 1), \quad a_1 = \beta^2 \lambda_2 + \sigma_3, \quad a_2 = \beta (\lambda_2 - \lambda_3) \quad e \quad a_3 = \lambda_2 + \beta^2 \lambda_3.$$

Integrando (4.21) em λ_0 e λ_1 , temos que

$$\begin{aligned} \pi(\beta, \lambda_2, \lambda_3 | \mathbf{X}, \mathbf{Y}) &\propto (\lambda_2 \lambda_3)^{-n/2} \left(\frac{D^{1/2}}{n^{1/2} a_1} \right) \left(\frac{D^{1/2}}{n^{1/2} [a_3 - \frac{a_2^2}{a_1}]} \right) \\ &\times \exp \left[-\frac{n}{2D} (a_1 S_{xx} - 2a_2 S_{xy} + a_3 S_{yy}) \right]. \end{aligned}$$

Finalmente integrando em λ_2 e λ_3 obtem-se a posteriori marginal de β , ou seja,

$$\pi(\beta | \mathbf{X}, \mathbf{Y}) \propto \left[\frac{\beta^2 + 1}{(S_{xx} + 2\beta S_{xy} + \beta^2 S_{yy})} \frac{\beta^2 + 1}{(\beta^2 S_{xx} - 2\beta S_{xy} + S_{yy})} \right]^{\frac{n-2}{2}}, \quad (4.22)$$

cujas moda satisfaz à equação

$$\sum_{i=0}^5 c_i \beta^i = 0,$$

com

$$\begin{aligned} c_5 &= c S_{yy} S_{xx} + 4c S_{yy} S_{xx} - c S_{xx} S_{yy}, \quad c = n - 2, \\ c_4 &= 8c S_{xy} S_{xx} - 8c S_{yy} S_{xy} - c S_{xy} S_{xx} + 2c S_{yy} S_{xy} - 2c S_{xx} S_{yy} + c S_{xy} S_{yy}, \\ c_3 &= -12c S_{xy}^2 + 3c S_{xx}^2 + 5S_{yy}^2 - c S_{xx} S_{yy} - c S_{yy} S_{xx}, \\ c_2 &= -10c S_{xx} S_{xy} + 10S_{xy} S_{yy}, \quad c_1 = 4c S_{xx} S_{yy} + 4c S_{xy}^2 - c S_{yy}^2 - c S_{xx}^2, \\ c_0 &= c S_{xy} (S_{xx} + S_{yy}). \end{aligned}$$

Lindley e El Sayad (1968), obtem através de métodos aproximados uma expressão próxima de (4.22) para a posteriori de β . O resultado seguinte apresenta uma aproximação para (4.22), para n grande.

Teorema 4.3.1 *A distribuição assintótica a posteriori de β é dada por*

$$\beta|\mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\hat{\beta}, \hat{\sigma}_\beta^2),$$

onde $\hat{\beta}$ é dado por (1.6) e

$$\hat{\sigma}_\beta^2 = \frac{(S_{xx}S_{yy} - S_{xy}^2)\hat{\beta}(\hat{\beta}^2 + 1)}{nS_{xy}\sqrt{\Delta}}, \quad (4.23)$$

com

$$\Delta = (S_{yy} - S_{xx})^2 + 4S_{xy}^2.$$

Prova: De (4.22) e (4.10) e como, para n grande $n - 2 \approx n$, podemos escrever com γ definido em (4.12), que

$$\begin{aligned} \log \pi(\beta|\mathbf{X}, \mathbf{Y}) &\propto -\frac{n}{2} \left\{ \log \left[(S_{xx} + S_{yy} - \gamma) - \frac{(S_{xx} - \gamma)}{\beta^2 + 1} (\beta - \hat{\beta})^2 \right] \right. \\ &\quad \left. + \log \left[\gamma + \frac{(S_{xx} - \gamma)}{\beta^2 + 1} (\beta - \hat{\beta})^2 \right] \right\}. \end{aligned}$$

Dada a consistência de $\hat{\beta}$ e o fato de que $\log(1 + x) \approx x$, para x próximo de 0, podemos escrever, para n grande,

$$\log \pi(\beta|\mathbf{X}, \mathbf{Y}) \propto -\frac{n}{2} \left\{ \frac{(S_{xx} - \gamma)}{\gamma} \frac{(S_{xx} + S_{yy} - 2\gamma)}{(S_{xx} + S_{yy} - \gamma)} \frac{(\beta - \hat{\beta})^2}{(\beta^2 + 1)} \right\}. \quad (4.24)$$

Maximizar (4.24) corresponde a minimizar $(\beta - \hat{\beta})^2/(\beta^2 + 1)$, de modo que $\hat{\beta}$ dado em (1.6) maximiza (4.24). A variância assintótica a posteriori de β , $\hat{\sigma}_\beta^2$, pode ser obtida a partir da segunda derivada de (4.24), ou seja,

$$\hat{\sigma}_\beta^2 = \left[\frac{-\partial^2 \log \pi(\beta|\mathbf{X}, \mathbf{Y})}{\partial \beta^2} \Big|_{\beta=\hat{\beta}} \right]^{-1}$$

$$\begin{aligned}
&= \frac{\gamma(S_{xx} + S_{yy} - \gamma)(\hat{\beta}^2 + 1)}{n(S_{xx} - \gamma)(S_{xx} + S_{yy} - 2\gamma)} \\
&= \frac{(S_{xx} + S_{yy} - S_{xy}^2)(\hat{\beta}^2 + 1)\hat{\beta}}{nS_{xy}\sqrt{\Delta}}, \tag{4.25}
\end{aligned}$$

concluindo-se assim a prova. $\square$

Quando $n \rightarrow \infty$, de (1.13), (4.13) e (4.14) tem-se que

$$\hat{\sigma}_\beta^2 \xrightarrow{P} \frac{\lambda_2 \lambda_3 (\beta^2 + 1)^2}{(\lambda_2 - \lambda_3)^2},$$

onde o lado direito é idêntico à variância assintótica de $\hat{\beta}$ que segue de (2.8). Este resultado estabelece uma conexão entre a variância da distribuição a posteriori de β e a variância da distribuição amostral do estimador de máxima verossimilhança (que também é $\hat{\beta}$) em grandes amostras. Lindley e El-Sayad (1968) sugerem que tal fato pode ocorrer mas não chegam a obter um resultado formal como a convergência acima.

Lindley e El-Sayad (1968), usando um enfoque diferente do proposto acima mostram que, para n grande,

$$\beta | \mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\hat{\beta}, \hat{\sigma}_L^2), \tag{4.26}$$

com

$$\hat{\sigma}_L^2 = \frac{\hat{\beta}^2(S_{xx}S_{yy} - S_{xy}^2)}{nS_{xy}^2}. \tag{4.27}$$

Note que as expressões (4.25) e (4.27) coincidem, pois podemos escrever

$$\hat{\sigma}_\beta^2 = \hat{\sigma}_L^2 \left[\frac{S_{xy}(\hat{\beta}^2 + 1)}{\sqrt{\Delta}\hat{\beta}} \right]^2,$$

e como

$$\hat{\beta} = \frac{S_{yy} - S_{xx} + \sqrt{\Delta}}{2S_{xy}},$$

temos, após algumas manipulações algébricas, que

$$\frac{S_{xy}(\hat{\beta}^2 + 1)}{\sqrt{\Delta}\hat{\beta}} = \frac{4S_{xy}^2 + (S_{xx} - S_{yy})^2 + (S_{xx} - S_{yy})\sqrt{\Delta}}{\sqrt{\Delta}(S_{xx} - S_{yy} + \sqrt{\Delta})} = 1,$$

com Δ dado no Teorema 4.3.1. Note de (4.16) e (4.25) que as expressões para as variâncias assintóticas nos Teoremas 4.2.1 e 4.3.1 são similares, com σ^2 do Teorema 4.2.1 substituído por γ no Teorema 4.3.1. No caso em que temos o modelo $y_i^* = \alpha_* + \beta_* x_i + e_i^*$, onde $\text{var}(e_i^*) = \sigma_e^2$, com $\phi = \sigma_e^2/\sigma_u^2 \neq 1$, temos que

$$\hat{\beta}_* = \hat{\beta}\phi^{1/2} \quad (4.28)$$

e

$$\hat{\sigma}_{\beta_*}^2 = \phi\hat{\sigma}_{\beta}^2, \quad (4.29)$$

onde $\hat{\beta}$ e $\hat{\sigma}_{\beta}^2$ são calculados com $\phi = 1$.

Neste caso, intervalo aproximado de $(1 - \frac{\alpha}{2})100\%$ de credibilidade para β_* é dado por

$$(\hat{\beta}_* \mp z_{\alpha/2}\hat{\sigma}_{\beta_*}). \quad (4.30)$$

Exemplo 4.2. Os dados apresentados a seguir são do Exemplo 1.3.1 em Fuller (1987). Representam contagens de faisões em Iowa em duas estações do ano. O tamanho da amostra é $n = 15$ e os pares observados (X_i, Y_i) são:

$(9, 0; 8, 0), (6, 6; 6, 0), (12, 3; 9, 8), (11, 9; 10, 8), (11, 9; 9, 7), (12, 0; 9, 3), (9, 6; 9, 2),$
 $(7, 5; 6, 9), (10, 9; 8, 1), (10, 4; 8, 7), (10, 2; 8, 7), (7, 4; 7, 4), (11, 0; 10, 1), (11, 8; 10, 0),$
 $(8, 2; 7, 3)$. Para estes dados, $S_{xx}^* = 3,3798$, $S_{yy}^* = 1,7262$, $S_{xy}^* = 2,1949$, $\phi = 1/6$ e $\hat{\beta}_* = 0,7516$. De (4.25) ou (4.27), temos que a variância a posteriori de β_* é dada por 0,00795 e de (4.30), um intervalo aproximado de 95% de credibilidade para β_* é $(0,5774 ; 0,9258)$. Pelo enfoque clássico dado em Fuller, Seção 1.3, sem

Figura 4.1: Posteriori normalizada de β_* , com σ^2 desconhecido.

a utilização de parametrização ortogonal, a variância assintótica de $\hat{\beta}_*$ é 0,00925 com um intervalo aproximado de 95% de confiança para β sendo dado por (0,5631 ; 0,9401). A Figura 4.1 mostra a posteriori (4.22) normalizada, que é levemente assimétrica.

Como visto acima, a utilização de uma parametrização ortogonal no caso em que ϕ é conhecido permite a obtenção de posteriores exatas para β . Lembramos que os resultados apresentados em Lindley e El-Sayad (1968) são aproximados e obtidos apenas para o caso $\mu = \alpha = 0$.

4.4 Coeficiente de atenuação k_x conhecido

Nesta Seção consideramos o caso em que o coeficiente de atenuação k_x é conhecido. Considere a verossimilhança (2.20) e novamente a identidade (4.1),

com

$$a_1 = \frac{\lambda_0 + \lambda_1 \beta^2}{k+1}, \quad a_2 = \beta \lambda_1 \quad e \quad a_3 = \lambda_1(k+1). \quad (4.31)$$

Substituindo (4.1) com a_0 , a_1 e a_2 dados em (4.31), em (2.20) e denotando

$$L(\beta, \boldsymbol{\lambda} | \mathbf{X}, \mathbf{Y}) = \exp \ell(\beta, \boldsymbol{\lambda}),$$

temos que

$$\begin{aligned} L(\beta, \boldsymbol{\lambda} | \mathbf{X}, \mathbf{Y}) &\propto (\lambda_0 \lambda_1)^{-n/2} \exp \left\{ -\frac{n}{2\lambda_0 \lambda_1} (a_1 S_{xx} - 2a_2 S_{xy} + a_3 S_{yy}) \right\} \\ &\times \exp \left\{ -\frac{n a_1}{2\lambda_0 \lambda_1} \left[\lambda_2 - \left(\bar{X} - \frac{a_2}{a_1} (\bar{Y} - \lambda_3) \right)^2 \right] \right\} \\ &\times \exp \left\{ \frac{-n(a_1 a_3 - a_2^2)}{2\lambda_0 \lambda_1 a_1} (\bar{Y} - \lambda_3)^2 \right\}. \end{aligned} \quad (4.32)$$

Consideramos inicialmente priori uniforme para λ_2 e λ_3 e priori não informativa (Jeffreys) para (λ_0, λ_1) , que segue da matriz de informação (2.21). Admitimos então, com $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$, que

$$\begin{aligned} p(\beta, \boldsymbol{\lambda}) &= p(\lambda_2, \lambda_3) p(\beta, \lambda_0, \lambda_1 | \lambda_2, \lambda_3) \\ &\propto (\lambda_0 \lambda_1)^{-1}, \end{aligned} \quad (4.33)$$

ou seja, a priori, (λ_2, λ_3) é independente de (λ_0, λ_1) . Combinando a priori (4.33) com a verossimilhança (4.32) e integrando em λ_2, λ_3 temos que a posteriori conjunta de $(\beta, \lambda_0, \lambda_1)$ é dada por

$$\begin{aligned} \pi(\beta, \lambda_0, \lambda_1 | \mathbf{X}, \mathbf{Y}) &\propto (\lambda_0 \lambda_1)^{-\frac{n+1}{2}} \exp \left\{ \frac{-n S_{xx}}{2(k+1)\lambda_1} \right\} \\ &\times \exp \left\{ -\frac{n}{2\lambda_0} \left[\frac{S_{xx} \beta^2}{k+1} - 2\beta S_{xy} + (k+1) S_{yy} \right] \right\}. \end{aligned} \quad (4.34)$$

Integrando (4.34) em λ_0 e λ_1 obtem-se a posteriori marginal de β ,

$$\pi(\beta | \mathbf{X}, \mathbf{Y}) \propto \left[1 + \frac{(\beta - \hat{\beta})^2}{(k+1)^2 \left(\frac{S_{yy}}{S_{xx}} - \frac{S_{xy}^2}{S_{xx}^2} \right)} \right]^{-\frac{n-1}{2}}, \quad (4.35)$$

que é proporcional à distribuição t -Student com $n-2$ graus de liberdade, centrada em $\hat{\beta}$ e com variância

$$\text{var}(\beta|\mathbf{X}, \mathbf{Y}) = \frac{n-2}{n-4}(k+1)^2 \left(\frac{S_{yy}}{S_{xx}} - \frac{S_{xy}^2}{S_{xx}^2} \right).$$

Um intervalo exato de $(1 - \frac{\alpha}{2})100\%$ de credibilidade para β é dado por

$$\left(\hat{\beta} \mp t_{\alpha}(k+1) \sqrt{\left(\frac{S_{yy}}{S_{xx}} - \frac{S_{xy}^2}{S_{xx}^2} \right) / (n-2)} \right), \quad (4.36)$$

que coincide com o intervalo obtido por Fuller (1987) usando argumentos um tanto heurísticos. Usando a priori MDIP (veja Apêndice A), que neste caso é dada por $p(\beta, \lambda) \propto (\lambda_2 \lambda_3)^{-1/2}$, obtem-se de forma análoga, uma posteriori t -Student com a mesma média mas com $n-1$ graus de liberdade.

Exemplo 4.3. Os dados deste exemplo aparecem em (Fuller, 1987, pg.5). Temos $n = 12$ pares (X_i, Y_i) , dados por $(3,9;3,6)$, $(2,9;2,5)$, $(4,4;3,9)$, $(5,9;5,5)$, $(5,4;4,9)$, $(4,2;4,5)$, $(2,3;2,9)$, $(4,5;5,2)$, $(3,5;2,7)$, $(6,0;5,8)$, $(3,3;4,1)$ e $(4,1;5,1)$, que supostamente representam uma amostra do modelo estrutural com $k_x = 0,85$. Para esta amostra, $S_{xx} = 1,2$, $S_{xy} = 0,9358$ e $S_{yy} = 1,0064$. Um intervalo de 95% de confiança para β , obtido por Fuller (1987) é $(0,5196 ; 1,3154)$, sendo também um intervalo de 95% de credibilidade obtido a partir de (4.36).

4.5 Inferência Bayesiana no modelo funcional

Conforme verificado no Capítulo 2, no caso do modelo funcional apresentado no Capítulo 1, a função de verossimilhança pode ser escrita como

$$\begin{aligned} L(\beta, \mathbf{x}) &\propto \frac{1}{\sigma_u^n} \exp \left\{ -\frac{1}{2\sigma_u^2} \sum_{i=1}^n (X_i - x_i)^2 \right\} \\ &\times \frac{1}{\sigma_e^n} \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{i=1}^n (Y_i - \beta x_i)^2 \right\}. \end{aligned} \quad (4.37)$$

Sendo σ_e^2 e σ_u^2 conhecidos, consideramos a priori não informativa

$$p(\beta, \mathbf{x}) \propto \text{constante}, \quad (4.38)$$

que coincide com a priori MDIP (veja o apêndice A). Esta priori não informativa é também considerada em Lindley (1966). Combinando a verossimilhança (4.37) com a priori (4.38) temos que a posteriori conjunta de $(\beta, \mathbf{x})$ é dada por

$$\pi(\beta, \mathbf{x} | \mathbf{X}, \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2\sigma_u^2} \sum_{i=1}^n (X_i - x_i)^2 - \frac{1}{2\sigma_e^2} \sum_{i=1}^n (Y_i - \beta x_i)^2 \right\},$$

que pode ser escrita como

$$\pi(\beta, \mathbf{x} | \mathbf{X}, \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2} F(\mathbf{x}, \beta) \right\}, \quad (4.39)$$

onde $F(\mathbf{x}, \beta)$ é dada em (1.7). Integrando a posteriori (4.39) em relação à x_i , $i = 1, \dots, n$, e usando (1.9), temos que

$$\pi(\beta | \mathbf{X}, \mathbf{Y}) \propto \frac{1}{\left(\frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right)^{n/2}} \exp \left\{ -\frac{n}{2} \left[\gamma + \frac{(S_{xx} - \gamma\sigma_u^2)}{\beta^2\sigma_u^2 + \sigma_e^2} (\beta - \hat{\beta})^2 \right] \right\}, \quad (4.40)$$

onde γ é o menor autovalor de (1.10) e $\hat{\beta}$ é dado em (1.6). Como a posteriori (4.40) tem um comportamento, complicado em relação aos momentos e outras características, Lindley (1966) sugere desprezar o fator fora da exponencial e considerar a densidade

$$\pi(\beta | \mathbf{X}, \mathbf{Y}) \propto \exp \left\{ -\frac{n}{2} \left[\gamma + \frac{(S_{xx} - \gamma\sigma_u^2)}{\beta^2\sigma_u^2 + \sigma_e^2} (\beta - \hat{\beta})^2 \right] \right\}. \quad (4.41)$$

Usando o enfoque considerado nos Teoremas 4.2.1 e 4.3.1, temos de (4.41), para n grande, que

$$\beta | \mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\hat{\beta}, \hat{\sigma}_\beta^2),$$

onde

$$\hat{\sigma}_\beta^2 = \frac{\hat{\beta}^2\sigma_u^2 + \sigma_e^2}{n(S_{xx} - \gamma\sigma_u^2)}.$$

Note que $\hat{\sigma}_\beta^2$ acima é idêntica à variância assintótica apresentada em (4.18) para o caso estrutural quando $\sigma_u^2 = \sigma_e^2 = \sigma^2$ é conhecido (neste caso (4.41) e (4.17) coincidem). Por outro lado, considerando agora a transformação ortogonal (2.51) com a verossimilhança (4.40), juntamente com a priori MDIP,

$$p(\beta, \boldsymbol{\lambda}) \propto \text{constante},$$

obtemos após integrar $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$ que

$$\pi(\beta|\mathbf{X}, \mathbf{Y}) \propto \exp \left\{ -\frac{n}{2} \left[\gamma + \frac{(S_{xx} - \gamma\sigma_u^2)}{\beta^2\sigma_u^2 + \sigma_e^2} (\beta - \hat{\beta})^2 \right] \right\}.$$

Note que neste caso o fator fora da exponencial que aparece em (4.40) não está presente. Portanto a transformação ortogonal (2.51) juntamente com a priori não informativa acima justifica exatamente a posteriori (4.41), sem a necessidade de se desprezar nenhum fator fora da exponencial. Usando (1.9) tem-se que $\pi(\beta|\mathbf{X}, \mathbf{Y})$ é idêntico à (4.41), com a mesma aproximação normal para n grande. Considerando agora que $\sigma_u^2 = \sigma_e^2 = \sigma^2$ é desconhecido, assumindo a priori

$$p(\beta, \sigma^2, \mathbf{x}) \propto \frac{1}{\sigma}, \quad (4.42)$$

considerada em Lindley (1966), e combinando (4.42) com (4.37), temos que a posteriori conjunta de $(\beta, \sigma^2, \mathbf{x})$ é dada por

$$\begin{aligned} \pi(\beta, \sigma^2, \mathbf{x}|\mathbf{X}, \mathbf{Y}) &\propto \frac{1}{\sigma^{2n+1}} \exp \left\{ -\frac{n}{2\sigma^2} \left[\gamma + \frac{(S_{xx} - \gamma\sigma_u^2)}{\beta^2\sigma_u^2 + \sigma_e^2} (\beta - \hat{\beta})^2 \right. \right. \\ &\quad \left. \left. + (\beta^2 + 1) \sum_{i=1}^n (x_i - \hat{x}_{i,\beta})^2 \right] \right\}. \end{aligned}$$

Integrando em x_i e σ temos,

$$\pi(\beta|\mathbf{X}, \mathbf{Y}) \propto \left(\frac{1}{\beta^2 + 1} \right)^{n/2} \left[\gamma + \frac{(S_{xx} - \gamma)}{\beta^2 + 1} (\beta - \hat{\beta})^2 \right]^{-n}. \quad (4.43)$$

Desprezando o termo $(\beta^2 + 1)^{-n/2}$ em (4.43) obtemos a posteriori

$$\pi(\beta|\mathbf{X}, \mathbf{Y}) \propto \left[\gamma + \frac{(S_{xx} - \gamma)}{\beta^2 + 1} (\beta - \hat{\beta})^2 \right]^{-n}. \quad (4.44)$$

Como

$$\log \pi(\beta|\mathbf{X}, \mathbf{Y}) \propto -n \log \left[\gamma + \frac{(S_{xx} - \gamma)}{\beta^2 + 1} (\beta - \hat{\beta})^2 \right],$$

usando a aproximação $\log(1 + x) \approx x$, justificada pela consistência de $\hat{\beta}$, temos para n grande, que

$$\beta|\mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\hat{\beta}, \hat{\sigma}_\beta^2),$$

onde $\hat{\beta}$ é dado em (1.6) e

$$\begin{aligned} \hat{\sigma}_\beta^2 &= \left[\frac{-\partial^2 \log \pi(\beta|\mathbf{X}, \mathbf{Y})}{\partial \beta^2} \Big|_{\beta=\hat{\beta}} \right]^{-1} \\ &= \frac{(\hat{\beta}^2 + 1)\gamma}{2n(S_{xx} - \gamma)}. \end{aligned} \quad (4.45)$$

Vamos procurar justificar agora a aproximação (4.44). Considerando a transformação ortogonal (2.51), temos que a função de verossimilhança é dada por

$$L(\beta, \boldsymbol{\lambda}) \propto (\lambda_0)^{-n} \exp \left\{ \frac{-1}{2\lambda_0} \left[\sum_{i=1}^n \left(X_i - \frac{\lambda_i}{\sqrt{1 + \beta^2}} \right)^2 + \sum_{i=1}^n \left(Y_i - \frac{\beta \lambda_i}{\sqrt{1 + \beta^2}} \right)^2 \right] \right\}, \quad (4.46)$$

onde $\lambda_0 = \sigma^2$ é desconhecido. Considerando a priori MDIP (veja apêndice A), que neste caso se reduz a

$$p(\beta, \boldsymbol{\lambda}) \propto \frac{1}{\lambda_0}, \quad (4.47)$$

temos de (4.46), após integrarmos em λ_i , $i = 1, \dots, n$, que a posteriori conjunta de (β, λ_0) , é dada por

$$\pi(\beta, \lambda_0|\mathbf{X}, \mathbf{Y}) \propto \frac{1}{\lambda_0^{\frac{n}{2}+1}} \exp \left\{ -\frac{1}{2\lambda_0} \left[\sum_{i=1}^n X_i^2 + \sum_{i=1}^n Y_i^2 - \sum_{i=1}^n \left(\frac{Y_i \beta + X_i}{\beta^2 + 1} \right)^2 \right] \right\}.$$

Note que neste caso não aparece nenhum fator fora da exponencial que dependa de β , como acontece em (4.44). Integrando em λ_0 e usando (1.9), temos que

$$\pi(\beta|\mathbf{X}, \mathbf{Y}) \propto \left[\gamma + \frac{(S_{xx} - \gamma)}{\beta^2 + 1} (\beta - \hat{\beta})^2 \right]^{-\frac{n}{2}}, \quad (4.48)$$

onde γ é o menor autovalor de (1.10) e $\hat{\beta}$ dado por (1.6). Pode-se mostrar usando os enfoques dados nos Teoremas 4.2.1 e 4.3.1, que uma aproximação normal para (4.48) tem média $\hat{\beta}$ e variância

$$\hat{\sigma}_\beta^2 = \frac{\gamma(\hat{\beta}^2 + 1)}{n(S_{xx} - \gamma)}.$$

4.6 Inferência Bayesiana no modelo funcional com a transformação polar

Nesta seção consideramos que a razão de variâncias $\phi = \sigma_e^2/\sigma_u^2$ é conhecida, de modo que, sem perda de generalidade, $\sigma_e^2 = \sigma_u^2 = \sigma^2$. Considerando a transformação polar definida na Seção 2.4.1, a função de verossimilhança que corresponde à amostra observada $(\mathbf{X}, \mathbf{Y})$ com $\mathbf{X} = (X_1, \dots, X_n)$ e $\mathbf{Y} = (Y_1, \dots, Y_n)$ pode ser escrita como

$$L(\boldsymbol{\omega}, \alpha, \sigma^2, \theta) \propto (\sigma)^{-2n} \exp \left\{ \frac{1}{2\sigma^2} \left[\sum_{i=1}^n (X_i - \omega_i \cos \theta)^2 + \sum_{i=1}^n (Y_i - \omega_i \sin \theta)^2 \right] \right\}, \quad (4.49)$$

onde $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)$. Consideramos a seguir posteriores para θ , nos casos σ^2 conhecido ou desconhecido. O enfoque pode ser estendido para o caso $\alpha \neq 0$ sem muitas dificuldades.

4.6.1 Variância σ^2 conhecida com parâmetro $\alpha = 0$

Supondo a priori que θ e ω são independentes e considerando priori uniforme para θ e priori de Jeffreys para ω (justificada pela ortogonalidade entre θ e ω), temos que

$$\begin{aligned} p(\theta, \omega) &\propto p(\omega)p(\theta|\omega) \\ &\propto \frac{1}{2\pi}, \quad 0 < \theta < 2\pi, \quad -\infty < \omega_i < \infty. \end{aligned} \quad (4.50)$$

Combinando (4.49) e (4.50), podemos escrever a posteriori conjunta de (ω, θ) como

$$\pi(\omega, \theta | \mathbf{X}, \mathbf{Y}) \propto (\sigma)^{-2n} c_1 c_2(\theta) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [\omega_i - (Y_i \sin \theta + X_i \cos \theta)]^2 \right\}, \quad (4.51)$$

onde

$$c_1 = \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i^2 + Y_i^2) \right\}, \quad c_2(\theta) = \exp \left\{ \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i \sin \theta + X_i \cos \theta)^2 \right\}.$$

Integrando em relação a ω , obtemos

$$\pi(\theta | \mathbf{X}, \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i \sin \theta - Y_i \cos \theta)^2 \right\}, \quad (4.52)$$

com $0 < \theta < 2\pi$. Derivando (4.52) com relação a θ , temos que a moda de θ é dada por

$$\text{tg}(2\hat{\theta}) = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2 - \sum_{i=1}^n Y_i^2}. \quad (4.53)$$

Uma expressão para a variância a posteriori de θ em grandes amostras, que denotamos por $\hat{\sigma}_\theta^2$ é dada por

$$\begin{aligned} \hat{\sigma}_\theta^2 &= \left[\frac{-\partial^2 \log \pi(\theta | \mathbf{X}, \mathbf{Y})}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} \right]^{-1} \\ &= \frac{\sigma^2}{n} \left[\cos(2\hat{\theta})(S_{xx} - S_{yy}) + 2\sin(2\hat{\theta})S_{xy} \right]^{-1}. \end{aligned} \quad (4.54)$$

Portanto, assintoticamente,

$$\theta|\mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\hat{\theta}, \hat{\sigma}_\theta^2), \quad (4.55)$$

e um intervalo aproximado de $(1 - \frac{\alpha}{2})100\%$ de credibilidade para θ com coeficiente de confiança $1 - \alpha$ é dado por

$$(\hat{\theta} \mp z_{\alpha/2}[\hat{\sigma}_\theta^2]^{1/2}), \quad (4.56)$$

onde $\hat{\sigma}_\theta^2$ é dado em (4.54).

4.6.2 Variância σ^2 desconhecida

Admitindo-se a priori não informativa

$$p(\omega, \theta, \sigma^2) \propto \frac{1}{\sigma}, \quad (4.57)$$

temos de forma análoga às derivações da seção anterior, que

$$\pi(\theta|\mathbf{X}, \mathbf{Y}) \propto \left[\sum_{i=1}^n (X_i \sin \theta - Y_i \cos \theta)^2 \right]^{-n/2}, \quad (4.58)$$

cujas moda é também $\hat{\theta}$ que segue de (4.53). Como na Seção anterior, calculando-se a derivada segunda de $\log \pi(\theta|\mathbf{X}, \mathbf{Y})$ com relação a θ , avaliada em $\hat{\theta}$, temos que variância a posteriori de θ é dada por

$$\begin{aligned} \hat{\sigma}_\theta^2 &= \frac{2A^2(\hat{\theta})}{n [A(\hat{\theta})A'(\hat{\theta}) - (A'(\hat{\theta}))^2]} \\ &= B(\hat{\theta}), \end{aligned} \quad (4.59)$$

onde

$$A(\theta) = \sum_{i=1}^n (X_i \sin \theta - Y_i \cos \theta)^2.$$

Segue de (4.53) e (4.59) que a distribuição assintótica a posteriori de θ é tal que

$$\theta|\mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\hat{\theta}, B(\hat{\theta})). \quad (4.60)$$

onde $B(\hat{\theta})$ é o estimador de máxima verossimilhança de $B(\theta)$ definido em (4.59). Um intervalo aproximado de $(1 - \frac{\alpha}{2})100\%$ de credibilidade para θ com coeficiente de confiança $1 - \alpha$ é dado por (4.60), de modo que

$$\left(\hat{\theta} \mp z_{\alpha/2} \sqrt{B(\hat{\theta})} \right), \quad (4.61)$$

onde $z_{\alpha/2}$ é como nos intervalos de credibilidade considerados anteriormente.

4.7 O Fator de Bayes a posteriori

Nesta seção, o fator de Bayes a posteriori (FBP) é considerado como uma alternativa para a comparação de vários modelos. O FBP é aplicado em uma extensão do modelo com erro nas variáveis definido no Capítulo 1 para s populações, sendo estudado no caso estrutural. Supõe-se que a razão de variâncias ϕ_i , $i = 1, \dots, s$, ou o coeficiente de atenuação k_{x_i} , $i = 1, \dots, s$, são conhecidos.

O FBP desenvolvido em Aitkin (1991) tem a vantagem de evitar o paradoxo de Lindley (Lindley, 1957). No caso de s populações obtém-se através do FBP uma estatística para testar hipóteses do tipo $H_0 : \beta_1 = \dots = \beta_s = \beta_0$ e $H_0 : \beta_1 = \dots = \beta_s = \beta$, onde β_0 é um valor especificado de β , supondo que na população i , $\sigma_{ei}^2 = \sigma_{ui}^2 = \sigma_i^2$, sem perda de generalidade, $i = 1, \dots, s$, ou com o coeficiente de atenuação k_{x_i} conhecido, $i = 1, \dots, s$, onde s é o número de populações.

