CALIBRAÇÃO COMPARATIVA
ESTRUTURAL E FUNCIONAL

Manuel Jesús Galea Rojas

TESE APRESENTADA
AO
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

- São Paulo, Outubro de 1995 -

CALIBRAÇÃO COMPARATIVA ESTRUTURAL E FUNCIONAL

Este exemplar corresponde à redação final
da tese devidamente corrigida e defendida
por Manuel Jesús Galea Rojas e aprovada
pela Comissão Julgadora.

São Paulo, 20 de outubro de 1995

Comissão Julgadora:

- Prof. Dr. Heleno Bolfarine (Orientador) - IME-USP
- Prof. Dr. Gilberto Alvarenga Paula - IME-USP
- Prof. Dr. Jorge Alberto Achcar - ICMSC-USP
- Prof. Dr. Renato Assunção - UFMG
- Profa. Dra. Clarice Garcia Borges Demétrio - ESALQ-USP

RESUMO

Neste trabalho estudamos uma sub-classe de modelos com erros nas variáveis, frequentemente usados para comparar instrumentos ou métodos de medição. Este problema é conhecido na literatura estatística como Calibração Comparativa.

Discutimos os modelos estrutural e funcional normais, com algumas extensões e os modelos estruturais t -elíptico e elíptico.

Usamos o algoritmo EM para obter os estimadores de máxima verossimilhança. Para cada modelo obtemos também, a matriz de covariância assintótica destes estimadores. Testes assintóticos são propostos usando principalmente a estatística de Wald.

Apresentamos alguns estudos de simulação e ilustramos a metodologia proposta com vários conjuntos de dados apresentados na literatura.

ABSTRACT

The aim of this thesis is to study an special sub-class of the more general models with measurement errors, specially designed to comparing measuring devices. This problem is known in the literature as Comparative Calibration.

Structural and functional normal models are considered, with extension to elliptical t and general elliptical models.

The EM algorithm is used to obtain the maximum likelihood estimators. Large sample testing statistics based on the Wald statistics are proposed.

Simulations studies are reported and we illustrate, based on published data, the proposed methodology.

AGRADECIMENTOS

Eu gostaria de agradecer ao Prof. Dr. Heleno Bolfarine pela confiança em mim depositada e especialmente pelo constante apoio durante o desenvolvimento deste trabalho.

A meus colegas de turma: Antonio, Claudia, Filidor, Miguel, Manoel e Oscar, pela amizade, companheirismo e boa disposição que sempre tiveram comigo.

A meus compatriotas Pilar e Reinaldo pelo permanente apoio, em especial durante meu primeiro ano de estadia no Brasil.

À minha família, em especial à minha esposa Patricia e meus filhos Rodrigo e Felipe, pelo carinho, compreensão e incentivo permanente.

Ao Governo do Chile, pelo apoio financeiro via "Beca Presidente de la República".

Às Universidades de Valparaíso e de São Paulo, a primeira me ofereceu a oportunidade de continuar minha formação acadêmica e a segunda me acolheu como mais um de seus alunos.

Ademais, gostaria de agradecer a todos aqueles, que colaboraram de uma forma ou de outra para a realização deste trabalho.

A meus pais Enrique e Lorenza

CONTEÚDO

CAPÍTULO 1: INTRODUÇÃO	1
1.1 O problema de calibração comparativa	1
1.2 O modelo estrutural	2
1.3 Definição de confiabilidade	10
1.4 Objetivos do trabalho	13
1.5 Organização do trabalho	14
CAPÍTULO 2: O MODELO ESTRUTURAL NORMAL	16
2.1 Introdução	16
2.2 O caso de dois instrumentos	18
2.3 O caso de três ou mais instrumentos	25
2.3.1 Estimação via o algoritmo EM	25
2.3.2 Distribuição assintótica dos EMV	30
2.3.3 Região de confiança assintótica para ρ	39
2.3.4 Testes de hipóteses	41
2.4 O modelo estrutural quando o instrumento de referência mede sem erro a característica de interesse	44
2.5 Ilustração numérica	49
CAPÍTULO 3: EXTENSÕES DO MODELO ESTRUTURAL NORMAL	56
3.1 Introdução	56
3.2 O caso de dois instrumentos	56
3.2.1 Os modelos	57
(i) O modelo estrutural repetido	57
(ii) O modelo ultraestrutural repetido	59
3.2.2 Estimação de máxima verossimilhança via algoritmo EM	60
3.2.3 A matriz de covariância assintótica dos EMV	65
3.2.4 Testes de hipóteses	68
3.3 O modelo estrutural normal com uma covariável	69
3.3.1 O modelo	70
3.3.2 Estimação via algoritmo EM	71
3.3.3 A matriz de informação	75
3.3.4 Ilustração numérica	79
3.3.5 Um estudo de simulação	81

CAPÍTULO 4: O MODELO ESTRUTURAL t -ELÍPTICO	83
4.1 Introdução	83
4.2 O caso de dois instrumentos	86
4.2.1 A matriz de informação no modelo estrutural elíptico	86
4.2.2 Modelo estrutural t -elíptico	90
(i) Estimação via algoritmo EM	90
(ii) Matriz de covariância assintótica do EMV	93
(iii) Testes de hipóteses	95
4.3 O caso de três ou mais instrumentos	97
4.3.1 Modelo estrutural t -elíptico	97
(i) Estimação usando o algoritmo EM	98
(ii) A matriz de informação esperada	101
4.3.2 Modelo estrutural elíptico	104
4.3.3 Ilustração numérica	107
CAPÍTULO 5: MODELO ELÍPTICO DE GRUBBS	112
5.1 Introdução	112
5.2 Estimação de máxima verossimilhança	114
5.2.1 Estimação no modelo normal	114
5.2.2 Estimação no modelo t -elíptico	116
5.3 A matriz de informação esperada	118
5.3.1 Modelo normal	120
5.3.2 Modelo t -elíptico	120
5.4 Testes de hipóteses	122
CAPÍTULO 6: UM MODELO FUNCIONAL NORMAL	126
6.1 Introdução	126
6.2 Estimação em presença de parâmetros incidentais	128
6.3 Estimação de máxima verossimilhança	131
6.3.1 Obtenção dos EMV	131
6.3.2 Propriedades assintóticas do EMV	133
6.3.3 A matriz de informação esperada	138
6.4 Estimadores de momentos	140
6.5 Teste da hipótese de igualdade das linhas de calibração	144
6.6 Extensão do modelo funcional	145
6.7 Ilustração numérica	146
6.8 Um estudo de simulação	148
APÊNDICE A: DISTRIBUIÇÕES ELÍPTICAS SIMÉTRICAS	150
APÊNDICE B: A DISTRIBUIÇÃO t MULTIVARIADA	159

APÊNDICE C: OBTENÇÃO DA MATRIZ DE INFORMAÇÃO NO MODELO NORMAL	164
APÊNDICE D: OBTENÇÃO DA MATRIZ DE INFORMAÇÃO NO MODELO ESTRUTURAL t -ELÍPTICO	171
REFERÊNCIAS	179

Capítulo 1

Introdução

1.1 O problema de calibração comparativa

Este trabalho trata da comparação de instrumentos ou métodos de medição, quando cada um deles é usado para medir uma característica num grupo comum de unidades experimentais. Este problema é conhecido, na literatura estatística como calibração comparativa. O interesse aqui é avaliar a qualidade de diferentes instrumentos de medição.

A necessidade de comparar instrumentos de medição que diferem em custo, velocidade ou outros fatores, tem aparecido com frequência em diversas áreas do conhecimento. Grubbs (1948, 1973), por exemplo, compara três cronômetros, Barnett (1969) dá um exemplo onde quatro combinações instrumento-operador concebidos para medir a capacidade vital num grupo de pacientes são avaliados. Leurgans (1980) compara dois métodos para medir a concentração da glicose no sangue. Vários outros exemplos, na área médica, são mencionados por Kelly (1984, 1985) e Kaaks et al. (1994). Já Jaech (1985) dá vários exemplos na área industrial, Fuller (1987) apresenta exemplos na área da Agricultura. Dunn (1992) dá aplicações em Psicologia e Educação.

Quando o processo de medição não é destrutivo todos os instrumentos podem ser usados em cada unidade experimental. Se o processo é destrutivo, assume-se que cada unidade (experimental) prove partes homogêneas, as quais são atribuídas ao acaso a cada um dos métodos de medição.

O termo “método” ou “instrumento” (de medição), refere-se a diferentes formas de medição de uma característica dada. O termo também pode ser aplicado quando a mesma forma de medição é feita por diferentes aparelhos ou diferentes operadores. Por exemplo, quando uma análise química é realizada em diferentes laboratórios, cada laboratório é identificado como um método ou instrumento de medição. Em alguns casos pode acontecer, também, que os diferentes métodos sejam um mesmo instrumento avaliado em diferentes tempos. Na prática a definição do método de medição terá influência sobre a construção do modelo estatístico.

Este trabalho trata, então, da estimação, sob o ponto de vista clássico, dos parâmetros do modelo estatístico. Testes de hipóteses sobre parâmetros do modelo são também considerados.

1.2 O modelo estrutural

Suponha que dispomos de p (≥ 2) instrumentos para medir uma característica ou resposta x num grupo de n unidades. Sejam

x_j : O verdadeiro valor da característica de interesse na unidade j ;

y_{ij} : Medida fornecida pelo instrumento i para a característica da unidade j , $i = 1, 2, \dots, p$
e $j = 1, 2, \dots, n$.

Um modelo frequentemente adotado na literatura; ver Jaech (1964); Cochran (1968); Barnett (1969) e Williams (1969), para este tipo de situação é

$$y_{ij} = \alpha_i + \beta_i x_j + \varepsilon_{ij}, \quad i = 1, 2, \dots, p, \quad j = 1, 2, \dots, n, \quad (1.1)$$

onde os erros aleatórios de medição ε_{ij} são não correlacionados e para cada i , $\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in}$ são tais que

$$E(\varepsilon_{ij}) = 0, \quad \text{Var}(\varepsilon_{ij}) = \sigma_i^2 \quad (1.2)$$

e

α_i : vício aditivo do instrumento i ;

β_i : vício multiplicativo ou relativo do instrumento i , $i = 1, 2, \dots, p$.

Este modelo é reconhecido como um caso especial de um modelo de regressão linear com erros nas variáveis (Kendall e Stuart, 1979 e Fuller, 1987). Na literatura sobre modelos com erros nas variáveis, consideram-se duas formas de modelagem das quantidades x_j :

- (i) As x_j 's são constantes desconhecidas, parâmetros;
- (ii) As x_j 's são uma amostra aleatória de uma população com média μ_x e variância σ_x^2 e são não correlacionadas com os ε_{ij} .

O modelo descrito por (i) e (1.1) é um modelo linear funcional com erros nas variáveis ou simplesmente modelo funcional. O modelo descrito por (ii) e (1.1) é denominado modelo linear estrutural com erros nas variáveis ou modelo estrutural. Já que a suposição (ii) é comumente adotada na literatura, na maior parte deste trabalho assumimos um modelo estrutural. Só no final da tese discutimos um modelo funcional. A seguir, então, estudamos algumas características do modelo estrutural.

Em notação matricial o modelo pode ser escrito como:

$$Y_j = \tilde{\alpha} + \tilde{\beta}x_j + \varepsilon_j = \tilde{\alpha} + Bu_j, \quad j = 1, 2, \dots, n, \quad (1.3)$$

onde $\tilde{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_p)'$, $\tilde{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ são vetores $p \times 1$, $B = [\tilde{\beta}, I_p]$ é uma matriz $p \times (p+1)$, $Y_j = (y_{1j}, y_{2j}, \dots, y_{pj})'$, $\varepsilon_j = (\varepsilon_{1j}, \varepsilon_{2j}, \dots, \varepsilon_{pj})'$ são vetores $p \times 1$, $u_j = (x_j, \varepsilon_j)'$ são vetores $(p+1) \times 1$ e I_p denota a matriz identidade de ordem p .