4.7.1 Introdução

Para esclarecer a notação usada nesta seção vamos supor que $\mathbf{y}$ tem uma densidade $f_h(\mathbf{y}|\boldsymbol{\theta}_h)$, $h = 1, 2$, que satisfaz às condições de regularidade, depende do parâmetro $\boldsymbol{\theta}_h$ de dimensão p_h , com função de verossimilhança $L_h(\boldsymbol{\theta}_h)$ baseada

no modelo M_h e

$$L_h^A = \int L_h(\theta_h) \pi_h(\theta_h | y) d\theta_h, \quad (4.62)$$

onde $\pi(\theta_h | y)$ é a densidade posteriori de θ_h dada por

$$\pi(\theta_h | y) = \frac{L_h(\theta_h) \pi_h(\theta_h)}{\int L_h(\theta_h) \pi_h(\theta_h) d\theta_h}. \quad (4.63)$$

De (4.62) e (4.63), tem-se que

$$\begin{aligned} L_h^A &= \frac{\int L_h^2(\theta_h) \pi_h(\theta_h) d\theta_h}{\int L_h(\theta_h) \pi_h(\theta_h) d\theta_h} \\ &= \frac{\ell_h^{(2)}}{\ell_h^{(1)}}, \end{aligned} \quad (4.64)$$

$h = 1, 2$. Definimos o fator de Bayes a posteriori como a razão das médias à posteriori, isto é

$$FBP = \frac{L_1^A}{L_2^A} = \frac{\ell_1^{(2)} \ell_2^{(1)}}{\ell_1^{(1)} \ell_2^{(2)}}. \quad (4.65)$$

O fator definido em (4.65) é uma medida subjetiva da evidência da amostra em favor de M_1 quando comparado com M_2 . Valores do FBP menores que 1/20, 1/100 ou 1/1000 serão interpretados, respectivamente, como evidência forte, muito forte e extremamente forte em favor do modelo M_2 . Deste modo o modelo M_1 que representa a hipótese nula, terá maior credibilidade quanto maior for o FBP.

Exemplo 4.4. Suponhamos que sob o modelo M_1 , $y \sim \mathcal{N}(\mu_1, \sigma^2)$ com μ_1 conhecido, e que sob o modelo M_2 , $y \sim \mathcal{N}(\mu_2, \sigma^2)$, com μ_2 desconhecido. A informação sobre μ_2 é vaga e modelada com a priori uniforme $\pi(\mu_2) = 1/(2C)$ no intervalo $(-C, C)$, com C grande. Então o fator de Bayes a priori (FBB) para comparar os modelos M_1 e M_2 é dado por

$$FBB = \frac{2C \phi(z) n^{\frac{1}{2}} / \sigma}{\Phi \left[n^{\frac{1}{2}} (\bar{y} + C) / \sigma \right] - \Phi \left[n^{\frac{1}{2}} (\bar{y} - C) / \sigma \right]},$$

onde $\phi(z)$ é a distribuição de probabilidade normal padrão e $\Phi(z)$ sua função de distribuição acumulada. Nota-se que o denominador tende a 1 quando n cresce, de

modo que o valor do FBB torna-se arbitrariamente grande quando $C \rightarrow +\infty$ ou $n \rightarrow +\infty$ para um valor de z fixado. Pode ocorrer então que o FBB se aproxime de $+\infty$ sem que isto signifique evidência forte para M_1 , mas uma forte rejeição da priori difusa associada ao modelo M_2 . Esta situação é conhecida na literatura como o paradoxo de Lindley.

4.7.2 A estatística da razão de verossimilhança e o FBP

Seja M_1 um sub-modelo regular de M_2 , ou seja, $M_1 \subset M_2$ e suponha que a função de verossimilhança associada a M_1 satisfaz às condições usuais de regularidade. Consideremos o tamanho da amostra como sendo grande em relação ao número de parâmetros, a priori $\pi(\theta_h)$ localmente uniforme numa vizinhança de $\hat{\theta}_h$ (Aitken, 1991), o estimador de máxima verossimilhança de θ_h . Seja I_h a matriz de informação de M_h , $h = 1, 2$. Então, para n grande,

$$L_h(\theta_h) = L_h(\hat{\theta}_h) \exp \left[-\frac{1}{2}(\theta_h - \hat{\theta}_h)' I_h(\theta_h - \hat{\theta}_h) \right],$$

de onde segue que

$$L_h^A = 2^{-p_h/2} L_h(\hat{\theta}_h).$$

Desta última equação e de (4.65) podemos escrever

$$FBP = 2^{\nu/2} \frac{L_1(\hat{\theta}_1)}{L_2(\hat{\theta}_2)}, \quad (4.66)$$

onde $\nu = p_2 - p_1$. Percebe-se de (4.66) que o fator de Bayes a posteriori é em grandes amostras uma verossimilhança penalizada, pois o segundo fator no lado direito de (4.66) é a razão de verossimilhanças (RV). Consequentemente,

$$-2 \log(FBP) = -2 \log(RV) - \nu \log(2). \quad (4.67)$$

Segue portanto que (4.67) pertence a classe das estatísticas da razão de verossimilhança penalizada (Aitken, 1991), ou seja, $\Lambda(m) = -2\log(RV) - \nu m$, com $m = 0,693$. Se $m = 0$, temos $\Lambda(m)$ se reduz a estatística da razão de verossimilhança maximizada usual, se $m = 2$, temos o critério de informação proposto por Akaike (1973) e se $m = 1$, temos a estatística "deviance" introduzida por Nelder e Wedderburn (1972) no modelo linear generalizado.

4.8 Comparação de s modelos estruturais via fator de Bayes a posteriori

O objetivo desta seção é derivar o FBP para a comparação de $\beta_1, \dots, \beta_s$ no caso em que o coeficiente de atenuação k_{x_i} é conhecido e também para o caso em que a razão de variâncias ϕ_i é conhecida, $i = 1, \dots, s$. As estatísticas apresentadas aqui possuem a vantagem da simplicidade no sentido de serem de fácil cálculo. Isto é vantajoso, haja visto que as estatísticas apresentadas no Capítulo 3, geralmente envolvem procedimentos iterativos.

4.8.1 Coeficiente de atenuação k_{x_i} conhecidos

A função de verossimilhança, $L(\theta)$, é dada por

$$L(\theta) \propto \prod_{i=1}^s (\lambda_{0i} \lambda_{1i})^{-n_i/2} \exp \left\{ \frac{-\sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2}{2\lambda_{1i}(k_i + 1)} \right\} \exp \left\{ \frac{-1}{2\lambda_{0i}} \left[\frac{\beta_i^2}{k_i + 1} \sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2 - 2\beta_i \sum_{j=1}^{n_i} (X_{ij} - \mu_i)(Y_{ij} - \beta_i \mu_i) + (k_i + 1) \sum_{j=1}^{n_i} (Y_{ij} - \beta_i \mu_i)^2 \right] \right\}, \quad (4.68)$$

onde $\theta = (\beta_1, \dots, \beta_s, \lambda_{01}, \dots, \lambda_{0s}, \lambda_{11}, \dots, \lambda_{1s}) = (\beta, \lambda_0, \lambda_1)$.

Inicialmente consideramos μ_i conhecido, $i = 1, \dots, s$, e admitimos a priori MDIP (veja Apêndice A) para θ , ou seja,

$$p(\theta) \propto \prod_{i=1}^s (\lambda_{0i} \lambda_{1i})^{-1/2}. \quad (4.69)$$

Combinando (4.68) e (4.69), podemos calcular $\ell_h^{(2)}$, $h = 1, 2$ definida em (4.65). Com este objetivo, usando a identidade (A.5) (ver apêndice A) e integrando em λ_1 , obtemos

$$\begin{aligned} \ell_h^{(2)} &\propto \int \left\{ \prod_{i=1}^s \Gamma\left(\frac{n_i}{2}\right) 2^{\frac{n_i h}{2}} h^{-\frac{n_i}{2}} \left[\sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2 \right]^{-\frac{n_i h}{2}} \right\} \\ &\times \left\{ \prod_{i=1}^s \lambda_{0i}^{\frac{n_i}{2}+1} \exp \left(-\frac{h}{2\lambda_{0i}} \left[1 + \frac{(\beta_i - \hat{\beta}_i)^2}{(n_i - 1)\Sigma_i} \right] \right) \right\} d\lambda_0 d\beta, \end{aligned}$$

onde Σ_i está definida em (A.5) no Apêndice A e $\hat{\beta}_i$ é dado em (3.71), $i = 1, \dots, s$.

Integrando esta última expressão com relação à λ_{0i} , $i = 1, \dots, s$, temos que,

$$\begin{aligned} \ell_h^{(2)} &\propto 2^{\frac{nh}{2}} h^{\frac{nh}{2}} \left\{ \prod_{i=1}^s \Gamma\left(\frac{n_i}{2}\right) \right\} \left\{ \prod_{i=1}^s \left[\sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2 \right]^{-\frac{n_i h}{2}} \right\} \\ &\times \int \prod_{i=1}^s \Gamma\left(\frac{n_i h}{2}\right) 2^{\frac{n_i h}{2}} h^{-\frac{n_i h}{2}} \left[1 + \frac{(\beta_i - \hat{\beta}_i)^2}{(n_i - 1)\Sigma_i} \right]^{-\frac{n_i h}{2}} d\beta. \end{aligned}$$

Usando a aproximação $1 + x \simeq e^x$, justificada pela consistência de $\hat{\beta}_i$ e usando a aproximação de Laplace (Tierney e Kadane, 1991), a última integral fica

$$\begin{aligned} \ell_h^{(2)} &\propto 2^{nh} h^{-nh} \left[\prod_{i=1}^s \Gamma^2\left(\frac{n_i h}{2}\right) \right] \left\{ \prod_{i=1}^s \left[\sum_i (X_{ij} - \mu_i)^2 \right]^{-\frac{n_i h}{2}} \right\} \\ &\times \prod_{i=1}^s [(n_i - 1)\Sigma_i n_i h]^{-1/2}, \quad (4.70) \end{aligned}$$

$h = 1, 2$.

Sob $H_0 : \beta_1 = \dots = \beta_s = \beta$, onde β é desconhecido, temos que a função de verossimilhança deste sub-modelo, é obtida a partir de (4.68) com β no lugar de

β_i . Combinando (4.68) com a priori (4.69) e integrando nos elementos de $\theta = (\beta, \lambda_0, \lambda_1)$, obtemos $\ell_h^{(1)}$, $h = 1, 2$, ou seja,

$$\ell_h^{(1)} \propto \int L(\beta, \lambda_0, \lambda_1) p(\theta) d\theta.$$

Assim, integrando inicialmente em λ_0 e λ_1 obtemos com $h = 1, 2$, que

$$\begin{aligned} \ell_h^{(1)} &\propto 2^{nh} h^{-nh} \left[\prod_{i=1}^s \Gamma^2\left(\frac{n_i h}{2}\right) \right] \left\{ \prod_{i=1}^s \left[\sum_{j=1}^s (X_{ij} - \mu_i)^2 \right]^{-\frac{n_i h}{2}} \right\} \\ &\times \int \prod_{i=1}^s \exp \left\{ -\frac{n_i h}{(n_i - 1) \Sigma_i} (\beta - \hat{\beta}_i)^2 \right\} d \end{aligned}$$

Usando a aproximação de Laplace mais uma vez, podemos escrever,

$$\begin{aligned} \ell_h^{(1)} &\propto 2^{nh} h^{-nh} \left[\prod_{i=1}^s \Gamma^2\left(\frac{n_i h}{2}\right) \right] \left\{ \prod_{i=1}^s \left[\sum_{j=1}^s (X_{ij} - \mu_i)^2 \right]^{-\frac{n_i h}{2}} \right\} \\ &\times \left\{ \prod_{i=1}^s [(n_i - 1) \Sigma_i]^{1/2} (n_i h)^{-1/2} \right\} \exp \left\{ -\frac{h}{2} \gamma \right\}, \end{aligned} \quad (4.71)$$

$h = 1, 2$, onde

$$\gamma = \sum_{i=1}^s \frac{n_i}{(n_i - 1) \Sigma_i} (\hat{\beta} - \hat{\beta}_i)^2$$

e

$$\hat{\beta} = n^{-1} \sum_{i=1}^s n_i \hat{\beta}_i. \quad (4.72)$$

O fator de Bayes a posteriori pode agora ser calculado, substituindo-se (4.70) e (4.71) em (4.65), obtendo-se

$$\begin{aligned} FBP &= \frac{\ell_1^{(2)} \ell_2^{(1)}}{\ell_1^{(1)} \ell_2^{(2)}} \\ &= 2^{\frac{s-1}{2}} \exp \left\{ -\frac{1}{2} \gamma \right\}. \end{aligned} \quad (4.73)$$

Sob a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta_0$, onde β_0 é conhecido obtem-se, por um caminho análogo, que o fator de Bayes posteriori é dado por

$$FBP = 2^{\frac{s}{2}} \exp \left\{ -\frac{1}{2} \gamma \right\}, \quad (4.74)$$

onde

$$\gamma = \sum_{i=1}^s \frac{n_i}{(n_i - 1)\Sigma_i} (\beta_0 - \hat{\beta}_i)^2,$$

que tem forma semelhante à estatística de Wald derivada em (3.85), mas com outra ponderação em $(\beta_0 - \hat{\beta}_i)^2$ e $\hat{\beta}_i$ calculado a partir da Identidade 1A (Apêndice A). Notamos também que obtém-se as mesmas expressões para o FBP usando a priori de Jeffreys para θ , ou seja, $p(\theta) \propto \prod_{i=1}^s (\lambda_{0i} \lambda_{1i})^{-1}$.

Considerando agora que μ_i e α_i são desconhecidos, a função de verossimilhança, $L(\theta)$, é dada por

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^s (\lambda_{0i} \lambda_{1i})^{-n_i/2} \exp \left\{ -\frac{1}{2\lambda_{0i} \lambda_{1i}} \left[\frac{(\lambda_{0i} + \lambda_{1i} \beta_i^2)}{k_i + 1} \sum_{j=1}^{n_i} (X_{ij} \lambda_{2i})^2 - 2\beta_i \lambda_{1i} \right. \right. \\ &\quad \times \left. \left. \sum_{j=1}^{n_i} (X_{ij} - \lambda_{2i})(Y_{ij} - \beta_i \lambda_{3i}) + \lambda_{1i}(k_i + 1) \sum_{j=1}^{n_i} (Y_{ij} - \lambda_{3i})^2 \right] \right\}, \end{aligned} \quad (4.75)$$

podendo com o uso de (4.1) ser colocada na forma

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^s (\lambda_{0i} \lambda_{2i})^{-n_i/2} \exp \left\{ -\frac{n_i}{2\lambda_{0i} \lambda_{1i}} (a_{1i} S_{xx}^{(i)} - 2a_{2i} S_{xy}^{(i)} + a_{3i} S_{yy}^{(i)}) \right\} \\ &\quad \times \exp \left\{ -\frac{n_i}{2\lambda_{0i} \lambda_{1i}} a_{1i} \left[\lambda_{2i} - \left(\bar{X}_i - \frac{a_{2i}}{a_{1i}} (\bar{Y}_i - \lambda_{3i}) \right) \right]^2 \right\} \\ &\quad \times \exp \left\{ -\frac{n_i}{2\lambda_{0i} \lambda_{1i} a_{1i}} (a_{1i} a_{3i} - a_{2i}^2) (\bar{Y}_i - \lambda_{3i})^2 \right\}. \end{aligned} \quad (4.76)$$

Considerando a priori de Jeffreys,

$$p(\beta, \lambda) \propto \prod_{i=1}^s (\lambda_{0i} \lambda_{1i})^{-1}, \quad (4.77)$$

combinando (4.76) com (4.77) e integrando em λ_{2i} e λ_{3i} temos que

$$\begin{aligned} \ell_h^{(2)} &\propto \prod_{i=1}^s \int \frac{1}{(\lambda_{0i} \lambda_{1i})^{\frac{n_i h - 1}{2} + 1}} \exp \left\{ -\frac{n_i h S_{xx}^{(i)}}{2(k_i + 1)\lambda_{0i}} \right\} \\ &\quad \times \exp \left\{ -\frac{n_i h}{2\lambda_{1i}} \left[\frac{S_{xx}^{(i)}}{(k_i + 1)} \beta_i^2 - 2S_{xy}^{(i)} \beta_i + (k_i + 1) S_{yy}^{(i)} \right] \right\} d\lambda_{0i} d\lambda_{1i} d\beta_i, \end{aligned}$$

e efetuando as integrações obtemos

$$\ell_h^{(2)} \propto 2^{nh} h^{-nh+s} \left\{ \prod_{i=1}^s \Gamma^2 \left(\frac{n_i h - 1}{2} \right) n^{-n_i h} \left(\frac{S_{xx}^{(i)} C_i}{k_i + 1} \right)^{-n_i h/2} (n_i h - 1)^{-1/2} \right\}, \quad (4.78)$$

onde $C_i = \frac{k_i + 1}{S_{xx}^{(i)}} (S_{xx}^{(i)} S_{yy}^{(i)} - S_{xy}^{(i)2})$. Sob a hipótese $H_0 : \beta_1 = \dots = \beta_s = \beta$, onde β é desconhecido, temos que a função de verossimilhança deste sub-modelo é obtida a partir de (4.76) com β no lugar de β_i . Combinando esta verossimilhança com a priori (4.77) e integrando em λ_{0i} e λ_{1i} obtemos

$$\ell_h^{(1)} \propto \prod_{i=1}^s \int \Gamma^2 \left(\frac{n_i h - 1}{2} \right) \left[\frac{n_i h S_{xx}^{(i)}}{2(k_i + 1)} \right]^{-\frac{n_i h - 1}{2}} (n_i h C_i / 2)^{-\frac{n_i h - 1}{2}} \left[1 + \frac{S_{xx}^{(i)} (\beta - \hat{\beta}_i)^2}{(k_i + 1) C_i} \right]^{-\frac{n_i h - 1}{2}} d\beta.$$

Usando a aproximação $1 + x \approx e^x$, considerando $n_i h - 1 \approx n_i h$ e resolvendo a integração em β pelo método de Laplace, obtemos

$$\ell_h^{(1)} \propto h^{-nh+s-\frac{1}{2}} 2^{nh} \left\{ \prod_{i=1}^s \left(\frac{S_{xx}^{(i)} C_i}{k_i + 1} \right)^{-\frac{n_i h - 1}{2}} n_i^{-n_i h} \Gamma^2 \left(\frac{n_i h - 1}{2} \right) \right\} \exp \left\{ -\frac{\gamma}{2} \right\}, \quad (4.79)$$

onde

$$\gamma = \sum_{i=1}^s n_i \frac{S_{xx}^{(i)} (\hat{\beta} - \hat{\beta}_i)^2}{(k_i + 1) C_i},$$

$$\hat{\beta} = \frac{\sum_{i=1}^s n_i t_i^{-2} \hat{\beta}_i}{\sum_{i=1}^s n_i t_i^{-2}}$$

e

$$t_i = (k_i + 1) \left[\frac{S_{yy}^{(i)}}{S_{xx}^{(i)}} - \left(\frac{S_{xy}^{(i)}}{S_{xx}^{(i)}} \right)^2 \right]^{1/2}.$$

De (4.78), (4.79) e (4.65) obtemos o fator de Bayes a posteriori que é dado por

$$FBP = 2^{\frac{s-1}{2}} \exp \left\{ -\frac{\gamma}{2} \right\}. \quad (4.80)$$

4.8.2 Razão de variâncias ϕ_i , $i = 1, \dots, s$ conhecidas

Considere sem perda de generalidade que $\phi_i = 1$, $i = 1, \dots, s$. A função de verossimilhança é dada por

$$L(\theta) \propto \prod_{i=1}^s (\lambda_{0i} \lambda_{1i})^{-n_i/2} \exp \left\{ -\frac{1}{2D_i} \left[(\lambda_{2i} \beta_i^2 + \lambda_{3i}) \sum_{j=1}^{n_i} (X_{ij} - \lambda_{0i})^2 - 2\beta_i (\lambda_{2i} - \lambda_{3i}) \right. \right. \\ \left. \left. \times \sum_{j=1}^{n_i} (X_{ij} - \lambda_{0i})(Y_{ij} - \lambda_{1i}) + (\lambda_{2i} + \beta_i^2 \lambda_{3i}) \sum_{j=1}^{n_i} (Y_{ij} - \lambda_{1i})^2 \right] \right\}, \quad (4.81)$$

onde $D_i = \lambda_{2i}\lambda_{3i}(1 + \beta_i^2)$. Considere que $\lambda_{0i} = \lambda_{1i} = 0$ e a priori

$$p(\theta) \propto \prod_{i=1}^s (\lambda_{2i}\lambda_{3i})^{-1}. \quad (4.82)$$

Combinando (4.81) e (4.82) e integrando em λ_{2i} e λ_{3i} podemos obter $\ell_h^{(2)}$, $h = 1, 2$ que é dado por

$$\begin{aligned} \ell_h^{(2)} &\propto \int \prod_{i=1}^s \Gamma^2\left(\frac{n_i h}{2}\right) \left(\frac{h}{2}\right)^{-n_i h} \left[\left(\frac{(\beta_i^2 S_{xx}^{(i)} - 2\beta_i S_{xy}^{(i)} + S_{yy}^{(i)})}{1 + \beta_i^2} \right) \left(\frac{(S_{xx}^{(i)} + 2\beta_i S_{xy}^{(i)} + \beta_i^2 S_{yy}^{(i)})}{1 + \beta_i^2} \right) \right]^{-\frac{n_i h}{2}} \\ &\times d\beta_1 \dots d\beta_s. \end{aligned}$$

Usando a igualdade de Sprent, mostrada no Apêndice A e a aproximação $1 + x \approx e^x$ e aproximando a última integral pelo método de Laplace (ver Apêndice A), obtemos

$$\ell_h^{(2)} \propto 2^{nh} h^{-nh - \frac{s}{2}} \left\{ \prod_{i=1}^s \gamma^2\left(\frac{n_i h}{2}\right) [\gamma_i f_i(\hat{\beta}_i)]^{-n_i h/2} \right\}, \quad (4.83)$$

onde $f_i(\cdot)$ e $\hat{\beta}_i$ são respectivamente dados por (3.78) e (3.71). Sob o modelo M_1 , integrando inicialmente em λ_{0i} e λ_{1i} temos que $\ell_h^{(1)}$, $h = 1, 2$ é dado por

$$\ell_h^{(1)} \propto \int \prod_{i=1}^s \Gamma^2\left(\frac{n_i h}{2}\right) \left(\frac{h}{2}\right)^{-n_i h} \left[\left(\frac{(\beta^2 S_{xx}^{(i)} - 2\beta S_{xy}^{(i)} + S_{yy}^{(i)})}{1 + \beta^2} \right) \left(\frac{(S_{xx}^{(i)} + 2\beta S_{xy}^{(i)} + \beta^2 S_{yy}^{(i)})}{1 + \beta^2} \right) \right]^{-\frac{n_i h}{2}} d\beta.$$

Usando a igualdade de Sprent (ver Apêndice A) e a aproximação $1 + x \approx e^x$ e integrando com o método de Laplace obtemos

$$\ell_h^{(1)} \propto h^{-nh} 2^{nh} \left\{ \prod_{i=1}^s \Gamma^2\left(\frac{n_i h}{2}\right) \gamma_i^{-\frac{n_i h}{2}} \right\} h^{-1/2} \left\{ \prod_{i=1}^s [f_i(\hat{\beta})]^{-\frac{n_i h}{2}} \right\}. \quad (4.84)$$

De (4.83), (4.84) e (4.65), obtem-se o fator de Bayes a posteriori, que é dado por

$$FBP = 2^{\frac{s-1}{2}} \left[\prod_{i=1}^s \left\{ \frac{f_i(\hat{\beta}_i)}{f_i(\hat{\beta})} \right\}^{\frac{n_i}{2}} \right] \exp \left\{ -\frac{1}{2} \gamma \right\}, \quad (4.85)$$

onde $\hat{\beta}$ minimiza

$$\sum_{i=1}^s n_i \frac{(S_{xx}^{(i)} - \gamma_i)}{\gamma_i (1 + \beta^2)} (\beta - \hat{\beta}_i)^2, \quad (4.86)$$

e γ_i segue da Identidade 1A (Apêndice A). O Teorema seguinte fornece algumas propriedades da estatística γ .

Teorema 4.8.1 *Valem os seguintes resultados:*

i) O estimador $\hat{\beta}$ que minimiza (4.86) é dado por

$$\hat{\beta} = \frac{\sum_{i=1}^s n_i \frac{S_{xx}^{(i)} - \gamma_i}{\gamma_i} \hat{\beta}_i}{\sum_{i=1}^s n_i \frac{S_{xx}^{(i)} - \gamma_i}{\gamma_i} - \gamma}, \quad (4.87)$$

onde γ_i e $\hat{\beta}_i$ seguem da Identidade 1A (Apêndice A), $i = 1, \dots, s$, e γ é a menor solução da equação

$$\left| \begin{pmatrix} \sum_{i=1}^s n_i \frac{S_{xx}^{(i)} - \gamma_i}{\gamma_i} & \sum_{i=1}^s n_i \frac{S_{xx}^{(i)} - \gamma_i}{\gamma_i} \hat{\beta}_i \\ \sum_{i=1}^s n_i \frac{S_{xx}^{(i)} - \gamma_i}{\gamma_i} \hat{\beta}_i & \sum_{i=1}^s n_i \frac{S_{xx}^{(i)} - \gamma_i}{\gamma_i} \hat{\beta}_i^2 \end{pmatrix} - \gamma \mathbf{I}_2 \right| = 0$$

e $\mathbf{I}_2$ é a matriz identidade de ordem 2;

ii)

$$\gamma = \sum_{i=1}^s n_i \frac{(S_{xx}^{(i)} - \gamma_i)}{\gamma_i (1 + \hat{\beta}_i^2)} (\hat{\beta} - \hat{\beta}_i)^2; \quad (4.88)$$

iii) $\gamma = 0$ se e somente se $\hat{\beta}_1 = \dots = \hat{\beta}_s = \hat{\beta}$.

Prova. Pode-se mostrar que

$$\sum_{i=1}^s n_i \frac{S_{xx}^{(i)} - \gamma_i}{\gamma_i (\beta^2 + 1)} (\beta - \hat{\beta}_i)^2 = \frac{1}{1 + \beta^2} \sum_{i=1}^s (Z_i - \beta H_i)^2, \quad (4.89)$$

onde

$$\begin{aligned} Z_i &= \sqrt{\frac{n_i (S_{xx}^{(i)} - \gamma_i)}{\gamma_i}} \hat{\beta}_i \\ H_i &= \sqrt{\frac{n_i (S_{xx}^{(i)} - \gamma_i)}{\gamma_i}}. \end{aligned} \quad (4.90)$$

Portanto, usando argumentos análogos aos apresentados em Sprent (1966) (ver (1.9) e Apêndice A) temos que

$$\sum_{i=1}^s n_i \frac{S_{xx}^{(i)} - \gamma_i}{\gamma_i (\beta^2 + 1)} (\beta - \hat{\beta}_i)^2 = \gamma + \frac{(S_{HZ} - \gamma)}{1 + \beta^2} (\beta - \hat{\beta})^2, \quad (4.91)$$

onde

$$\begin{aligned}\hat{\beta} &= \frac{S_{HZ}}{S_{HH} - \gamma}, \\ S_{HZ} &= \sum_{i=1}^s H_i Z_i, \\ S_{HH} &= \sum_{i=1}^s H_i^2,\end{aligned}\tag{4.92}$$

de onde (i) segue. A parte ii) é uma consequência direta de (4.91). Para provar iii), suponha inicialmente que $\gamma = 0$. De (4.87), (4.88) e (4.90) temos então que

$$\left(\sum_{i=1}^s H_i^2 \hat{\beta}_i^2 \right) \left(\sum_{i=1}^s H_i^2 \right) = \left(\sum_{i=1}^s H_i^2 \hat{\beta}_i^2 \right)^2, \tag{4.93}$$

ou seja,

$$\begin{aligned}\sum_{i=1}^s H_i^4 \hat{\beta}_i^2 + 2 \sum_{\{(i,j): i < j\}} H_i^2 H_j^2 \hat{\beta}_i \hat{\beta}_j &= \sum_{i=1}^s H_i^4 \hat{\beta}_i^2 + \sum_{\{(i,j): i \neq j\}} H_i^2 H_j^2 \hat{\beta}_j^2 \\ &= \sum_{i=1}^s H_i^4 \hat{\beta}_i^2 + \sum_{\{(i,j): i < j\}} H_i^2 H_j^2 (\hat{\beta}_i^2 + \hat{\beta}_j^2),\end{aligned}$$

de modo que

$$2 \sum_{\{(i,j): i < j\}} H_i^2 H_j^2 \hat{\beta}_i \hat{\beta}_j = \sum_{\{(i,j): i < j\}} H_i^2 H_j^2 (\hat{\beta}_i^2 + \hat{\beta}_j^2).$$

Portanto, temos que

$$\sum_{\{(i,j): i < j\}} H_i^2 H_j^2 (\hat{\beta}_i - \hat{\beta}_j)^2 = 0,$$

de onde conclui-se, juntamente com (4.87), que $\hat{\beta}_1 = \dots = \hat{\beta}_s = \hat{\beta}$. Se $\hat{\beta}_1 = \dots = \hat{\beta}_s = \hat{\beta}$, segue de (4.88) que $\gamma = 0$. $\square$

Note que o FBP depende dos dados somente através de γ e é máximo quando $\gamma = 0$, o que é compatível com o resultado (iii) do Teorema 4.8.1. É também interessante notar a similaridade de γ com a estatística de Wald derivada para o caso estrutural. Contudo a estatística de Wald é baseada no estimador de máxima verossimilhança enquanto que o FBP é baseado em $\hat{\beta}$ que segue do Teorema 4.8.1.

Exemplo 4.5. Considere novamente os dados apresentados em Madansky (1959), já utilizados no Exemplo 4.1. Consideramos o modelo da Seção 3.6.2 com $s =$

2. Neste exemplo, utilizamos o teste da razão de verossimilhança, Wald e o Fator de Bayes a posteriori para verificar se é razoável a suposição de que os parâmetros de inclinação das duas retas associadas aos dois tratamentos são iguais. Transformando-se os dados originais de modo que $\phi_i = \sigma_{ui}^2 = \sigma_{ei}^2 = \sigma_i^2$, é desconhecido $i = 1, 2$, temos que

$$\mathbf{B}_1 = \begin{bmatrix} S_{xx}^{(1)} & S_{xy}^{(1)} \\ S_{xy}^{(1)} & S_{yy}^{(1)} \end{bmatrix} = \begin{bmatrix} 13,7536 & 4,8755 \\ 4,8755 & 35,3669 \end{bmatrix},$$

$$\mathbf{B}_2 = \begin{bmatrix} S_{xx}^{(2)} & S_{xy}^{(2)} \\ S_{xy}^{(2)} & S_{yy}^{(2)} \end{bmatrix} = \begin{bmatrix} 79,2224 & 29,2312 \\ 29,2312 & 58,9691 \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} S_{HH} & S_{HZ} \\ S_{HZ} & S_{ZZ} \end{bmatrix} = \begin{bmatrix} 28,9653 & 28,7442 \\ 28,7442 & 58,2268 \end{bmatrix},$$

de onde obtemos $\hat{\beta}_1 = 4,6482$, $\hat{\beta}_2 = 0,71187$, $\gamma_1 = 12,7047$, $\gamma_2 = 38,1601$, $\gamma = 11,3426$, $\hat{\beta}_C = 1,6311$ (o estimador que segue de (4.72)), e o de máxima verossimilhança $\hat{\beta}_M = 0.7015$ obtido pela solução numérica de uma equação similar a (3.91). Considerando o modelo estrutural temos de (4.85), após centrarmos as variáveis X_i e Y_i , que o $FBP = 0,0428$, a estatística de Wald, $W_a = 0,9162$ e a estatística da razão de verossimilhança, $W = 7,1197$. Estes valores mostram que a evidência em favor de H_1 é forte quando consideramos o FBP e a estatística da razão de verossimilhança. A estatística de Wald W_a apresenta evidência em favor da aceitação da hipótese nula, $H_0 : \beta_1 = \beta_2$.

Ilustração numérica 4.1

Neste caso foram gerados 1000 amostras de tamanhos $n_1 = n_2 = n$ de acordo com o modelo estrutural com k_{x_i} conhecido, onde $k_{x_1} = k_{x_2} = 2/3$, $\mu_1 = 30,0$, $\mu_2 = 40,0$, $\alpha_1 = 0,5$, $\alpha_2 = 3,0$, $\beta_1 = \beta_2 = 1,5$, $\sigma_{u1}^2 = \sigma_{u2}^2 = 1,0$. Para cada

Tabela 4.2: Aceitação de H_0 em 1000 amostras

α	n	N_1	$\hat{E}(FBP)$
0,10	05	870	0,80865
	10	902	0,92350
	20	980	0,98294
	30	983	0,98004
	50	988	0,99192
	100	991	1,01136
0,05	05	870	0,80865
	10	945	0,92350
	20	966	0,98294
	30	967	0,98004
	50	978	0,99192
	100	981	1.01136.

amostra gerada calculou-se o valor da estatística dada em (4.80) para o caso em que μ_i e α_i são desconhecidos. Apresenta-se na Tabela 4.2 o número de vezes N_1 nas 1000 amostras em que aceita-se a hipótese $H_0 : \beta_1 = \beta_2 = \beta$, β desconhecido, para os níveis $\alpha = 0,10$ e $0,05$. Apresenta-se também um estimador para $E(FBP)$, baseado nas 1000 simulações realizadas, ou seja, $\hat{E}(FBP) = \sum_{i=1}^{1000} FBP_i / 1000$.

□□□

Capítulo 5

Previsão do valor real

5.1 Introdução

Neste capítulo propomos um método alternativo para a estimação dos parâmetros do modelo estrutural através do algoritmo EM (Dempster et al., 1977). O enfoque proporciona também a possibilidade da previsão do valor real e desconhecido x_i . Fuller (1987) considera a estimação de x_i para alguns modelos especiais. Conforme veremos adiante a utilização do algoritmo EM possibilita a estimação (previsão) de x_i sob várias suposições e também sob modelos mais gerais do que aqueles considerados nos capítulos anteriores. Utilizando a transformação polar para o caso funcional considerada no Capítulo 2, estimadores de máxima verossimilhança são propostos para (x_i, y_i) . Neste caso, resultados de estudos de simulação mostram que o erro quadrático médio do estimador de máxima verossimilhança cresce com x_i . Estimadores da variância do estimador que levam em consideração tal fato são propostos. Tais estimativas parecem ser mais realistas que a proposta de Fuller (1987).

Consideramos também um enfoque Bayesiano para estimar x_i no modelo fun-

cional, para os casos em que as variâncias são conhecidas e desconhecidas. Tal enfoque justifica o estimador proposto por Fuller (1987) que utiliza argumentos um tanto heurísticos.

A Secção 5.2 é dedicada a utilização do algoritmo EM no modelo estrutural. A estimação de (x_i, y_i) no modelo funcional é considerada na Secção 5.3.

5.2 O algoritmo EM

Em situações onde a obtenção dos estimadores de máxima verossimilhança é complicada, como no caso do MEV, a utilização do algoritmo EM pode se apresentar uma alternativa a ser considerada. Isto é decorrência do fato de que a maximização da verossimilhança completa (que inclui os "valores" não observados) é em geral mais simples que a maximização da verossimilhança observada (sem os valores não observados). O algoritmo EM consiste basicamente de duas etapas. Na etapa E, a esperança condicional da verossimilhança completa é estimada em cada ciclo do algoritmo, onde os parâmetros desconhecidos são substituídos pelos seus estimadores obtidos na etapa anterior. No caso da família exponencial regular, a estimação da verossimilhança completa se reduz à estimação das estatísticas suficientes, o que em geral facilita a implementação do algoritmo. Na etapa M a verossimilhança completa é maximizada com as estatísticas suficientes substituídas pelas suas estimativas obtidas na etapa E. Em muitos casos, a implementação da etapa M pode envolver procedimentos iterativos adicionais, o que acontece quando a maximização da verossimilhança completa envolve equações que não apresentam soluções explícitas. Como veremos no caso de uma e s variáveis regressoras com uma estrutura de covariância especial, tal fato não apresenta dificuldades pois a

maximização da verossimilhança completa é bastante simples.

Na seção 5.2.3, consideramos a utilização do algoritmo EM no modelo com réplicas considerado na literatura por vários autores, entre eles Madansky (1959), Dolby (1976) e Chan e Mak (1979). Chan e Mak (1979) mostram que o estimador de máxima verossimilhança de β é dado pela solução de uma equação polinomial do quarto grau. Mostramos após uma série de manipulações algébricas que a implementação do algoritmo EM é bastante similar ao caso de uma variável auxiliar discutida na Seção 5.2.1. Também a implementação da etapa M do algoritmo EM não necessita de procedimentos iterativos adicionais. Um fato importante é que programas de computação escritos para o caso de uma variável auxiliar da seção 5.2.1 podem ser usados com pequenas modificações para o caso mais geral do modelo com réplicas. Na Seção 5.2.2 apresentamos um modelo com s variáveis regressoras.

5.2.1 O Modelo Estrutural com uma variável regressora

Com a suposição usual de que $Cov(u_i, x_i) = 0$, $i = 1, \dots, n$ temos que a distribuição conjunta de (Y_i, X_i, x_i) é dada por

$$\begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix} \sim \mathcal{N}_3 \left[\begin{pmatrix} \alpha + \beta\mu \\ \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \beta^2\tau^2 + \sigma_e^2 & \beta\tau^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 + \sigma_u^2 & \tau^2 \\ \beta\tau^2 & \tau^2 & \tau^2 \end{pmatrix} \right], \quad (5.1)$$

de onde segue que a densidade conjunta de (X_i, Y_i) é tal que

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \sim \mathcal{N}_2 \left[\begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \beta^2\tau^2 + \sigma_e^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 + \sigma_u^2 \end{pmatrix} \right], \quad (5.2)$$

$i = 1, \dots, n$. De (5.1), segue que a verossimilhança dos dados completos, ou seja, (Y_i, X_i, x_i) , $i = 1, \dots, n$, pertence à família exponencial regular. Tal fato facili-

ta em geral, a implementação do algoritmo EM. Conforme visto anteriormente, existem duas etapas neste algoritmo. A primeira etapa calcula a esperança da log-verossimilhança dado os dados observados (X_i, Y_i) , $i = 1, \dots, n$, e os parâmetros. Como a verossimilhança completa é da família exponencial regular, é suficiente calcular o valor esperado das estatísticas suficientes (Dempster et al., 1977) que neste caso dependem de x_i somente através de x_i e x_i^2 , $i = 1, \dots, n$. Seja $\theta = (\mu, \alpha, \beta, \tau^2, \sigma_e^2, \sigma_u^2)$. Pode-se mostrar, usando propriedades da distribuição normal multivariada, que a etapa E do algoritmo é definida pelas equações

$$\begin{aligned}\hat{x}_i &= E(x_i | Y_i, X_i, \theta) \\ &= \mu + \frac{\tau^2}{\Delta} [\beta \sigma_u^2 (Y_i - \alpha - \beta \mu) + \sigma_e^2 (X_i - \mu)]\end{aligned}\quad (5.3)$$

e

$$\begin{aligned}\widehat{x_i^2} &= E(x_i^2 | Y_i, X_i, \theta) \\ &= \tau^2 \left[1 - \frac{\tau^2 (\beta^2 \sigma_u^2 + \sigma_e^2)}{\Delta} \right] + \hat{x}_i^2,\end{aligned}\quad (5.4)$$

onde $\hat{x}_i^2$ é o quadrado de $\hat{x}_i$ dado em (5.3), $i = 1, \dots, n$, e

$$\Delta = (\beta^2 \sigma_u^2 + \sigma_e^2) \tau^2 + \sigma_e^2 \sigma_u^2.$$

Fuller (1987) propôs um preditor não iterativo que é equivalente à (5.3). Note que se $\sigma_u^2 = 0$, então $X_i = x_i$ ou seja, x_i é observado sem erro e neste caso,

$$\hat{x}_i = X_i, \quad (5.5)$$

$i = 1, \dots, n$. Dado (X_i, Y_i, θ) segue de (5.3) e (5.4) que a variância condicional de x_i é dada por

$$\begin{aligned}var(x_i | X_i, Y_i, \theta) &= \tau^2 \left[1 - \frac{\tau^2 (\beta^2 \sigma_u^2 + \sigma_e^2)}{\Delta} \right] \\ &= \frac{\sigma_e^2 \sigma_u^2 \tau^2}{(\beta^2 \sigma_u^2 + \sigma_e^2) \tau^2 + \sigma_e^2 \sigma_u^2}.\end{aligned}\quad (5.6)$$

Note que se $\sigma_e^2 = \sigma_u^2 = \sigma^2$, sem perda de generalidade (como vimos na seção 2.2), segue de (5.6) que

$$\text{var}(x_i | X_i, Y_i, \theta) = \frac{\sigma^2}{\beta^2 + 1 + \frac{\sigma^2}{\tau^2}},$$

$i = 1, \dots, n$, resultado que também aparece em Fuller (1987). Um estimador de (5.6) pode ser obtido substituindo os parâmetros desconhecidos pelos estimadores obtidos no último ciclo do algoritmo EM. Na etapa M, que é a etapa seguinte do algoritmo, maximiza-se a verossimilhança completa que segue de (5.1), em relação aos parâmetros desconhecidos, ou seja, maximiza-se a função de log-verossimilhança

$$\begin{aligned} \ell = & -(3n/2) \log(2\pi) - \frac{n}{2} (\log \tau^2 + \log \sigma_e^2 + \log \sigma_u^2) - \frac{1}{2} \left\{ \frac{1}{\sigma_u^2} \sum_{i=1}^n (X_i - \mu)^2 \right. \\ & - \frac{2}{\sigma_u^2} \sum_{i=1}^n (X_i - \mu)(x_i - \mu) - \frac{2\beta}{\sigma_e^2} \sum_{i=1}^n (Y_i - \alpha - \beta\mu)(x_i - \mu) \\ & \left. + \frac{1}{\sigma_e^2} \sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2 + \frac{(\tau^2 \sigma_u^2 \beta^2 + \tau^2 \sigma_e^2 + \sigma_u^2 \sigma_e^2)}{\sigma_e^2 \sigma_u^2 \tau^2} \sum_{i=1}^n (x_i - \mu)^2 \right\}. \quad (5.7) \end{aligned}$$

Esta maximização produz os valores dos parâmetros a serem utilizados na próxima iteração do algoritmo. Conforme verificado em Dempster et al. (1977), cada iteração do algoritmo aumenta a verossimilhança observada (incompleta). Supondo que σ_e^2 e σ_u^2 são conhecidos e diferenciando (5.7) em relação a $\mu, \alpha, \beta, \tau^2$ obtem-se as seguintes equações de verossimilhança:

$$\hat{\mu} = n^{-1} \sum_{i=1}^n x_i = \bar{x}, \quad (5.8)$$

$$\hat{\tau}^2 = n^{-1} \sum_{i=1}^n (x_i - \hat{\mu})^2, \quad (5.9)$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{x} \quad (5.10)$$

e

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (5.11)$$

Nota-se que no passo M, os estimadores dos parâmetros desconhecidos dados acima, dependem de x_i que não são observáveis, $i = 1, \dots, n$. Além disso os estimadores de α e β são os estimadores de mínimos quadrados ordinários mas dependendo de x_i , $i = 1, \dots, n$. Note também que se $\sigma_u^2 = 0$, ou seja, não havendo erros na variável x , então (5.5) é válida e neste caso, o algoritmo EM converge em uma iteração para os estimadores de mínimos quadrados usuais, como era de se esperar. Neste modelo os ciclos do algoritmo EM percorrem as equações (5.3)-(5.4) e (5.8)-(5.11), com x_i e x_i^2 de (5.8)-(5.11) substituídos em cada ciclo por $\hat{x}_i$ e $\widehat{x_i^2}$ de (5.3)-(5.4), respectivamente. Portanto, x_i é naturalmente estimado (predito) pelo algoritmo EM. Se $\phi = \sigma_e^2/\sigma_u^2$ for conhecido (podemos tomar $\phi = 1$, sem perda de generalidade, ou seja, $\sigma_e^2 = \sigma_u^2 = \sigma^2$) temos na etapa M ainda as equações (5.8)-(5.11) e a equação adicional

$$\begin{aligned} \hat{\sigma}^2 = \frac{1}{2n} \left\{ \sum_j (X_j - \hat{\mu})^2 - 2 \sum_j (X_j - \hat{\mu})(x_j - \hat{\mu}) - \frac{2\hat{\beta}}{\phi} \sum_j (Y_j - \hat{\alpha} - \hat{\beta}\hat{\mu})(x_j - \hat{\mu}) \right. \\ \left. + \phi^{-1} \sum_j (x_j - \hat{\mu})^2 \right\}. \end{aligned} \quad (5.12)$$

O algoritmo EM pode também ser implementado sem dificuldades para os casos:

i) razão de variâncias ϕ conhecida e ii) coeficiente de atenuação k_x conhecido onde $k_x = \tau^2/(\sigma_u^2 + \tau^2)$.

Ilustração numérica 5.1. Um estudo de simulação foi feito para ilustrar a performance do algoritmo EM considerando o caso σ_e^2 e σ_u^2 conhecido. O modelo simulado é o MEV estrutural definido no Capítulo 1 com $\alpha = \mu = 0,0$, $\beta = 1,5$, $\sigma_e^2 = \sigma_u^2 = 3,0$ e $\tau^2 = 0,1$. Os valores não observados de x_1 e x_2 que seguem da simulação são iguais a $-1,7993$ e $-3,1462$. O algoritmo EM foi inicializado com os estimadores de mínimos quadrados ordinários. A tabela 5.1 mostra 5 iterações do algoritmo que neste caso converge razoavelmente rápido. $\square$

Tabela 5.1: O Algoritmo EM. Um exemplo numérico

Iteração	$\hat{\beta}$	$\hat{x}_1$	$\hat{x}_2$
0	1,271853	-1,486004	-2,829326
1	1,424431	-1,488152	-2,834033
2	1,431906	-1,487365	-2,832573
3	1,433115	-1,487184	-2,832317
4	1,433347	-1,487150	-2,832276
5	1,433389	-1,487144	-2,832269

Exemplo 5.1. Considere novamente os dados do Exemplo 4.1, onde $\sigma_u^2 = 50$ e $\sigma_e^2 = 7500$, de modo que $\phi = \sigma_e^2/\sigma_u^2 = 150$. A tabela 5.2 apresenta os estimadores de β, τ^2, x_1, x_2 e o valor da função de log-verossimilhança observada (que pode ser obtida a partir de (5.2)), em cada iteração. As variâncias σ_u^2 e σ_e^2 são conhecidas com os valores acima citados. Iniciando o algoritmo EM com os estimadores de mínimos quadrados na iteração "0", observe na tabela 5.2 o resultado das demais iterações.

A tabela 5.3 é análoga à anterior, exceto que o valor que inicializa o algoritmo EM está bem afastado dos estimadores de mínimos quadrados. Nestas duas situações consideradas, o algoritmo converge razoavelmente rápido para o mesmo conjunto de valores. Outros valores iniciais para o algoritmo foram analisados e o máximo da função de log-verossimilhança observada não se modificou, ou seja, a função de log-verossimilhança parece ser unimodal e isto facilita a convergência do algoritmo. $\square$

Tabela 5.2: O Algoritmo EM. Dados de um ensaio de dureza Brinell

Iteração	$\hat{\beta}$	$\hat{\tau}^2$	$\hat{x}_1$	$\hat{x}_2$	log(Veros.)
0	3,288200	714,193600	232,165983	232,369778	-433,524731
1	3,440454	675,385336	232,410824	232,581889	-433,416841
3	3,473891	665,500338	232,460663	232,625789	-433,412540
5	3,475303	665,072702	232,473138	232,636810	-433,412349
7	3,475364	665,054118	232,473678	232,637288	-433,412349
10	3,475367	665,053354	232,473702	232,637309	-433,412349
15	3,475367	665,053347	232,473702	232,637330	-433,412349
20	3,475367	665,053347	232,473702	232,637309	-433,412349
30	3,475367	665,053347	232,473702	232,637309	-433,412349

Tabela 5.3: O Algoritmo EM. Dados de um ensaio de dureza Brinell

Iteração	$\hat{\beta}$	$\hat{\tau}^2$	$\hat{x}_1$	$\hat{x}_2$	log(Veros.)
0	2,356700	0,856700	263,315886	262,901057	-718,179167
1	40,542375	1,715780	263,044785	262,901057	-704,582433
3	57,178899	2,554373	262,826053	262,585543	-695,385384
5	39,932059	5,671063	262,113456	261,771555	-683,506195
7	22,801219	18,055800	260,176098	259,683616	-656,540395
10	5,686869	275,013124	245,089448	244,512728	-503,175865
15	3,476708	664,646539	232,485558	232,6477930	-433,415957
20	3,475367	665,053188	232,473707	232,637313	-433,412349
30	3,475367	665,053347	232,473702	232,637309	-433,412349

Exemplo 5.2. Nesta ilustração, os dados analisados aparecem em Fuller (1987), Exercício 17 do Cap. 1. Temos $n = 18$ pares (X_i, Y_i) onde (X) é a concentração de Fósforo na parte inorgânica do solo e (Y) o conteúdo de Fósforo observado no milho. Os valores observados foram : $(64;1,18)$, $(60;1,18)$, $(71;2,02)$, $(61;1,26)$, $(54;2,39)$, $(77;1,64)$, $(81;3,22)$, $(93;3,33)$, $(93;3,55)$, $(51;3,69)$, $(76;3,45)$, $(96;4,91)$, $(77;4,91)$, $(93;4,75)$, $(95;4,91)$, $(54;1,70)$, $(68;5,27)$ e $(99;5,56)$. Consideramos que σ_u^2 é conhecido e igual a 0.25. Uma vez que σ_e^2 é desconhecido, a etapa M do algoritmo EM foi implementada usando as equações (5.8)-(5.11) e a equação adicional

$$\hat{\sigma}_e^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}x_i)^2,$$

que corresponde ao estimador de máxima verossimilhança de σ_e^2 quando x_i é conhecido, $i = 1, \dots, n$. A tabela 5.4 mostra os estimadores de $\beta, \tau^2, \sigma_e^2, x_1$ e x_2 . Os valores que iniciam o algoritmo dados na iteração "0" correspondem aos estimadores de mínimos quadrados. O comportamento do algoritmo EM para outros valores iniciais que consideramos é bastante similar ao da tabela 5.4. Isto nos leva a concluir que é bastante razoável a suposição de que a função de log-verossimilhança para os dados é unimodal. Este exemplo ilustra também o fato de que o algoritmo EM pode convergir razoavelmente rápido em situações onde o tamanho da amostra não é muito grande. $\square$

No caso do modelo (Wong, 1991),

$$\begin{aligned} X_{ij} &= x_{ij} + u_{ij}, \\ Y_{ij} &= y_{ij} + e_{ij}, \\ y_{ij} &= \alpha_i + \beta_i x_{ij}, \end{aligned} \tag{5.13}$$

com as suposições usuais do MEV,

$$e_{ij} \sim \mathcal{N}(0, \sigma_e^2), u_{ij} \sim \mathcal{N}(0, \sigma_u^2), x_{ij} \sim \mathcal{N}(0, \tau^2),$$

Tabela 5.4: O Algoritmo EM. Dados do conteúdo de Fósforo

Iteração	$\hat{\beta}$	$\hat{\tau}^2$	$\hat{\sigma}_e^2$	$\hat{x}_1$	$\hat{x}_2$	log(Veros.)
0	6,736378	2,223966	148,501908	1,400639	1,362481	-62,755557
1	7,318497	2,047070	139,780905	1,428573	1,385584	-62,183554
3	7,558834	1,981983	136,180307	1,440719	1,395629	-61,981073
5	7,585954	1,974897	135,774001	1,442113	1,396782	-61,976171
7	7,589113	1,974075	135,726682	1,442275	1,396917	-61,973933
10	7,589514	1,973971	135,720669	1,442296	1,396934	-61,973649
15	7,589531	1,973966	135,720420	1,442297	1,396935	-61,973637
20	7,589531	1,973966	135,720241	1,442297	1,396935	-61,973637
30	7,589531	1,973966	135,720241	1,442297	1,396935	-61,973637

$i = 1, \dots, s$, $j = 1, \dots, n_i$, e supondo sem perda de generalidade que $\sigma_e^2 = \sigma_u^2 = \sigma^2$, com σ^2 desconhecido pode-se mostrar que as equações da etapa E são dadas por

$$\begin{aligned}\hat{x}_{ij} &= E(x_{ij}|Y_{ij}, X_{ij}, \theta) \\ &= \mu_i + \frac{\tau^2}{\Delta_2} [\beta_i(Y_{ij} - \alpha_i - \beta_i\mu_i) + (X_{ij} - \mu_i)]\end{aligned}\quad (5.14)$$

e

$$\begin{aligned}\widehat{x_i^2} &= E(x_{ij}^2|Y_{ij}, X_{ij}, \theta) \\ &= \frac{\tau^2\sigma^2}{\Delta_2} + (\hat{x}_{ij})^2,\end{aligned}\quad (5.15)$$

onde $\Delta_2 = (\beta_i^2 + 1)\tau^2 + \sigma^2$ e $(\hat{x}_{ij})^2$ é o quadrado de $\hat{x}_{ij}$ dado em (5.14). Dado X_{ij} , Y_{ij} , $i = 1, \dots, s$, $j = 1, \dots, n_i$ e $\theta = (\mu_1, \dots, \mu_s, \beta_1, \dots, \beta_s, \sigma^2)$, segue de (5.15) que a variância condicional de x_{ij} é

$$var(x_{ij}|X_{ij}, Y_{ij}, \theta) = \frac{\tau^2\sigma^2}{\Delta_2}, \quad (5.16)$$

$i = 1, \dots, s, j = 1, \dots, n_i.$

Quanto às equações da etapa M, não é difícil verificar que estas são dadas por

$$\begin{aligned}\hat{\beta}_i &= \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(Y_{ij} - \bar{Y}_i)}{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}, \\ \hat{\alpha}_i &= \bar{Y}_i - \hat{\beta}_i \bar{x}_i, \\ \hat{\mu}_i &= n_i^{-1} \sum_{j=1}^{n_i} x_{ij} = \bar{x}_i, \\ \hat{\tau}^2 &= n^{-1} \sum_{i=1}^s \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2,\end{aligned}\tag{5.17}$$

$i = 1, \dots, s$, e

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{2n} \sum_{i=1}^s \left\{ (\hat{\beta}_i^2 + 1) \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2 - 2 \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)(Y_{ij} - \hat{\alpha}_i) - 2 \hat{\beta}_i \sum_{j=1}^{n_i} (Y_{ij} - \hat{\alpha}_i \right. \\ &\quad \left. - \hat{\beta}_i \hat{\mu}_i)(x_{ij} - \hat{\mu}_i) + \sum_{j=1}^{n_i} (Y_{ij} - \hat{\alpha}_i - \hat{\beta}_i \hat{\mu}_i)^2 \right\}.\end{aligned}\tag{5.18}$$

5.2.2 Um modelo com s variáveis regressoras

Nesta seção consideramos o algoritmo EM para o caso de s variáveis regressoras para uma estrutura de covariância especial. Conforme veremos abaixo, neste caso a etapa M pode também ser implementada sem a necessidade de procedimentos iterativos adicionais. Contudo, para o caso de estruturas de covariâncias mais gerais acreditamos que procedimentos iterativos adicionais possam ser necessários para a implementação da etapa M. Uma versão multivariada funcional deste modelo é considerada em Gleser (1981). No enfoque considerado por Gleser e outros autores, os estimadores de máxima verossimilhança são obtidos a partir de autovalores e autovetores de certas matrizes que dependem dos dados. Consideramos

o MEV estrutural com s variáveis regressoras, dado pelas equações

$$\begin{aligned} X_{ij} &= x_{ij} + u_{ij}, \\ y_i &= \sum_{j=1}^s \beta_j x_{ij}, \\ Y_i &= y_i + e_i, \end{aligned} \quad (5.19)$$

$i = 1, \dots, s, j = 1, \dots, n$, com as suposições de que

$$\begin{pmatrix} x_{ij} \\ u_{ij} \\ e_{ij} \end{pmatrix} \sim \mathcal{N}_3 \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right],$$

$i = 1, \dots, s, j = 1, \dots, n$. Então o vetor aleatório $\mathbf{X}_i = (x_{1i}, \dots, x_{si}, X_{1i}, \dots, X_{si}, Y_i)'$, $i = 1, \dots, n$, tem distribuição normal $(2s + 1)$ -variada, com vetor de médias $\mathbf{0}$ e matriz de covariâncias dada por

$$\Sigma = \left(\begin{array}{cc|c} \mathbf{I}_s & \mathbf{I}_s & \boldsymbol{\beta} \\ \hline \mathbf{I}_s & 2\mathbf{I}_s & \boldsymbol{\beta} \\ \hline \boldsymbol{\beta}' & \boldsymbol{\beta}' & \boldsymbol{\beta}'\boldsymbol{\beta} + 1 \end{array} \right) \sigma^2 = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \sigma^2, \quad (5.20)$$

onde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_s)$ e $\mathbf{I}_s$ a matriz identidade de ordem s . Note que neste caso a distribuição dos dados completos pertence à família exponencial regular. No caso em que $s = 2$, (5.20) fica

$$\Sigma = \left(\begin{array}{cccc|c} 1 & 0 & 1 & 0 & \beta_1 \\ 0 & 1 & 0 & 1 & \beta_2 \\ 1 & 0 & 2 & 0 & \beta_1 \\ 0 & 1 & 0 & 2 & \beta_2 \\ \hline \beta_1 & \beta_2 & \beta_1 & \beta_2 & \beta_1^2 + \beta_2^2 + 1 \end{array} \right) \sigma^2.$$

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$, $\mathbf{X}_i^{(1)} = (x_{1i}, \dots, x_{si})$ e $\mathbf{X}_i^{(2)} = (X_{1i}, \dots, X_{si}, Y_i)$, $i = 1, \dots, n$. Usando algumas propriedades da distribuição normal multivariada, pode-se mostrar após

algumas manipulações algébricas, que a etapa E do algoritmo EM tem equações dadas por

$$\begin{aligned} E(\mathbf{X}_i^{(1)}|\mathbf{X}_i^{(2)}, \boldsymbol{\theta}) &= \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_i^{(2)} \\ &= \frac{1}{2}(\mathbf{I}_s - \frac{\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\beta} + 2})\mathbf{X}_i^* + \frac{Y_i\boldsymbol{\beta}}{\boldsymbol{\beta}'\boldsymbol{\beta} + 2}, \end{aligned} \quad (5.21)$$

onde $\mathbf{X}_i^* = (X_{1i}, \dots, X_{si})$, $\boldsymbol{\Sigma}_{12}$ e $\boldsymbol{\Sigma}_{22}$ definidos por (5.20),

$$\boldsymbol{\Sigma}_{22}^{-1} = \begin{pmatrix} \frac{1}{2}(\mathbf{I}_s + \frac{\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\beta} + 2}) & \frac{-\mathbf{I}_s\boldsymbol{\beta}}{\boldsymbol{\beta}'\boldsymbol{\beta} + 2} \\ \frac{-\mathbf{I}_s\boldsymbol{\beta}}{\boldsymbol{\beta}'\boldsymbol{\beta} + 2} & \frac{2}{\boldsymbol{\beta}'\boldsymbol{\beta} + 2} \end{pmatrix}$$

e

$$V(\mathbf{X}_i^{(1)}|\mathbf{X}_i^{(2)}, \boldsymbol{\theta}) = \sigma^2 \frac{1}{2}(\mathbf{I}_s - \frac{\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\beta} + 2}). \quad (5.22)$$

As equações (5.21)-(5.22) formam a etapa E do algoritmo EM. No caso $s = 2$, estas equações são dadas por

$$\begin{aligned} \hat{x}_{1i} &= E(x_{1i}|X_{1i}, X_{2i}, Y_i, \boldsymbol{\theta}) \\ &= \frac{1}{\Delta} [(\beta_2^2 + 2)X_{1i} - \beta_1\beta_2X_{2i} + 2\beta_1Y_i], \end{aligned}$$

$$\begin{aligned} \hat{x}_{2i} &= E(x_{2i}|X_{1i}, X_{2i}, Y_i, \boldsymbol{\theta}) \\ &= \frac{1}{\Delta} [-\beta_1\beta_2X_{1i} + (\beta_1^2 + 2)X_{2i} + 2\beta_2Y_i], \end{aligned}$$

$$V(x_{1i}|X_{1i}, X_{2i}, Y_i, \boldsymbol{\theta}) = \frac{\sigma^2}{\Delta}(\beta_2^2 + 2),$$

$$V(x_{2i}|X_{1i}, X_{2i}, Y_i, \boldsymbol{\theta}) = \frac{\sigma^2}{\Delta}(\beta_1^2 + 2)$$

e

$$Cov(x_{1i}, x_{2i}|X_{1i}, X_{2i}, Y_i, \boldsymbol{\theta}) = -\frac{\sigma^2}{\Delta}\beta_1\beta_2,$$

$i = 1, \dots, n$ com $\Delta = 2\beta_1^2 + 2\beta_2^2 + 4$.

Seja L a verossimilhança completa. Para uma amostra de tamanho n ,

$$L = \prod_{i=1}^n (2\pi)^{-(2s+1)} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)})' \Sigma^{-1} \begin{pmatrix} \mathbf{X}_i^{(1)} \\ \mathbf{X}_i^{(2)} \end{pmatrix} \right\}, \quad (5.23)$$

onde

$$|\Sigma| = (\sigma^2)^{2s+1}$$

e

$$\Sigma^{-1} = \sigma^{-2} \left(\begin{array}{cc|c} 2\mathbf{I}_s + \beta\beta' & -\mathbf{I}_s & -\mathbf{I}_s\beta \\ -\mathbf{I}_s & \mathbf{I}_s & \mathbf{0} \\ \hline -\beta'\mathbf{I}_s & \mathbf{0} & 1 \end{array} \right).$$

Note, como no caso univariado, que $|\Sigma|$ é independente de β . Conforme veremos, isto simplifica bastante a implementação da etapa M. A implementação desta etapa do algoritmo depende da obtenção dos estimadores de máxima verossimilhança de $\beta_1, \dots, \beta_s$ e σ^2 que seguem da verossimilhança completa (5.23). Estes estimadores são obtidas maximizando (5.23), ou seja, derivando (5.23) em relação à β e igualando estas derivadas a zero obtemos o sistema de equações que pode ser colocada na forma

$$\mathbf{x}'\mathbf{x}\beta = \mathbf{x}'\mathbf{Y}, \quad (5.24)$$

onde $\mathbf{x} = (x_1, \dots, x_s)'$, $\mathbf{Y} = (Y_1, \dots, Y_n)'$,

$$\mathbf{x}'\mathbf{x} = \begin{bmatrix} \sum_i x_{1i}^2 & \sum_i x_{1i}x_{2i} & \dots & \sum_i x_{1i}x_{si} \\ \sum_i x_{2i}x_{1i} & \sum_i x_{2i}^2 & \dots & \sum_i x_{2i}x_{si} \\ \vdots & \vdots & \dots & \vdots \\ \sum_i x_{si}x_{1i} & \sum_i x_{si}x_{2i} & \dots & \sum_i x_{si}^2 \end{bmatrix} \quad \mathbf{x}'\mathbf{Y} = \begin{bmatrix} \sum_i x_{1i}Y_i \\ \sum_i x_{2i}Y_i \\ \vdots \\ \sum_i x_{si}Y_i \end{bmatrix}.$$

De forma análoga, o estimador de σ^2 é dado por

$$\hat{\sigma}^2 = \frac{1}{n(2s+1)} \sum_{i=1}^s (\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)})' \mathbf{C}^{-1} \begin{pmatrix} \mathbf{X}_i^{(1)} \\ \mathbf{X}_i^{(2)} \end{pmatrix}, \quad (5.25)$$

onde $\mathbf{C} = \sigma^2 \boldsymbol{\Sigma}^{-1}$.