Note que

$$E(u_j) = \begin{pmatrix} \mu_x \\ O_p \end{pmatrix} \quad \text{e} \quad \text{Var}(u_j) = \begin{pmatrix} \sigma_x^2 & O_p' \\ O_p & D(\phi) \end{pmatrix}, \quad (1.4)$$

onde $D(\phi) = \text{Diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$, $\phi = (\phi_1, \dots, \phi_p)' = (\sigma_1^2, \dots, \sigma_p^2)'$, O_p é um vetor nulo $p \times 1$ e A' indica a transposta de A .

De (1.3) e (1.4) temos então:

$$E(Y_j) = \tilde{\alpha} + \tilde{\beta}\mu_x = \tilde{\mu} = \tilde{\mu}(\tilde{\theta}) \quad \text{e} \quad \text{Var}(Y_j) = D(\phi) + \sigma_x^2 \tilde{\beta}\tilde{\beta}' = \tilde{V} = \tilde{V}(\tilde{\theta}), \quad (1.5)$$

com $\tilde{\theta} = (\mu_x, \tilde{\alpha}', \tilde{\beta}', \sigma_x^2, \phi')'$.

A seguir provamos que $\tilde{\theta}$ não é identificável.

Definição 1.1 *Seja Y um vetor aleatório observável com distribuição $F(y, \theta)$; $\theta \in \Theta$ e $y \in \mathcal{Y}$.*

(i) *Diz-se que o ponto $\theta_* \in \Theta$ é identificável por $F(y, \theta_*)$ se não existe $\theta \in \Theta$ tal que*

$$F(y, \theta_*) = F(y, \theta) \quad \text{para algum } y \in \mathcal{Y}. \quad (1.6)$$

(ii) *O espaço paramétrico, Θ ou o modelo F , diz-se identificável se cada $\theta \in \Theta$ é identificável. Neste caso diremos que θ é identificável. Caso contrário, diremos que θ não é identificável.*

Note que o parâmetro θ não será identificável se existem $\theta_* \neq \theta_{**} \in \Theta$ tais que

$$F(y; \theta_*) = F(y; \theta_{**}), \quad \forall y \in \mathcal{Y}. \quad (1.7)$$

Lema 1.1 *Se no modelo estrutural (1.3) u_j segue uma distribuição elíptica com média e variância dadas por (1.4), então $\tilde{\theta}$ não é identificável.*

Prova: Como u_j segue uma distribuição elíptica, temos que (ver Apêndice A) Y_j tem uma distribuição elíptica com média e variância dadas por (1.5).

Seja $\bar{\theta} = (\bar{\mu}_x, \bar{\alpha}', \bar{\beta}', \bar{\sigma}_x^2, \bar{\phi}')'$, onde $\bar{\mu}_x = (\mu_x + d)/c$, $\bar{\alpha} = \tilde{\alpha} - d\tilde{\beta}$, $\bar{\beta} = c\tilde{\beta}$, $\bar{\sigma}_{\frac{\mu}{\chi}}^2 = \sigma_x^2/c$, $\bar{\phi} = \phi$ e $c, d \neq 0 \in \mathbb{R}$. Logo, é fácil ver que

$$E_{\bar{\theta}}(Y_j) = E_{\tilde{\theta}}(Y_j) = \tilde{\mu},$$

$$\text{Var}_{\bar{\theta}}(Y_j) = \text{Var}_{\tilde{\theta}}(Y_j) = \tilde{V}.$$

Ou seja, existem $\theta_* \neq \theta_{**} \in \Theta$ tais que

$$F(y; \theta_*) = \overline{f}(y; \theta_{**}) , \quad \forall y \in \mathcal{Y},$$

onde, neste caso, F é uma distribuição elíptica com média $\tilde{\mu}$ e matriz de covariância $\tilde{V}$.

Exemplo 1.1 Considere o modelo (1.3) com $p = 3$. Sejam

$$\tilde{\theta}_* = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)'$$

$$\tilde{\theta}_{**} = (2, -2, -2, -2, 2, 2, 2, 0.25, 1, 1)'.$$

De (1.5) é fácil ver que ambos valores de $\tilde{\theta}$ geram a mesma média e a mesma matriz de covariâncias de Y_j :

$$E(Y_j) = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} , \quad \text{Var}(Y_j) = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} , \quad j = 1, 2, \dots, n.$$

Logo, $\tilde{\theta}$ não é identificável.

Uma das conseqüências de $\tilde{\theta}$ não ser identificável é dada no lema a seguir.

Lema 1.2 *No modelo estrutural (1.3), se $\tilde{\theta}$ não é identificável, então não existe um estimador consistente para $\tilde{\theta}$.*

Prova: Se $\tilde{\theta}$ não é identificável, então existem $\theta_* \neq \theta_{**}$ tais que

$$F(y; \theta_*) = F(y; \theta_{**}) , \quad \forall y \in \mathcal{Y},$$

onde F é a distribuição de Y_j , $j = 1, 2, \dots, n$.

Se $\hat{\theta}_n$ é um estimador de $\tilde{\theta}$ baseado numa amostra de tamanho n , então a distribuição de $\hat{\theta}_n$ sob θ_* e θ_{**} é a mesma, logo, o limite de $\hat{\theta}_n$ (em probabilidade ou com probabilidade 1) quando $n \mapsto \infty$, se existe, deve ser idêntico. Mas se $\hat{\theta}_n$ é um estimador consistente

para $\tilde{\theta}$, então $\hat{\theta}_n \mapsto \theta_*$ sob θ_* e $\hat{\theta}_n \mapsto \theta_{**}$ sob θ_{**} e como $\theta_* \neq \theta_{**}$; $\hat{\theta}_n$ não pode ser um estimador consistente para $\tilde{\theta}$.

Do lema temos que falta de identificabilidade implica falta de estimadores consistentes, o que tem sérias implicações na teoria assintótica de estimação e testes de hipóteses. Contudo, $\tilde{\theta}$ ser identificável não implica que exista um estimador consistente para $\tilde{\theta}$. Para um exemplo interessante ver Grabielsen (1978).

Uma das formas de contornar o problema de identificabilidade é colocar restrições sob $\tilde{\theta}$.

Neste caso, Barnett (1969) assume que existe um método de medição, que ele chama instrumento de referência, que mede x sem vício. Especificamente, supondo que o instrumento da referência é o número 1, Barnett assume que

$$\alpha_1 = 0 \quad \text{e} \quad \beta_1 = 1. \quad (1.8)$$

Assim, o modelo proposto por Barnett (1969) é dado por

$$Y_j = \begin{pmatrix} 0 \\ \alpha \end{pmatrix} + \begin{pmatrix} 1 \\ \beta \end{pmatrix} x_j + \varepsilon_j = \mathbf{a} + \mathbf{b}x_j + \varepsilon_j = \mathbf{a} + Bu_j, \quad j = 1, 2, \dots, n; \quad (1.9)$$

onde

$$\mathbf{a} = \begin{pmatrix} 0 \\ \alpha \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ \beta \end{pmatrix}, \quad \alpha = (\alpha_2, \dots, \alpha_p)', \\ \beta = (\beta_2, \dots, \beta_p)', \quad B = (\mathbf{b}, I_p) \quad \text{e} \quad Y_j; \varepsilon_j \quad \text{e} \quad u_j \quad \text{como em (1.3).}$$

Seja $\gamma = (\mu_x, \alpha', \sigma_x, \beta', \phi)'$.

Para $p = 3$, Barnett (1969) consegue estimativas explícitas para os elementos de γ , usando o método dos momentos. No entanto, pode ocorrer que algumas estimativas das variâncias (σ_x^2, ϕ) sejam negativas. Barnett (1969), também assume normalidade dos u_j 's e logo normalidade dos Y_j 's; neste caso se as estimativas das variâncias são não-negativas os estimadores propostos por Barnett correspondem aos de máxima verossimilhança. A

possibilidade dos estimadores propostos por Barnett (1969) serem negativos é contornada, no caso $p = 3$, por Carter (1981), sob normalidade, usando estimação de máxima verossimilhança restrita.

Para $p > 3$ não existem expressões explícitas para os estimadores de máxima verossimilhança de γ . Devido a isto, Barnett (1969) propõe estimadores de momentos. Barnett também dá a matriz de variâncias e covariância assintótica dos estimadores para $p = 3$.

Fuller (1987) apresenta um tratamento bastante completo de modelos com erros nas variáveis, do qual o modelo (1.9) é um caso especial. No caso normal, Fuller (1987) dá um procedimento iterativo para obter os EMV e sua matriz de variância assintótica.

Algumas vezes o modelo (1.9) é referido como a versão (parametrização) de regressão do modelo estrutural.

Theobald e Mallison (1978) usam uma reparametrização do modelo (1.9) proposto por Barnett (1969), em termos de análise de fatores. Especificamente, seja

$$x_j = \mu_x + \sigma_x F_j, \quad j = 1, 2, \dots, n; \quad (1.10)$$

onde F_j tem média zero e variância 1. Substituindo em (1.9) fica

$$Y_j = \mu + \lambda F_j + \varepsilon_j = \mu + L v_j, \quad (1.11)$$

com $\mu = a + b\mu_x$, $\lambda = \sigma_x b$, $L = [\lambda, I_p]$ e $v_j = \begin{pmatrix} F_j \\ \varepsilon_j \end{pmatrix}$, $j = 1, 2, \dots, n$.

Note que

$$E(v_j) = O_{(p+1) \times 1} \quad \text{e} \quad \text{Var}(v_j) = \begin{pmatrix} 1 & 0_p \\ 0 & D(\phi) \end{pmatrix} \quad (1.12)$$

Seja $\theta = (\mu', \lambda', \phi')'$.

Se v_j segue uma distribuição normal, o modelo (1.11) corresponde a um modelo de análise de fator com um fator comum e Theobald e Mallison (1978) usam pacotes desenvolvidos para análise de fatores para obter o EMV de θ . A matriz de variâncias assintótica do EMV de θ é estimada usando um resultado de Lawley e Maxwell (1971). Testes de hipóteses são abordados através da razão de verossimilhança.

A reparametrização de Theobald e Mallison (1978) tem vantagem do ponto de vista computacional e a partir dela é possível obter estimativas dos parâmetros do modelo (1.9), pois $\gamma = g(\theta)$, onde g segue de (1.11). Com efeito, temos

- (i) Se $\hat{\theta}$ denota o EMV de θ , pela propriedade de invariância, o EMV de γ é dado por $\hat{\gamma} = g(\hat{\theta})$;
- (ii) Seja $I_F(\theta)$ a matriz de informação, por observação, na parametrização de análise de fator. Notando que

$$\theta = a(\gamma) = (\mu_x, \alpha' + \beta' \mu_x, \sigma_x, \sigma_x \beta', \phi')',$$

segue que a matriz de informação, por observação no modelo (1.9) é dada por

$$I_R(\gamma) = A' I_F(a(\gamma)) A, \quad (1.13)$$

onde

$$A = \left[\left(\frac{\partial a_i(\gamma)}{\partial \gamma_j} \right) \right] = \begin{pmatrix} A_{11} & A_{12} & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & I_p \end{pmatrix}, \quad \text{com}$$

$$A_{11} = \begin{pmatrix} 1 & 0 \\ \beta & I_{p-1} \end{pmatrix}, \quad A_{12} = \begin{pmatrix} 0 & 0 \\ 0 & \mu_x I_{p-1} \end{pmatrix}, \quad A_{22} = \begin{pmatrix} 1 & 0 \\ \beta & \sigma_x I_{p-1} \end{pmatrix}.$$

- (iii) Se $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N_{3p}(0; I_F^{-1}(\theta))$; então

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{D} N_{3p}(0, I_R^{-1}(\gamma)) \quad (1.14)$$

onde $I_R^{-1}(\gamma) = G I_F^{-1}(a(\gamma)) G'$ e

$$G = \left[\left(\frac{\partial g_i(\theta)}{\partial \theta_j} \right) \right] = \begin{pmatrix} G_{11} & G_{12} & 0 \\ 0 & G_{22} & 0 \\ 0 & 0 & I_p \end{pmatrix}_{3p \times 3p}, \quad \text{com}$$

$$G_{11} = \begin{pmatrix} 1 & 0 \\ -\beta & I_{p-1} \end{pmatrix}, \quad G_{12} = (\mu_x / \sigma_x) \begin{pmatrix} 0 & 0 \\ \beta & -I_{p-1} \end{pmatrix},$$

$$G_{22} = (1 / \sigma_x) \begin{pmatrix} \sigma_x & 0 \\ -\beta & I_{p-1} \end{pmatrix}.$$

De (i), (ii) e (iii) segue que podemos trabalhar com a parametrização ou modelo (1.11), a qual tem vantagens computacionais, como por exemplo, a implementação do algoritmo EM, e a partir dela obter resultados para o modelo de Barnett (1969), (1.9).