No caso particular em que $s = 2$, (5.24) se reduz ao sistema

$$\beta_1 \sum_i x_{1i}^2 + \beta_2 \sum_i x_{1i}x_{2i} = \sum_i x_{1i}Y_i$$

$$\beta_1 \sum_i x_{1i}x_{2i} + \beta_2 \sum_i x_{2i}^2 = \sum_i x_{2i}Y_i.$$

Então, resolvendo em $\boldsymbol{\beta}$ o sistema de equações (5.24) tem-se

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{Y}. \quad (5.26)$$

Note que $\hat{\boldsymbol{\beta}}$ em (5.26) é semelhante à solução das equações normais no modelo de regressão clássica, ou seja, quando $\mathbf{X}_i^{(1)}$ é observado sem erros. Quando $\mathbf{X}_i^{(1)}$ é observado sem erros, a solução de (5.26) não necessita de procedimentos iterativos adicionais, reduzindo-se aos mínimos quadrados usuais de $\boldsymbol{\beta}$. Os ciclos do algoritmo EM ocorrem entre as equações (5.21)-(5.22) e (5.25)-(5.26) até alcançar a convergência. Se não há erros em $\mathbf{X}_i^{(1)}$, a convergência ocorre em apenas um ciclo, como já é esperado. O algoritmo para o modelo desta seção pode ser facilmente estendido para os casos mais gerais onde α e μ são desconhecidos.

Podemos obter uma generalização do modelo acima, assumindo que $\text{var}(e_i) = \text{var}(x_{ij}) = \text{var}(u_{ij}) = \sigma_i^2 = \phi_i \sigma^2$, com ϕ_i conhecido, $i = 1, \dots, s$. Consideramos o modelo transformado

$$y'_i = \alpha' + \sum_{j=1}^s \beta_j x'_{ij} + e'_i,$$

$$X'_{ij} = x'_{ij} + u'_{ij},$$

onde $y'_i = y_i/\sqrt{\phi_i}$, $x'_{ij} = x_{ij}/\sqrt{\phi_i}$, $e'_i = e_i/\sqrt{\phi_i}$, $X'_{ij} = X_{ij}/\sqrt{\phi_i}$, $u'_{ij} = u_{ij}/\sqrt{\phi_i}$, $i = 1, \dots, s$, e $j = 1, \dots, n$, de modo que, as equações das etapas E e M são como dadas acima com y_i , X_{ij} , e_i , u_{ij} , α , x_{ij} substituídos respectivamente por y'_i , X'_{ij} , e'_i , u'_{ij} , α' e x'_{ij} .

5.2.3 O modelo com réplicas

Nesta seção, r réplicas de (X_{ij}, Y_{ij}) , $i = 1, \dots, n$ e $j = 1, \dots, r$, são feitas para os pares não observáveis (x_i, y_i) , $i = 1, \dots, n$. Este modelo é definido pelas equações

$$\begin{aligned} Y_{ij} &= \alpha + \beta x_i + e_{ij}, \\ X_{ij} &= x_i + u_{ij}, \end{aligned} \tag{5.27}$$

onde x_i , e_{ij} e u_{ij} são mutuamente independentes com $x_i \sim \mathcal{N}(\mu, \tau^2)$, $e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ e $u_{ij} \sim \mathcal{N}(0, \sigma_u^2)$, $j = 1, \dots, r$, $i = 1, \dots, n$. Este modelo é considerado por vários autores, dentre os quais destacamos Madansky (1959), Dolby (1976) e Chan e Mak (1979) onde é mostrado que o estimador de máxima verossimilhança de β é a raiz de uma equação polinomial do quarto grau e onde também se obtém os estimadores de máxima verossimilhança de μ , α , τ^2 , σ_e^2 e σ_u^2 . Neste caso não necessitamos de informações adicionais sobre os parâmetros para tornar o modelo identificável. Mostraremos que a etapa M do algoritmo EM não precisa de procedimentos iterativos adicionais e que os estimadores dos parâmetros α e β nesta etapa são simplesmente os estimadores de mínimos quadrados usuais quando x_i , $i = 1, \dots, n$, são conhecidos. A etapa E pode ser facilmente implementada uma vez que estamos lidando com a distribuição normal multivariada. No caso $r = 1$, temos o modelo usual com uma variável regressora discutido na Seção 5.2.1. Segue-se de (5.27) que a distribuição conjunta de $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ir})'$ e $\mathbf{X}_i = (X_{i1}, \dots, X_{ir})'$ (dados incompletos) é dada por

$$\begin{pmatrix} \mathbf{Y}_i \\ \mathbf{X}_i \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} \mathbf{1}_r(\alpha + \beta\mu) \\ \mathbf{1}_r\mu \end{pmatrix}; \mathbf{V} \right], \tag{5.28}$$

$i = 1, \dots, n$, onde $\mathbf{1}_r$ é um vetor de uns de dimensão r e a matriz $\mathbf{V}$ de dimensão $2r \times 2r$ é tal que

$$\mathbf{V} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \beta\tau^2\mathbf{J}_r \\ \beta\tau^2\mathbf{J}_r & \boldsymbol{\Sigma}_2 \end{pmatrix},$$

com $\mathbf{J}_r = \mathbf{1}_r\mathbf{1}_r'$ a matriz de uns de dimensão $r \times r$ e $\boldsymbol{\Sigma}_1$ e $\boldsymbol{\Sigma}_2$ dados, respectivamente, por

$$\boldsymbol{\Sigma}_1 = \sigma_e^2\mathbf{I}_r + \beta^2\tau^2\mathbf{J}_r \text{ e } \boldsymbol{\Sigma}_2 = \sigma_u^2\mathbf{I}_r + \tau^2\mathbf{J}_r,$$

onde $\mathbf{I}_r$ é a matriz identidade de dimensão r . Portanto, $\boldsymbol{\Sigma}_1$ é uma matriz de dimensão $r \times r$ onde os elementos na diagonal são iguais a $\sigma_e^2 + \beta^2\tau^2$ e os elementos fora da diagonal são iguais a $\beta^2\tau^2$. Por outro lado, $\boldsymbol{\Sigma}_2$ é uma matriz $r \times r$ onde os elementos da diagonal são iguais a $\sigma_u^2 + \tau^2$ e os elementos fora da diagonal são iguais a τ^2 . Após algumas manipulações matriciais obtemos

$$\mathbf{V}^{-1} = \frac{1}{\Delta} \begin{bmatrix} (r\tau^2 + \sigma_u^2)\mathbf{I}_r + \beta^2\tau^2\sigma_u^2\sigma_e^2(r\mathbf{I}_r - \mathbf{J}_r) & -\beta\tau^2\mathbf{J}_r \\ -\beta\tau^2\mathbf{J}_r & (r\beta^2\tau^2 + \sigma_e^2)\mathbf{I}_r + \sigma_e^2\tau^2\sigma_u^2(r\mathbf{I}_r - \mathbf{J}_r) \end{bmatrix},$$

onde

$$\Delta = r\tau^2\sigma_e^2 + r\beta^2\sigma_u^2\tau^2 + \sigma_e^2\sigma_u^2.$$

Com as suposições feitas nesta seção, segue-se que a distribuição conjunta de $(\mathbf{Y}_i, \mathbf{X}_i, x_i)$ (dados completos) é tal que

$$\begin{pmatrix} \mathbf{Y}_i \\ \mathbf{X}_i \\ x_i \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} (\alpha + \beta\mu)\mathbf{1}_r \\ \mu\mathbf{1}_r \\ \mu \end{pmatrix}, \boldsymbol{\Sigma} \right], \quad (5.29)$$

$i = 1, \dots, s$, onde

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \beta\tau^2\mathbf{J}_r & \beta\tau^2\mathbf{1}_r \\ \beta\tau^2\mathbf{J}_r & \boldsymbol{\Sigma}_2 & \tau^2\mathbf{1}_r \\ \beta\tau^2\mathbf{1}_r' & \tau^2\mathbf{1}_r' & \tau^2 \end{pmatrix}.$$

Então, para escrever a função de verossimilhança dos dados completos, precisamos inverter a matriz Σ . Após longas manipulações algébricas, segue que

$$|\Sigma| = \tau^2 \sigma_e^{2r} \sigma_u^{2r},$$

que novamente não depende de β , e

$$\Sigma^{-1} = \frac{1}{\sigma_e^2 \sigma_u^2 \tau^2} \begin{pmatrix} \tau^2 \sigma_u^2 \mathbf{I}_r & 0 & -\beta \sigma_u^2 \tau^2 \mathbf{1}_r \\ 0 & \tau^2 \sigma_e^2 \mathbf{I}_r & -\tau^2 \sigma_e^2 \mathbf{1}_r \\ -\beta \sigma_u^2 \tau^2 \mathbf{1}_r & -\tau^2 \sigma_e^2 \mathbf{1}_r & \Delta \end{pmatrix},$$

com $\Delta = r\tau^2 \sigma_e^2 + r\beta^2 \sigma_u^2 \tau^2 + \sigma_e^2 \sigma_u^2$. Portanto a função de verossimilhança dos dados completos é proporcional à

$$\left(\frac{1}{\tau^2 \sigma_e^{2r} \sigma_u^{2r}} \right)^{n/2} \exp\left(-\frac{Q}{2}\right), \quad (5.30)$$

onde

$$\begin{aligned} Q &= \sum_{i=1}^n (\mathbf{Y}_i - (\alpha + \beta\mu)\mathbf{1}_r)' (\mathbf{X}_i' - \mu\mathbf{1}_r', x_i - \mu) \Sigma^{-1} \begin{pmatrix} \mathbf{Y}_i - (\alpha + \beta\mu)\mathbf{1}_r \\ \mathbf{X}_i - \mu\mathbf{1}_r \\ x_i - \mu \end{pmatrix} \\ &= \frac{1}{\sigma_e^2} \left\{ \sum_{i=1}^n (\mathbf{Y}_i - (\alpha + \beta\mu)\mathbf{1}_r)' (\mathbf{Y}_i - (\alpha + \beta\mu)\mathbf{1}_r) - 2\beta(x_i - \mu)\mathbf{1}_r' (\mathbf{Y}_i - (\alpha + \beta\mu)\mathbf{1}_r) \right. \\ &\quad \left. + r\beta^2 \sum_{i=1}^n (x_i - \mu)^2 \right\} + \frac{1}{\sigma_u^2} \left\{ \sum_{i=1}^n (\mathbf{X}_i - \mu\mathbf{1}_r)' (\mathbf{X}_i - \mu\mathbf{1}_r) \right. \\ &\quad \left. - 2 \sum_{i=1}^n (x_i - \mu)\mathbf{1}_r' (\mathbf{X}_i - \mu\mathbf{1}_r) + r \sum_{i=1}^n (x_i - \mu)^2 \right\} + \frac{1}{\tau^2} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

De (5.29) vemos que a função de verossimilhança dos dados completos pertence à família exponencial regular e isto simplifica a aplicação do algoritmo EM. Por outro lado, a função de verossimilhança dos dados incompletos que segue de (5.28) é difícil de ser maximizada. A etapa E do algoritmo EM é implementada usando as

propriedades da distribuição normal multivariada. Após algum trabalho algébrico segue de (5.29) que a etapa E é definida pelas equações

$$\hat{x}_i = \mathbb{E}(x_i | \mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\theta}) = \mu \frac{\tau^2}{\Delta} \left(\beta \sigma_u^2 \sum_{j=1}^n (Y_{ij} - r\alpha - r\beta\mu) + \sigma_e^2 \sum_{j=1}^n (X_{ij} - r\mu) \right) \quad (5.31)$$

e

$$\widehat{x_i^2} = \mathbb{E}(x_i^2 | \mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\theta}) = \frac{\sigma_e^2 \sigma_u^2 \tau^2}{\Delta} + (\hat{x}_i)^2,$$

$i = 1, \dots, n$, onde $\Delta = r\tau^2\sigma_e^2 + r\beta^2\sigma_u^2\tau^2 + \sigma_e^2\sigma_u^2$ e $\boldsymbol{\theta} = (\beta, \mu, \alpha, \tau^2, \sigma_e^2, \sigma_u^2)$. Estimadores para $\text{var}(x_i | \mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\theta})$ também podem ser obtidos sem dificuldades a partir das equações acima. Diferenciando (5.30) em relação à $\boldsymbol{\theta}$, obtemos os seguintes estimadores de máxima verossimilhança que são utilizados na etapa M,

$$\begin{aligned} \hat{\mu} &= n^{-1} \sum_{i=1}^n x_i = \bar{x}, \\ \hat{\alpha} &= \bar{Y} - \hat{\beta}\bar{x}, \\ \hat{\beta} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - r\bar{Y})}{r \sum_{i=1}^n (x_i - \bar{x})^2}, \\ \hat{\tau}^2 &= n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2, \\ \hat{\sigma}_e^2 &= (nr)^{-1} \sum_{i=1}^n (\mathbf{Y}_i - (\alpha + \beta x_i)\mathbf{1}_r)'(\mathbf{Y}_i - (\alpha + \beta x_i)\mathbf{1}_r) \end{aligned} \quad (5.32)$$

e

$$\hat{\sigma}_u^2 = (nr)^{-1} \sum_{i=1}^n (\mathbf{X}_i - \mathbf{1}_r x_i)'(\mathbf{X}_i - \mathbf{1}_r x_i),$$

onde $Y_{i.} = \sum_{j=1}^r Y_{ij}$ e $\bar{Y} = (nr)^{-1} \sum_{i=1}^n Y_{i.}$. Os estimadores de α e β dados em (5.32) são justamente os estimadores de mínimos quadrados no modelo de regressão convencional, ou seja, quando x_{ij} são observados sem erros. Mais uma vez ocorre que a implementação da etapa M não necessita de procedimentos iterativos adicionais. Programas de computador para a implementação dos algoritmos desta Seção e da Seção 5.2.1 são bastante similares.

Considere agora o caso especial do modelo dado em (5.27), onde $\sigma_e^2 = \beta^2 \sigma_u^2$.

Tal relação é sugerida pela idéia de que a escala dos valores de y possa estar relacionada com a escala dos valores de x . Este caso particular possui aplicações em ensaios biológicos (Dolby, 1987). As equações das etapas E e M do algoritmo EM são as mesmas do modelo mais geral considerado acima, com $\beta^2 \sigma_u^2$ no lugar de σ_e^2 , sendo que na etapa M, os estimadores de máxima verossimilhança de β e σ_u^2 são dados por

$$\hat{\beta} = \sqrt{\frac{S_{YY}}{rS_{xY} + S_{xX}}}$$

e

$$\hat{\sigma}_u^2 = \frac{1}{2rn} \left\{ \beta^{-2} S_{YY} - 2\beta^{-1} S_{xY} + rS_{xX} \right\},$$

onde $S_{YY} = \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{1}_r)' (\mathbf{Y}_i - \mathbf{1}_r \bar{y})$, $S_{xX} = \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{1}_r x_i)' \mathbf{1}_r x_i$ e $S_{xY} = \sum_{i=1}^n (\mathbf{X}_i - \mathbf{1}_r x_i)' (\mathbf{X}_i - \mathbf{1}_r x_i)$.

Ilustração numérica 5.2 A performance do algoritmo EM na estimação dos parâmetros desconhecidos que surgem no modelo com réplicas é aqui ilustrada usando uma simulação Monte Carlo. Uma amostra de tamanho $n = 20$ com $r = 20$ réplicas foi gerada de acordo com o modelo dado em (5.27), com $\mu = 10, 0$, $\alpha = 40, 0$, $\beta = 1, 5$, $\tau^2 = \sigma_u^2 = 4, 0$ e $\sigma_e^2 = 0, 1$. Para esta amostra, $\sum_{i=1}^n \sum_{j=1}^r X_{ij} = 7889, 7$, $\sum_{i=1}^n \sum_{j=1}^r X_{ij}^2 = 79492, 8$, $\sum_{i=1}^n \sum_{j=1}^r Y_{ij} = 43811, 8$ e $\sum_{i=1}^n \sum_{j=1}^r Y_{ij}^2 = 2406304, 5$. O algoritmo foi inicializado considerando a iteração zero como $\hat{x}_i = X_i$ e $\hat{x}_i^2 = (3/2)\hat{x}_i$, $i = 1, \dots, n$ (veja na tabela os valores correspondentes a iteração 0). A Tabela 5.5 mostra os estimadores dos parâmetros a cada iteração do algoritmo. Como se vê, a convergência até cinco casas decimais ocorre até a iteração 20. Nota-se também o comportamento da função de log-verossimilhança observada, que começa em -2137,24 e cresce até 614,64. Além disso, os estimadores dos seis parâmetros estão

Tabela 5.5: O Algoritmo EM. O modelo com réplicas

Iteração	$\hat{\beta}$	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\tau}^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\log(\text{Veros.})$
0	0,76755	19,72429	39,62523	8,00354	3,99395	99,36604	-2137,24
1	0,79259	19,10665	39,62523	7,55010	3,96610	87,51785	-2084,47
3	0,81893	18,49654	39,61285	6,61234	3,96494	66,01810	-2029,30
5	0,87717	17,27975	39,60736	6,16624	3,96464	56,42233	-1906,79
10	1,07177	14,19074	39,55547	4,13300	3,96161	19,34158	-1474,41
15	1,41864	10,80100	39,46190	2,36470	3,95007	1,04976	-302,76
20	1,53756	9,86215	39,60100	1,99820	3,98520	0,10590	614,74
30	1,53756	9,86215	39,60100	1,99820	3,98520	0,10590	614,74

razoavelmente próximos dos valores reais correspondentes, segundo os quais os dados foram simulados. $\square$

5.3 Um enfoque Bayesiano

Nesta seção consideramos a estimação Bayesiana da variável não observável na situação onde x_i é considerada fixada ou seja, temos o MEV funcional apresentado no Capítulo 1. Neste modelo, a função de verossimilhança correspondente aos dados observados é dada por

$$L(\mathbf{x}, \alpha, \beta, \sigma_e^2, \sigma_u^2) \propto \frac{1}{\sigma_e^n} \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 \right\} \frac{1}{\sigma_u^n} \exp \left\{ -\frac{1}{2\sigma_u^2} \sum_{i=1}^n (X_i - x_i)^2 \right\}. \quad (5.33)$$

Suponha que α é conhecido (ou que $\alpha = 0$, sem perda de generalidade). Quando n é grande, conforme sugerido por Lindley e El-Sayad (1968), resultados aproximados

para o caso em que α desconhecido são obtidos centrando-se as variáveis X_i e Y_i , ou seja, considerando-se $X_i - \bar{X}$ e $Y_i - \bar{Y}$ em lugar de X_i e Y_i , $i = 1, \dots, n$. Consideramos inicialmente o caso em que σ_e^2 e σ_u^2 são considerados conhecidos e como priori para $(\mathbf{x}, \beta)$, tomamos

$$p(\mathbf{x}, \beta) \propto \text{constante},$$

que é a priori MDIP (ver Apêndice A). De (1.7) podemos escrever que

$$\frac{1}{\sigma_e^2} \sum_{i=1}^n (Y_i - \beta x_i)^2 + \frac{1}{\sigma_u^2} \sum_{i=1}^n (X_i - x_i)^2 = \sum_{i=1}^n \frac{(Y_i - \beta X_i)^2}{\sigma_u^2 \beta^2 + \sigma_e^2} + \left(\frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right) \sum_{i=1}^n (x_i - \hat{x}_{i,\beta})^2, \quad (5.34)$$

onde $\hat{x}_{i,\beta}$, $i = 1, \dots, n$, é dado por (1.8). O primeiro termo do lado direito de (5.34) pode ser escrito como em (1.9), onde $\hat{\beta}$ é dado por (1.10) e γ é o menor autovalor de (1.11), com

$$S_{xx} = \sum_{i=1}^n X_i^2, \quad S_{yy} = \sum_{i=1}^n Y_i^2 \quad \text{e} \quad S_{xy} = \sum_{i=1}^n X_i Y_i.$$

Portanto, a densidade a posteriori de $\mathbf{x}$ é dada por

$$\pi(\mathbf{x}|\mathbf{X}, \mathbf{Y}) \propto \int \exp \left[-\frac{1}{2} \frac{S_{xx} - \lambda \sigma_u^2}{\sigma_e^2 + \sigma_u^2 \beta^2} (\beta - \hat{\beta})^2 - \frac{1}{2} \left(\frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right) \sum_{i=1}^n (x_i - \hat{x}_{i,\beta})^2 \right] d\beta.$$

Usando o método de Laplace (veja Apêndice A), com

$$f(\mathbf{x}, \beta) = \exp \left[-\frac{1}{2} \left(\frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right) \sum_{i=1}^n (x_i - \hat{x}_{i,\beta})^2 \right],$$

e

$$-nh(\mathbf{x}, \beta) = \frac{S_{xx} - \lambda \sigma_u^2}{2(\sigma_e^2 + \sigma_u^2 \beta^2)} (\beta - \hat{\beta})^2,$$

tem-se, para n grande, que a posteriori marginal de $\mathbf{x}$ é dada por

$$\pi(\mathbf{x}|\mathbf{X}, \mathbf{Y}) \propto \exp \left[-\frac{1}{2} \left(\frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right) \sum_{i=1}^n (x_i - \hat{x}_{i,\beta})^2 \right], \quad (5.35)$$

onde

$$\hat{x}_{i,\beta} = \frac{\hat{\beta} \sigma_u^2 Y_i + \sigma_e^2 X_i}{\hat{\beta}^2 \sigma_u^2 + \sigma_e^2}, \quad (5.36)$$

com $\hat{\beta}$ dado em (1.10). De (5.35), temos que

$$x_i | \mathbf{X}, \mathbf{Y} \sim \mathcal{N} \left(\hat{x}_{i,\hat{\beta}}, \frac{1}{\frac{\hat{\beta}^2}{\sigma_e^2} + \frac{1}{\sigma_u^2}} \right), \quad (5.37)$$

$i = 1, \dots, n$. Fuller (1987) usando um argumento um tanto heurístico propõe o estimador (5.36) com variância dada em (5.37). O argumento acima mostra que $\hat{x}_{i,\hat{\beta}}$ dado em (5.36) é em grandes amostras, um estimador de Bayes de x_i e inferência sobre x_i deve ser baseada na distribuição normal dada em (5.37).

Exemplo 5.3. Considere os dados do Exemplo 1.3.2 de Fuller (1987). Segue de (5.37) que $\hat{x}_{5,\hat{\beta}} = 9,6974$ com $\widehat{var}(x_5 | \mathbf{X}, \mathbf{Y}) = 0,1830$. Um intervalo de 95% de credibilidade para x_5 baseado na posteriori (5.37) é então dado por (8,8590;10,5359). $\square$

No caso em que apenas a razão de variâncias $\phi = \sigma_e^2 / \sigma_u^2$ é conhecida, podemos considerar sem perda de generalidade (como foi visto no Capítulo 2) que $\sigma_e^2 = \sigma_u^2 = \sigma^2$ é desconhecido, ou de forma equivalente que a razão de variâncias $\phi = \sigma_e^2 / \sigma_u^2 = 1$. Como priori para $(\mathbf{x}, \beta)$ adotamos a priori não informativa dada em Lindley (1966) por

$$p(\mathbf{x}, \beta, \sigma) = p(\mathbf{x} | \beta, \sigma) p(\beta, \sigma) \propto \frac{1}{\sigma}. \quad (5.38)$$

Usando (1.8) e (1.9) com $\sigma_e^2 = \sigma_u^2 = \sigma^2$ segue que a posteriori marginal do vetor $\mathbf{x}$ é dada por

$$\begin{aligned} \pi(\mathbf{x} | \mathbf{X}, \mathbf{Y}) \propto & \int \int \frac{1}{\sigma^{2n+1}} \exp \left\{ -\frac{1}{2\sigma^2} \left[\gamma + \frac{(S_{xx} - \gamma\sigma^2)}{\beta^2 + 1} (\beta - \hat{\beta})^2 \right. \right. \\ & \left. \left. + (\beta^2 + 1) \sum_{i=1}^n (x_i - \hat{x}_{i,\beta})^2 \right] \right\} d\sigma d\beta. \end{aligned}$$

Usando a aproximação de Laplace para proceder a integração em β , temos, para n grande, que

$$\pi(\mathbf{x} | \mathbf{X}, \mathbf{Y}) \propto \int \sigma^{-(2n+1)} \exp \left[-\frac{1}{2\sigma^2} \left(\gamma + (\hat{\beta}^2 + 1) \sum_{i=1}^n (x_i - \hat{x}_{i,\hat{\beta}})^2 \right) \right] d\sigma.$$

Resolvendo-se a integral em σ segue que a densidade posteriori marginal de $\mathbf{x}$ é dada por

$$\pi(\mathbf{x}|\mathbf{X}, \mathbf{Y}) \propto \left[\gamma + (\hat{\beta}^2 + 1) \sum_{i=1}^n (x_i - \hat{x}_{i,\hat{\beta}})^2 \right]^{-n}. \quad (5.39)$$

Usando alguns resultados da distribuição t-multivariada (ver por exemplo, Zellner, 1971) temos, de (5.39), que a posteriori marginal de x_i , $i = 1, \dots, n$, tem distribuição t com $\nu = 2n - 1$ graus de liberdade e centrada em $\hat{x}_i$, ou seja,

$$\pi(x_i|\mathbf{X}, \mathbf{Y}) \propto \left[1 + \frac{\hat{\beta}^2 + 1}{\gamma} (x_i - \hat{x}_{i,\hat{\beta}})^2 \right]^{\frac{-(\nu+1)}{2}}, \quad (5.40)$$

$i = 1, \dots, n$. De (5.40), obtem-se um intervalo credibilidade para x_i ,

$$\left(\hat{x}_i \mp t_\alpha \sqrt{\frac{\gamma}{\nu(\hat{\beta}^2 + 1)}} \right), \quad (5.41)$$

onde t_α é o percentil $(1 - \alpha)/2$ da distribuição t com ν graus de liberdade.

Exemplo 5.4. Considerando novamente os dados do Exemplo 4.2, temos que $\hat{x}_{5,\hat{\beta}} = 11,5299$, $S_{xx} = 50,6973$, $S_{yy} = 25,8924$, $S_{xy} = 32,933$, $\phi = 1/6$, $\hat{\beta} = 0,7516$, $n = 15$. Para $\alpha = 0,05$, $t_\alpha = 2,045$, de modo que aplicando (5.41) obtem-se um intervalo aproximado de 95% de credibilidade para x_5 que é dado por (10,6789;12,3808). $\square$

5.4 Estimação de máxima verossimilhança.

Nesta seção consideramos a estimadores de máxima verossimilhança para x_i e y_i , juntamente com uma expressão para as variâncias das estimativas. Será usada a transformação polar considerada na seção 2.4. Muitos autores como por exemplo, Florens et al. (1974) e Madansky (1959) utilizam a transformação polar no modelo funcional. Com a parametrização (2.41), a função de verossimilhança

é proporcional à (2.42). Derivando (2.42) em relação à ω_i e θ encontramos as equações

$$\hat{\omega}_i = X_i \cos \hat{\theta} + Y_i \sin \hat{\theta}, \quad (5.42)$$

$i = 1, \dots, n$, e

$$\tan(2\hat{\theta}) = \frac{2S_{xy}}{S_{xx} - S_{yy}}. \quad (5.43)$$

Então os estimadores de máxima verossimilhança de x_i e y_i são dados por

$$\hat{x}_i = \hat{\omega}_i \cos \hat{\theta} \quad (5.44)$$

e

$$\hat{y}_i = \hat{\omega}_i \sin \hat{\theta}, \quad (5.45)$$

$i = 1, \dots, n$. Com a transformação polar (2.41), o estimador de máxima verossimilhança de β é $\hat{\beta} = \tan(\hat{\theta})$. Não é difícil ver que o estimador $\hat{\beta}$ é o mesmo que o apresentado em Fuller (1987) e Sprent (1966) (dado por (1.6)). O estudo de simulação apresentado a seguir mostra que o estimador de máxima verossimilhança de x_i apresenta um erro quadrático médio que cresce com x_i .

Ilustração numérica 5.3. Nesta ilustração, 1000 amostras de tamanhos $n = 10$ e $n = 50$ foram geradas satisfazendo ao MEV funcional dado por (1.3), com $\alpha = 0, 0$, $\beta = 1, 0$ e $\sigma_e^2 = \sigma_u^2 = 4, 0$. Consideramos o problema de estimar x_4 e x_8 , sabendo a priori (antes de gerar as 1000 amostras) que $x_4 = 67, 86$ e $x_8 = 16, 50$. Em cada amostra gerada calculamos o estimador de máxima verossimilhança de x_i dado por (5.44). A Tabela 5.6 mostra o erro quadrático médio estimado ($\widehat{EQM}$) de x_4 e x_8 calculado sobre as 1000 amostras simuladas, ou seja, $\widehat{EQM}(\hat{x}) = \sum (x - \hat{x})^2 / 1000$. Observa-se na Tabela 5.6 que valores grandes de x são estimados com erro quadrático médio maior. Notamos este mesmo comportamento em muitas outras simulações realizadas. Notamos também que o mesmo ocorre com y_i . Portanto o

Tabela 5.6: Erro Quadrático Médio de x_4 e x_8 .

n	$\widehat{EQM}(x_4)$	$\widehat{EQM}(x_8)$
10	2,388	2,024
50	2,192	2,017

estudo de simulação feito mostra que a variância dos estimadores $\hat{x}_i$ e $\hat{y}_i$ deve levar este fato em consideração. $\square$

Calculando a matriz de informação de Fischer associada à $(\mathbf{x}, \beta)$, que neste caso é diagonal, invertendo-a e usando o método delta, propomos estimar a variância para $\hat{x}_i$ por

$$\widehat{var}_f(\hat{x}_i) = \widehat{var}(\hat{x}_i) = \sigma^2 \hat{f}_{x_i}, \quad (5.46)$$

onde

$$\hat{f}_{x_i} = \cos^2 \hat{\theta} + \hat{\omega}_i^2 \frac{\sin^2 \hat{\theta}}{\sum_{i=1}^n \hat{\omega}_i^2},$$

$i = 1, \dots, n$. Como $\cos^2 \theta = (1 + \tan^2(\theta))^{-1}$, temos que quando n é grande, (5.46) e o estimador que segue de (5.37) (dad em Fuller (1987)) estão bem próximos. De forma análoga, propomos para a variância de $\hat{y}_i$, a quantidade

$$\widehat{var}_f(\hat{y}_i) = \widehat{var}(\hat{y}_i) = \sigma^2 \hat{f}_{y_i}, \quad (5.47)$$

onde

$$\hat{f}_{y_i} = \sin^2 \hat{\theta} + \hat{\omega}_i^2 \frac{\cos^2 \hat{\theta}}{\sum_{i=1}^n \hat{\omega}_i^2}.$$

Ilustração numérica 5.4. Com os parâmetros da Ilustração numérica 5.3, a Tabela 5.7 compara o EQM dado na tabela 5.6 com os valores da variância (5.46) e da variância dada em Fuller (1987). Seja $\hat{v}_f(\hat{x}_i) = \sigma^2 \hat{f}_{x_i}$ e $\hat{v}_F = \sigma^2 / (\hat{\beta}^2 + 1)$ (Fuller, 1987), dado em (5.37). $\square$

O estudo de simulação apresentado a seguir na Tabela 5.8 mostra a cobertura dos intervalos de confiança

$$\hat{x}_i \mp z_\alpha \sigma \hat{f}_{x_i}^{1/2} \quad (5.48)$$

e

$$\hat{y}_i \mp z_\alpha \sigma \hat{f}_{y_i}^{1/2}, \quad (5.49)$$

onde z_α denota o percentil $(1 - \alpha)/2$ da distribuição normal padrão. O enfoque acima pode ser empregado em modelos mais gerais, como por exemplo, quando o intercepto α é desconhecido.

Ilustração numérica 5.5. Um estudo Monte Carlo para estudar a probabilidade de cobertura dos intervalos dados por (5.48) e (5.49) é feito considerando-se 1000 amostras geradas nas condições da Ilustração numérica 5.4. Em cada amostra verificamos se os parâmetros x_5 e y_5 pertencem ou não aos intervalos (5.48) e (5.49). A Tabela 5.8 mostra o número de vezes que isto ocorre, ficando este bem próximo dos valores nominais $\alpha = 0,95$ e $0,90$. $\square$

Exemplo 5.5. Neste exemplo consideramos os dados do Exemplo 1.3.1 dado em (Fuller,1987). Neste caso, $\gamma = 1/6$, $\bar{X} = 10,046$, $\bar{Y} = 8,666$, $S_{xx} = 50,6973$, $S_{xy} = 32,933$ e $S_{yy} = 25,8924$. Neste exemplo o intercepto α é um parâmetro desconhecido, o que torna mais trabalhoso a obtenção de f_{x_i} e f_{y_i} . Após longas manipulações algébricas, segue-se que

$$\hat{x}_5 = 11,5299, \hat{y}_5 = 9,7822, \hat{f}_{x_5} = 5,4494 \text{ e } \hat{f}_{y_5} = 6,77328.$$

De (5.48) e (5.49) temos que intervalos aproximados com 95% de confiança para x_5 e y_5 são dados por (10,6101;12,4497) e (7,4947;12,0697). $\square$

Expressões para f_{x_i} e f_{y_i} quando α é desconhecido e $\phi \neq 1$ são dadas, respec-

tivamente, por

$$f_{x_j} = (b^{-1} + b_{jj})\cos^2\theta + 2d_{j3}\sin\theta\cos\theta + t^{-1}\omega_j^2 c_{33}\sin^2\theta \quad (5.50)$$

e

$$f_{y_j} = (b^{-1} + b_{jj})\sin^2\theta + 2d_{j1}\sin\theta - 2d_{j3}\sin\theta\cos\theta + t^{-1}[c_{11} + \omega_j(2c_{13}\cos\theta + c_{33}\omega_j\cos^2\theta)], \quad (5.51)$$

onde

$$b_{jj} = d_{j1}\frac{\sin\theta}{b\phi} - d_{j2}\frac{\alpha\sin\theta}{b\phi\sigma_u^2} + \frac{d_{j3}i_{\omega_j\theta}^*}{b},$$

$$d_{jk} = (bt)^{-1} \left[\phi^{-1}\sin\theta(c_{1k} - \alpha\sigma_u^2 c_{2k}) + i_{\omega_j\theta}^* c_{3k} \right],$$

com b e $i_{\omega_j\theta}^*$ dados em (2.43) e c_{ij} sendo os elementos da matriz dos cofatores de $\mathbf{E}$, onde $\mathbf{E}$ é obtida invertendo-se $\mathbf{I}(\gamma)$ dada em (2.43), ou seja,

$$\mathbf{I}^{-1}(\gamma) = \sigma_u^2 \left(\begin{array}{c|c} \mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' & -\mathbf{F}\mathbf{E}^{-1} \\ \hline -\mathbf{F}\mathbf{E}^{-1} & \mathbf{E}^{-1} \end{array} \right),$$

com $t = |\mathbf{E}|$. □

Podemos também construir regiões de confiança para (x_i, y_i) .

□□□

Tabela 5.7: Erro quadrático médio e variâncias de $\hat{x}_4$ e $\hat{x}_8$.

n	$\widehat{EQM}(\hat{x}_4)$	$\widehat{EQM}(\hat{x}_8)$	$\hat{v}_f(\hat{x}_4)$	$\hat{v}_f(\hat{x}_8)$	$\hat{v}_F$
10	2,388	2,024	2,349	2,031	2,0
50	2,192	2,017	2,131	2,012	2,0

Tabela 5.8: Número de vezes em 1000 que o I.C. contém o parâmetro

Tamanho da amostra	Nível de confiança	I.C. para x_5	I.C. para y_5
(n)	(%)	$\hat{x}_5 \pm \hat{f}_{x_5}^{1/2} z_{\frac{\alpha}{2}}$	$\hat{y}_5 \pm \hat{f}_{y_5}^{1/2} z_{\frac{\alpha}{2}}$
10	95	958	958
10	99	993	994
20	95	943	939
20	99	986	987
30	95	962	956
30	99	983	984
50	95	952	955
50	99	991	988

Apêndice A

Identities, prior MDIP and orthogonal parametrization

A.1 Identities

Aim of this section is that some chapters do not become long and that the demonstrations of some identities used in the text are facilitated, by putting here the most used ones.

A.1.1 Identity 1A

$$\begin{aligned} A_i &= \sum_j (Y_{ij} - \beta_i x_{ij})^2 + \sum_j (X_{ij} - x_{ij})^2 \\ &= U_i + (\beta_i^2 + 1) \sum_j (x_{ij} - \hat{x}_{ij})^2 \\ &= \gamma_i + \frac{(S_{xx}^{(i)} - \gamma_i)}{\beta_i^2 + 1} (\beta_i - \hat{\beta}_i)^2 + (\beta_i^2 + 1) \sum_j (x_{ij} - \hat{x}_{ij})^2 \end{aligned}$$

onde γ_i é o menor autovalor de $\mathbf{B}_i = \begin{pmatrix} S_{xx}^{(i)} & S_{xy}^{(i)} \\ S_{xy}^{(i)} & S_{yy}^{(i)} \end{pmatrix}$ com $S_{yy}^{(i)} = \sum_{j=1}^{n_i} Y_{ij}^2$, $S_{xy}^{(i)} = \sum_{j=1}^{n_i} X_{ij}Y_{ij}$, $S_{xx}^{(i)} = \sum_{j=1}^{n_i} X_{ij}^2$ e

$$U_i = \frac{S_{yy}^{(i)} - 2\beta_i S_{xy}^{(i)} + \beta_i^2 S_{xx}^{(i)}}{\beta_i^2 + 1}, \quad (\text{A.1})$$

$i = 1, \dots, s$.

Prova:

$$\begin{aligned} A_i &= \sum_j (Y_{ij} - \beta_i x_{ij})^2 + \sum_j (X_{ij} - x_{ij})^2 \\ &= \sum_j Y_{ij}^2 + \sum_j X_{ij}^2 + \sum_j [x_{ij}^2(\beta_i^2 + 1) - 2x_{ij}(\beta_i Y_{ij} + X_{ij})] \\ &= S_{yy}^{(i)} + S_{xx}^{(i)} + (\beta_i^2 + 1) \sum_j \left[x_{ij} - \left(\frac{\beta_i Y_{ij} + X_{ij}}{\beta_i^2 + 1} \right) \right]^2 - (\beta_i^2 + 1) \sum_j \left(\frac{\beta_i Y_{ij} + X_{ij}}{\beta_i^2 + 1} \right)^2 \\ &= S_{yy}^{(i)} + S_{xx}^{(i)} + (\beta_i^2 + 1) \sum_j (x_{ij} - \hat{x}_{ij})^2 - \frac{\beta_i^2 S_{yy}^{(i)} + 2\beta_i S_{xy}^{(i)} + S_{xx}^{(i)}}{\beta_i^2 + 1} \\ &= U_i + (\beta_i^2 + 1) \sum_j (x_{ij} - \hat{x}_{ij})^2. \end{aligned}$$

Para concluir a prova, basta mostrar que U_i dada em (A.1) é igual a

$$U_i = \gamma_i + \frac{(S_{xx}^{(i)} - \gamma_i)}{\beta_i^2 + 1} (\beta_i - \hat{\beta}_i)^2, \quad (\text{A.2})$$

onde γ_i são os menores autovalores de $\mathbf{B}_i$, $i = 1, \dots, s$. Para ver isto, note que U_i dada em (A.1) é um quociente de duas formas quadráticas, digamos, $U_i = \frac{Q_{1i}}{Q_{2i}}$, $i = 1, \dots, s$, de modo que podemos escrever

$$\begin{aligned} Q_{1i} - \gamma_i Q_{2i} &= (S_{xx}^{(i)} - \gamma_i) \left[\beta_i^2 - \frac{2\beta_i S_{xy}^{(i)}}{S_{xx}^{(i)} - \gamma_i} + \frac{S_{yy}^{(i)} - \gamma_i}{S_{xx}^{(i)} - \gamma_i} \right] \\ &= (S_{xx}^{(i)} - \gamma_i) (\beta_i - \hat{\beta}_i)^2, \end{aligned} \quad (\text{A.3})$$

com

$$\hat{\beta}_i = \frac{S_{xy}^{(i)}}{S_{xx}^{(i)} - \gamma_i}$$

e

$$\frac{S_{yy}^{(i)} - \gamma_i}{S_{xx}^{(i)} - \gamma_i} = \frac{S_{xy}^{(i)^2}}{(S_{xx}^{(i)} - \gamma_i)^2},$$

$i = 1, \dots, s$. Esta última igualdade implica em $|\mathbf{B}_i - \gamma_i \mathbf{I}_2| = 0$, onde $\mathbf{B}_i$ é dado em (??) e $\mathbf{I}_2$ é a matriz identidade de ordem 2, ou seja, γ_i é autovalor de $\mathbf{B}_i$. De (A.3) tem-se que

$$\frac{Q_{1i} - \gamma_i Q_{2i}}{Q_{2i}} = \frac{(S_{xx}^{(i)} - \gamma_i)(\beta_i - \hat{\beta}_i)^2}{Q_{2i}},$$

$i = 1, \dots, s$, obtendo-se (A.2). $\square$

Se em (A.2) tomarmos γ_i como o menor autovalor de $\mathbf{B}_i$ e $\beta_i = \hat{\beta}_i$, obtemos $\min_{\beta_i} U_i$. Esta identidade é usada na Sessão 4.8.1. Uma versão univariada é usada na seção 1.1.

A.1.2 Identidade 2A

$$\begin{aligned} n_i = & \frac{1}{2\hat{\lambda}_{0i}\hat{\lambda}_{1i}} \left[\frac{\hat{\lambda}_{0i} + \hat{\lambda}_{1i}\hat{\beta}_i^2}{k_i + 1} \sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2 - 2\hat{\beta}_i\hat{\lambda}_{1i}(X_{ij} - \hat{\beta}_i\mu_i) + \right. \\ & \left. + \hat{\lambda}_{1i}(k_i + 1) \sum_{j=1}^{n_i} (Y_{ij} - \hat{\beta}_i\mu_i)^2 \right], \end{aligned} \quad (\text{A.4})$$

onde $\hat{\beta}_i$, $\hat{\lambda}_{0i}$ e $\hat{\lambda}_{1i}$ são dados na Sessão 3.7.1.

Prova:

Os estimadores de máxima verossimilhança $\hat{\beta}_i$, $\hat{\lambda}_{1i}$ e $\hat{\lambda}_{0i}$, obtidos na Sessão 3.7.1, são dados por

$$\hat{\beta}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}(X_{ij} - \mu_i) - (k_i + 1)\mu_i \sum_{j=1}^{n_i} Y_{ij}}{[S_{xx}^{(i)}/(k_i + 1)] + 2\mu_i \sum_{j=1}^{n_i} (X_{ij} - \mu_i) + n_i\mu_i^2}, \quad (\text{A.5})$$

$$\hat{\lambda}_{1i} = \frac{S_{xx}^{(i)}}{k_i + 1} \quad (\text{A.6})$$

e

$$\hat{\lambda}_{0i} = (k_i + 1)^{-1} \hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (1 + k_i) S_{yy}^{(i)}, \quad (\text{A.7})$$

$i = 1, \dots, s$. Substituindo (A.5)-(A.7) no lado direito de (A.4), obtem-se

$$\frac{n_i}{2} \frac{(1+k_i)^{-1} \hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (1+k_i) S_{yy}^{(i)} + \frac{\hat{\beta}_i^2}{(1+k_i)} S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (1+k_i) S_{yy}^{(i)}}{(1+k_i)^{-1} \hat{\beta}_i^2 S_{xx}^{(i)} - 2\hat{\beta}_i S_{xy}^{(i)} + (1+k_i) S_{yy}^{(i)}} = n_i,$$

o que prova (A.4). $\square$

Esta identidade é usada para a obtenção da estatística da razão de verossimilhança em (3.116).

A.1.3 Identidade 3A

$$\frac{\beta_i^2}{k_i + 1} \sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2 - 2\beta_i \sum_{j=1}^{n_i} (X_{ij} - \mu_i)(Y_{ij} - \beta_i \mu_i) + (k_i + 1)(Y_{ij} - \beta_i \mu_i)^2 = \frac{(\beta_i - \hat{\beta}_i)^2}{(n_i - 1)\Sigma_i} + 1, \quad (\text{A.8})$$

$i = 1, \dots, s$.

Prova: Esta identidade é obtida completando-se os quadrados de seu lado esquerdo, de tal forma que se isole $(\beta_i - \hat{\beta}_i)^2$. A expressão final de Σ_i usada na Seção 4.9 é a seguinte:

$$\Sigma_i = \frac{\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{\left[\sum_{j=1}^{n_i} (X_{ij} - \mu_i) Y_{ij} + \mu_i (k_i + 1) \sum_{j=1}^{n_i} Y_{ij} \right]^2}{\sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2 + 2\mu_i (k_i + 1) \sum_{j=1}^{n_i} (X_{ij} - \mu_i) + n_i} [\mu_i (k_i + 1)]^2}{(n_i - 1)k_i^2 \left[\sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2 + 2\mu_i (k_i + 1) \sum_{j=1}^{n_i} (X_{ij} - \mu_i) + n_i [\mu_i (k_i + 1)]^2 \right]},$$

$i = 1, \dots, s$.

A.2 "Maximal data information priori"

Nesta seção consideramos a priori "Maximal data information priori" (MDIP) introduzida por Zellner (1977, 1990) que é usada em algumas seções do Capítulo 4. Esta priori maximiza o valor esperado do logaritmo da função de verossimilhança $\ell(\theta|y)$ calculada com relação à priori $p(\theta)$.

A entropia negativa de uma função de densidade $g(y, \theta)$, onde y corresponde aos dados e θ representa o parâmetro desconhecido, é dada por

$$-H = \int \int g(y, \theta) \log g(y, \theta) dy d\theta.$$

Sendo $g(y, \theta) = p(\theta)f(y|\theta)$, podemos escrever

$$-H = \int I(\theta)p(\theta)d\theta + \int p(\theta) \log p(\theta)d\theta, \quad (\text{A.9})$$

com

$$I(\theta) = \int f(y|\theta) \log f(y|\theta) dy = \mathbb{E}(\log f(y|\theta)). \quad (\text{A.10})$$

Em (A.9), $-H$ é uma medida da informação de $g(y, \theta)$ que é descrita em duas parcelas. A primeira, $\int I(\theta)p(\theta)d\theta$, é a informação média da densidade $f(y|\theta)$; e a segunda parcela, $\int p(\theta) \log p(\theta)d\theta$, é a informação contida na priori $p(\theta)$.

Para obter-se a MDIP normalizada, denotada por $p_*(\theta)$, maximiza-se

$$G(p(\theta)) = \int I(\theta)p(\theta)d\theta - \int p(\theta) \log p(\theta)d\theta, \quad (\text{A.11})$$

que é a diferença entre as parcelas de (A.9), ou seja, $G(p(\theta))$ representa a informação média contida nos dados $I(\theta)$ menos a informação contida na priori $p(\theta)$, onde $p(\theta)$ está sujeita à restrição $\int p(\theta)d\theta = 1$. Desta forma obtem-se uma solução única, $p_*(\theta)$.

Outra maneira de escrever (A.11) é

$$G(p(\theta)) = \int \left[r(\theta|y) \log \left\{ \frac{\ell(\theta|y)}{p(\theta)} \right\} d\theta \right] h(y) dy, \quad (\text{A.12})$$

onde $r(y|\theta) = h(y)r(\theta|y)$ é a posteriori de θ , $h(y)$ a densidade marginal de y e $\ell(\theta|y) = \log f(\theta|y)$ é a função de log-verossimilhança. Portanto, (A.12) mostra que $G(p(\theta))$ é uma média do logaritmo da razão de verossimilhança a priori, $\ell(\theta|y)/p(\theta)$.

Zellner, (1971, 1977, 1990) mostra que maximizando $G(p(\theta))$ em (A.11) em relação à $p(\theta)$ com a restrição $\int p(\theta)d\theta = 1$ obtém-se, como solução, a priori

$$p_*(\theta) = c \exp \{I(\theta)\}, \quad a < \theta < b, \quad (\text{A.13})$$

com $I(\theta)$ dada em (A.10) e $c^{-1} = \int \exp \{I(\theta)\} d\theta$ como a constante de normalização. Se $I(\theta) = \text{constante}$, então $p_*(\theta)$ é a densidade uniforme. E se g é a família de densidades da forma $g[(y - \theta)/\sigma]$ (locação-escala), então

$$p_*(\theta, \sigma) \propto \frac{1}{\sigma},$$

com $a_1 < \theta < b_1$ e $0 < a_2 < \sigma < b_2$.

A.2.1 Priori MDIP para o MEV estrutural e funcional

Considere inicialmente o MEV estrutural definido no Capítulo 1 com $\sigma_e^2 = \sigma_u^2 = \sigma^2$. Sendo σ^2 conhecido, temos, usando a parametrização ortogonal (2.7) em (2.1), que a distribuição de (X_i, Y_i) é dada por

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} \mu \\ \alpha + \beta\mu \end{pmatrix}, \begin{pmatrix} \frac{\lambda_2 + \beta^2\sigma^2}{\beta^2 + 1} & \frac{\beta(\lambda_2 - \sigma^2)}{\beta^2 + 1} \\ \frac{\beta(\lambda_2 - \sigma^2)}{\beta^2 + 1} & \frac{\beta^2\lambda_2 + \sigma^2}{\beta^2 + 1} \end{pmatrix} \right],$$

$i = 1, \dots, s$. De (2.4), obtemos

$$\mathbb{E} \left\{ n^{-1} \ell(\beta, \lambda | \mathbf{X}, \mathbf{Y}) \right\} = -\log 2\pi - (1/2) \log \sigma^2 - (1/2) \log \lambda_2 - 1.$$

Segue então de (A.13) que

$$p_*(\beta, \lambda) = c \lambda_2^{-1/2},$$

ou seja,

$$p_*(\beta, \lambda) \propto \lambda_2^{-1/2}.$$

De forma análoga, considerando σ^2 desconhecido, temos que

$$p_*(\beta, \lambda) \propto (\lambda_2 \lambda_3)^{-1/2}.$$

Se o coeficiente de atenuação, k_x , é conhecido, usando a parametrização ortogonal (2.18)-(2.19) para o caso em que μ e α são desconhecidos, obtem-se a log-verossimilhança (2.20). Segue então que

$$\mathbb{E} \left\{ n^{-1} \ell(\lambda, \beta | \mathbf{X}, \mathbf{Y}) \right\} = -(1/2) \log(\lambda_0 \lambda_1) - 2,$$

de modo que

$$p_*(\lambda, \beta) = c(\lambda_0 \lambda_1)^{-1/2},$$

ou seja,

$$p_*(\lambda, \beta) \propto (\lambda_0 \lambda_1)^{-1/2}. \quad (\text{A.14})$$

De forma similar, considerando k_x , μ e α conhecidos, $p_*(\lambda, \beta)$ é idêntico à (A.14).

No caso funcional, é fácil ver que se σ_e^2 e σ_u^2 são conhecidos, $p_*(\beta, \lambda) \propto \text{constante}$ e se $\sigma_e^2 = \sigma_u^2 = \sigma^2$ é desconhecido, $p_*(\beta, \lambda) \propto 1/\sigma^2$.

A.3 O método de Laplace para aproximação de Integrais

Seja h uma função derivável de um parâmetro m -dimensional θ , tal que $-h$ tenha máximo em $\hat{\theta}$. O método de Laplace é usado para aproximar integrais da forma

$$I = \int f(\theta) \exp[-nh(\theta)] d\theta,$$

após uma expansão de h e f em séries de Taylor em torno de $\hat{\theta}$ (Kass et al., 1990).

Considerando inicialmente o caso uni-paramétrico, o método de Laplace fornece para I a aproximação

$$\hat{I} \approx \sqrt{2\pi n^{-1}} \sigma f(\hat{\theta}) \exp[-nh(\hat{\theta})],$$

onde $\sigma = [h''(\hat{\theta})]^{-\frac{1}{2}}$. No caso m-paramétrico, ou seja, $\theta \in \mathbb{R}^m$, temos que

$$\hat{I} \approx (2\pi)^{m/2} \left\{ \det [nD^2h(\hat{\theta})] \right\}^{-\frac{1}{2}} f(\hat{\theta}) e^{-nh(\hat{\theta})},$$

onde $\hat{\theta}$ maximiza $-h(\theta)$ e $D^2h(\hat{\theta})$ é a matriz hessiana de h avaliada em $\hat{\theta}$.

A precisão desta aproximação é estudada em Kass, Tierney e Kadane (1990). Um caso especial da aproximação de Laplace (Tierney et al., 1989) aproxima integrais da forma $\int \exp[nh(\theta)] d\theta$. Veja em Achcar (1993), uma interessante utilização deste método.

A.4 Parâmetros ortogonais

Definição: Seja $\theta = (\psi, \xi)$ onde ψ é um escalar e ξ um vetor de dimensão p .

Seja $\ell(\theta)$ o logaritmo da função de verossimilhança dos dados e

$$\begin{aligned} \mathbf{J}(\theta) &= -\frac{\partial^2 \ell(\theta)}{\partial \theta^2} \\ &= \begin{pmatrix} J_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{pmatrix}, \end{aligned}$$

onde $J_{ik} = -\partial^2 \ell(\theta) / \partial \theta_i \partial \theta_k$. Os parâmetros ψ e ξ são ortogonais se:

$$\mathbf{I}_{12} = E_{\theta}(\mathbf{J}_{12}) = 0,$$

para todo θ .

A ortogonalidade é uma forma de controlar a influência do parâmetro ξ na estimação do parâmetro de interesse ψ (Cox e Reid, 1987). Quando ψ e ξ não são

ortogonais Cox e Reid sugerem utilizar a parametrização $\lambda = f(\psi, \xi)$, onde λ é ortogonal a ψ . Neste trabalho, esta parametrização é obtida através da solução de um sistema de equações diferenciais (Cox e Reid, 1987), apresentado a seguir.

A.4.1 Construção

Seja $\theta = (\psi, \xi)$ com ψ escalar sendo o parâmetro de interesse e $\xi = (\xi_1, \dots, \xi_q)$ o vetor de parâmetros de perturbação, com $q = p-1$. A função de log-verossimilhança é originalmente escrita em termos de $(\psi, \xi_1, \dots, \xi_q)$ e o objetivo é reparametrizar os parâmetros obtendo-se $(\psi, \lambda_1, \dots, \lambda_q)$, de modo que ψ seja ortogonal a $\lambda = (\lambda_1, \dots, \lambda_q)$ e o jacobiano da transformação $(\psi, \xi) \rightarrow (\psi, \lambda)$ seja não nulo. Considerando $\xi_1 = \xi_1(\psi, \lambda)$, ..., $\xi_q = \xi_q(\psi, \lambda)$, podemos escrever

$$\ell(\psi, \xi) = \ell^*(\psi, \xi_1(\psi, \lambda), \dots, \xi_q(\psi, \lambda)),$$

e usando a regra da cadeia, obtemos

$$\frac{\partial^2 \ell}{\partial \psi \partial \lambda_t} = \sum \frac{\partial^2 \ell^*}{\partial \psi \partial \xi_s} \frac{\partial \xi_s}{\partial \lambda_t} + \sum \frac{\partial^2 \ell^*}{\partial \xi_r \partial \xi_s} \frac{\partial \xi_s}{\partial \lambda_t} \frac{\partial \xi_r}{\partial \psi} + \sum \frac{\partial \ell^*}{\partial \xi_r} \frac{\partial^2 \xi_r}{\partial \psi \partial \lambda_t},$$

$t = 1, \dots, q$. Tomando o valor esperado na expressão acima, obtem-se

$$\sum \frac{\partial \xi_s}{\partial \lambda_t} \left(i_{\psi \xi_s}^* + \sum i_{\xi_r \xi_s}^* \frac{\partial \xi_r}{\partial \psi} \right) = 0,$$

$t = 1, \dots, q$, com $i_{\xi_r \xi_s}^* = \mathbb{E}(-\partial^2 \ell^*(\psi, \lambda) / \partial \xi_r \partial \xi_s)$. Portanto, $\lambda = (\lambda_1, \dots, \lambda_q)$ é obtido através da solução do sistema

$$\sum_{r=1}^q i_{\xi_r \xi_s}^* \frac{\partial \xi_r}{\partial \psi} = -i_{\psi \xi_s}^*, \quad (\text{A.15})$$

com $s = 1, \dots, q$. Dessa forma, o conjunto original (ψ, ξ) é transformado para (ψ, λ) com ψ ortogonal a λ_t , $t = 1, \dots, q$. Quando o sistema (A.15) admite solução, esta

é em geral arbitrária e é interessante escolher λ com uma interpretação conveniente. Denominamos a transformação $(\psi, \xi) \rightarrow (\psi, \lambda)$ que satisfaz (A.15) de transformação ortogonal. Os parâmetros desta transformação possuem algumas propriedades importantes:

- Os estimadores de máxima verossimilhança de β , $\hat{\beta}$ e de λ , $\hat{\lambda}$ são assintoticamente independentes.
- O desvio padrão assintótico do estimador de β não depende de λ ser ou não fixado.
- Sendo $\hat{\beta}_\lambda$ o estimador de máxima verossimilhança de β dado λ , então $\hat{\beta}_\lambda - \hat{\beta} = O_p(1/n)$. O mesmo vale para $\hat{\lambda}_\beta$, o estimador de máxima verossimilhança de λ dado β .
- Se $\hat{\beta}_\lambda = \hat{\beta}$ para todo λ fixado, então λ e β são ortogonais.

□□□

Apêndice B

Programas computacionais

Aqui está uma relação de programas em FORTRAN IV, utilizados neste trabalho, descritos sem as instruções do sistema operacional local.

B.0.2 Programa arazão

Este programa foi utilizado nas ilustrações numéricas 5.3, 5.4, 5.5 e nos Exemplos 5.1, 5.3, 5.4 e 5.5.

```
$SET AUTOBIND
00000100 $BIND=FROM BIB/IMSL/2,BIB/IMSL/3,BIB/IMSL/4
000002000 INTEGER NR,N,M,N1,N2,N3,N4
00000300 REAL X(60),Y(60),XMI(60),YMI(60),E(7500),W(60),R(100),L(60),G(60)
00000400 REAL SUM,XBAR,SUMD,SXX,SUM1,YBAR,SUM2,SY,Y,SUM3,SXY,BET,TET
00000500 REAL SUM4,SUM5,SUM8,SIGU,C,F1,XHAT,YHAT,H1,LFI,SUM10,F4,F5,F8,T
00000600 REAL LFS,LHI,LHS,LBI,LBS,LMI,LMS,BMQYX,BMQXY
00000700 REAL EQFX8,EQFX4,EQX8,EQX4,XHATF8,XHATF4,XHAT8,XHAT4
00000800 REAL EQMFX8,EQMFX4,EQMX8,EQMX4
00000900 DOUBLE PRECISION RXMI5,RYMI5,RLFI,RLFS,RLHI,RLHS,RLBI,RLBS,RB
00001000 DOUBLE PRECISION SEED,DSEED
00001100 READ (5,/) B,N,M,NR
00001200 SEED = 13*ATAN(1)*4
00001300 N1 = 0
00001400 N2 = 0
00001500 N3 = 0
00001600 N4 = 0
00001700 EQFX8 = 0.0
00001800 EQFX4 = 0.0
00001900 EQX8 = 0.0
00002000 EQX4 = 0.0
00002100 DSEED = 123457.DO
00002200 CALL GGUBT(DSEED,NR,R)
00002300 DO 1 I=2,N+1
00002400 XMI(I-1) = 100*R(I-1) + 10
00002500 YMI(I-1) = B * XMI(I-1)
00002600 1 CONTINUE
00002700 PRINT */ , XMI,YMI
00002800 T = ATAN(B)
```



```

00002810 SUM10 = 0.0
00002820 DO 6 I = 2,N+1
00002825 G(I-1) = XMI(I-1)*COS(T)
00002830 L(I-1) = G(I-1)*G(I-1)
00002840 SUM10 = SUM10 + L(I-1)
00002850 6 CONTINUE
00002860 F4 = COS(T)**2 + (G(4)*SIN(T))**2 /SUM10
00002870 F5 = COS(T)**2 + ( G(5)*SIN(T) )**2/SUM10
00002872 F8 = COS(T)**2 + ( G(8)*SIN(T) )**2/SUM10
00002880 SEED = SEED*1.8
00002900 2 DO 3 J = 1,M
00003000 SEED = 13*17*SEED
00003100 CALL GGNML(SEED,N+10,E)
00003200 DO 4 I=2,N+1
00003300 4 Y(I-1) = YMI(I-1) + 2*E(I+4)
00003400 SEED = (ATAN(1)*4)**0.5 +SEED
00003500 CALL GGNML(SEED,N+10,E)
00003600 DO 5 I = 2,N+1
00003700 5 X(I-1) = XMI(I-1) + 2*E(I+4)
00003800 PRINT */ , SUM10,F4,F5,F8
00003900 C CALCULO DAS COVARIANCIAS AMOSTRAIS
00004000 SUM = 0.0
00004100 SUMD = 0.0
00004200 SUM1 = 0.0
00004300 SUM2 = 0.0
00004400 SUM3 = 0.0
00004500 SUM4 = 0.0
00004600 SUM5 = 0.0
00004700 SUM8 = 0.0
00004800 DO 23 I = 1,N
00004900 SUM = SUM + X(I)
00005000 SUMD = SUMD + X(I)**2
00005100 SUM1 = SUM1 +Y(I)
00005200 SUM2 = SUM2 + Y(I)**2
00005300 SUM3 = SUM3 + X(I)*Y(I)
00005400 23 CONTINUE
00005500 SXX = SUMD / FLOAT(N-1)
00005600 SXY = SUM3 /FLOAT(N-1)
00005700 SYX = SUM2 / FLOAT(N-1)
00005800 XBAR = SUM / FLOAT(N)
00005900 YBAR = SUM1 / FLOAT(N)
00006000 BET = ( SYX-SXX + ( (SYY-SXX)**2 + 4*(SXY**2) )**0.5 ) / (2*SXY)
00006100 BMQYX = SXY/SXX
00006200 BMQXY = SXY/SYY
0006300 TET = ATAN(BET)
0006400 C PRINT */ ,BET,TET,BMQYX,BMQXY,SXX,SXY,SYX,SUM,SUM1
0006500 DO 38 I = 2,N+1
0006600 W(I-1) = X(I-1)*COS(TET) + Y(I-1)*SIN(TET)
00006700 SUM4 = SUM4 + W(I-1)
00006800 SUM5 = SUM5 + W(I-1)**2
00006900 SUM8=SUM8+(Y(I-1)-W(I-1)*SIN(TET))**2+(X(I-1)-W(I-1)*COS(TET))**2
00007000 38 CONTINUE
00007100 SIGU =(1/FLOAT(N-2))*SUM8
00007200 SIGU = 4.0
00007300 C = SUM5
00007400 F1 = ( COS(TET) )**2 + ( ( W(5)*SIN(TET) )**2)/C
00007500 C NOTE QUE W(I) EH O EST.MAX.VER. DE G(I) , I=1,...,N.
00007550 XHAT = W(5)*COS(TET)
00007600 XHAT8 = W(8)*COS(TET)
00007700 XHAT4 = W(4)*COS(TET)
00007800 YHAT = W(5)*SIN(TET)
00007900 H1
00008000 LFI = XHAT - 1.96*((SIGU*F1)**0.5)
00008100
00008200 XHATF4 = ( X(4) + BET*Y(4) ) / ( 1 + BET**2 )
00008300 EQFX8 = EQFX8 + ( XHATF8 - XMI(8) )**2

```

```

00008400 EQFX4 = EQFX4 + ( XHATF4 - XMI(4) )**2
00008500 EQX8 = EQX8 + ( XHAT8 - XMI(8) )**2
00008600 EQX4 = EQX4 + ( XHAT4 - XMI(4) )**2
00008700 LFS = XHAT + 1.96*((SIGU*F1)**0.5)
00008800 LHI = YHAT - 1.96*((SIGU*H1)**0.5)
00008900 LHS = YHAT + 1.96*((SIGU*H1)**0.5)
00009000 LBI = BET - 1.96 * ( (SIGU/C)**0.5 ) / ( COS(TET) )**2
00009100 LBS = BET + 1.96 * ( (SIGU/C)**0.5 ) / ( COS(TET) )**2
00009200 LMI = (1 - 1.96/FLOAT(N)) * SIGU
00009300 RXMI5 = XMI(5)
00009400 RYMI5 = YMI(5)
00009500 RLFI = LFI
00009600 RLFS = LFS
00009700 RLHI = LHI
00009800 RLHS = LHS
00009900 RLBI = LBI
00010000 RLBS = LBS
00010100 RB = B
00010200 LMI = (1 - 1.96/FLOAT(N)) * SIGU
00010300 LMS = (1 + 1.96/FLOAT(N)) * SIGU
00010400 C PRINT *, SIGU,SUM8,RLFI,RLFS,RLHI,RLBI,RLBS,F1,H1,LFI,RB
00010500 IF( ( RX@Q5
00010600 N1 = N1 + 1
00010700 55 IF ( (RYMI5 .GT. RLHS ) .OR. ( RYMI5 .LT. RLHI) ) GO TO 65
00010800 N2 = N2 + 1
00010900 65 IF ((RB .GT. RLBS ) .OR. (RB .LT. RLBI)) GO TO 70
00011000 C EM TODOS OS INTERVALOS ( PARA XMI(5),YMI(5) E B A INSTRUCAO
IF EH
00011100 C VERDADE SE O PARAMETRO CAI FORA DO INTERVALO DADO PELA AMOSTRA,
ISTO
00011200 C EH: OS CONTADORES N1,N2 E N3 MEDEM AS VEZES QUE O PAR ESTA
NO IC
00011300 N3 = N3 + 1
00011400 C A VAR TX SERVE APENAS PARA O PROG IR A INTRUCAO SEGUINTE
00011500 70 N4 = N4 + 1
00011600 C PRINT *, N1,N2,N3,J,M,BET,SIGU,F1,H1,C,NR
00011700 3 CONTINUE
00011800 EQMFX8 = EQFX8/1000
00011900 EQMFX4 = EQFX4/1000
00012000 EQMX8 = EQX8/1000
00012100 EQMX4 = EQX4/1000
00012200 WRITE(6,75) N1,N2,N3,N4,N,M,NR
00012300 75 FORMAT(4X,I5,4X,I5,4X,I5,4X,I5,4X,I5,4X,I5,4X,I5)
00012400 PRINT *, EQMFX8,EQMX8,EQMFX4,EQMX4
00012500 PRINT *, N,F4,F5,F8
00012550 STOP
00012600 END
00012700 C DADOS DE ENTRADA
00012800 C 1.0,50,1000

```

B.0.3 Programa fabe

Este program foi utilizado nas ilustrações numéricas 3.1 e 3.4.

```

$SET AUTOBIND LINEINFO
00000100 $BIND=FROM BIB/IMSL/1,BIB/IMSL/2,BIB/IMSL/3,BIB/IMSL/4 00000200
INTEGER N,N1,N2,N3,M
00000300 REAL PSI1(100),PSI2(100),X1(100),X2(100),Y1(100),Y2(100),E(1000)
00000400 REAL LD10,LD20,LD11,LD21,LT10,LT20,LT11,LT21,LR2,LR1,MI1,MI2
00000500 REAL LAMB1,LAMB2,LAMB0,LAMBZ1,M2,M1
00000600 DOUBLE PRECISION SEED
00000700 READ (5,/) N,TAU,B0,B1,B2,MI1,MI2,AL1,AL2,M,SIG1,SIG2,C,C1,C3
00000800 C C EH O VALOR QUIQUADRADO COM 1 G.L. E C1 NIVEL C FBP = FATOR
DE BAYES A PORTERIORI.