Quando $p = 2$, o modelo (1.9) (ou (1.11)) corresponde ao modelo de regressão linear simples com erros nas variáveis, extensivamente discutido na literatura, ver Fuller (1987), Wong (1989) e Arellano e Bolfarine (1995). Um fato conhecido é que este modelo não é identificável.

Exemplo 1.2 Considere os u_j 's como no Lema 1.1 e os seguintes valores de $\gamma = (\mu_x, \alpha_2, \sigma_x, \beta_2, \phi_1, \phi_2)$

$$\gamma_* = (1, 2, 1, 1, 3, 7)' \quad \text{e} \quad \gamma_{**} = (1, 1, 1/\sqrt{2}, 2, 7/2, 6)' .$$

De (1.9) é fácil ver que ambos valores de γ geram a mesma média e a mesma matriz de covariâncias de Y_j :

$$E(Y_j) = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad \text{e} \quad \text{Var}(Y_j) = \begin{pmatrix} 4 & 1 \\ 1 & 8 \end{pmatrix}, \quad j = 1, 2, \dots, n,$$

de onde segue que γ não é identificável.

Para contornar esta dificuldade vamos supor que a razão das variâncias dos erros de medição ϕ_2/ϕ_1 é conhecida. Sem perda de generalidade, podemos tomar $(\phi_2/\phi_1) = 1$; isto é, $\phi_1 = \phi_2 = \phi$, digamos. Para outras restrições de identificabilidade no caso $p = 2$, ver Fuller (1987), Shyr e Gleser (1986) e Jaech (1985).

Neste caso o modelo fica:

$$Y_j = \begin{pmatrix} 0 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} 1 \\ \beta_2 \end{pmatrix} x_j + \varepsilon_j = \mathbf{a} + B u_j, \quad j = 1, 2, \dots, n, \quad (1.15)$$

onde, agora, $\mathbf{a} = (0, \alpha_2)'$, $B = (\mathbf{b}, I_2)$, $\mathbf{b} = (1, \beta_2)'$, $Y_j = (y_{1j}, y_{2j})'$, $\varepsilon_j = (\varepsilon_{1j}, \varepsilon_{2j})'$ e $u_j = (x_j, \varepsilon_j')'$, com

$$E(u_j) = \begin{pmatrix} \mu_x \\ 0 \\ 0 \end{pmatrix}, \quad \text{Var}(u_j) = \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & \phi \end{pmatrix} \quad (1.16)$$

e $\gamma = (\mu_x, \alpha_2, \sigma_x, \beta_2, \phi)'$.

Também podemos usar a parametrização de análise de fatores utilizada por Theobald e Mallison (1978) para $p \geq 3$. Com efeito, para $p = 2$, temos

$$Y_j = \mu + \lambda F_j + \varepsilon_j = \mu + L v_j, \quad j = 1, 2, \dots, n, \quad (1.17)$$

com

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_x \\ \alpha_2 + \beta_2 \mu_x \end{bmatrix}, \quad \lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} \sigma_x \\ \sigma_x \beta_2 \end{bmatrix}$$

$$L = \begin{pmatrix} \lambda_1 & 1 & 0 \\ \lambda_2 & 0 & 1 \end{pmatrix} \quad \text{e} \quad v_j = \begin{pmatrix} F_j \\ \varepsilon_j \end{pmatrix} \quad \text{com}$$

$$E(v_j) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \text{Var}(v_j) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & \phi \end{pmatrix} \quad \text{e} \quad \theta = (\mu_1, \mu_2, \lambda_1, \lambda_2, \phi)'. \quad (1.18)$$

Aqui, como no caso $p \geq 3$, podemos usar qualquer parametrização do modelo (1.15) ou (1.17), pois valem também os fatos similares aos indicados em (i), (ii) e (iii) para o caso $p \geq 3$.

1.3 Definição de confiabilidade

Um conceito importante na comparação de instrumentos de medição é a confiabilidade.

Definição 1.2 A confiabilidade do i -ésimo instrumento, denotada por ρ_i , define-se por:

$$\rho_i = \beta_i^2 \sigma_x^2 / (\beta_i^2 \sigma_x^2 + \sigma_i^2) \quad (1.19)$$

ou

$$\rho_i = \lambda_i^2 / (\lambda_i^2 + \sigma_i^2), \quad i = 1, 2, \dots, p. \quad (1.20)$$

Agora, como $\text{Cov}(y_{ij}, x_j) = \beta_i \sigma_x^2$, $\text{Var}(x_j) = \sigma_x^2$ e $\text{Var}(y_{ij}) = \beta_i^2 \sigma_x^2 + \sigma_i^2$, temos que ρ_i corresponde ao quadrado do coeficiente de correlação entre y_{ij} e x_j . Ou seja, é uma

medida tipo coeficiente de determinação usada em regressão linear e pode ser interpretada como a proporção da variação observada “explicada” pela i -ésima linha de calibração e logicamente $0 \leq \rho_i \leq 1$, $i = 1, 2, \dots, p$.

Uma outra medida ou índice usado para avaliar a qualidade de um instrumento ou método de medição é a precisão. Se conhecemos os parâmetros α_i e β_i , então as observações y_{ij} , $j = 1, \dots, n$, no instrumento i podem ser calibradas como

$$y_{ij}^* = \frac{1}{\beta_i}(y_{ij} - \alpha_i) = x_j + \frac{1}{\beta_i}\varepsilon_{ij} = x_j + \varepsilon_{ij}^*,$$

as quais são “não viciadas” para x_j , $j = 1, 2, \dots, n$. Os erros de medição ε_{ij}^* de y_{ij}^* têm variância σ_i^2/β_i^2 . Já que depois da calibração todos os instrumentos medem x_j sem vício, uma definição natural da precisão seria, ver Shyr e Gleser (1986) e Cochran (1968):

Definição 1.3 A precisão do instrumento i , denotado por π_i , define-se como:

$$\pi_i = 1/\text{Var}(\varepsilon_{ij}^*) = \beta_i^2/\sigma_i^2 = \lambda_i^2/\sigma_x^2\sigma_i^2, \quad i = 1, 2, \dots, p; \quad \beta_1 = 1. \quad (1.21)$$

Theobald e Mallison (1978) usam a “precisão relativa”, denotada por τ_i , definida por:

$$\tau_i = \pi_i\sigma_x^2 = \beta_i^2\sigma_x^2/\sigma_i^2 = \lambda_i^2/\sigma_i^2, \quad i = 1, 2, \dots, p; \quad \beta_1 = 1. \quad (1.22)$$

No entanto, Draper e Guttman (1975, 1976), definem a precisão relativa como

$$\delta_i \equiv \frac{1}{\tau_i} = \sigma_i^2/\beta_i^2\sigma_x^2 = \sigma_i^2/\lambda_i^2, \quad i = 1, 2, \dots, p. \quad (1.23)$$

Note que

$$\rho_i = \sigma_x^2\pi_i/(\sigma_x^2\pi_i + 1) = \tau_i/(\tau_i + 1) = 1/(1 + \delta_i), \quad i = 1, 2, \dots, p \quad (1.24)$$

e que

$$\rho_i = \rho_k \Leftrightarrow \pi_i = \pi_k \Leftrightarrow \tau_i = \tau_k \Leftrightarrow \delta_i = \delta_k, \quad \forall i \neq k = 1, 2, \dots, p. \quad (1.25)$$

Das medidas anteriores, usaremos preferencialmente a confiabilidade, pois esta não depende das unidades de medição dos x_j e é mais fácil de ser interpretada. Contudo, de

(1.25) vemos que comparar confiabilidades é equivalente a comparar precisões ou precisões relativas.

Em geral o problema de comparar p instrumentos reduz-se a fazer inferências sobre os vícios (α_i, β_i) e confiabilidades (ρ_i) . É necessário considerar o vício e a confiabilidade simultaneamente, pois um valor alto de ρ_i (próximo de 1) implica só que y_i (medida feita pelo instrumento i) é uma medida confiável de $\alpha_i + \beta_i x$. O instrumento i forneceria medidas “exatas” da característica x só se ρ_i fosse próximo de um e se ele não fosse viciado ($\alpha_i = 0, \beta_i = 1$).

Uma hipótese inicial de interesse, usando o modelo (1.9), poderia ser:

$$H_{01} : \alpha_2 = \alpha_3 = \dots = \alpha_p = 0, \quad \beta_2 = \dots = \beta_p = 1, \quad \rho_1 = \dots = \rho_p$$

ou, equivalentemente, usando o modelo (1.11),

$$H_{01} : \mu_1 = \mu_2 = \dots = \mu_p, \quad \lambda_1 = \lambda_2 = \dots = \lambda_p, \quad \rho_1 = \dots = \rho_p .$$

Ou seja, estamos interessados em testar se os p instrumentos medem a característica x sem vício e com a mesma confiabilidade.

Outra hipótese que pode ser testada é:

$$H_{02} : \alpha_2 = \dots = \alpha_p = 0, \quad \beta_2 = \dots = \beta_p = 1$$

ou

$$H_{02} : \mu_1 = \mu_2 = \dots = \mu_p, \quad \lambda_1 = \lambda_2 = \dots = \lambda_p .$$

Aqui queremos testar se os instrumentos medem x sem vício.

Para testar se os instrumentos são igualmente confiáveis, a hipótese seria

$$H_{03} : \rho_1 = \rho_2 = \dots = \rho_p .$$

Dependente do estudo, também poderia ser interessante testar

$$H_{04} : \beta_2 = \dots = \beta_p = 1 \quad \text{ou} \quad H_{04} : \lambda_1 = \lambda_2 = \dots = \lambda_p .$$

Neste caso, se H_{04} não é rejeitada, teríamos o modelo introduzido por Grubbs (1948, 1973), que tem sido usado para comparar instrumentos de medição na área industrial.

Logicamente, dependendo do problema, poderia ser de interesse testar outro tipo de hipótese, embora as mencionadas acima sejam de interesse geral.

1.4 Objetivos do trabalho

O objetivo deste trabalho é a inferência estatística no modelo estrutural (1.9) ou (1.11) para $p \geq 3$. A idéia é propor uma forma alternativa para a obtenção dos estimadores de máxima verossimilhança (EMV) via algoritmo EM (Dempster et al., 1977; Little e Rubin, 1987), que em geral é de simples implementação, e também desenvolver testes para hipóteses de interesse. Além disso, procuramos uma modelagem “robusta” alternativa ao modelo normal usando a distribuição t -elíptica. Recentemente, a distribuição t tem sido sugerida como alternativa à distribuição normal (Lange et al., 1989) por ter as caudas mais pesadas e, portanto, “acomoda” possíveis “outliers” presentes nos dados. Neste sentido, o uso da distribuição t forneceria um método robusto de análise de dados. Contudo, o caso $p = 2$ também é considerado. Finalmente estudamos o modelo funcional normal discutido por Kimura (1992). Assim, os objetivos específicos deste trabalho são:

- (i) No modelo estrutural normal, para $p \geq 3$, apresentar uma metodologia alternativa à atual (Barnett, 1969, Theobald e Mallison, 1978 e Fuller, 1987), baseada no algoritmo EM para obter os EMV. Obter a matriz de informação utilizando a verossimilhança completa (ou não observada) usada na etapa M do algoritmo. Na implementação do algoritmo EM e na obtenção da matriz de informação usamos a parametrização de análise de fator do modelo.
- (ii) Estender o modelo estrutural normal de modo que permita a incorporação de informação auxiliar. Especificamente, a idéia é incorporar uma covariável relacionada

com a característica x .