```



```

00001000 SEED = SEED**0.5
00001100 N1 = 0
00001200 N2 = 0
00001300 N3 = 0
00001400 N4 = 0
00001500 N5 = 0
00001600 N6 = 0
00001700 N7 = 0
00001800 N8 = 0
00001900 N9 = 0
00002000 N10 = 0
00002100 N11 = 0
00002200 N12 = 1
00002300 N13 = 1
00002400 C INICIO DAS M=1110 SIMULACOES.
00002500 DO 3 J = 1,M
00002600 S1 =0.0
00002700 S2 =0.0
00002800 S3 =0.0
00002900 S4 =0.0
00003000 S5 =0.0
00003100 S6 =0.0
00003200 S7 =0.0
00003300 S8 =0.0
00003400 S9 =0.0
00003500 S10=0.0
00003600 SEED = SEED + ATAN(1)
00003700 CALL GGNML(SEED,N+10,E)
00003800 DO 4 I=2,N+1
00003900 PSI1(I-1) = MI1+SQRT(TAU)*E(I-1)
00004000 4 CONTINUE
00004100 SEED = SEED + 4*ATAN(1)
00004200 CALL GGNML(SEED,N+10,E)
00004300 DO 5 I = 2,N+1
00004400 PSI2(I-1) = MI2+SQRT(TAU)*E(I-1)
00004500 5 CONTINUE
00004600 SEED = DSQRT(SEED)
00004700 CALL GGNML(SEED,N+10,E)
00004800 DO 6 I = 2,N+1
00004900 X1(I-1)=PSI1(I-1) + SQRT(SIG1)*E(I-1)
00005000 6 CONTINUE
00005100 SEED = SEED - 0.66666666
00005200 CALL GGNML(SEED,N+10,E)
00005300 DO 8 I = 2,N+1
00005400 X2(I-1) = PSI2(I-1) + SQRT(SIG2)*E(I-1)
00005500 8 CONTINUE
00005600 SEED = SEED + 0.5555555
00005700 CALL GGNML(SEED,N+10,E)
00005800 DO 10 I = 2,N+1
00005900 Y1(I-1) = AL1+B1*PSI1(I-1)+ SQRT(SIG1)*E(I-1)
00006000 10 CONTINUE
00006100 SEED = SEED + ATAN(1)
00006200 CALL GGNML(SEED,N+10,E)
00006300 DO 12 I = 2,N+1
00006400 Y2(I-1) = AL2+B2*PSI2(I-1)+ SQRT(SIG2)*E(I-1)
00006500 12 CONTINUE
00006600 C PRINT */, X1,Y1,X2,Y2
00006700 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 1 E 2
00006800 DO 20 I = 1,N
00006900 S1= S1 + X1(I)
00007000 S2= S2 + X1(I)**2
00007100 S3= S3 + Y1(I)
00007200 S4= S4 + Y1(I)**2
00007300 S5= S5 + X1(I)*Y1(I)
00007400 S6= S6 + X2(I)
00007500 S7= S7 + X2(I)**2
00007600 S8= S8 + Y2(I)
00007700 S9= S9 + Y2(I)**2

```

```

00007800 S10=S10 + X2(I)*Y2(I)
00007900 20 CONTINUE
00008000 SXX1 = ( S2 - S1**2/FLOAT(N) ) / FLOAT(N)
00008100 SYY1 = ( S4 - S3**2/FLOAT(N) ) / FLOAT(N)
00008200 SXY1 = ( S5 - S1*S3/FLOAT(N) ) / FLOAT(N)
00008300 XB1 = S1/FLOAT(N)
00008400 YB1 = S3/FLOAT(N)
00008500 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 2
00008600 SXX2 = ( S7 - S6**2/FLOAT(N) ) / FLOAT(N)
00008700 SYY2 = ( S9 - S8**2/FLOAT(N) ) / FLOAT(N)
00008800 SXY2 = ( S10 - S6*S8/FLOAT(N) ) / FLOAT(N)
00008900 XB2 = S6/FLOAT(N)
00009000 YB2 = S8/FLOAT(N)
00009300 C TERMINOU CALCULO DE COVARIANCIAS DAS POP. 1 E 2
00009400 BET1 = (SYY1-SXX1 + ( (SYY1-SXX1)**2+4*(SXY1**2))**.5)/(2*SXY1)
00009500 BET2 = (SYY2-SXX2 + ( (SYY2-SXX2)**2+4*(SXY2**2))**.5)/(2*SXY2)
00009600 C TESTE DE HIPOTESE H0: BETA = B0, B0 CONHECIDO IGUAL A 1 (B1
OU B2)
00009700 LD10 = SXX1+2*B0*SXY1+ (B0**2)*SYY1
00009800 LD20 = SXX2+2*B0*SXY2+ (B0**2)*SYY2
00009900 LT10 = (B0**2)*SXX1 - 2*B0*SXY1 + SYY1
00010000 LT20 = (B0**2)*SXX2 - 2*B0*SXY2 + SYY2
00010100 LT11 = (BET1**2)*SXX1-2*BET1*SXY1+SYY1
00010200 LT21 = (BET2**2)*SXX2-2*BET2*SXY2+SYY2
00010300 LD11 = SXX1 + 2*BET1*SXY1 + (BET1**2)*SYY1
00010400 LD21 = SXX2 + 2*BET2*SXY2 + (BET2**2)*SYY2
00010500 D0 = B0**2 + 1
00010600 D1 = BET1**2 + 1
00010700 D2 = BET2**2 + 1
00010800 R10 = (LT10*LD10)*(LT20*LD20)/(D0**4)
00010900 R11 = (LT11*LD11)*(LT21*LD21)/( (D1**2)*(D2**2) )
00011000 RRR = R10/R11
00011100 LR2 = N * ALOG(RRR)
00011200 R0 = LT10*LD10 / (D0**2)
00011300 R1 = LT11*LD11 / (D1**2)
00011400 P0 = LT20*LD20 / (D0**2)
00011600 RR2 = P0/P1
00011800 RR2 = N*ALOG(RR2)
00011900 F1 = SXX1 +SYY1
00012000 F2 = SXX2 +SYY2
00012100 DETB1 = SXX1*SYY1 - SXY1**2
00012200 DETB2 = SXX2*SYY2 - SXY2**2
00012300 CCC1 = F1**2 - 4*DETB1
00012400 CCC2 = F2**2 - 4*DETB2
00012500 LAMB1 = ( F1 - SQRT(CCC1) )/2
00012600 LAMB2 = ( F2 - SQRT(CCC2) )/2
00012700 COX1 = N* (SXX1 - LAMB1 )/LAMB1
00012800 COX2 = N* (SXX2 - LAMB2 )/LAMB2
00012900 LAMB0 = (1/D0) * ( COX1*((B0-BET1)**2) + COX2*((B0-BET2)**2)
)
00013000 Q10 = D0/LD10
00013100 Q20 = D0/LD20
00013200 Q12 = D1/LD11
00013300 Q22 = D2/LD21
00013400 M2 = LAMB0 - 2*LOG2 - N* ( ALOG(Q10/Q12) + ALOG(Q20/Q22) )
00013500 LAMBZ1 =(1/D0)* COX1* ( (B0-BET1)**2)
00013600 M1 = LAMBZ1 - LOG2 - N* ALOG(Q10/Q12)
00013700 LR1 = N * ALOG( RR )
00013800 WC1 = (1-(3./FLOAT(N)))*LR1
00013900 WC2 = (1-(3./FLOAT(N)))*RR*WB1
00014000 WB1 = LR1/(1+2.5/FLOAT(N))
00014100 WB2 = LR2/(1+2.5/FLOAT(N))
00014020 BB1 = N*((LD11-LT11)**2)/( (LD11*LT11)*(D1**2) )
01014100 BB2 = N*((LD21-LT21)**2)/( (LD21*LT21)*(D2**2) )

```



```

00014200 W2 =( (BET1-B0)**2 ) *BB1 + { (BET2-B0)**2 } *BB2
00014300 WW2 = { (BET1-BET2)**2 } / { (1/BB1) + (1/BB2) }
00014400 W1 =( (BET1-B0)**2 ) *BB1
00014500 C A VARIABEL A2 EH -2*LOG( FATOR DE BAYES A POSTERIORI )
00014600 A2= LR2 - ALOG(2)
00014700 A1= LR1- ALOG(2)
00014800 C2= -2*ALOG(C1)
00014900 C PRINT *, LR2,LR1,A2,A1,W2,W1
00015000 IF( LR2 .GT. C3 ) GO TO 50
00015100 N1 = N1+1
00015200 50 IF ( W2 .GT. C3 ) GO TO 60
00015300 N2 = N2 + 1
00015400 60 IF ( WB2 .GT. C3 ) GO TO 70
00015500 N3 = N3 + 1
00015600 70 IF ( WB1 .GT. C ) GO TO 80
00015700 N4 = N4 + 1
00015800 80 IF ( LR1 .GT. C ) GO TO 85
00015900 N5 = N5 + 1
00016000 85 IF ( W1 .GT. C ) GO TO 90
00016100 N6 = N6 + 1
00016200 90 IF ( A1 .GT. C ) GO TO 92
00016300 N7 = N7 + 1
00016400 92 IF ( A1 .GT. C2 ) GO TO 94
00016500 N8 = N8 + 1
00016600 94 IF ( WW2 .GT. C ) GO TO 110
00016700 N9 = N9 + 1
00016800 110 IF ( M2 .GT. C3 ) GO TO 120
00016900 N10 = N10 + 1
00017000 120 IF ( WC2 .GT. C3 ) GO TO 130
00017100 N11 = N11 + 1
00017200 130 IF ( WC1 .GT. C ) GO TO 140
00017300 N12 = N12 + 1
00017400 140 IF ( M1 .GT. C2 ) GO TO 150
00017500 N13 = N13 + 1
00017600 150 GO TO 3
00017700 3 CONTINUE
00017800 PRINT *, N,TAU,B0,B1,B2,MI1,MI2,AL1,AL2,M,SIG1,SIG2,C,C1,C2,C3
00017900 PRINT *, N7,N8,N9,N10,N11,N12,N13
00018000 WRITE (6,95) N1,N2,N3,N,C,C1,C2
00018100 95 FORMAT(4X,I5,4X,I5,4X,I5,4X,I5,2X,F5.2,2X,F5.2,2X,F5.2)
00018200 WRITE (6,97) N4,N5,N6
00018300 97 FORMAT(4X,I5,4X,I5,4X,I5)
00018400 STOP
00018500 END

```

B.0.4 Programa fabek

Este programa foi utilizado nas ilustrações numéricas 3.3 e 3.5.

```

00000050 $SET AUTOBIND LINEINFO
00000100 $BIND=FROM BIB/IMSL/1,BIB/IMSL/2,BIB/IMSL/3,BIB/IMSL/4
00000200 INTEGER N,N1,N2,N3,M
00000300 REAL PSI1(100),PSI2(100),X1(100),X2(100),Y1(100),Y2(100),E(500)
00000400 REAL K1,K2,KX1,KX2,L01,L02,L11,L12,LA1,LA2,ZM,LR1,LR2
00000500 REAL LAMB1,LAMB2,LRC1,LRC2,NSIG,LZ2,LZC1
00000600 DOUBLE PRECISION SEED
00000700 READ (5,/) N,TAU,KX1,KX2,AL1,AL2,MI1,MI2,B1,B2,B0,NSIG,C1,C2,M
00000800 C C1 EH O VALOR QUIQUADRADO COM 1 G.L. E C1 NIVEL.SIG. NSIG.
00000900 C FB1 E FB2 FATOR DE BAYES A POST.,W1,W2 ESTAT. DE WALD E LR1,LR2
A
00001000 C RAZAO DE VEROSS.,FBQUI1,FBQUI2 APROX. QUI-QU@R@DO PARA O FAT
DE
00001100 C BAYES POST.. LRC1 E LRC2 A RAZAO DE VEROSS. CORRIGIDA COM
O FATOR DE

```

```

00001200 C BARTLETT.
00001300 C ESTE PROGRAMA FAZ 1000 SIMULACOES PARA O MEV COM COEFICIEN-
00001400 C TE DE ATENUACAO CONHECIDO. OS INDICES 1 E 2 SAO PARA AS 2 POPULACOES
00001500 C A HIPOTESE DESTE PROGRAMA EH B1=B2=B0 (2 POP.) E B1=B0 (1 POP.)
00001600 SEED = SEED**0.5
00001700 N1 = 0
00001800 N2 = 0
00001900 N3 = 0
00002000 N6 = 0
00002300 N7 = 0
00002400 N8 = 0
00002500 N9 = 0
00002600 N10 = 0
00002700 N11 = 0
00002800 N12 = 0
00002900 N13 = 0
00003000 N14 = 0
00003010 N15 = 0
00003020 N16 = 0
00003030 N17 = 0
00003040 N18 = 0
00003050 K1 = (1-KX1)/KX1
00003100 K2 = (1-KX2)/KX2
00003200 FBS1 = 0.0
00003300 FBS2 = 0.0
00003400 C INICIO DAS M=1000 SIMULACOES.
00003500 DO 3 J = 1,M
00003600 S1 = 0.0
00003700 S2 = 0.0
00003800 S3 = 0.0
00003900 S4 = 0.0
00004000 S5 = 0.0
00004100 S6 = 0.0
00004200 S7 = 0.0
00004300 S8 = 0.0
00004400 S9 = 0.0
00004500 S10 = 0.0
00004600 SEED = SEED + ATAN(1)
00004700 CALL GGNML(SEED,N+10,E)
00004800 DO 4 I = 2,N+1
00004900 PR01(I-1) = MI1+SQRT(TAU)*E(I-1)
00005000 4 CONTINUE
00005100 SEED = SEED + 4*ATAN(1)
00005200 CALL GGNML(SEED,N+10,E)
00005300 DO 5 I = 2,N+1
00005400 PSI2(I-1) = MI2+SQRT(TAU)*E(I-1)
00005500 5 CONTINUE
00005600 SEED = DSQRT(SEED)
00005700 CALL GGNML(SEED,N+10,E)
00005800 DO 6 I = 2,N+1
00005900 X1(I-1)=PSI1(I-1) + SQRT(K1*TAU)*E(I-1)
00006000 6 CONTINUE
00006100 SEED = SEED - 0.66666666
00006200 CALL GGNML(SEED,N+10,E)
00006300 DO 8 I = 2,N+1
00006400 X2(I-1) = PSI2(I-1) + SQRT(K2*TAU)*E(I-1)
00006500 8 CONTINUE
00006600 SEED = SEED + 0.55555555
00006700 CALL GGNML(SEED,N+10,E)
00006800 DO 10 I = 2,N+1
00006900 Y1(I-1) = AL1+B1*PSI1(I-1)+ SQRT(K1*TAU)*E(I-1)
00007000 10 CONTINUE
00007100 SEED = SEED + ATAN(1)
00007200 CALL GGNML(SEED,N+10,E)
00007300 DO 12 I = 2,N+1
00007400 Y2(I-1) = AL2+B2*PSI2(I-1)+ SQRT(K2*TAU)*E(I-1)
00007500 12 CONTINUE
00007600 C PRINT */, X1,Y1,X2,Y2
00007700 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 1 E 2
00007800 DO 20 I = 1,N

```

```

00007900 S1= S1 + X1(I)
00008000 S2= S2 + X1(I)**2
00008100 S3= S3 + Y1(I)
00008200 S4= S4 + Y1(I)**2
00008300 S5= S5 + X1(I)*Y1(I)
00008400 S6= S6 + X2(I) 00008500 S7= S7 + X2(I)**2
00008600 S8= S8 + Y2(I)
00008700 S9= S9 + Y2(I)**2
00008800 S10=S10 + X2(I)*Y2(I)
00008900 20 CONTINUE
00009000 SXX1 = S2 / FLOAT(N)
00009100 SYY1 = S4 / FLOAT(N)
00009200 SXY1 = S5 / FLOAT(N)
00009300 XB1 = S1/FLOAT(N)
00009400 YB1 = S3/FLOAT(N)
00009500 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 2
00009600 SXX2 = S7 / FLOAT(N)
00009700 SYY2 = S9 / FLOAT(N)
00009800 SXY2 = S10 / FLOAT(N)
00009900 XB2 = S6/FLOAT(N)
00010000 C PRINT *,S1,S2,S3,S4,S5,S6,S7,S8,S9,S10,N
00010100 C PRINT *,SXX1,SXX2,SYY1,SYY2,SXY1,SXY2,XB1,XB2,YB1
00010200 YB2 = S8/FLOAT(N)
00010300 C TERMINOU CALCULO DE COVARIANCIAS DAS POP. 1 E 2
00010400 LA1 = KX1*SXX1
00010500 LA2 = KX2*SXX2
00010600 BET1 = SXY1/ LA1
00010700 BET2 = SXY2/ LA2
00010800 C PRINT *,BET1,BET2
00010900 C TESTE DE WALD A SEGUIR
00011000 L11 = KX1*(BET1**2)*SXX1 - 2*BET1*SXY1 + (1/KX1)*SYY1
00011100 L12 = KX2
00011200 A1 = LA1/L11
0 0011300 A2 = LA2/L12
00011400 W1 =N* ( (BET1-B0)**2 )*A1
00011500 W2 = W1 + N*( (BET2-B0)**2 )*A2
00011600 C TESTE DA RAZAO DE VEROSSIMILHANCA A SEGUIR
00011700 L01 = KX1*(B0**2)*SXX1 - 2*B0*SXY1 + (1/KX1)*SYY1
00011800 L02 = KX2*(B0**2)*SXX2 - 2*B0*SXY2 + (1/KX2)*SYY2
00011900 C PRINT *,L01,L02,L11,L12
00012000 ZM1 = L01/L11
00012100 ZM2 = L02/L12
00012200 LR2 = N*ALOG(ZM1) + N*ALOG(ZM2)
00012300 LR1 = N*ALOG(ZM1)
00012400 LZ2 = N * ALOG(L02/L12)
00012500 LRC2= (1-1.5/N)*LR2
00012600 LRC1= (1-1.5/N)*LR1
00012700 LZC1 = (1 - 1.5/N)*LZ2
00012800 VCLR1 = (N-2)*LR1/N
00012810 VCBLR1 = ( 1 + (0.5/N) ) * LR1
00012820 VCLZ1 = (N-2)*LZ2/N
00012830 VCBLZ1 = ( 1 + (0.5/N) ) * LZ2
00012840 VCLR2 = (N-2)*LR2/N
00012850 VCBLR2 = VCLR2/(1- 0.5/N)
00012860 C FATOR DE BAYES A POSTERIORI
00012900 SIGMA1 = ( SXX1*SYY1 - SXY1**2 ) / ( ( KX1*SXX1)**2 )
00013000 SIGMA2 = ( SXX2*SYY2 - SXY2**2 ) / ( ( KX2*SXX2)**2 )
00013100 LAMB1 = N*( (BET1 - B0)**2 ) / SIGMA1
00013200 LAMB2 = N*( (BET2 - B0)**2 ) / SIGMA2
00013300 FB1 = SQRT(2.0) * EXP ( -0.5*(LAMB2) )
00013400 FB2 = 2 * EXP ( -0.5*(LAMB1+LAMB2) )
00013500 FBS1 = FBS1 + FB1
00013600 FBS2 = FBS2 + FB2
00013700 FBQUI1 = LAMB2 - ALOG(2.0)

```

```

00013800 FBQUI2 = (LAMB1+LAMB2) - 2*ALOG(2.0)
00013900 IF ( LR2 .GT. C2 ) GO TO 60
00014000 N1 = N1+1
00014100 60 IF ( LRC2 .GT. C2 ) GO TO 62
00014200 N2= N2 + 1
00014210 62 IF ( VCLR2 .GT. C2 ) GO TO 64
00014220 N13 = N13 + 1
00014230 64 IF ( VCBLR2 .GT. C2 ) GO TO 70
00014240 N14= N14 + 1
00014300 70 IF ( W2 .GT. C2 ) GO TO 80
00014400 N3 = N3 + 1
00014500 80 IF ( FB2 .LE. NSIG ) GO TO 90
00014600 N4 = N4 + 1
00014700 90 IF ( FBQUI2 .GT. C2 ) GO TO 100
00014800 N5 = N5 + 1
00014900 100 IF (FBQUI1 .GT. C1 ) GO TO 110
00015000 N6 = N6 + 1
00015100 110 IF (FB2 .GT. C2 ) GO TO 112
00015200 N7 = N7 + 1
00015300 112 IF ( VCLR1 .GT. C1 ) GO TO 114
00015310 N15 = N15 + 1
00015320 114 IF ( VCBLR1 .GE. C1 ) GO TO 116
00015330 N16 = N16 + 1
00015340 116 IF ( VCLZ1 .GT. C1 ) GO TO 118
00015350 N17 = N17 + 1
00015360 118 IF ( VCBLZ1 .GT. C1 ) GO TO 120
00015370 N18 = N18 + 1
00015380 120 IF ( W1 .GT. C1 ) GO TO 130
00015400 N8 = N8 + 1
00015500 130 IF ( FB1 .LE. NSIG ) GO TO 140
00015600 N9 = N9 + 1
00015700 140 IF ( FBQUI1 .GT. C1 ) GO TO 155
00015800 N10 = N10 + 1
00015900 155 IF ( LZ2 .GT. C1 ) GO TO 160
00016000 N11 = N11 + 1
00016100 160 IF ( LZC1 .GT. C1 ) GO TO 170
00016200 N12 = N12 + 1
00016300 170 GO TO 3
00016400 3 CONTINUE
00016500 XFB1 = FBS1/FLOAT(M)
00016600 XFB2 = FBS2/FLOAT(M)
00016700 PRINT */,N,TAU,KX1,KX2,AL1,AL2,MI1,MI2,B1,B2,B0,NSIG,C1,C2
00016800 PRINT */,N1,N2,N13,N14,N3,N4,N5
00016900 PRINT */,N6,N11,N12,N7,N13,N14,N15,N16,N17,N18,N8,N9,N10
00017000 PRINT */,XFB1,XFB2
00017100 STOP END

```

B.0.5 Programa Sibet1

Este programa foi utilizado nas ilustração numérica 3.5.

```

$SET AUTOBIND LINEINFO
00000200 $BIND=FROM BIB/IMSL/1,BIB/IMSL/2,BIB/IMSL/3,BIB/IMSL/4
00000300 INTEGER N,N1,N2,N3,M
00000400 REAL PSI1(100),PSI2(100),X1(100),X2(100),Y1(100),Y2(100),E(500)
00000500 REAL K1,K2,KX1,KX2,L01,L02,L11,L12,LA1,LA2,ZM,LR1,LR2
00000600 REAL LAMB1,LAMB2,LRC1,LRC2,NSIG,LZ2,LZC1
00000700 DOUBLE PRECISION SEED
00000800 READ (5,/) N,TAU,KX1,KX2,AL1,AL2,MI1,MI2,B1,B2,B0,NSIG,C1,C2,M
00000900 READ (5,/) PREC
00001000 C C1 EH O VALOR QUIQUADRADO COM 1 G.L. E C1 NIVEL.SIG. NSIG.
00001100 CE FB2 FATOR DE BAYES A POST.,W2 ESTAT. DE WALD
00001200 C RAZAO DE VEROSS.,FBQUI1,FBQUI2 APROX. QUI-QUADRADO PARA O FAT
DE

```



```

00001300 C BAYES POST.. LRC1 E LRC2 A RAZAO DE VEROSS. CORRIGIDA COM
0 FATOR DE
00001400 C BARTLETT. LAMBDA01=LAMBDA02 E B TEM UMA FORMA EXPLICITA
00001500 C ESTE ARQUIVO CHAMA-SE SIBET1, COMPILA COM FITA/COMPILA E EXECCUTA-SE
00001600 C COM EXECFABEK.
00001610 C ESTE PROGRAMA FAZ 1000 SIMULACOES PARA O MEV COM COEFICIEN-
00001700 C TE DE ATENUACAO CONHECIDO. OS INDICES 1 E 2 SAO PARA AS 2 POPULACOES
00001800 C PRINT */ , N1,N2,N3,N4,N5,N6,N7
00001900 C A HIPOTESE DESTES PROGRAMA EH B1=B2=B (2 POP.)
00002000 SEED
00002100 N1 = 0
00002200 N2 = 0
00002300 N3 = 0
00002400 N4 = 0
00002500 N5 = 0
00002600 N6 = 0
00002700 N7 = 0
00002800 N8 = 0
00002900 N9 = 0
00003000 N10 = 0
00003100 N11 = 0
00003200 N12 = 0
00003300 N13 = 0
00003400 N14 = 0
00003500 N15 = 0
00003600 N16 = 0
00003700 N17 = 0
00003800 N18 = 0
00003900 K1 = (1-KX1)/KX1
00004000 K2 = (1-KX2)/KX2
00004100 C PRINT */ , K1,K2
00004200 FBS1 = 0.0
00004300 FBS2 = 0.0
00004400 C INICIO DAS M=1000 SIMULACOES.
00004500 DO 3 J = 1,M
00004600 S1 =0.0
00004700 S2 =0.0
00004800 S3 =0.0
00004900 S4 =0.0
00005000 S5 =0.0
00005100 S6 =0.0
00005200 S7 =0.0
00005300 S8 =0.0
00005400 S9 =0.0
00005500 S10=0.0
00005600 SEED = SEED + ATAN(1)
00005700 CALL GGNML(SEED,N+10,E)
00005800 DO 4 I=2,N+1
00005900 PSI1(I-1) = MI1+SQRT(TAU)*E(I-1)
00006000 4 CONTINUE
00006100 SEED = SEED + 4*ATAN(1)
00006200 CALL GGNML(SEED,N+10,E)
00006300 DO 5 I = 2,N+1
00006400 PSI2(I-1) = MI2+SQRT(TAU)*E(I-1)
00006500 5 CONTINUE
00006600 SEED = DSQRT(SEED)
00006700 CALL GGNML(SEED,N+10,E)
00006800 DO 6 I = 2,N+1
00006900 X1(I-1)=PSI1(I-1) + SQRT(K1*TAU)*E(I-1)
00007000 6 CONTINUE
00007100 SEED = SEED - 0.66666666
00007200 DO 8 I = 2,N+1
00007300 X2(I-1) = PSI2(I-1) + SQRT(K2*TAU)*E(I-1)
00007400 8 CONTINUE
00007500 SEED = SEED + 0.55555555
00007600 CALL GGNML(SEED,N+10,E)
00007700 DO 10 I = 2,N+1
00007800 Y1(I-1) = AL1+B1*PSI1(I-1)+ SQRT(K1*TAU)*E(I-1)
00007900 10 CONTINUE
00008000 SEED = SEED + ATAN(1)