- (iii) Devido à inferência estatística baseada na distribuição normal ser “vulnerável” a “outliers”, a idéia é substituir a distribuição normal pela distribuição t multivariada que é “mais resistente” à presença de observações extremas nos dados. O algoritmo EM é utilizado para encontrar os EMV.
- (iv) Complementar alguns resultados dados por Kimura (1992) sob o modelo funcional normal, apresentando EMV, como também sua distribuição assintótica. Além disso, apresentamos uma estimativa consistente para a matriz de covariância assintótica dos EMV.

1.5 Organização do trabalho

Este trabalho consta de seis capítulos e quatro apêndices.

O Capítulo 2 está dedicado ao modelo estrutural normal. Depois de estabelecer alguns resultados para o caso $p = 2$, a implementação do algoritmo EM e a obtenção da matriz de informação, usando a verossimilhança completa, é discutida para o caso $p \geq 3$. A estatística de Wald é proposta para testar hipóteses de interesse.

No Capítulo 3, estendemos o modelo estrutural normal (1.9), para incorporar uma covariável linearmente relacionada com a característica de interesse x . Estimação de parâmetros via algoritmo EM e testes de hipóteses são discutidos. Também incluímos um pequeno estudo de simulação.

O Capítulo 4 está dedicado ao modelo estrutural t -elíptico. Estabelecemos alguns resultados para modelos elípticos em geral e discutimos a estimação de parâmetros e testes de hipóteses no modelo t multivariado.

Devido ao modelo de Grubbs ser amplamente usado na área industrial, no Capítulo 5 estudamos este modelo. Obtemos a distribuição assintótica dos EMV em qualquer modelo

elíptico e discutimos a implementação do algoritmo EM no modelo t e no modelo normal.

No Capítulo 6 estudamos um modelo funcional normal discutido por Kimura (1992). Um tipo de algoritmo EM é usado para encontrar os EMV. Utilizando alguns resultados de Mak (1982), obtemos a matriz de covariância assintótica dos EMV e propomos um estimador consistente desta matriz. Estimadores de momento são também considerados, assim como suas propriedades assintóticas.

Finalmente, nos apêndices, resumimos alguns resultados da literatura utilizados na tese e também apresentamos lemas usados para obter as matrizes de informação.

Capítulo 2

O modelo estrutural normal

2.1 Introdução

Este capítulo está dedicado à análise estatística do modelo estrutural (1.9) (ou (1.11)) supondo que a característica de interesse x , bem como os erros de medição dos instrumentos, seguem uma distribuição normal.

Este modelo foi proposto por Barnett (1969) (ver também Jaech, 1964). Barnett consegue estimativas explícitas para os parâmetros do modelo (γ) no caso $p = 3$ pelo método dos momentos. No entanto, pode acontecer que algumas estimativas das variâncias (σ_x^2, ϕ) sejam negativas. Ou seja, os estimadores propostos por Barnett são EMV só se seus valores pertencem ao espaço paramétrico, ver Bickel e Doksum (1977, p. 106). Barnett também dá a matriz de covariância assintótica dos estimadores. Carter (1981) obtém, para $p = 3$, estimativas não negativas das variâncias usando EMV restrita.

Para $p > 3$, os EMV não têm forma explícita. Devido a isto, Barnett (1969) propõe estimadores tipo momentos. Theobald e Mallison (1978) reparametrizam o modelo proposto por Barnett na forma de análise de fatores, modelo (1.11), e usam pacotes desenvolvidos para análise de fatores para obter os EMV. A matriz de covariância assintótica dos EMV é estimada usando um resultado devido a Lawley e Maxell (1971). Testes de hipóteses são abordados por razão de verossimilhança.

Fuller (1987) apresenta um modelo com erros nas variáveis do qual o modelo proposto

por Barnett (1969) é um caso especial. Para esse modelo multivariado, Fuller (1987) desenvolve um procedimento iterativo para obter os EMV e também dá uma expressão explícita para a matriz de covariância assintótica do EMV. Mas, para utilizar as técnicas propostas por Fuller (1987) e Theobald e Mallison (1978), é necessário dispor de pacotes especializados, pois sua implementação não é simples. Além disso, para o leitor interessado principalmente nas aplicações, a metodologia desenvolvida por Fuller é um tanto complicada de ser entendida, como é notado também por Kimura (1992).

Devido a isto, nossa idéia é considerar o modelo proposto por Barnett (1969) (1.9) (ou equivalentemente (1.11)) e utilizar, para $p \geq 3$, o algoritmo EM (Dempster et al., 1977, Little e Rubin, 1987) para encontrar os EMV. Este algoritmo é de fácil implementação, especialmente na parametrização de análise de fatores, e de pouco custo computacional. Além disso, produz estimativas não negativas das variâncias (σ_x^2, ϕ) . Neste capítulo a implementação do algoritmo EM e a obtenção da matriz de informação é ilustrada na versão de análise de fatores do modelo. Os EMV e sua matriz de covariância assintótica para os parâmetros do modelo na versão de regressão são obtidos usando os resultados da Seção 1.2.

O caso $p = 2$, que corresponde ao modelo de regressão linear simples com erros nas variáveis, tem sido tratado extensivamente na literatura, ver Fuller (1987) e referências ali citadas. Wong (1989) e Arellano e Bolfarine (1995) estudam este modelo usando uma parametrização ortogonal (Cox e Reid, 1987). Na Seção 2.2 analisamos, em forma resumida, este caso, usando a parametrização de análise de fatores. Apresentamos os EMV e sua distribuição assintótica. Também discutimos teste de hipóteses.

A Seção 2.3 é dedicada ao caso de três ou mais instrumentos. Apresentamos a implementação do algoritmo EM e a obtenção da matriz de informação na parametrização de análise de fatores do modelo. A matriz de covariância assintótica do EMV, obtida nas duas versões do modelo. Consideramos também testes de hipóteses usando a estatística

de Wald.

Finalmente, na Seção 2.4 apresentamos a análise estatística do modelo estrutural quando o instrumento de referência mede sem erro a característica de interesse x .

2.2 O caso de dois instrumentos

A idéia aqui é apresentar alguns resultados utilizando principalmente a parametrização de análise de fatores do modelo.

Consideremos novamente o modelo (1.15):

$$Y_j = a + Bu_j, \quad j = 1, 2, \dots, n. \quad (2.1)$$

onde, agora, $u_j, j = 1, 2, \dots, n$, são vetores

$$\text{iid } N_3 \left(\begin{pmatrix} \mu_x \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & \phi \end{pmatrix} \right). \quad (2.2)$$

De (2.2) temos que

$$Y_j \sim N_2(\mu(\gamma); V(\gamma)), \quad (2.3)$$

independente, com $\mu(\gamma) = a + b\mu_x$, $V(\gamma) = \sigma_x^2 b b' + \phi I_2$ e a, b, B e γ como em (1.15).

Equivalentemente, se considerarmos a versão de análise de fatores do modelo (1.17),

$$Y_j = \mu + Lv_j, \quad j = 1, 2, \dots, n. \quad (2.4)$$

Assumimos que $v_j, 1 \leq j \leq n$ são

$$\text{iid } N_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & \phi \end{pmatrix} \right), \quad (2.5)$$

de onde segue que

$$Y_j \sim N_2(\mu(\theta); V(\theta)) \quad (2.6)$$

com $\mu(\theta) = \mu$, $V(\theta) = \lambda \lambda' + \phi I_2$ e μ, λ, L e θ como em (1.17).

De (2.6) segue que o logaritmo da função de verossimilhança, denotado por $\ell(\theta)$, é dado por:

$$\ell(\theta) = \text{cte.} - \frac{n}{2} \ln |V| - \frac{1}{2} \sum_{j=1}^n (y_j - \mu)' V^{-1} (y_j - \mu) , \quad (2.7)$$

onde $\mu = \mu(\theta)$ e $V = V(\theta)$.

É fácil ver que (Graybill, 1983)

$$|V| = c\phi^2 \quad \text{e} \quad V^{-1} = (1/\phi)I_2 - (1/c\phi^2)\lambda\lambda' , \quad (2.8)$$

com $c = 1 + (\lambda'\lambda)/\phi$.

De (2.7) e usando (2.8) temos que os EMV são

$$\left. \begin{aligned} \hat{\mu}_i &= \bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij} , \quad i = 1, 2; \\ \hat{\lambda}_1 &= \left[\frac{S_{11} - S_{22} + 2b_2 S_{12}}{1 + b_2^2} \right]^{1/2} , \quad \hat{\lambda}_2 = b_2 \hat{\lambda}_1 ; \\ \hat{\phi} &= S_{22} - b_2 S_{12} \end{aligned} \right\} , \quad (2.9)$$

onde

$$b_2 = \left[\frac{S_{22} - S_{11}}{2S_{12}} \right] + \left\{ \left[\frac{S_{22} - S_{11}}{2S_{12}} \right]^2 + 1 \right\}^{1/2} \quad \text{e}$$

$$S_{ik} = \frac{1}{n} \sum_{j=1}^n (y_{ij} - \bar{y}_i)(y_{kj} - \bar{y}_k) ; \quad i, k = 1, 2.$$

Agora, como

$$\gamma = g(\theta) = (\mu_1, \mu_2 - (\lambda_2/\lambda_1)\mu_1, \lambda_1, \lambda_2/\lambda_1, \phi)' , \quad (2.10)$$

o EMV de $\gamma = (\mu_x, \alpha_2, \sigma_x, \beta_2, \phi)'$ é dado por $\hat{\mu}_x = \hat{\mu}_1$, $\hat{\alpha}_2 = \hat{\mu}_2 - b_2 \hat{\mu}_x$, $\hat{\beta}_2 = b_2$, $\hat{\sigma}_x = \hat{\lambda}_1$ e $\hat{\phi}$ como em (2.9).

No lema seguinte apresentamos a matriz de informação esperada.

Lema 2.1 A matriz de informação esperada, denotada por $I(\theta)$, para θ no modelo (2.4) - (2.5), é dada por:

$$I(\theta) = nI_1(\theta) = n \begin{bmatrix} I_{\mu\mu} & 0 & 0 \\ 0 & I_{\lambda\lambda} & I_{\lambda\phi} \\ 0 & I_{\phi\lambda} & i_{\phi\phi} \end{bmatrix}, \quad (2.11)$$

onde $I_{\mu\mu} = V^{-1}$, $I_{\lambda\lambda} = (1/\phi)(1 - c^{-1})I_2 - (1/c\phi^2)(1 - 2c^{-1})\lambda\lambda'$, $I_{\lambda\phi} = (1/c^2\phi^2)\lambda$, $i_{\phi\phi} = (1 + c^{-2})/2\phi^2$, $I_{\phi\lambda} = I'_{\lambda\phi}$ e V^{-1} e c como em (2.8).

Prova: Usando a fórmula dada por Dolby (1976),

$$-E \left(\frac{\partial^2 \ell(\theta)}{\partial \theta_i \partial \theta_j} \right) = n \left\{ \frac{1}{2} \text{tr} \left(V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} \frac{\partial V}{\partial \theta_j} \right) + \frac{\partial \mu'}{\partial \theta_i} V^{-1} \frac{\partial \mu}{\partial \theta_j} \right\},$$

e utilizando (2.8), o resultado segue, depois de alguma álgebra. Aqui, $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)'$ $= (\mu_1, \mu_2, \lambda_1, \lambda_2, \phi)'$.

A seguir, no Teorema 2.1, apresentamos a distribuição assintótica do EMV de θ e γ .

Teorema 2.1 Sob o modelo (2.4) e (2.5):

(i) $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N_5(0; \Omega_\theta)$, onde

$$\Omega_\theta = \begin{bmatrix} V & 0 & 0 \\ 0 & (1/c_1)I_2 - \frac{1}{2}\lambda\lambda' & -(\frac{\phi}{c-1})\lambda \\ 0 & -(\frac{\phi}{c-1})\lambda' & \frac{2\phi^2}{2\phi^2} \end{bmatrix} \quad (2.12)$$

com $c_1 = (1 - c^{-1})/\phi$.

(ii) $\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{D} N_5(0; \Omega_\gamma)$, onde

$$\Omega_\gamma = \begin{bmatrix} \sigma_x^2 + \phi & -\phi\beta_2 & 0 & 0 & 0 \\ \frac{\mu_x^2\phi^2c}{\sigma_x^4} + \phi(1 + \beta_1^2) & \frac{\mu_x\phi^2c\beta_2}{\sigma_x^3(1 + \beta_2^2)} & -\mu_x\phi^2c & 0 & 0 \\ \frac{1}{c_1} + \frac{\sigma_x^2}{2} & -\frac{\beta_2}{\sigma_x c_1} & -\frac{\phi\sigma_x}{c-1} & 0 & 0 \\ \frac{1 + \beta_2^2}{c_1\sigma_x^2} & 0 & 0 & 2\phi^2 & 0 \end{bmatrix}, \quad (2.13)$$

com $c = 1 + \sigma_x^2(1 + \beta_2^2)/\phi$, $c_1 = (1 - c^{-1})/\phi = \sigma_x^2(1 + \beta_2^2)/\phi^2c$. Aqui, $\xrightarrow{D}$ indica convergência em distribuição quando $n \rightarrow \infty$.

Prova: (i) Das propriedades do EMV $\hat{\theta}$, no caso regular (Sen e Singer, 1993), sabemos que $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N_5(0; I_1^{-1}(\theta))$; e usando resultados de inversa de matrizes partidas (Graybill, 1983), temos que $I_1^{-1}(\theta) = \Omega_\theta$.

(ii) Por (2.10) sabemos que $\gamma = g(\theta)$ e usando a parte (i) segue que, por uso do método delta,

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{D} N_5(0, G\Omega_\theta G') ,$$

onde

$$G = [(g_{ij})] = \left[\left(\frac{\partial g_i(\theta)}{\partial \theta_j} \right) \right] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\beta_2 & 1 & \mu_x \beta_2 / \sigma_x & -\mu_x / \sigma_x & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -\beta_2 / \sigma_x & 1 / \sigma_x & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.14)$$

e o resultado segue, pois $\Omega_\gamma = G\Omega_\theta G'$.

Notemos que $\sqrt{n}(\hat{\alpha}_2 - \alpha_2, \hat{\beta}_2 - \beta_2, \hat{\phi} - \phi) \xrightarrow{D} N_3(0, \Gamma)$, onde $\Gamma = \Gamma_{\theta\theta}$ dada no Teorema 1.3.1 de Fuller (1987).

Neste caso, $p = 2$, como $\beta_2 = 1$ (ou $\lambda_1 = \lambda_2$) $\Leftrightarrow \rho_1 = \rho_2$, temos que H_{01} é equivalente a H_{02} e que H_{03} é equivalente a H_{04} , ver Seção 1.3. Assim, usando os resultados anteriores, Teorema 2.1, podemos testar estas hipóteses utilizando, por exemplo, a estatística de Wald, em qualquer parametrização do modelo. Como ilustração, consideremos H_{02} : $\alpha_2 = 0, \beta_2 = 1$. Notemos que H_{02} pode ser escrita como $H_{02} : C\gamma = \gamma_0$, onde

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \text{ de posto } 2 \text{ e } \gamma_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} .$$

A estatística de Wald para testar H_{02} é dada por

$$W_a = n(C\hat{\gamma} - \gamma_0)'(C\hat{\Omega}_\gamma C')^{-1}(C\hat{\gamma} - \gamma_0) \xrightarrow{D} \chi^2(2) , \quad (2.15)$$

sob H_{02} , onde $\hat{\Omega}_\gamma$ é a matriz Ω_γ avaliada no EMV $\hat{\gamma}$.

Uma outra parametrização que tem sido especialmente útil para obter inferências sobre

o parâmetro β_2 é a parametrização ortogonal obtida por Wong (1989) e dada por:

$$\mu_1, \mu_2, \lambda_1^2 + \lambda_2^2 + \phi, \phi \text{ e } \frac{\lambda_2}{\lambda_1} (= \beta_2). \quad (2.16)$$

Aqui β_2 é ortogonal a $(\mu_1, \mu_2, \lambda_1^2 + \lambda_2^2 + \phi, \phi)$, no sentido de Cox e Reid (1987). Com esta parametrização Wong (1989) e Arellano e Bolfarine (1995) calculam o fator de Bartlett para melhorar o teste da razão de verossimilhança para testar $H : \beta_2 = \beta_{02}$ com β_{02} conhecido. Logo, podemos usar esses resultados para testar $H_{03} : \beta_2 = 1$.

Também, Arellano e Bolfarine (1994) usam o teste escore para testar $H : \beta_2 = \beta_{02}$. Em nosso caso $\beta_{02} = 1$ e a estatística escore, S_R , para testar $H_{03} : \beta_2 = 1$, é dada por

$$S_R = n(S_{22} - S_{11})^2 / 4(d_1' S d_1)(d_2' S d_2), \quad (2.17)$$

onde $d_1 = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]'$, $d_2 = [-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]'$ e $S = [(S_{ij})]$.

Arellano e Bolfarine (1994) mostram que, sob H_{03} ,

$$\frac{S_R}{n} \sim \text{Beta} \left(\frac{1}{2}, \frac{n-2}{2} \right),$$

uma distribuição beta com parâmetros $\frac{1}{2}$ e $\frac{n-2}{2}$. Daqui segue que, sob H_{03} , a estatística

$$F_S = (n-2) \left(\frac{n}{S_R} - 1 \right)^{-1} = (n-2) \left[\frac{S_R/n}{1 - \frac{S_R}{n}} \right] \sim F(1, n-2), \quad (2.18)$$

uma distribuição F com 1 e $n-2$ graus de liberdade.

Por outro lado, temos que, ver (2.6),

$$Y_j \sim N_2 \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right), \quad j = 1, 2, \dots, n. \quad (2.19)$$

onde $\sigma_{11} = \lambda_1^2 + \phi$, $\sigma_{12} = \sigma_{21} = \lambda_1 \lambda_2$ e $\sigma_{22} = \lambda_2^2 + \phi$. Daqui vemos que $\sigma_{11} = \sigma_{22} \Leftrightarrow \lambda_1 = \lambda_2$ (ou $\beta_2 = 1$); logo H_{03} é equivalente a $H : \sigma_{11} = \sigma_{22}$ e para testar H podemos usar o teste proposto por Pitman (1939). Com efeito, seja

$$X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} y_{1j} \\ y_{2j} \end{pmatrix} = T Y_j, \quad j = 1, 2, \dots, n. \quad (2.20)$$

com $T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Logo, temos que

$$X_j \sim N_2 \left(\begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1 - \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} + \sigma_{22} + 2\sigma_{12} & \sigma_{11} - \sigma_{22} \\ \sigma_{11} - \sigma_{22} & \sigma_{11} + \sigma_{22} - 2\sigma_{12} \end{pmatrix} \right). \quad (2.21)$$

Seja ρ_{12} o coeficiente de correlação entre x_{1j} e x_{2j} . É fácil ver que testar a hipótese $H : \sigma_{11} = \sigma_{22}$ é equivalente a testar a hipótese $H : \rho_{12} = 0$ (Pitman, 1939). Agora sabemos que o teste da razão de verossimilhança para testar $H : \rho_{12} = 0$ é dado por

$$t_n^2 = (n-2) \left[\frac{\hat{\rho}_{12}^2}{1 - \hat{\rho}_{12}^2} \right] \sim F(1, n-2). \quad (2.22)$$

Seja S_x a matriz de covariâncias amostrais de $X_1, X_2, \dots, X_n$. Então por (2.20) temos que

$$S_x = TST' = \begin{bmatrix} S_{11} + S_{22} + 2S_{12} & S_{11} - S_{22} \\ S_{11} - S_{22} & S_{11} + S_{22} - 2S_{12} \end{bmatrix}. \quad (2.23)$$

Assim, o EMV de ρ_{12} , $\hat{\rho}_{12}$ ao quadrado fica

$$\begin{aligned} \hat{\rho}_{12}^2 &= (S_{11} - S_{22})^2 / (S_{11} + S_{22} + 2S_{12})(S_{11} + S_{22} - 2S_{12}) \\ &= (S_{11} - S_{22})^2 / 4(d_1' S d_1)(d_2' S d_2) \\ &= \frac{S_R}{n}, \end{aligned} \quad (2.24)$$

ou seja,

$$t_n^2 = (n-2) \left[\frac{\frac{S_R}{n}}{1 - \frac{S_R}{n}} \right]. \quad (2.25)$$

Note que $t_n^2 = F_s$ dado em (2.18). Isto é, a estatística escore para testar H_{03} tem distribuição exata, sob H_{02} , e coincide com o teste da razão de verossimilhança para testar $H : \rho_{12} = 0$. Logo, temos um teste exato para testar igualdade dos coeficientes de compatibilidade; $H_{03} : \rho_1 = \rho_2$.

Finalmente, também é possível obter um teste exato para testar $H_{01} : \mu_1 = \mu_2$, $\lambda_1 = \lambda_2$. Com efeito, note que testar H_{01} é equivalente a testar $H : \mu_1 = \mu_2, \sigma_{11} = \sigma_{22}$.

Assim, podemos usar o teste proposto por Bradley e Blackwood (1989) para testar H . Como X_j tem uma distribuição normal, de (2.20) segue que

$$E(x_{2j}|x_{1j}) = (\mu_1 - \mu_2) + \left[\frac{\sigma_{11} - \sigma_{22}}{\sigma_{x1}^2} \right] (x_{1j} - (\mu_1 + \mu_2)) = \varphi_0 + \varphi_1 x_{1j}, \quad (2.26)$$

onde $\sigma_{x1}^2 = \text{Var}(x_{1j})$, $\varphi_0 = (\mu_1 - \mu_2) - \left[\frac{\sigma_{11} - \sigma_{22}}{\sigma_{x1}^2} \right] (\mu_1 + \mu_2)$ e $\varphi_1 = \left[\frac{\sigma_{11} - \sigma_{22}}{\sigma_{x1}^2} \right]$.

Daqui vemos que testar H (ou H_{01}) é equivalente a testar a hipótese $H : \varphi_0 = \varphi_1 = 0$. Ou seja, podemos testar que os instrumentos medem sem vício a característica de interesse e são igualmente confiáveis, usando resultados de análise de regressão linear simples. Note também que $\sigma_{11} = \sigma_{22} \Leftrightarrow \varphi_1 = 0 \Leftrightarrow \rho_{12} = 0$. Estes resultados foram utilizados por Blackwood e Bradley (1991) no modelo de Grubbs para testar H_{01} . As estatísticas para testar H_{01} e H_{03} são dados por (Seber, 1977):

$$F_{01} = \frac{\hat{\varphi}'(X'X)\hat{\varphi}(n-2)}{SQE} \sim F(2, n-2), \quad \text{sob } H_{01} \quad (2.27)$$

e

$$F_{02} = (n-2)\hat{\varphi}_1^2 \sum_{j=1}^n (x_{1j} - \bar{x}_1)^2 / SQE \sim F(1, n-2), \quad \text{sob } H_{03}, \quad (2.28)$$

onde

$$\hat{\varphi} = \begin{bmatrix} \hat{\varphi}_0 \\ \hat{\varphi}_1 \end{bmatrix}, \quad \hat{\varphi}_0 = \bar{x}_2 - \hat{\varphi}_1 \bar{x}_1,$$

$$\hat{\varphi}_1 = \sum_{j=1}^n (x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2) / \sum_{j=1}^n (x_{1j} - \bar{x}_1)^2,$$

$$SQE = \sum_{j=1}^n (x_{2j} - \bar{x}_2)^2 - \hat{\varphi}_1^2 \sum_{j=1}^n (x_{1j} - \bar{x}_1)^2,$$

$$\bar{x}_1 = \frac{1}{n} \sum_{j=1}^n x_{1j}, \quad \bar{x}_2 = \frac{1}{n} \sum_{j=1}^n x_{2j} \quad \text{e}$$

$$X' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \end{bmatrix}.$$

Note que $F_{02} = t_n^2 = F_s$.

2.3 O caso de três ou mais instrumentos

Nesta seção, discutimos estimação e testes de hipóteses no modelo estrutural (1.9) no caso $p \geq 3$. Em primeiro lugar, discutimos a implementação do algoritmo EM na versão de análise de fatores do modelo, para encontrar os EMV e depois estudamos suas propriedades assintóticas. Também discutimos testes de hipóteses baseados em amostras grandes.