```

```

00008200 CALL GGNML(SEED,N+10,E)
00008300 DO 12 I = 2,N+1
00008400 Y2(I-1) = AL2+B2*PSI2(I-1)+ SQRT(K2*TAU)*E(I-1)
00008500 12 CONTINUE
00008600 C PRINT *, X1,Y1,X2,Y2
00008700 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 1 E 2
00008800 DO 20 I = 1,N
00008900 S1= S1 + X1(I)
00009000 S2= S2 + X1(I)**2
00009100 S3= S3 + Y1(I)
00009200 S4= S4 + Y1(I)**2
00009300 S5= S5 + X1(I)*Y1(I)
00009400 S6= S6 + X2(I)
00009500 S7= S7 + X2(I)**2
00009600 S8= S8 + Y2(I)
00009700 S9= S9 + Y2(I)**2
00009800 S10=S10 + X2(I)*Y2(I)
00009900 20 CONTINUE
00010000 SXX1 = S2 / FLOAT(N)
00010100 SYY1 = S4 / FLOAT(N)
00010200 SXY1 = S5 / FLOAT(N)
00010300 XB1 = S1/FLOAT(N)
00010400 YB1 = S3/FLOAT(N)
00010500 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 2
00010600 SXX2 = S7 / FLOAT(N)
00010700 SYY2 = S9 / FLOAT(N)
00010800 SXY2 = S10 / FLOAT(N)
00010900 XB2 = S6/FLOAT(N)
00011000 C PRINT *,S1,S2,S3,S4,S5,S6,S7,S8,S9,S10,N,M
00011100 C PRINT *,SXX1,SXX2,SYY1,SYY2,SXY1,SXY2,XB1,XB2,YB1
00011200 YB2 = S8/FLOAT(N)
00011300 C TERMINOU CALCULO DE COVARIANCIAS DAS POP. 1 E 2
00011400 LA1 = KX1*SXX1
00011500 LA2 = KX2*SXX2
00011600 BET1 = SXY1/ LA1
00011700 BET2 = SXY2/LA2
00011800 C PRINT *,BET1,BET2
00011900 C TESTE DA RAZAO DE VEROSSIMILHANCA
00012000 L11 = KX1*(BET1**2)*SXX1 - 2*BET1*SXY1 + (1/KX1)*SYY1
00012100 L12 = KX2*(BET2**2)*SXX2 - 2*BET2*SXY2 + (1/KX2)*SYY2
00012200 C PRINT *,LA1,LA2,BET1,BET2,L11,L12
00012300 C CALL BVERO (N,N8,KX1,KX2,K1,K2,SXX1,SXY1,SYY1,SXX2,SXY2,SYY2,IVO,
00012400 C *PREC,B,ITER)
00012500 C IF (IVO.EQ. 0 ) GO TO 3
00012600 B =(SXY1+SXY2)/(KX1*(SXX1+SXX2))
00012700 C PRINT *,B0,BET1,BET2,B,SXX1,SXX2,SXY1,SXX2,SXY2,SYY2,N8,KX1,KX2
00012800 C PRINT *, K1,K2,M
00012900 C B =(1 + (N+3)*(N+5)/( (N-2)*(N-4) ) ) *B
00013000 C B = 2*B
00013100 C PRINT *,B
00013200 L01 = KX1*(B**2)*SXX1 - 2*B*SXY1 + (1/KX1)*SYY1
00013300 L02 = KX2*(B**2)*SXX2 - 2*B*SXY2 + (1/KX2)*SYY2
00013400 ZM1 = L01/L11
00013500 ZM2 = L02/L12
00013600 RM1 = N * ALOG(ZM1) + N*ALOG(ZM2)
00013700 ERM1=ERM1+RM1
00013750 C CORRECAO DE BARTLETT D PARA RM1
00013800 C CALCULO DE D COM N1=N2=N
00013900 BD1 = LA1/L01
00014000 BD2 = LA2/L02
00014100 AD = `*B*B
00014200 DD1 = (3.0/N) - 2.0*(BD1+BD2)/AD
00014300 C PRINT *,L01,L02,ZM1,ZM2,RM1,ZM2,RM1,BD1,BD2,AD
00014400 C NAS CORRECOES DE BARTLETT OS ESTIMADORES SAO SOB H ZERO

```

```

00014500 DD2 = (1./3.)* ( (1/AD)**2 ) * N *(LA1**2+LA2**2)
00014600 DD3 = (5./6)* ( (1/AD)**2 ) * N* ( BD1**2 + BD2**2 )
00014700 D = DD1 - DD2 + DD3
00014800 C VEROSSIMILHANCA CORRIGIDA PELO FATOR D DE BARTLETT
00014900 RM4 = RM1/(1+D/2)
00015000 ERM4=ERM4+RM4
00015010 C VEROSSIMILHANCA CONDICIONAL PROFILE
00015100 RM5 = ( 1 - 2/N ) * RM1
00015200 C VEROSSIMILHANCA CONDICIONAL PROFILE COM CORRECAO DE BARTLETT
00015300 RM6 = RM5/D
00015400 C FATOR DE BAYES POSTERIORI
00015500 SIGMA1 = { SXX1*SY11 - SXY1**2 } /({ ( KX1*SXX1)**2 }*(N-1))
00015600 SIGMA2 = { SXX2*SY22 - SXY2**2 } /({ ( KX2*SXX2)**2 }*(N-1))
00015700 BBA = (BET1+BET2)/2
00015800 G1 =N*{(BBA-BET1)**2}/((N-1)*SIGMA1)
00015900 G2 =N*{(BBA-BET2)**2}/ FBP =
00015600 EFBP=EFBP+FBP
00016210 FBQUI2 = G - 0.5*ALOG(2.0)
00016300 EFBQUI2=EFBQUI2+FBQUI2
00016310 C TESTE DE WALD
00016400 BW1=LA1/L11
00016500 BW2=LA2/L12
00016600 W2 = ( N*(BET1-BET2)**2 ) / ( (1/BW1) + (1/BW2) )
00016700 C PRINT */, DD1,DD2,DD3,D,RM4,RM5,RM6,SIGMA1,SIGMA2,BBA,G1,G2,G
00016800 C PRINT */, FBP,FBQUI2,W2
00016900 IF ( RM1 .GT. C1 ) GO TO 70
00017000 N1 = N1+1
00017100 70 IF ( RM4 .GT. C1 ) GO TO 80
00017200 N2= N2 + 1
00017300 80 IF ( RM5 .GT. C1 ) GO TO 90
00017400 N3 = N3 + 1
00017500 90 IF ( RM6 .GT. C1 ) GO TO 100
00017600 N4= N4 + 1
00017700 100 IF ( W2 .GT. C1 ) GO TO 110
00017800 N5 = N5 + 1
00017900 110 IF ( FBP .LE. NSIG ) GO TO 120
00018000 N6 = N6 + 1
00018100 120 IF ( FBQUI2 .GT. C1 ) GO TO 130
00018200 N7 = N7 + 1
00018300 C PRINT */, N1,N2,N3,N4,N5,N6,N7,N,M,C1
00018400 130 GO TO 3
00018500 3 CONTINUE
00018600 ERM1=ERM1/FLOAT(M)
00018610 ERM4=ERM4/FLOAT(M)
00018620 EFBP=EFBP/FLOAT(M)
00018630 EFBQUI2=EFBQUI2/FLOAT(M)
00018640 PRINT */,N,TAU,KX1,KX2,AL1,AL2,MI1,MI2,B1,B2,B0,NSIG,C1,C2
00018700 PRINT */,N1,N2,N3,N4,N5,N6,N7,N8,N
00018800 PRINT */,ERM1,ERM4
00018810 PRINT */,EFBP,EFBQUI2
00018820 STOP
00018900 END

```

B.0.6 Programa Sibet2

Este programa foi utilizado nas ilustrações numéricas 3.7 e 4.3

```

$SET AUTOBIND LINEINFO
00000200 $BIND=FROM BIB/IMSL/1,BIB/IMSL/2,BIB/IMSL/3,BIB/IMSL/4
00000300 INTEGER N,N1,N2,N3,M
00000400 REAL PSI1(100),PSI2(100),X1(100),X2(100),Y1(100),Y2(100),E(500)
00000500 REAL K1,K2,KX1,KX2,L01,L02,L11,L12,LA1,LA2,ZM,LR1,LR2
00000600 REAL LAMB1,LAMB2,LRC1,LRC2,NSIG,LZ2,LZC1
00000700 DOUBLE PRECISION SEED

```

```

00000800 READ (5,/) N,TAU,AL1,AL2,MI1,MI2,B1,B2,NSIG,C1,C2,M
00000900 READ (5,/) KX1,KX2
00001000 C C1 EH O VALOR QUIQUADRADO COM 1 G.L. E C1 NIVEL.SIG. NSIG.
00001100 CE FB2 FATOR DE BAYES A POST.,W2 ESTAT. DE WALD
00001200 C RAZAO DE VEROSS.,FBQUI1,FBQUI2 APROX. QUI-QUADRADO PARA O FAT
DE
00001300 C BAYES POST.. LRC1 E LRC2 A RAZAO DE VEROSS. CORRIGIDA COM
O FATOR DE
00001400 C BARTLETT. LAMBDA01=LAMBDA02 E B TEM UMA FORMA EXPLICITA
00001500 C ESTE ARQUIVO CHAMA-SE SIBET2, COMPILA COM FITA/COMPILA, EXECCUTA-SE
00001600 C COM EXECFABEK E A SAIDA EH SIBET2/SAI
00001700 C ESTE PROGRAMA FAZ 1000 SIMULACOES PARA O MEV COM COEFICIEN-
00001800 C TE DE ATENUACAO CONHECIDO. OS INDICES 1 E 2 SAO PARA AS 2 POPULACOES
00001900 C PRINT */ , N1,N2,N3,N4,N5,N6,N7
00002000 C A HIPOTESE DESTE PROGRAMA EH B1=B2=B (2 POP.)
00002100 SEED = SEED**0.5
00002200 N1 = 0
00002300 N2 = 0
00002400 N3 = 0
00002500 N4 = 0
00002600 N5 = 0
00002700 N6 = 0
00002800 N7 = 0
00002900 N8 = 0
00003000 N9 = 0
00003100 N10 = 0
00003200 N11 = 0
00003300 N12 = 0
00003400 N13 = 0
00003500 N14 = 0
00003600 N15 = 0
00003700 N16 = 0
00003800 N17 = 0
00003900 N18 = 0
00004000 K1 = (1-KX1)/KX1
00004100 K2 = (1-KX2)/KX2
00004200 C PRINT */ , K1,K2
00004300 EFBP = 0.0
00004400 EFBQUI2=0.0
00004500 C INICIO DAS M=1000 SIMULACOES.
00004600 DO 3 J = 1,M
00004700 S1 =0.0
00004800 S2 =0.0
00004900 S3 =0.0
00005100 S4 =0.0
00005100 S5 =0.0
00005200 S6 =0.0
00005300 S7 =0.0
00005400 S8 =0.0
00005500 S9 =0.0
00005600 S10=0.0
00005700 SEED = SEED + ATAN(1)
00005800 CALL GGNML(SEED,N+10,E)
00005900 DO 4 I=2,N+1
00006000 PSI1(I-1) = MI1+SQRT(TAU)*E(I-1)
00006100 4 CONTINUE
00006200 SEED = SEED + 4*ATAN(1)
00006300 CALL GGNML(SEED,N+10,E)
00006400 DO 5 I = 2,N+1
00006500 PSI2(I-1) = MI2+SQRT(TAU)*E(I-1)
00006600 5 CONTINUE
00006700 SEED = DSQRT(SEED)
00006800 CALL GGNML(SEED,N+10,E)
00006900 DO 6 I = 2,N+1
00007000 X1(I-1) = PSI1(I-1) + SQRT(K1*TAU)*E(I-1)
00007100 6 CONTINUE
00007200 SEED = SEED - .44444444
00007300 CALL GGNML(SEED,N+10,E)
00007400 DO 8 I = 2,N+1

```

```

00007500 X2(I-1) = PSI2(I-1) + SQRT(K2*TAU)*E(I-1)
00007600 8 CONTINUE
00007700 SEED = SEED + 0.5555555
00007800 CALL GGNML(SEED,N+10,E)
00007900 DO 10 I = 2,N+1
00008000 Y1(I-1) = AL1+B1*PSI1(I-1)+ SQRT(K1*TAU)*E(I-1)
00008100 10 CONTINUE
00008200 SEED = SEED + ATAN(1)
00008300 CALL GGNML(SEED,N+10,E)
00008400 DO 12 I = 2,N+1
00008500 Y2(I-1) = AL2+B2*PSI2(I-1)+ SQRT(K2*TAU)*E(I-1)
00008600 12 CONTINUE
00008700 C PRINT *, X1,Y1,X2,Y2
00008800 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 1 E 2
00008900 DO 20 I = 1,N
00009000 S1= S1 + X1(I)
00009100 S2= S2 + X1(I)**2
00009200 S3= S3 + Y1(I)
00009300 S4= S4 + Y1(I)**2
00009400 S5= S5 + X1(I)*Y1(I)
00009500 S6= S6 + X2(I)
00009600 S7= S7 + X2(I)**2
00009700 S8= S8 + Y2(I)
00009800 S9= S9 + Y2(I)**2
00009900 S10=S10 + X2(I)*Y2(I)
00010000 20 CONTINUE
00010100 SXX1 =( S2 - (S1**2)/FLOAT(N) ) / FLOAT(N)
00010200 SYY1 =( S4 - (S3**2)/FLOAT(N) ) / FLOAT(N)
00010300 SXY1 =( S5 - S1*S3/FLOAT(N) ) / FLOAT(N)
00010400 XB1 = S1/FLOAT(N)
00010500 YB1 = S3/FLOAT(N)
00010600 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 2
00010700 SXX2 =( S7 - (S6**2)/FLOAT(N) ) / FLOAT(N)
00010800 SYY2 =( S9 - (S8**2)/FLOAT(N) ) / FLOAT(N)
00010900 SXY2 =( S10 - S6*S8/FLOAT(N) ) / FLOAT(N)
00011000 XB2 = S6/FLOAT(N)
00011100 C PRINT *,S1,S2,S3,S4,S5,S6,S7,S8,S9,S10,N,M
00011200 C PRINT *,SXX1,SXX2,SYY1,SYY2,SXY1,SXY2,XB1,XB2,YB1
00011300 YB2 = S8/FLOAT(N)
00011400 LA1 = KX1*SXX1
00011600 LA2 = KX2*SXX2
00011700 BET1 = SXY1/ LA1
00011800 BET2 = SXY2/ LA2
00011900 C PRINT *,BET1,BET2
00012000 C TESTE DA RAZAO DE VEROSSIMILHANCA
00012100 L11 = KX1*(BET1**2)*SXX1 - 2*BET1*SXY1 + (1/KX1)*SYY1
00012200 L12 = KX2*(BET2**2)*SXX2 - 2*BET2*SXY2 + (1/KX2)*SYY2
00012300 C PRINT *,LA1,LA2,BET1,BET2,L11,L12
00012400 C CALL BVERO (N,N8,KX1,KX2,K1,K2,SXX1,SXY1,SYY1,SXX2,SXY2,SYY2,IVO,
00012500 C *PREC,B,ITER)
00012600 C IF (IVO .EQ. 0 ) GO TO 3
00012700 B =(SXY1+SXY2)/(KX1*(SXX1+SXX2))
00012800 C PRINT *,B0,BET1,BET2,B,SXX1,SXX2,SXY1,SXX2,SXY2,SYY2,N8,KX1,KX2
00012900 C PRINT *, K1,K2,M
00013000 C B =(1 + (N+3)*(N+5)/( (N-2)*(N-4) ) ) *B
00013100 C B = 2*B
00013200 C PRINT *,B
00013300 L01 = KX1*(B**2)*SXX1 - 2*B*SXY1 + (1/KX1)*SYY1
00013400 L02 = KX2*(B**2)*SXX2 - 2*B*SXY2 + (1/KX2)*SYY2
00013500 ZM1 = L01/L11
00013600 ZM2 = L02/L12
00013700 RM1 = N * ALOG(ZM1) + N*ALOG(ZM2)
00013800 ERM1=ERM1+RM1
00013900 C CORRECAO DE BARTLETT D PARA RM1

```

```

00014000 C CALCULO DE D COM N1=N2=N
00014100 BD1 = LA1/L01
00014200 BD2 = LA2/L02
00014300 AD = N*(BD1+BD2)
00014400 AA1 = L01+LA1*(B**2)
00014500 AA2 = L02+LA2*(B**2)
00014600 DD1 = 4/FLOAT(N)
00014700 DD2 = 3/FLOAT(N)
00014800 DD31= (LA1**2)/AA1 + (LA2**2)/AA2
00014900 DD3 = ((B**2)/FLOAT(N))*DD31
00015000 FF1 = L01**2 + (LA1**2)*(B**4)
00015100 FF2 = L02**2 + (LA2**2)*(B**4)
00015200 DD4 = (1.0/(2.0*N))*(FF1/(AA1**2)+FF2/(AA2**2) )
00015300 GG1 = ( (LA1*B)**2)/(L01*AA1)
00015400 GG2 = ( (LA2*B)**2)/(L02*AA2)
00015500 DD5 = (1/AD)*(GG1+GG2)
00015600 DD61= (B/AA1)**2
00015700 DD62= (B/AA2)**2
00015800 DD6 = (1/AD)*(LA1*DD61 + LA2*DD62)
00015900 DD7 = (N/(2*(AD**2)))*(BD1**2 + BD2**2)
00016000 D = DD1 + DD2 - DD3 - DD4 + DD5 - DD6 + DD7
00016100 C PRINT *,L01,L02,ZM1,ZM2,RM1,ZM2,RM1,BD1,BD2,AD
00016300 C VEROSSIMILHANCA CORR@GIDA PELO FATOR D D@ BARTLETT
00016800 RM4 = RM1/(1+D/2)
00016900 ERM4=ERM4+RM4
00017000 C VEROSSIMILHANCA CONDICIONAL PROFILE
00017100 RM5 = ( 1 - 2/N ) * RM1
00017200 C VEROSSIMILHANCA CONDICIONAL PROFILE COM CORRECAO DE BARTLETT
00017300 RM6 = RM5/D
00017400 C FATOR DE BAYES POSTERIORI
00017500 TT1=((K1+1)**2)* (SXX1*SY1-SXY1**2)/(SXX1**2)
00017600 TT2=((K2+1)**2)* (SXX2*SY2-SXY2**2)/(SXX2**2)
00017700 BBAYS=((BET1/TT1)+(BET2/TT2))/((1/TT1)+(1/TT2))
00017800 G1 = N*( (BBAYS-BET1)**2 ) / TT1
00017900 G2 = N*( (BBAYS-BET2)**2 ) / TT2
00018000 G = G1 + G2
00018100 C
00018110 FBP = SQRT(2.0)*EXP(-0.5*G)
00018200 EFBP =EFBP +FBP
00018250 FBQUI2 = G - 0.5*ALOG(2.0)
00018300 EFBQUI2=EFBQUI2+FBQUI2
00018350 C TESTE DE WALD
00018400 BW1=LA1/L11
00018500 BW2=LA2/L12
00018600 W2 = ( N*(BET1-BET2)**2 ) / ( (1/BW1) + (1/BW2) )
00018700 C PRINT *, DD1,DD2,DD3,D,RM4,RM5,RM6,SIGMA1,SIGMA2,BBA,G1,G2,G
00018800 C PRINT *, FBP,FBQUI2,W2
00018900 IF ( RM1 .GT. C1 ) GO TO 70
00019000 N1 = N1+1
00019100 70 IF ( RM4 .GT. C1 ) GO TO 80
00019200 N2= N2 + 1
00019300 80 IF ( RM5 .GT. C1 ) GO TO 90
00019400 N3 = N3 + 1
00019500 90 IF ( RM6 .GT. C1 ) GO TO 100
00019600 N4= N4 + 1
00019700 100 IF ( W2 .GT. C1 ) GO TO 110
00019800 N5 = N5 + 1
00019900 110 IF ( FBP .LE. NSIG ) GO TO 120
00020000 N6 = N6 + 1
00020100 120 IF ( FBQUI2 .GT. C1 ) GO TO 130
00020200 N7 = N7 + 1
00020300 C PRINT *, N1,N2,N3,N4,N5,N6,N7,N,M,C1
00020400 130 GO TO 3
00020500 3 CONTINUE
00020600 ERM1=ERM1/FLOAT(M)
00020700 ERM4=ERM4/FLOAT(M)

```



```

00020800 EFBP=EFBP/(M*1.0)
00020810 EFBQUI2=EFBQUI2/FLOAT(M)
00020820 PRINT *,N,TAU,KX1,KX2,AL1,AL2,MI1,MI2,B1,B2,B0,NSIG,C1,C2
00020900 PRINT *,N1,N2,N3,N4,N5,N6,N7,N8,N
00021000 PRINT *,ERM1,ERM4
00021100 PRINT *,EFBP,EFBQUI2
00021110 STOP
00021200 END

```

B.0.7 Programa Sieswv

Este programa foi utilizado no exemplo 4.5

```

      SETAUTOBINDLINEINFO
00000100 BIND=FROM BIB/IMSL/1,BIB/IMSL/2,BIB/IMSL/3,BIB/IMSL/4
00000200 INTEGER N,I
00000300 REAL X1(100),X2(100),Y1(100),Y2(100)
00000400 DOUBLE PRECISION SEED
00000600 READ (5,/) N
00000700 READ (5,/) (X1(I),Y1(I),I=1,N)
00000800 READ (5,/) (X2(I),Y2(I),I=1,N)
00000900 READ (5,/) XB1,YB1,XB2,YB2
00000910 C CASO ESTRUTURAL, RAZ. VAR=150.
00001000 C WALD E' A ESTAT. DE WALD E RVER A RAZ. DE VER. N1=N2=N.
00001100 C ESTE ARQUIVO CHAMA-SE SIESWV, COMPILA COM FITA/COMPILA E EXECUTA-SE
00001200 C COM EXCEX.
00001300 C ESTE PROGRAMA E' UM EXEMPLO COM DADOS DE MADANSKY, 1959
00001400 C NO CASO ESTRUTURAL. OS INDICES 1 E 2 SAO PARA AS 2 POPULACOES
00001500 C A HIPOTESE DESTE PROGRAMA EH B1=B2=B (2 POP.), B DESCONHECIDO.
00001600 C OS DADOS ESTAO TRANSFORMADOS PARA QUE A RAZ. VAR SEJA = 1.(DIV
150
00001700 C OS VALORES DE SYE E POR RAIZ DE 150 OS VALORES DE SXY)
00001800 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 1 E 2
00002000 DO 7 I=1,N
00002010 X1(I)=X1(I)-XB1
00002020 X2(I)=X2(I)-XB2
00002030 Y1(I)=Y1(I)-YB1
00002040 Y2(I)=Y2(I)-YB2 7 CONTINUE
00002060 S1=0.0
00002100 S2=0.0
00002200 S3=0.0
00002300 S4=0.0
00002400 S5=0.0
00002500 S6=0.0
00002600 S7=0.0
00002700 S8=0.0
00002800 S9=0.0
00002900 S10=0.0
00003000 DO 20 I = 1,N
00003100 S1= S1 + X1(I)
00003200 S2= S2 + X1(I)**2
00003300 S3= S3 + Y1(I)
00003400 S4= S4 + Y1(I)**2
00003500 S5= S5 + X1(I)*Y1(I)
00003600 S6= S6 + X2(I)
00003700 S7= S7 + X2(I)**2
00003800 S8= S8 + Y2(I)
00003900 S9= S9 + Y2(I)
00004000 S10= S10 + X2(I)*Y2(I)
00004100 20 CONTINUE
00004200 SXX1 = S2 / FLOAT(N)
00004300 SYY1 = S4 / (FLOAT(N)*150.)
00004400 SXY1 = S5 / (FLOAT(N)*SQRT(150.))
00004500 XB1 = S1/FLOAT(N)

```

```

00004600 YB1 = S3/FLOAT(N)
00004700 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 2
00004800 SXX2 = S7 / FLOAT(N)
00004900 SYY2 = S9 / (FLOAT(N)*150.)
00005000 SKY2 = S10 / (FLOAT(N)*SQRT(150.))
00005100 XB2 = S6/FLOAT(N)
00005200 C PRINT *,S1,S2,S3,S4,S5,S6,S7,S8,S9,S10,N,M
00005300 C PRINT *,SXX1,SXX2,SYY1,SYY2,SKY1,SKY2,XB1,XB2,YB1
00005400 YB2 = S8/FLOAT(N)
00005500 C TERMINOU CALCULO DE COVARIANCIAS DAS POP. 1 E 2
00005600 C RAZAO DE VEROSSIMILHANCA (B FOI OBTIDO EM OUTRO PROGRAMA)
00005700 B=0.7115
00005800 BET1 = ((SYY1-SXX1)+SQRT((SYY1-SXX1)**2 + 4*(SKY1**2)))/(2*SKY1)
00005900 BET2 = ((SYY2-SXX2)+SQRT((SYY2-SXX2)**2 + 4*(SKY2**2)))/(2*SKY2)
00006000 F1B1= SXX1+2*BET1*SKY1+(BET1**2)*SYY1
00006100 F1B2= SXX2+2*BET2*SKY2+(BET2**2)*SYY2
00006200 F1B = SXX1+2*B*SKY1+(B**2)*SYY1
00006300 F2B = SXX2+2*B*SKY2+(B**2)*SYY2
00006400 G1B1=(BET1**2)*SXX1-2*BET1*SKY1+SYY1
00006500 G1B = (B**2)*SXX1-2*B*SKY1+SYY1
00006600 G2B
00006700 FG10= (G1B*F1B)/((B**2+1)**2)
00006900 FG20= (G2B*F2B)/((B**2+1)**2)
00007000 FG11= G1B1*F1B1/((BET1**2+1)**2)
00007100 FG21= G1B2*F1B2/((BET2**2+1)**2)
00007200 CC1=FG10/FG11
00007210 CC2=FG20/FG21
00007220 RVER = N*(ALOG(CC2)+ ALOG(CC1))
00007300 C TESTE DE WALD
00007400 F1B1=F1B1/(1+BET1**2)
00007410 F1B2=F1B2/(1+BET2**2)
00007420 G1B1=G1B1/(1+BET1**2)
00007430 G1B2=G1B2/(1+BET2**2)
00007440 BB1=N*((F1B1-G1B1)**2)/(F1B1*G1B1)
00007500 BB1=BB1/((BET1**2+1)**2)
00007600 BB2=N*((F1B2-G1B2)**2)/(F1B2*G1B2)
00007700 O G1B1=G1B1/(1+BET1**2)
00007743
      WALD=((BET1-BET2)**2)/((1/BB1)+(1/BB2))
00007900 PRINT *,N,BET1,BET2,BB1,BB2
00008000 PRINT *,RVER,WALD
00008100 PRINT *,SXX1,SKY1,SYY1
00008200 PRINT *,SXX2,SKY2,SYY2,B
00008300 STOP
00008400 END
00008500

```

B.0.8 Programa Sies

Este programa foi utilizado no exemplo 4.5.

```

      SET AUTOBINDLINEINFO
00000100 BIND=FROM BIB/IMSL/1,BIB/IMSL/2,BIB/IMSL/3,BIB/IMSL/4
00000200 INTEGER N,I
00000300 REAL X1(100),X2(100),Y1(100),Y2(100)
00000400 DOUBLE PRECISION SEED
00000600 READ (5,/) N
00000700 READ (5,/) (X1(I),Y1(I),I=1,N)
00000800 READ (5,/) (X2(I),Y2(I),I=1,N)
00000900 READ (5,/) XB1,YB1,XB2,YB2
00000910 C CASO ESTRUTURAL, RAZ. VAR=150.
00001000 C WALD E' A ESTAT. DE WALD E RVER A RAZ. DE VER. N1=N2=N.

```

```

00001100 C ESTE ARQUIVO CHAMA-SE SIESWV, COMPILA COM FITA/COMPILA E EXECUTA-SE
00001200 C COM EXCEX.
00001300 C ESTE PROGRAMA E' UM EXEMPLO COM DADOS DE MADANSKY, 1959
00001400 C NO CASO ESTRUTURAL. OS INDICES 1 E 2 SAO PARA AS 2 POPULACOES
00001500 C A HIPOTESE DESTE PROGRAMA EH B1=B2=B (2 POP.), B DESCONHECIDO.
00001600 C OS DADOS ESTAO TRANSFORMADOS PARA QUE A RAZ. VAR SEJA = 1.(DIV
150
00001700 C OS VALORES DE SYY E POR RAIZ DE 150 OS VALORES DE SXY)
00001800
C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 1 E 2
00002000 DO 7 I=1,N
00002010 X1(I)=X1(I)-XB1
00002020 X2(I)=X2(I)-XB2
00002030 Y1(I)=Y1(I)-YB1
00002040 Y2(I)=Y2 7 CONTINUE
00002060 S1=0.0
00002100 S2=0.0
00002200 S3=0.0
00002300 S4=0.0
00002400 S5=0.0
00002500 S6=0.0
00002600 S7=0.0
00002700 S8=0.0
00002800 S9=0.0
00002900 S10=0.0
00003000 DO 20 I = 1,N
00003100 S1= S1 + X1(I)
00003200 S2= S2 + X1(I)**2
00003300 S3= S3 + Y1(I)
00003400 S4= S4 + Y1(I)**2
00003500 S5= S5 + X1(I)*Y1(I)
00003600 S6= S6 + X2(I)
00003700 S7= S7 + X2(I)**2
00003800 S8= S8 + Y2(I)
00003900 S9= S9 + Y2(I)
00004000 S10= S10 + X2(I)*Y2(I)
00004100 20 CONTINUE
00004200 SXX1 = S2 / FLOAT(N)
00004300 SYY1 = S4 / (FLOAT(N)*150.)
00004400 SXY1 = S5 / (FLOAT(N)*SQRT(150.))
00004500 XB1 = S1/FLOAT(N)
00004600 YB1 = S3/FLOAT(N)
00004700 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 2
00004800 SXX2 = S7 / FLOAT(N)
00004900 SYY2 = S9 / (FLOAT(N)*150.)
00005000 SXY2 = S10 / (FLOAT(N)*SQRT(150.))
00005100 XB2 = S6/FLOAT(N)
00005200 C PRINT *,S1,S2,S3,S4,S5,S6,S7,S8,S9,S10,N,M
00005300 C PRINT *,SXX1,SXX2,SYY1,SYY2,SXY1,SXY2,XB1,XB2,YB1
00005400 YB2 = S8/FLOAT(N)
00005500 C TERMINOU CALCULO DE COVARIANCIAS DAS POP. 1 E 2
00005600 C RAZAO DE VEROSSIMILHANCA (B FOI OBTIDO EM OUTRO PROGRAMA)
00005700 B=0.7115
00005800 BET1 = ((SYY1-SXX1)+SQRT((SYY1-SXX1)**2 + 4*(SXY1**2)))/(2*SXY1)
00005900 BET2 = ((SYY2-SXX2)+SQRT((SYY2-SXX2)**2 + 4*(SXY2**2)))/(2*SXY2)
00006000 F1B1= SXX1+2*BET1*SXY1+(BET1**2)*SYY1
00006100 F1B2= SXX2+2*BET2*SXY2+(BET2**2)*SYY2
00006200 F1B = SXX1+2*B*SXY1+(B**2)*SYY1
00006300 F2B = SXX2+2*B*SXY2+(B**2)*SYY2
00006400 G1B1=(BET1**2)*SXX1-2*BET1*SXY1+SYY1
00006500 G1B = (B**2)*SXX1-2*B*SXY1+SYY1
00006600 G2B
00006700 FG10= (G1B*F1B)/((B**2+1)**2)
00006900 FG20= (G2B*F2B)/((B**2+1)**2)
00007000 FG11= G1B1*F1B1/((BET1**2+1)**2)
00007100 FG21= G1B2*F1B2/((BET2**2+1)**2)
00007200 CC1=FG10/FG11

```

```

00007210 CC2=FG20/FG21
00007220 RVER = N*(ALOG(CC2)+ ALOG(CC1))
00007300 C TESTE DE WALD
00007400 F1B1=F1B1/(1+BET1**2)
00007410 F1B2=F1B2/(1+BET2**2)
00007420 G1B1=G1B1/(1+BET1**2)
00007430 G1B2=G1B2/(1+BET2**2)
00007440 BB1=N*((F1B1-G1B1)**2)/(F1B1*G1B1)
00007500 BB1=BB1/((BET1**2+1)**2)
00007600 BB2=N*((F1B2-G1B2)**2)/(F1B2*G1B2)
00007700 O G1B1=G1B1/(1+BET1**2)
00007743
      WALD=((BET1-BET2)**2)/((1/BB1)+(1/BB2))
00007900 PRINT *,N,BET1,BET2,BB1,BB2
00008000 PRINT *,RVER,WALD
00008100 PRINT *,SXX1,SXY1,SY11
00008200 PRINT *,SXX2,SXY2,SY22,B
00008300 STOP
00008400 END
00008500

```

B.0.9 Programa Sifuex

Este programa foi utilizado no exemplo 4.5.

```

      SET AUTOBINDLINEINFO
00000100 BIND=FROM BIB/IMSL/1,BIB/IMSL/2,BIB/IMSL/3,BIB/IMSL/4
00000200 INTEGER N,I
00000300 REAL X1(100),X2(100),Y1(100),Y2(100)
00000400 DOUBLE PRECISION SEED
00000500 READ (5,/) N
00000600 READ (5,/) (X1(I),Y1(I),I=1,N)
00000700 READ (5,/) (X2(I),Y2(I),I=1,N)
00000800 READ (5,/) XB1,YB1,XB2,YB2
00000900 C FBP EH O FATOR DE BAYES A POST. CASO FUNCIONAL RAZ. VAR=150
00001000 C ESTE ARQUIVO CHAMA-SE SIFUEX, COMPILA COM FITA/COMPILA E EXECUTA-SI
00001100 C COM EXCEX. FBQUI E' A APROXIMACAO QUI-QUADRADO PARA O FBP.
00001200 C ESTE PROGRAMA E' UM EXEMPLO COM DADOS DE MADANSKY, 1959
00001300 C NO CASO FUNBOBAL. OS INDICES 1 E 2 SAO PARA AS 2 POPULACOES
00001400 C A HIPOTESE DESTE PROGRAMA EH B1=B2=B (2 POP.), B DESCONHECIDO.
00001500 C OS DADOS ESTAO TRANSFORMADOS PARA QUE A RAZ. VAR SEJA = 1.
00001600 SEED = SEED**0.5
00001700 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 1 E 2
00001800 DO 7 I=1,N
00001900 X1(I)=X1(I)-XB1 X2(I)=X2(I)-XB2
00002100 Y1(I)=Y1(I)-YB1 Y2(I)=Y2(I)-YB2
00002300 7 CONTINUE
00002400 S1=0.0
00002500 S2=0.0
00002600 S3=0.0
00002700 S4=0.0
00002800 S5=0.0
00002900 S6=0.0
00003000 S7=0.0
00003100 S8=0.0
00003200 S9=0.0
00003300 S10=0.0
00003400 DO 20 I = 1,N
00003500 S1= S1 + X1(I)
00003600 S2= S2 + X1(I)**2
00003700 S3= S3 + Y1(I)
00003800 S4= S4 + Y1(I)**2
00003900 S5= S5 + X1(I)*Y1(I)
00004000 S6= S6 + X2(I)

```

```

00004100 S7= S7 + X2(I)**2
00004200 S8= S8 + Y2(I)
00004300 S9= S9 + Y2(I)**2
00004400
20 CONTINUE
00004600 RPX1 = S2 / FLOAT(N)
00004700 SYY1 = S4 / (FLOAT(N)*150.)
00004800 SXY1 = S5 / (FLOAT(N)*SQRT(150.))
00004900 XB1 = S1/FLOAT(N)
00005000 YB1 = S3/FLOAT(N)
00005100 C CALCULO DAS COVARIANCIAS AMOSTRAIS DA POPULACAO 2
00005200 SXX2 = S7 / FLOAT(N)
00005300 SYY2 = S9 / (FLOAT(N)*150.)
00005400 SXY2 = S10 / (FLOAT(N)*SQRT(150.))
00005500 XB2 = S6/FLOAT(N)
00005600 C PRINT *,S1,S2,S3,S4,S5,S6,S7,S8,S9,S10,N,M
00005700 C PRINT *,SXX1,SXX2,SYY1,SYY2,SXY1,SXY2,XB1,XB2,YB1
00005800 YB2 = S8/FLOAT(N)
00005900 C TERMINOU CALCULO DE COVARIANCIAS DAS POP. 1 E 2
00006000 C FATOR DE BAYES POSTERIORI PARA O CASO FUNCIONAL.
00006100 GAM1=((SXX1+SYY1)- SQRT((SXX1-SYY1)**2+4*(SXY1**2)))/2
00006200 GAM2=((SXX2+SYY2)- SQRT((SXX2-SYY2)**2+4*(SXY2**2)))/2
00006300 BET1 = SXY1/(SXX1-GAM1)
00006400 BET2 = SXY2/(SXX2-GAM2)
00006500 SG1 = (SXX1-GAM1)/GAM1
00006600 SG2 = (SXX2-GAM2)/GAM2
00006700 SHH = N*(SG1+SG2)
00006710 SHZ = N*(SG1*BET1+SG2*BET2)
00006720 SZZ = N*(SG1*(BET1**2)+SG2*(BET2**2))
00006730 G = ( (SZZ+SHH)-SQRT((SZZ-SHH)**2+4*(SHZ**2)))/2.
00006740 BBAYS = SHZ/(SHH-G)
00006800 FBP = SQRT(2.0)* EXP((-0.5)*G)
00007200 FBQUI = G - 0.5*ALOG(2.0)
00007300 PRINT *,N,GAM1,GAM2,G,BET1,BET2
00007400 PRINT *, X1,Y1
00007500 PRINT *,X2,Y2,XB1,XB2,YB1,YB2
00007600 PRINT *,G,FBP,FBQUI
00007700 PRINT *,SXX1,SXX2,SXY1,SXY2,SYY1,SYY2
00007800 PRINT *,BBAYS,SHH,SHZ,SZZ
00007810 STOP
00007900 END
00008000

```

B.0.10 Programa Hbsm22

Este programa foi utilizado nas ilustrações numéricas 3.2 e 3.6.

```

?BEGIN JOB HB/EVM;CLASS=71
00000010 ?COMPILE AA FORTRAN LIBRARY
00000020 ?FORTRAN DATA
00000030 SETAUTOBIND
00000035 BIND = FROM BIB/IMSL/1,BIB/IMSL/2,BIB/IMSL/3,BIB/IMSL/4
00000040 REAL XG1(100),Y1(100),XM1(100),E(99),XG2(100),XME11(100)
00000045 REAL Y2(100),XM2(100),XME12(100),XME1(100),XME2(100)
00000057 REAL*8 SEED
00000060 N=50
00000070 C ?FILE FILE6(KIND=DISK,TITLE=HJ1,PROTECTION=PROTECTED)
00000073 NIT=1000
00000075 BET1=1.5
00000080 BET2=1.5
00000085 BET0=1.5
00000087 AMX1=30.0
00000090 AMX2=40.0

```

```

000000092 SEED=1.3
000000095 SIGX1=2.0
000000100 SIGX=2.0
000000102 SIGX2=2.0
000000105 SIGU1=1.0
000000110 SIGU2=1.0
000000115 SIGE1=1.0
000000120 SIG0=1.0
000000122 SIGE2=1.0
000000125 IAV=0
000000126 IA1=0
000000127 IAFB=0
000000128 IA3=0
000000129 IA4=0
000000130 AKX=SIGX1/(SIGX1+SIGU1)
000000133 AKX1=(1-AKX)/AKX
000000135 DO 70 J=1,NIT
000000136 SEED=SEED*J*0.000187
000000137 CALL GGNML(SEED,N+3,E)
000000138 DO 1 I=2,N+1
000000140 1 XM1(I-1)=AMX1+E(I)*(SIGX1**0.5)
000000160 SEED=SEED*1.3
000000170 CALL GGNML(SEED,N+3,E)
000000175 DO 3 I=2,N+1
000000180 3 XG1(I-1)=XM1(I-1)+E(I)*(SIGU1**0.5)
000000185 SEED=SEED*1.5
000000190 CALL GGNML(SEED,N+3,E)
000000195 DO 5 I=2,N+1
000000200 5 Y1(I-1)=0.5+BET1*XM1(I-1)+E(I)*(SIGE1**0.5)
000000205 C PRINT*,XM1,XG1,Y1
000000210 SEED=SEED*1.8
000000212 CALL GGNML(SEED,N+3,E)
000000215 DO 2 I=2,N+1
000000220 2 XM2(I-1)=AMX2+E(I)*(SIGX2**0.5)
000000225 SEED=SEED*1.3
000000230 CALL GGNML(SEED,N+3,E)
000000235 DO 4 I=2,N+1
000000240 4 XG2(I-1)=XM2(I-1)+E(I)*(SIGU2**0.5)
000000245 SEED=SEED*1.5
000000250 CALL GGNML(SEED,N+3,E)
000000255 DO 6 I=2,N+1
000000260 6 Y2(I-1)=3.0+BET2*XM2(I-1)+E(I)*(SIGE2**0.5)
000000265 C PRINT*,XM2,XG2,Y2
000000270 S1X=0.0
000000360 S1Y=0.0
000000362 S2Y=0.0
000000363 SXY=0.0
000000365 S2X=0.0
000000370 DO 7 I=1,N
000000375 S2Y=S2Y+Y1(I)**2
000000382 SXY=SXY+XG1(I)*Y1(I)
000000385 S1Y=S1Y+Y1(I)
000000387 7 S2X=S2X+XG1(I)**2
000000390 XBAR1=S1X/(N*1.0)
000000395 YBAR1=S1Y/(N*1.0)
000000400 C PRINT*,XBAR1,BETE1,S1X,S1Y,S2X,S2Y,SXY,YBAR1
000000410 S21=0.0
000000415 S2Y2=0.0
000000417 S22=0.0
000000420 S23=0.0
000000425 STY2=0.0
000000427 DO 8 I=1,N
000000430 S21=S21+XG2(I)
000000435 S2Y2=S2Y+Y2(I)**2
000000437 S22=S22+XG2(I)*Y2(I)
000000440 STY2=STY2+Y2(I)
000000443 8 S23=S23+XG2(I)**2
000000445 BETE2=S22/S23

```



```

00000455 XBAR2=S21/(N*1.0)
00000460 C PRINT*,BETE2,SIGXE2,XBAR2,S2Y,S21,S22,S23
00000461 S2Y2=S2Y2/(N*1.0)-((STY2/N*1.0)**2)
00000462 S22=S22/(N*1.0)-(XBAR2*(STY2/N*1.0))
00000463 S23=S23/(N*1.0)-(XBAR2**2)
00000464 S2Y=S2Y/(N*1.0)-((S1Y/N*1.0)**2)
00000465 S2X=S2X/(N*1.0)-(XBAR1**2)
00000466 SXY=SXY/(N*1.0)-((S1Y/N*1.0)*XBAR1)
00000467 C PRINT*,BETE2,S1Y,S2Y,S1X,S2X,XBAR1,SXY
00000468 BETE=SXY/(AKX*S2X)
00000470 ANU=AKX*(BETO**2)*S2X-2*BETO*SXY+(AKX1+1)*S2Y
00000480 ADU=AKX*(BETE**2)*S2X-2*BETE*SXY+(AKX1+1)*S2Y
00000490 W=N*ALOG(ANU/ADU)
00000500 AL1=SIGX1
00000510 AL0=(AKX1+1)*SIGE1+AKX1*AL1*(BETO**2)
00000515 AUX=((AL0**2)+(AL1**2)*(BETO**2))/(N*((AL0+AL1*(BETO**2))**2))
00000520 AUX=-AUX+(2.0/N*1.0)+AL1*(BETO**2)/(N*(AL0+AL1*(BETO**2)))
00000530 WST=W*(1-AUX)
00000540 BLO=AKX*(BETE**2)*(S2X)-2*BETE*(SXY)+(AKX1+1)*(S2Y)
00000560 BL1=AKX*(S2X)
00000563 AUX=((BLO**2)+(BL1**2)*(BETO**2))/(N*((BLO+BL1*(BETO**2))**2))
00000566 AUX=-AUX+(2.0/N*1.0)+BL1*(BETO**2)/(N*(BLO+BL1*(BETO**2)))
00000569 WST1=W*(1-AUX)
00000570 C PRINT*,ANU,ADU,W,AUX,WST1,BETE,BLO,BL1,BETO,AKX,AKX1
00000573 SW=SW+W
00000575 STW=STW+WST
00000580 STW1=STW1+WST1
00000585 IF(W.LT. 2.706) GO TO 30
00000600 IW=IW+1
00000610 30 IF(WST.LT. 2.706) GO TO 31
00000620 IWST=IWST+1
00000630 31 IF(WST1.LT. 2.706) GO TO 70
00000640 IWST1=IWST1+1
00000650 70 AUX=0.0
00000800 SW=SW/NIT
00000805 STW=STW/NIT
00000810 STW1=STW1/NIT
00000815 PRINT*,SW,STW,STW1,IW,IWST,IWST1
00000820 STOP
00001100 END
00001110 ?RUN AA
00001200 ?FILE FILE6(KIND=DISK,TITLE=HJ1,PROTECTION=PROTECTED)
00001205 ?DATA
00001210 ?END JOB

```

B.0.11 Programa Hbsm60

Este programa foi utilizado nas ilustrações numéricas 5.2 e 5.3.

```

?BEGIN JOB HB/EVM;CLASS=72
00000010 ?COMPILE AA FORTRAN LIBRARY
00000020 ?FORTRAN DATA
00000030 SETAUTOBIND
00000035 BIND = FROM BIB/IMSL/1,BIB/IMSL/2,BIB/IMSL/3,BIB/IMSL/4
00000040 REAL S1X(100),S2X(100),S1Y(100),S2Y(100),XME2(100)
00000050 REAL X(100,100),Y(100,100),XM(500),E(100),XME(100)
00000055 REAL*8 SEED
00000060 N=20
00000070 R=40
00000075 BET=1.5
00000080 ALF=40.0
00000085 C ?FILE FILE6(KIND=DISK,TITLE=HBF1,PROTECTION=PROTECTED)
00000087 AMX=10.0
00000090 SEED=0.9

```

```

00000095 SIGX=4.0
00000100 SIGU=0.1
00000110 SIGE=4.0
00000120 CALL GGNML(SEED,N+3,E)
00000130 DO 1 I=2,N+1
00000140 1 XM(I-1)=AMX+E(I)*(SIGX**0.5)
00000160 DO 3 I=1,N
00000180 SEED=SEED*1.3
00000182 CALL GGNML(SEED,R+3,E)
00000183 DO 5 J=1,R
00000184 5 X(I,J)=XM(I)+E(J+1)*(SIGU**0.5)
00000185 3 CONTINUE
00000187 DO 7 I=1,N
00000200 SEED=SEED*1.5
00000202 CALL GGNML(SEED,R+3,E)
00000203 DO 9 J=1,R
00000204 9 Y(I,J)=ALF+BET*XM(I)+E(J+1)*(SIGE**0.5)
00000205 7 CONTINUE
00000207 PRINT*/,X(1,1),Y(2,2)
00000210 DO 11 I=1,N
00000310 S2X(I)=0.0
00000354 S1Y(I)=0.0
00000356 11 S2Y(I)=0.0
00000358 DO 13 I=1,N
00000375 DO 15 J=1,R SX(I)=S2X(I)+X(I,J)**2
00000382 S1Y(I)=S1Y(I)+Y(I,J)
00000384 15 S2Y(I)=S2Y(I)+Y(I,J)**2
00000386 13 CONTINUE
00000390 TX=0.0
00000392 TX2=0.0
00000394 TY=0.0
00000396 TY2=0.0
00000398 DO 19 I=1,N
00000407 TX=TX+S1X(I)
00000409 TX2=TX2+S2X(I)
00000410 TY=TY+S1Y(I)
00000415 19 TY2=TY2+S2Y(I)
00000420 WRITE(6,/) (S1X(I),I=1,N)
00000435 WRITE(6,/) (S1Y(I),I=1,N)
00000436 PRINT*/,TX,TX2,TY,TY2
00000437 DO 18 I=1,N
00000440 18 XME(I)=S1X(I)/N
00000445 DO 17 I=1,N
00000450 17 XME2(I)=(3/2)*(PMQ(@)*@2)
00000455 YBAR=TY/(N*R)
00000495 SYX=TY2-N*R*(YBAR**2)
00000500 DO 20 I1=1,30
00000515 TXE=0.0
00000520 TXE2=0.0
00000525 SXX=0.0
00000530 SXY=0.0
00000533 C PRINT*/,BETE,SIGXE,XBARE
00000535 DO 23 I=1,N
00000540 TXE=TXE+XME(I)
00000545 TXE2=TXE2+XME2(I)
00000550 SXX=SXX+S1X(I)*XME(I)
00000555 23 SXY=SXY+S1Y(I)*XME(I)
00000560 XBARE=TXE/N
00000565 SYXM=0.0
00000570 DO 25 I=1,N
00000575 25 SYXM=SYXM+S1Y(I)*(XME(I)-XBARE)
00000580 SXE2=TXE2-N*(XBARE**2)
00000585 BETE=(SXY-N*R*XBARE*YBAR)/(R*SXE2)
00000590 ALFE=YBAR-BETE*XBARE
00000595 SIGXE=SXE2/N

```

```

00000600 SIGUE=(TX2-2*SXX+R*TXE2)/(N*R)
00000605 SIGEE=(SYY-2*BETE*SXYM+R*(BETE**2)*SXE2)/(N*R)
00000610 DELT1=R*SIGXE*SIGEE+R*(BETE**2)*SIGUE*SIGXE+SIGUE*SIGEE
00000795 C PRINT*,TXE2,BETE,ALFE,N,R,SIGXE,SIGEE,SIGUE
00000797 DO 32 I=1,N
00000800 XME(I)=SIGXE*(SIGUE*BETE*(S1Y(I)-R*ALFE-R*BETE*XBARE))/DELT1
00000810 XME(I)=XBARE+XME(I)+SIGXE*(SIGEE*(S1X(I)-R*XBARE))/DELT1
00000820 32 XME2(I)=(SIGXE*SIGUE*SIGEE/DELT1)+(XME(I)**2)
00000825 SH1=0.0
00000835 SC1=0.0
00000837 DO 51 I=1,N
00000840 DO 53 J=1,R
00000845 SH1=SH1+(1/SIGUE)*((X(I,J)-XBARE)**2)
00000850 SH1=SH1+(1/SIGEE)*((Y(I,J)-ALFE-BETE*XBARE)**2)
00000855 AUX1=(1/SIGUE)*(X(I,J)-XBARE)
00000860 AUX1=AUX1+BETE*(1/SIGEE)*(Y(I,J)-ALFE-BETE*XBARE)
00000865 53 SC1=(AUX1**2)
00000870 51 CONTINUE
00000875 A=R*(1/SIGUE)+R*(BETE**2)*(1/SIGUE)+(1/SIGXE)
00000880 VL=-(N/2)*ALOG(SIGXE)-(R*N/2)*ALOG(SIGUE)-(N*N/2)*ALOG(SIGEE)
00000885 VL=VL-(1/2)*SH1+(1/(2*A))*SC1
00000890 PRINT*,VL,BETE,ALFE,SIGXE,SIGEE,SIGUE,XBARE
00000900 20 CONTINUE
00000930 STOP
00001100 END
00001110 ?RUN AA
00001200 ?DATA
00001210 ?END JOB

```

Aqui termina a apresentação de alguns programas computacionais que foram utilizados nas ilustrações numéricas e alguns exemplos dos Capítulos 3, 4 e 5.

Bibliografia

- [1] Achcar, J. A. (1993). Inferences for the Birnbaum-Saunders fatigue life model using Bayesian methods. *Comput. Stat. of Data Anal.* **84** 710-716.
- [2] ADCOCK, R. J. (1877). Note on the method of least squares. *The Analyst* **4**, 183-184.
- [3] ADCOCK, R. J. (1878). A problem in least squares. *The Analyst* **5**, 53-54.
- [4] AKAIKE, H. (1973). Information and the extension of the maximum likelihood principle In Proc. 2end. *Int. Symp. Information Theory* (eds. B.N. Petior and F. Csaki) pp 267-281. *Budapesti Akademiai Kiado.*
- [5] ALLEN, R. G. D. (1939). The assumptions of linear regression. *Econometrika* **6**, 199-204.
- [6] ANDERSON, T. W. (1976). Estimating of linear functional relationships: Approximate distributions and conections with simultaneous equations in econometrics. *J. Roy. Statist. Soc. Ser. B* 1-36.
- [7] ANDERSON, T. W. (1984). An introduction to Multivariate Analysis. *John Wiley - New York.*
- [8] ANDERSON, T. W. (1984). Estimating linear statistical relationships. *Ann. Statist.* **12**, 1-45.
- [9] ANDERSON, T. W. and SAWA, T. (1973). Distributions of estimates of coefficients of a single equation in a simultaneous system and their assymptotic expansions. *Econometrika* **41**, 683-714.
- [10] AITKEN, M. (1991). Posterior Bayes Factors. *J. R. Statist. Soc. B* **53**, 11-142.
- [11] ARMSTRONG, B. (1985). Measurement error in the generalized linear model. *Comm. Statist. B* **14**, 529-544.
- [12] BANSAL, N. K.(1990). Comparison of the Slope Parameters in Errors-in-Variables Model. *Commun. Statist. - Theory Meth.*, **19**(7), 2705-2716.
- [13] BARTLETT, M. S. (1949). Fitting a straight line when both variables are subject to error. *Biometrics* **5**, 207-212.
- [14] BARTLETT, M. S.(1955). Approximate confidence intervals. *Biometrika* **42**, 201.

- [15] BARNETT, V. D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics* **25**, 129-142.
- [16] BERKSON, J. (1950). Are there two regressions? *J. Amer. Statist. Assoc.* **45**, 164-180.
- [17] BICKEL, P. J. and RITOV, Y. (1987). Efficient estimation in the errors-in-variables model. *Ann. Statist.* **13**, 515-540.
- [18] BOGGS, P. T. et al (1988). A computational examination of orthogonal distance regression. *J. Econometrics* **38**, 169-201.
- [19] BOLFARINE, H. and CORDANI, L. K. (1990). Estimation of a structural regression model with known reliability ratio. *Technical Report RT-MAE-9021*, 12p IME-USP.
- [20] BOLFARINE, H., RODRIGUES, J. e CORDANI, L. K. (1992). O modelo de regressão com erros nas variáveis. *Texto de minicurso do X SINAPE*.
- [21] BROWN, M. L. (1982). Robust line estimation with errors in both variables. *J. Am. Statist.* **77**, 71-79.
- [22] BROWN, R. L. (1957). The bivariate structural relation. *Biometrika* **44**, 84-96.
- [23] BROWN, R. L. and Fereday, F. (1958). Multivariate linear structural relation. *Biometrika* **45**, 136-153.
- [24] BUSE, A. (1982). The Likelihood ratio, Wald, and Lagrange Multiplier tests: an expository note. *The American Statistician* **36**, 153-157.
- [25] CARROLL, R. J. and GALLO, P. (1982). Some aspects of robustness in functional errors in variables regression models. *Comm. Statist. A* **11**, 2573-2585.
- [26] CHAN, L.K. and MAK, T.K. (1979). Maximum likelihood estimation of a linear structural relationship with replication. *J. Roy. Statist. Soc. Ser. B* **41**, 263-268.
- [27] COCHRAN, W. G. (1968). Errors of measurement in statistics. *Technometrics* **10**, 637-666.
- [28] CORDEIRO, G. (1992). Introdução à teoria de verossimilhança. *Texto de minicurso do X SINAPE*, 1-173.
- [29] CORDEIRO, G.M. and FERRARI, S. L. P. (1991). A modified score statistic having chi-squared distribution to order n^{-1} . *Biometrika* **78** 573-82.
- [30] COX, D. R. and REID, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B* **49** 1-39.
- [31] CREASY, M. A. (1957). Confidence limits for the gradient in the linear structural relationship. *J. Roy. Statist. Soc. Ser. B* **18**, 65-69.
- [32] DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.

- [33] DOLBY, G. R. (1976). The ultrastructural relation: a synthesis of the functional and structural relations. *Biometrika* **63**, 39-50.
- [34] DRION, E. F. (1951). Estimation of the parameters of a straight line and of the variances of the variables, if they are both subject to error. *Indag. Math.* **13**, 256-260.
- [35] EHRENBERG, A. S. C. (1982). A primer in data reduction. *John Wiley, Chichester*.
- [36] FLORENS, J. P., MOUCHART, M. and RICHARD, J. F. (1974). Bayesian inference in error-in-variable models. *Journal of Multivariate Analysis*. **4**, 419-452.
- [37] FULLER, W. (1987). Measurement error models. *John Wiley*.
- [38] GEARY, R. C. (1949). Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown. *Econometrica* **17**, 30-58.
- [39] GLESER, L. J. (1985). A note on G.R. Dolby's Ultrastructural model. *Biometrika* **72**, 117-124.
- [40] GLESER, L. J. and HWANG, J. T. (1987). The nonexistence of $100(1 - \alpha)\%$ confidence sets of finite expected diameter in errors-in-variables and related models. *Ann. Statist.* **15**, 1351-1362.
- [41] GLESER, L. J. and WATSON, G. S. (1973). Estimation of linear transformation. *Biometrika* **60**, 525-534.
- [42] HOOPER, J. W. and THEIL, H. (1958). The extension of Wald's method of fitting straight lines to multiple regression. *Revue Inst. Int. Stat.* **26**, 37-47.
- [43] KASS, R. E., Tierney, L. and Kadane, J.B. (1990). The validity of posterior expansions based Laplace's method, in: Essays in honor of George A. Barnard, J. Hodges (Ed. North-Holland, Amsterdam).
- [44] KASS, R. E., Tierney, L. and Kadane, J.B. (1990). Laplace method in bayesian analysis. *Unpublished report*.
- [45] KENDALL, M. G. (1951). Regression. Structure and functional relationship I. *Biometrika* **38**, 11-25.
- [46] KENDALL, M. G. (1952). Regression. Structure and functional relationship II. *Biometrika* **39**, 96-108.
- [47] KENDALL, M. G. and Stuart, A. (1961). The advanced theory of statistics Vol.II. *Griffin, London*.
- [48] KIEFER, J. (1964). Review of Kendall and Stuart's Advanced theory of statistics Vol.II. *Ann. Math. Statist.* **35**, 1371-1380.
- [49] KIEFER, J. and WOLFOWITZ A. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27**, 887-906.

- [50] KUMMEL, C. H. (1879). Reduction of observed equations which contain more than one observed quantity. *The Analyst* **6**, 97-105.
- [51] LAWLEY, D. N. (1955). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **14**, 112-142.
- [52] LINDLEY, D. V. (1947). Regression lines and the linear functional relationship. *Suppl. J. Roy. Statist. Soc.* **9**, 218-244.
- [53] LINDLEY, D.V. (1953). Estimation of a functional relationship. *Biometrika* **40**, 47-49.
- [54] LINDLEY, D.V. (1957). A statistical paradox. *Biometrika* **44**, 187-192.
- [55] LINDLEY, D. V. (1966). Discussion on a generalized least squares approach to linear functional relationship. *J. Roy. Statist. Soc. Ser. B* **28**, 279-297.
- [56] LINDLEY, D. V. and EL-SAYAD, G. M. (1968). The bayesian estimation of linear functional relationship. *J. Roy. Statist. Soc. Ser. B* **30**, 190-202.
- [57] LOUIS, T. A. (1982). Finding the observed information matrix when using EM algorithm. *J. Roy. Statist. Soc. Ser. B* **44**, 226-233.
- [58] MADANSKY, A. (1959). The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Assoc.* **54**, 173-205.
- [59] MURRAY, A. (1991). Posterior bayes factors. *J. Roy. Statist. Soc. Ser. B* **53**, 111-142.
- [60] MORAN, P. A. P. (1971). Estimating functional and structural relationships. *J. Multivariate Anal.* **1**, 232-255.
- [61] MORTON, R. (1981). Efficiency of estimating equations and the use of pivots. *Biometrika* **68**, 227-233.
- [62] NASCIMENTO, J. A. (1983). Métodos de estimação consistente não bayesiano para o modelo com erro nas variáveis. *Tese de Mestrado CCEN-DE-UFPE*.
- [63] NEYMAN, J. (1951). Existence of consistence estimates of the directional parameter in a linear structural relation between two variables. *Ann. Math. Statist.* **22**, 497-512.
- [64] NEYMAN, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika* **16**, 1-32.
- [65] NEYMAN, J. and Scott, E. L. (1951). On certain methods of estimating the linear structural relation. *Ann. Math. Statist.* **22**, 352-361. Correction, *Ann. Math. Statist.*(1952) **23**, 115.
- [66] PATEFIELD, W. M. (1978). The unreplicated ultra-estructural relation: large sample properties. *Biometrika* **65**, 535-540.
- [67] PATINO-LEAL, H. (1979). Advances in theory of the error in variables model with applications in chemical engineering. *Ph.D Thesis, University of Waterloo*.

- [68] REILLY, P. M. and PATINO-LEAL (1981). A bayesian study of the error-in-variables model. *Technometrics* **23**, 221-231.
- [69] PEARSON, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2**, 559-572.
- [70] REIERSOL, O. (1945). Confluence analysis by means of instrumental sets of variables. *Ark. Mat. Astronomique Fisycs* **32**, 1-119.
- [71] REIERSOL, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica* **18**, 375-389.
- [72] REILLY, P. and PATINO-LEAL, H. (1981). A Bayesian study of the error in variables models. *Technometrics* **23**, 221-231.
- [73] RODRIGUES, J. (1990). A bayesian analysis of the generalized least squares procedure for functional relationships Statistics. (*no prelo*).
- [74] RODRIGUES, J. and CORDANI, L. K. (1990). A note on likelihood estimation of a simple regression model with measurement error, via the orthogonal parametrization. *South African Stat. J.* **24**, 177-183.
- [75] ROOS, C.E. (1937). A general invariant criterion of fit for lines and planes where all variates are subject to error. *Metron* **13**, 3-20.
- [76] SARGAN, J. D. (1959). The estimation of relationships with autocorrelated residuals by use of instrumental variables. *J. Roy. Statist. Soc. Ser. B* **21**, 91-105.
- [77] SMITH, A. M. F. and SPIEGELHALTER, D. J. (1980). Bayes factors and choice criteria for linear models. *J. Roy. Statist. Soc. Ser. B* **42**, 190-202.
- [78] SOLARI, M. E. (1969). The maximum likelihood solution of the problem of estimating a linear functional relationship. *J. Roy. Statist. Soc. Ser. B* **31**, 372-375.
- [79] SPRENT, P. (1966). A generalized least squares approach to linear functional relationships (with discussion). *J. Roy. Statist. Soc. Ser. B* **28**, 278-297.
- [80] SPRENT, P. (1990). Some historical of functional and structural relationships. *Contemporary Mathematics* **112**, 3-15.
- [81] STEFANSKI, L. A. and CARROL, R. J. (1985). Covariate measurement error in logistic regression. *Ann. Statist.* **12**, 1335-1351.
- [82] STEFANSKI, L. A. and CARROL, R. J. (1985). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703-716.
- [83] TIERNEY, L. and KADANE, J. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Stat. Assoc.* **81**, 82-86.
- [84] TIERNEY, L. and KADANE, J. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Stat. Assoc.* **84**, 710-716.

- [85] THEIL, H. (1950). A rank invariant method of linear and polynomial regression analysis I,II,III. *Indag. Math.* , 85-91, 173-177, 467-482.
- [86] THEIL, H. and VAN YZERAN. J. (1956). On the efficiency of Wald's method of straight lines. *Rev. Inst. Internat. Statist.* **24**, 17-26.
- [87] THOMAS, E.V. (1991). Errors-in-variables estimation in multivariate calibration. *Technometrics* **33**, 405-413.
- [88] WALD, A. (1940). The fitting of straight lines if both variables are subject to error. *Ann. Math. Statist.* **11**, 284-300.
- [89] WELLMAN, J. M. and GUNST, R. F. (1991). Influence diagnostics for linear measurement error models. *Biometrika* **78**, 373-380.
- [90] WILLIAMS, E. J. (1955). Significance tests for discriminant functions and linear functional relationships. *Biometrika* **42**, 360-381.
- [91] WHITTEMORE, A. S. (1989). Errors-in-variables regression using Stein estimates. *Amer. Statist.* **43**, 226-228.
- [92] WONG, M. Y. (1989). Likelihood estimation of a simple regression model when both variables have error. *Biometrika* **76**, 141-148.
- [93] WONG, M. Y. (1991). Bartlett adjustment to the likelihood ratio statistic for testing several slopes. *Biometrika* **78**, 221-224.
- [94] ZELLNER, A. (1971). An introduction to Bayesian inference in econometrics. *John Wiley. New York*.
- [95] ZELLNER, A. (1977). Maximal data information prior distributions, in A. Aykac and C. Brumat, eds., New developments in the applications of Bayesian meth, *Amsterdam: North-Holland publishing co.*, 211-232 and reprint in A. Zellner (1984).
- [96] ZELLNER, A (1990). Bayesian methods and entropy in Economics and Econometrics, in W.T. Grandy, Jr. and L.H. Schick (eds.). Maximum entropy and Bayesian methods. *Dordrecht, Netherlands: Kluwer Academics Publishers*, 17-31.