Consideremos o modelo (1.9):

$$Y_j = a + Bu_j, \quad j = 1, 2, \dots, n,$$

$$u_j : 1 \leq j \leq n \text{ são iid } N_{p+1} \left(\begin{pmatrix} \mu_x \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & D(\phi) \end{pmatrix} \right), \quad (2.29)$$

com Y_j e u_j como em (1.3), a e B como em (1.9) e $D(\phi)$ como em (1.4). Daqui temos que

$$Y_j \sim N_p(\mu(\gamma), V(\gamma)), \quad j = 1, 2, \dots, n, \quad (2.30)$$

onde $\mu(\gamma) = a + b\mu_x$ e $V(\gamma) = \sigma_x^2 b b' + D(\phi)$, com $b = (1, \beta')'$ e $\gamma = (\mu_x, \alpha', \sigma_x, \beta', \phi')'$.

Na versão de análise de fatores do modelo teríamos:

$$Y_j = \mu + Lv_j, \quad j = 1, 2, \dots, n,$$

$$v_j : 1 \leq j \leq n, \text{ são iid } N_{p+1} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 & 0 \\ 0 & D(\phi) \end{pmatrix} \right) \quad (2.31)$$

com μ , L e v_j , $1 \leq j \leq n$, como em (1.11). De (2.31) temos que:

$$Y_j, \quad 1 \leq j \leq n, \text{ são iid } N_p(\mu(\theta), V(\theta)), \quad (2.32)$$

onde $\mu(\theta) = \mu$ e $V(\theta) = \lambda \lambda' + D(\phi)$, com $\theta = (\mu', \lambda', \phi')'$.

Handwritten note:
 $\theta = (\mu', \lambda', \phi')'$
 $\lambda \in \mathbb{R}^{p+1 \times 1}$

2.3.1 Estimação via o algoritmo EM

Aqui mostramos como obter os EMV de θ e γ usando o algoritmo EM. A implementação do algoritmo é ilustrada na parametrização de análise de fatores do modelo. Como é

conhecido, o algoritmo EM (Dempster et al., 1977; Little e Rubin, 1987) aumenta os dados $Y = (Y_1, \dots, Y_n)$ por adição de dados hipotéticos, ou não observáveis, $F = (F_1, F_2, \dots, F_n)$ de tal forma que o EMV de θ baseado em $Z = (Y, F)$ seja fácil de calcular. Dada a estimativa $\theta^{(m)}$, na iteração m , a iteração $m + 1$ do EM consiste de duas etapas; uma etapa E de esperança e uma etapa M de maximização. Na etapa E calculamos o valor esperado do logaritmo da função de verossimilhança completa $\ell(\theta|Z)$ com respeito à distribuição condicional de F dado Y e $\theta^{(m)}$. Na etapa M maximizamos a função resultante com respeito a θ , e obtemos assim uma nova estimativa $\theta^{(m+1)}$. Cada iteração do EM incrementa o logaritmo da função de verossimilhança observada $\ell(\theta|Y)$; $\ell(\theta^{(m)}|Y) \leq \ell(\theta^{(m+1)}|Y)$.

Seja $Z_j = \begin{bmatrix} F_j \\ Y_j \end{bmatrix}$, $j = 1, 2, \dots, n$.

De (2.31) temos que

$$Z_j, j = 1, 2, \dots, n \text{ são iid } N_{p+1}(\mu_Z, \Sigma_Z), \quad (2.33)$$

onde

$$\mu_Z = \begin{bmatrix} 0 \\ \mu \end{bmatrix} \quad \text{e} \quad \Sigma_Z = \begin{pmatrix} 1 & \lambda' \\ \lambda & \lambda\lambda' + D(\phi) \end{pmatrix}.$$

Daqui segue que o logaritmo da função de verossimilhança completa, denotado por $\ell(\theta|Z)$, é dado por

$$\ell(\theta|Z) = \text{cte.} - \frac{n}{2} \ln |\Sigma_Z| - \frac{1}{2} \sum_{j=1}^n (Z_j - \mu_Z)' \Sigma_Z^{-1} (Z_j - \mu_Z), \quad (2.34)$$

onde

$$|\Sigma_Z| = \prod_{i=1}^p \phi_i \quad \text{e} \quad \Sigma_Z^{-1} = \begin{pmatrix} c & -\lambda' D^{-1}(\phi) \\ -D^{-1}(\phi) \lambda & D^{-1}(\phi) \end{pmatrix}, \quad (2.35)$$

com $c = 1 + \sum_{i=1}^p (\lambda_i^2 / \phi_i)$.

Logo, usando (2.35), temos que

$$\begin{aligned} \ell(\theta|Z) = & \text{cte.} - \frac{n}{2} \sum_{i=1}^p \ln \phi_i - \frac{c}{2} \sum_{j=1}^n F_j^2 \\ & + \sum_{j=1}^n \sum_{i=1}^p F_j \lambda_i (y_{ij} - \mu_i) / \phi_i - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^p \left(\frac{1}{\phi_i} \right) (y_{ij} - \mu_i)^2. \end{aligned} \quad (2.36)$$

Etapa E: Calculamos o valor esperado:

$$\begin{aligned}
 E_F(\ell(\theta|Z)|Y, \theta^{(m)}) &= \text{cte.} - \frac{n}{2} \sum_{i=1}^p \ln \phi_i^{(m)} - \frac{c^{(m)}}{2} \sum_{j=1}^n F_j^{2(m)} \\
 &\quad + \sum_{j=1}^n \sum_{i=1}^p F_j^{(m)} \lambda_i^{(m)} (y_{ij} - \mu_i^{(m)}) / \phi_i^{(m)} \\
 &\quad - \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^p \left(\frac{1}{\phi_i^{(m)}} \right) (y_{ij} - \mu_i^{(m)})^2,
 \end{aligned} \tag{2.37}$$

onde

$$\begin{aligned}
 F_j^{(m)} &= E(F_j|Y_j, \theta^{(m)}) = \frac{1}{c^{(m)}} \sum_{i=1}^p \left(\frac{\lambda_i^{(m)}}{\phi_i^{(m)}} \right) (y_{ij} - \mu_i^{(m)}), \\
 F_j^{2(m)} &= E(F_j^2|Y_j, \theta^{(m)}) = \frac{1}{c^{(m)}} + (F_j^{(m)})^2, \\
 c^{(m)} &= 1 + \sum_{i=1}^p \left(\frac{\lambda_i^{(m)2}}{\phi_i^{(m)}} \right).
 \end{aligned} \tag{2.38}$$

Note que (2.36) segue do fato de que $Z_j \sim N_{p+1}(\mu_Z, \Sigma_Z)$, dado em (2.33).

Etapa M: Nesta etapa determinamos $\theta^{(m+1)}$, maximizando (2.37). Derivando com respeito a $\theta = \theta^{(m)}$, obtemos as seguintes estimativas:

$$\begin{aligned}
 \lambda_i^{(m+1)} &= \sum_{j=1}^n F_j^{(m)} (y_{ij} - \bar{y}_i) / \sum_{j=1}^n (F_j^{(m)} - \bar{F}^{(m)})^2, \\
 \mu_i^{(m+1)} &= \bar{y}_i - \bar{F}^{(m)} \lambda_i^{(m+1)}, \\
 \phi_i^{(m+1)} &= \frac{1}{n} \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 - \frac{1}{n} \lambda_i^{(m+1)2} \sum_{j=1}^n (F_j^{(m)} - \bar{F}^{(m)})^2 \\
 &= \frac{1}{n} \sum_{j=1}^n [(y_{ij} - \bar{y}_i) - \lambda_i^{(m+1)} (F_j^{(m)} - \bar{F}^{(m)})]^2, \quad i = 1, 2, \dots, p,
 \end{aligned} \tag{2.39}$$

onde

$$\begin{aligned}
 \bar{F}^{(m)} &= \frac{1}{n} \sum_{j=1}^n F_j^{(m)} \quad \text{e} \quad \bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij}, \quad i = 1, 2, \dots, p, \\
 \sum_{j=1}^n (F_j^{(m)} - \bar{F}^{(m)})^2 &= \sum_{j=1}^n F_j^{2(m)} - n(\bar{F}^{(m)})^2.
 \end{aligned}$$

Note que, em cada iteração, os estimadores das variâncias $(\phi^{(m+1)})$ são não-negativos.

Enfatizamos que o procedimento de estimação proposto por Barnett (1969), como também

$$-F_j^{(m)} \lambda_i^{(m)} + \sum_j \frac{1}{\phi_i^{(m)}} (y_{ij} - \mu_i^{(m)})$$

o procedimento iterativo proposto por Theobald e Mallison (1978), podem produzir estimativas negativas das variâncias. Note também que não precisamos de um procedimento iterativo para obter o máximo na etapa M do algoritmo. A implementação do algoritmo anterior é extremamente simples e computacionalmente conveniente (econômico), mesmo que necessitemos um grande número de iterações para atingir a convergência.

(i) Estimativas iniciais

Para iniciar o algoritmo EM, ou seja, dar um valor a $\theta^{(0)}$, podemos usar o método proposto por Barnett (1969), descrito a seguir. Cabe mencionar, no entanto, que faremos uso da parametrização de análise de fatores do modelo. Barnett trabalha com a versão de regressão. Consideremos subconjuntos de três instrumentos, digamos os instrumentos h, i e $k \in \{1, 2, \dots, j\}$, com $h \neq i \neq k$. Seja

$$S(h, i, k) = \begin{bmatrix} S_{hh} & S_{hi} & S_{hk} \\ S_{ih} & S_{ii} & S_{ik} \\ S_{kh} & S_{ki} & S_{kk} \end{bmatrix} \quad (2.40)$$

e

$$V(h, i, k) = \begin{bmatrix} \lambda_h^2 + \phi_h & \lambda_h \lambda_i & \lambda_h \lambda_k \\ \lambda_i \lambda_h & \lambda_i^2 + \phi_i & \lambda_i \lambda_k \\ \lambda_k \lambda_h & \lambda_k \lambda_i & \lambda_k^2 + \phi_k \end{bmatrix}. \quad (2.41)$$

De $V(h, i, k) = S(h, i, k)$, temos

$$\begin{aligned} \tilde{\lambda}_h &= \sqrt{S_{hi}S_{hk}/S_{ik}}, \\ \tilde{\lambda}_i &= \sqrt{S_{hi}S_{ik}/S_{hk}}, \\ \tilde{\lambda}_k &= \sqrt{S_{ik}S_{hk}/S_{hi}} \end{aligned} \quad (2.42)$$

e

$$\begin{aligned} \tilde{\phi}_h &= S_{hh} - S_{hi}S_{hk}/S_{ik}, \\ \tilde{\phi}_i &= S_{ii} - S_{hi}S_{ik}/S_{hk}, \\ \tilde{\phi}_k &= S_{kk} - S_{ik}S_{hk}/S_{hi}. \end{aligned} \quad (2.43)$$

Desta forma podemos ter $p^* = \frac{1}{2}(p-1)(p-2)$ estimativas para cada $\lambda_i, \phi_i, i = 1, 2, \dots, p$.

Sejam $\tilde{\lambda}_{i(j)}$ e $\tilde{\phi}_{i(j)}$ a j -ésima estimativa de λ_i e ϕ_i usando a (2.42) e (2.43), respectivamente, $i = 1, 2, \dots, p$ e $j = 1, 2, \dots, p^*$. Então podemos tomar como estimativas iniciais:

$$\begin{aligned}\lambda_i^{(0)} &= \frac{1}{p^*} \sum_{j=1}^{p^*} \tilde{\lambda}_{i(j)} , \\ \phi_i^{(0)} &= \frac{1}{p^*} \sum_{j=1}^{p^*} \tilde{\phi}_{i(j)} , \quad i = 1, 2, \dots, p.\end{aligned}\tag{2.44}$$

Para μ_i podemos tomar:

$$\mu_i^{(0)} = \bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij} , \quad i = 1, 2, \dots, p.\tag{2.45}$$

(ii) Critérios de convergência

Existem vários critérios de convergência utilizados na literatura. Aqui citamos três deles.

Sejam $\theta^{(m)}$ a estimativa de θ na m -ésima iteração, $m = 0, 1, 2, \dots$

Como critério de convergência do algoritmo EM podemos usar:

$$\|\theta^{(m)} - \theta^{(m-1)}\| < \varepsilon ,\tag{2.46}$$

onde $\|a\|$ indica a norma do vetor a e $\varepsilon > 0$, ou alternativamente

$$\frac{\|\theta^{(m)} - \theta^{(m-1)}\|}{\|\theta^{(m)}\|} < \varepsilon .\tag{2.47}$$

Finalmente, se $\ell(\theta|Y)$ é o logaritmo da função de verossimilhança observada, podemos dizer que o algoritmo EM converge na m -ésima iteração se

$$\ell(\theta^{(m)}|Y) - \ell(\theta^{(m-1)}|Y) < \varepsilon .\tag{2.48}$$

2.3.2 Distribuição assintótica dos EMV

Nesta seção analisamos a distribuição assintótica dos EMV de θ , γ e ρ . Primeiro obtemos a matriz de informação esperada na parametrização de análise de fatores do modelo. Para isto usaremos a função de verossimilhança completa.

Sejam Z_j , $1 \leq j \leq n$, vetores aleatórios iid com densidade $f(z; \theta)$, $\theta \in \Theta \subset \mathbb{R}^r$. Suponha que Z_j contém uma parte observada e uma não observada (ou “missing”). Especificamente, suponha que $Z_j = (Z_{mj}, Z_{0j})$, onde Z_{mj} não é observável de dimensão q e Z_{0j} é observável de dimensão p . Então, podemos escrever:

$$f(z_j; \theta) = f_0(z_{0j}; \theta) f_m(z_{mj}; \theta | z_{0j}) , \quad (2.49)$$

onde $f_0(z_{0j}; \theta)$ é a densidade Z_{0j} é $f_m(z_{mj}; \theta | z_{0j})$ é a densidade condicional de Z_{mj} , dado Z_{0j} .

Daqui segue que a função de verossimilhança completa é dada por

$$\ell(\theta | Z) = \sum_{j=1}^n \ell_{cj} = \ell(\theta | Z_0) + \ell(\theta; Z_0 | Z_m) , \quad (2.50)$$

onde $\ell(\theta | Z_0) = \sum_{j=1}^n \ell_{0j}$ e $\ell(\theta; Z_0 | Z_m) = \sum_{j=1}^n \ell_{mj}$, com

$$\ell_{0j} = \ln f(z_{0j}; \theta), \quad \ell_{mj} = \ln f_m(z_{mj}; \theta | z_{0j}) \quad \text{e} \quad \ell_{cj} = \ell_{0j} + \ell_{mj},$$

$j = 1, 2, \dots, n$.

Aqui, $Z = (Z_1, \dots, Z_n)$, $Z_0 = (Z_{01}, \dots, Z_{0n})$ e $Z_m = (Z_{m1}, \dots, Z_{mn})$.

A correspondente função escore é dada por

$$S_c(\theta | Z) = S_c(\theta | Z_0) + S_c(\theta; Z_0 | Z_m) , \quad (2.51)$$

onde

$$S_c(\theta | Z) = \frac{\partial \ell(\theta | Z)}{\partial \theta} = \sum_{j=1}^n \frac{\partial (\ell_{cj})}{\partial \theta}$$

$$S_c(\theta|Z_0) = \frac{\partial \ell(\theta|Z_0)}{\partial \theta} = \sum_{j=1}^n \frac{\partial(\ell_{0j})}{\partial \theta} \quad e$$

$$S_c(\theta; Z_0|Z_m) = \frac{\partial \ell(\theta; Z_0|Z_m)}{\partial \theta} = \sum_{j=1}^n \frac{\partial(\ell_{mj})}{\partial \theta} .$$

Derivando (2.50) duas vezes com respeito a θ temos:

$$\sum_{j=1}^n \frac{\partial^2(\ell_{cj})}{\partial \theta \partial \theta'} = \sum_{j=1}^n \frac{\partial^2(\ell_{0j})}{\partial \theta \partial \theta'} + \sum_{j=1}^n \frac{\partial^2(\ell_{mj})}{\partial \theta \partial \theta'} . \quad (2.52)$$

Tomando a esperança na distribuição condicional de Z_m dado Z_0 , obtemos

$$\sum_{j=1}^m \frac{\partial^2(q_j(\theta|Z_{0j}))}{\partial \theta \partial \theta'} = \sum_{j=1}^n \frac{\partial^2(\ell_{0j})}{\partial \theta \partial \theta'} + \sum_{j=1}^n \frac{\partial^2(h_j(\theta|Z_{0j}))}{\partial \theta \partial \theta'} , \quad (2.53)$$

onde

$$q_j(\theta|Z_{0j}) = E(\ell_{cj}|Z_{0j}) \quad e$$

$$h_j(\theta|Z_{0j}) = E(\ell_{mj}|Z_{0j}) , \quad j = 1, 2, \dots, n.$$

Note que estamos supondo satisfeitas as usuais condições de regularidade (Sen e Singer, 1993), de modo que as derivadas com respeito a θ podem ser passadas através do sinal integral. Finalmente, de (2.53), temos:

$$J_c(\theta|Z_0) = J_0(\theta|Z_0) + J_m(\theta|Z_0) , \quad (2.54)$$

onde

$$J_c(\theta|Z_0) = -\frac{\partial^2 Q(\theta|Z_0)}{\partial \theta \partial \theta'}$$

é chamada informação completa e

$$Q(\theta|Z_0) = \sum_{j=1}^n q_j(\theta|Z_{0j}) ,$$

$$J_0(\theta|Z_0) = -\frac{\partial^2 \ell(\theta|Z_0)}{\partial \theta \partial \theta'} = -\sum_{j=1}^n \frac{\partial^2(\ell_{0j})}{\partial \theta \partial \theta'}$$

é a informação observada e

$$J_m(\theta|Z_0) = -\frac{\partial^2 H(\theta|Z_0)}{\partial\theta\partial\theta'}$$

é denominada de Informação Perdida, ver Orchard e Woodburg (1972), com

$$H(\theta|Z_0) = \sum_{j=1}^n h_j(\theta|Z_{0j}) .$$

Ou seja, de (2.54) temos que a matriz de informação observada é dada por:

$$J_0(\theta|Z_0) = J_c(\theta|Z_0) - J_m(\theta|Z_0) . \quad (2.55)$$

Louis (1982) mostra que

$$J_m(\theta|Z_0) = E\{S_c(\theta|Z)S'_c(\theta|Z)|Z_0\} - S_c(\theta|Z_0)S'_c(\theta|Z_0) . \quad (2.56)$$

Agora, como $S_c(\hat{\theta}|Z_0) = 0$, onde $\hat{\theta}$ é o EMV de θ , (2.55) fica:

$$J_0(\hat{\theta}|Z_0) = J_c(\hat{\theta}|Z_0) - E\{S_c(\theta|Z)S'_c(\theta|Z)|Z_0\}_{\theta=\hat{\theta}} , \quad (2.57)$$

a qual pode ser útil para efeitos de cálculo quando o algoritmo EM é usado para encontrar o EMV.

Finalmente, tomando a esperança em (2.55) sobre Z_0 , obtemos uma expressão análoga para a matriz de informação esperada por elemento:

$$I_0(\theta) = I_c(\theta) - I_m(\theta) \quad (2.58)$$

com

$$I_c(\theta) = -E\left\{\frac{\partial^2 q_1(\theta|Z_{01})}{\partial\theta\partial\theta'}\right\} \quad \text{e} \quad I_m(\theta) = -E\left\{\frac{\partial^2 h_1(\theta|Z_{01})}{\partial\theta\partial\theta'}\right\} .$$

Alternativamente, notemos que:

$$I_0(\theta) = -E\left\{\frac{\partial^2(\ell_{01})}{\partial\theta\partial\theta'}\right\} . \quad (2.59)$$

A seguir calculamos as quantidades $q_j(\theta|Z_0)$ e $h_j(\theta|Z_0)$ no caso normal.

Lema 2.2 Seja $Z_j = (Z_{mj}; Z_{0j})'$, $1 \leq j \leq n$, vetores aleatórios iid $N_r(\mu, \Sigma)$. Partamos μ e Σ em concordância com Z_j como

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad e \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Então, temos:

$$(i) \quad q_j(\theta|Z_{0j}) = c_q - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} (Z_{0j} - \mu_1)' \Sigma_{22}^{-1} \Sigma_{21} \Sigma^{12} (Z_{0j} - \mu_1) - \frac{1}{2} (Z_{0j} - \mu_1)' \Sigma^{22} (Z_{0j} - \mu_1); \quad (2.60)$$

$$(ii) \quad h_j(\theta|Z_{0j}) = c_h - \frac{1}{2} \ln |\Sigma_{11.2}|, \quad (2.61)$$

onde

$$\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}, \quad \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

e c_q, c_h constantes que não dependem de θ .

Prova: (i) Como $Z_j \sim N(\mu, \Sigma)$, $j = 1, 2, \dots, n$, temos que

$$\ell_{cj} = \ln f(z_j; \theta) = c_z - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} (Z_j - \mu)' \Sigma^{-1} (Z_j - \mu), \quad (2.62)$$

com c_z constante que não depende de θ . Notemos que

$$\begin{aligned} (Z_j - \mu)' \Sigma^{-1} (Z_j - \mu) &= (Z_{mj} - \mu_1)' \Sigma^{11} (Z_{mj} - \mu_1) + (Z_{mj} - \mu_1)' \Sigma^{12} (Z_{0j} - \mu_2) \\ &+ (Z_{0j} - \mu_2)' \Sigma^{21} (Z_{mj} - \mu_1) + (Z_{0j} - \mu_2)' \Sigma^{22} (Z_{0j} - \mu_2). \end{aligned} \quad (2.63)$$

Por outro lado, sabemos que

$$Z_{mj}|Z_{0j} \sim N_q(\mu_m, \Sigma_{11.2}), \quad j = 1, 2, \dots, n, \quad (2.64)$$

onde $\mu_m = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (Z_{0j} - \mu_2)$ e $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$.

Logo, tomando a esperança na distribuição condicional de $Z_{mj}|Z_{0j}$ em (2.62) e utilizando (2.63) e (2.64) obtemos, depois de alguma álgebra, $E(\ell_{cj}|Z_0) = q_j(\theta|Z_{0j})$ dada em (2.60).

(ii) Como $Z_{mj}|Z_{0j} \sim N_q(\mu_m, \Sigma_{11.2})$, $j = 1, 2, \dots, n$, temos que, ver (2.50),

$$\ell_{mj} = \ln(2\pi)^{-q/2} - \frac{1}{2} \ln |\Sigma_{11.2}| - \frac{1}{2} (Z_{mj} - \mu_m)' \Sigma_{11.2}^{-1} (Z_{mj} - \mu_m). \quad (2.65)$$

Assim, obtemos

$$E(\ell_{mj}|Z_{0j}) = -\frac{q}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_{11.2}| - \frac{1}{2} \text{tr}(\Sigma_{11.2}^{-1} \Sigma_{11.2}) = c_h - \frac{1}{2} \ln |\Sigma_{11.2}|$$

dada em (2.61).

Voltando ao nosso problema, temos o seguinte resultado:

Lema 2.3 *Sejam $\{Z_j\}_{j=1}$ como em (2.33). Então,*

$$(i) \quad q_j(\theta|Y_j) = c_q - \frac{1}{2} \ln |D(\phi)| + \frac{1}{2c} [\lambda' D^{-1}(\phi)(Y_j - \mu)]^2 - \frac{1}{2} (Y_j - \mu)' D^{-1}(\phi)(Y_j - \mu) \quad (2.66)$$

$$(ii) \quad h_j(\theta|Y_j) = c_h + \frac{1}{2} \ln c, \quad (2.67)$$

onde $c = 1 + \lambda' D^{-1}(\phi)\lambda$.

Prova: Neste caso, temos que $Z_{mj} = F_j$ de dimensão $q = 1$ e $Z_{0j} = Y_j$ de dimensão p ; $r = p + 1$; $\mu_1 = 0$, $\mu_2 = \mu$ e

$$\Sigma = \Sigma_Z = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 1 & \lambda' \\ \lambda & \lambda\lambda' + D(\phi) \end{bmatrix}.$$

(i) De (2.35) segue que $|\Sigma| = |D(\phi)|$, $\Sigma^{22} = D^{-1}(\phi)$ e $\Sigma^{12} = -\lambda' D^{-1}(\phi)$. Agora, notando que $\Sigma_{22}^{-1} \Sigma_{21} \Sigma^{12} = -c^{-1} D^{-1}(\phi) \lambda \lambda' D^{-1}(\phi)$ e usando a parte (i) do Lema 2.2, o resultado segue.

(ii) Basta notar que $|\Sigma_{11.2}| = |\Sigma| |\Sigma_{22}| = c^{-1}$ e usar a parte (ii) do Lema 2.2.

No próximo resultado apresentamos a matriz de informação na parametrização de análise de fatores do modelo estrutural normal.

Lema 2.4 *Sob o modelo (2.31), temos que a matriz de informação esperada completa, por observação, denotada por $I_{Fc}(\theta)$, é dada por*

$$I_{Fc}(\theta) = I_{F0}(\theta) + I_{Fm}(\theta) , \quad (2.68)$$

onde:

$$I_{F0}(\theta) = \begin{bmatrix} I_{\mu\mu} & 0 & 0 \\ 0 & I_{\lambda\lambda} & I_{\lambda\phi} \\ 0 & I_{\phi\lambda} & I_{\phi\phi} \end{bmatrix} , \quad (2.69)$$

com

$$\begin{aligned} I_{\mu\mu} &= (\lambda\lambda' + D(\phi))^{-1} , \quad I_{\lambda\lambda} = (1 - c^{-1})D^{-1}(\phi) - c^{-1}(1 - 2c^{-1})M , \\ I_{\lambda\phi} &= c^{-1}D^{-2}(\phi)D(\lambda) - c^{-2}MD^{-1}(\phi)D(\lambda) , \\ I_{\phi\phi} &= \frac{1}{2}D^{-2}(\phi) - c^{-1}D^{-3}(\phi)D(\lambda)D(\lambda) + \frac{c^{-2}}{2}D(\lambda)D^{-1}(\phi)MD^{-1}(\phi)D(\lambda) , \\ I_{\phi\lambda} &= I'_{\lambda\phi} , \quad M = D^{-1}(\phi)\lambda\lambda'D^{-1}(\phi) \quad e \\ D^{-i}(\phi) &= \text{Diag}(\phi_1^{-i}, \phi_2^{-i}, \dots, \phi_p^{-i}) , \quad i = 1, 2, 3; \end{aligned}$$

é a matriz de informação esperada por elemento e

$$I_{Fm}(\theta) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & H_{\lambda\lambda} & H_{\lambda\phi} \\ 0 & H_{\phi\lambda} & H_{\phi\phi} \end{bmatrix} , \quad (2.70)$$

com

$$\begin{aligned} H_{\lambda\lambda} &= -c^{-1}(V^{-1} - c^{-1}M) ; \\ H_{\lambda\phi} &= c^{-1}V^{-1}D^{-1}(\phi)D(\lambda) ; \\ H_{\phi\phi} &= -c^{-1}D^{-3}(\phi)D(\lambda)D(\lambda) + \frac{c^{-2}}{2}D(\lambda)D^{-1}(\phi)MD^{-1}(\phi)D(\lambda) ; \\ H_{\phi\lambda} &= H'_{\lambda\phi} \quad e \quad V = \lambda\lambda' + D(\lambda) \end{aligned}$$

é a matriz de informação perdida por observação.

Prova: Ver apêndice C.

Um fato interessante de ser salientado é que a informação perdida para μ é zero.

A seguir, apresentamos a distribuição assintótica do EMV de θ , $\hat{\theta}$.

Teorema 2.2 *Sob o modelo (2.31), temos que*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N_{3p}(0, \Omega_\theta) ,$$

onde

$$\Omega_\theta = \begin{bmatrix} V & 0 & 0 \\ 0 & \Omega_\lambda & \Omega_{\lambda\phi} \\ 0 & \Omega_{\phi\lambda} & \Omega_\phi \end{bmatrix} , \quad (2.71)$$

com

$$V = \lambda\lambda' + D(\phi) ,$$

$$\Omega_\lambda = D(\phi) + D(\phi)D(\varphi) + \frac{1}{2}[\lambda\lambda' - \omega^{-1}(\lambda * \varphi)(\lambda * \varphi)'] ,$$

$$\Omega_{\lambda\phi} = -2D(\phi)D(\lambda)D(\varphi) + \omega^{-1}(\lambda * \varphi)(\phi * \psi)' ,$$

$$\Omega_\phi = 2\tau_T D(\phi)D(\varphi)D(\phi) - 2\omega^{-1}(\phi * \psi)(\phi * \psi)' ,$$

$$\Omega_{\phi\lambda} = \Omega'_{\lambda\phi} , \quad \varphi = (\varphi_1, \varphi_2, \dots, \varphi_p)' , \quad \psi = (\psi_1, \dots, \psi_p)' ,$$

$$\varphi_i = (\tau_T - 2\tau_i)^{-1} , \quad \psi_i = \tau_i/(\tau_T - 2\tau_i) = \tau_i\varphi_i , \quad i = 1, 2, \dots, p,$$

$$\omega = 1 + \frac{1}{\tau_T} \lambda' D(\psi) D^{-1}(\phi) \lambda = 1 + \tau_T^{-1} \sum_{i=1}^p \left\{ \frac{\tau_i^2}{\tau_T - 2\tau_i} \right\} ,$$

$$\tau_T = \sum_{i=1}^p \tau_i ,$$

onde estamos supondo que $\tau_T \neq 2\tau_i$, $\forall i = 1, 2, \dots, p$; τ_i , $i = 1, 2, \dots, p$, é a precisão relativa definida em (1.22) e $a * b = (a_1 b_1, \dots, a_p b_p)'$ é o produto de Hadamard entre os vetores a e b .

Prova: Sabemos que (Sen e Singer, 1993)

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N_{3p}(0, I_{F_0}^{-1}(\theta)) .$$

Agora, usando os Teoremas 8.2.1 e 8.9.3. de Graybill (1983), temos depois de alguma álgebra que $I_{F_0}^{-1}(\theta) = \Omega_\theta$, dada em (2.71). Lawley e Maxwell (1971) obtêm (2.71) usando resultados gerais de análise de fatores.

De (2.71) segue que $\hat{\mu}$ e $(\hat{\lambda}, \hat{\phi})$ são assintoticamente independentes.

A matriz de informação esperada para γ na versão de regressão do modelo estrutural pode ser obtida usando (1.13).

Seja $\hat{\gamma}$ o EMV de γ , então temos o seguinte resultado:

Corolário 2.1 *Sob o modelo (2.29) temos*

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{\mathcal{D}} N_{3p}(0, \Omega_\gamma) ,$$

onde $\Omega_\gamma = G\Omega_\theta G'$, com G dada em (1.14) e

$$\theta = a(\gamma) = (\mu_2, \alpha' + \beta' \mu_x, \sigma_x, \sigma_x \beta', \phi')' .$$

Prova: Basta notar que $\gamma = g(\theta)$ e usar o método delta, ver (1.14).

Seja $\gamma = (\alpha'_*, \beta'_*, \phi')$, onde $\alpha_* = \begin{bmatrix} \mu_x \\ \alpha \end{bmatrix}$ e $\beta_* = \begin{pmatrix} \sigma_x \\ \beta \end{pmatrix}$. Logo, podemos escrever:

$$\Omega_\gamma = \begin{bmatrix} \Omega_{\alpha_*} & \Omega_{\alpha_*\beta_*} & \Omega_{\alpha_*\phi} \\ & \Omega_{\beta_*} & \Omega_{\beta_*\phi} \\ \text{simetria} & & \Omega_\phi \end{bmatrix} , \quad (2.72)$$

com

$$\Omega_{\alpha_*} = G_{11}V G'_{11} + G_{12}\Omega_\lambda G'_{12}$$

$$\Omega_{\alpha_*\beta_*} = G_{12}\Omega_\lambda G'_{22} , \quad \Omega_{\alpha_*\phi} = G_{12}\Omega_{\lambda\phi}$$

$$\Omega_{\beta_*} = G_{22}\Omega_\lambda G'_{22} , \quad \Omega_{\beta_*\phi} = G_{22}\Omega_{\lambda\phi} ,$$

$V, \Omega_\lambda, \Omega_{\lambda\phi}$ e Ω_ϕ como em (2.71) e G_{11}, G_{12} e G_{22} como em (1.14).

Sejam

$$\theta^* = (\mu', \lambda', \rho')' \quad \text{e} \quad \gamma^* = (\alpha', \beta', \rho')' , \quad (2.73)$$

com $\rho = (\rho_1, \rho_2, \dots, \rho_p)'$.

Corolário 2.2 (i) *Sob o modelo (2.31), temos*

$$\sqrt{n}(\hat{\theta}^* - \theta^*) \xrightarrow{\mathcal{D}} N_{3p}(0, \Omega_{\theta^*}) ,$$

onde

$$\Omega_{\theta^*} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Omega_\lambda & \Omega_{\lambda\rho} \\ 0 & \Omega_{\rho\lambda} & \Omega_\rho \end{bmatrix}, \quad (2.74)$$

com

$$\Omega_\mu = \lambda\lambda' + D(\phi), \quad \Omega_{\lambda\rho} = \Omega'_{\rho\lambda} = \Omega_\lambda G_1 + \Omega_{\lambda\phi} G_2;$$

$$\Omega_\rho = G_1 \Omega_\lambda G_1 + G_2 \Omega_{\rho\lambda} G_1 + G_1 \Omega_{\lambda\rho} G_2 + G_2 \Omega_\phi G_2;$$

$$G_1 = 2\text{Diag}(\lambda_1^{-1}\rho_1(1-\rho_1), \dots, \lambda_p^{-1}\rho_p(1-\rho_p));$$

$$G_2 = -\text{Diag}(\lambda_1^{-2}\rho_1^2, \dots, \lambda_p^{-2}\rho_p^2),$$

matrizes diagonais de ordem p ; Ω_λ , $\Omega_{\lambda\phi}$ e Ω_ϕ como em (2.71).

(ii) Sob o modelo (2.29), temos que

$$\sqrt{n}(\hat{\gamma}^* - \gamma^*) \xrightarrow{\mathcal{D}} N_{3p-2}(0; \Omega_{\gamma^*}),$$

onde

$$\Omega_\gamma = \begin{bmatrix} \Omega_\alpha & \Omega_{\alpha\beta} & \Omega_{\alpha\rho} \\ \text{simetria} & \Omega_\beta & \Omega_{\beta\rho} \\ & & \Omega_\rho \end{bmatrix}, \quad (2.75)$$

com

$$\Omega_\alpha = G_1 \Omega_{\alpha^*} G_1', \quad \Omega_{\alpha\beta} = G_1 \Omega_{\alpha^*\beta^*} G_1';$$

$$\Omega_{\alpha\rho} = G_1 \Omega_{\alpha^*\beta^*} G_2' + G_1 \Omega_{\alpha^*\phi^*} G_3';$$

$$\Omega_\beta = G_1 \Omega_{\beta^*} G_1', \quad \Omega_{\beta\rho} = G_1 \Omega_{\beta^*} G_2' + G_1 \Omega_{\beta^*\phi^*} G_3';$$

$$\Omega_\rho = G_2 \Omega_{\beta^*} G_2' + G_3 \Omega_{\phi\beta^*} G_2' + G_2 \Omega_{\beta^*\phi^*} G_3' + G_3 \Omega_{\phi^*} G_3';$$

$$G_1 = (0, I_{p-1}), \text{ uma matriz } (p-1) \times p;$$

$$G_2 = (g_{21}, G_{22}), \text{ uma matriz } p \times p, \text{ com}$$

$$g_{21} = 2\sigma_x^{-1}(\rho, (1-\rho_1), \dots, \rho_p(1-\rho_p))' \text{ um vetor } p \times 1,$$

$$G_{22} = \begin{pmatrix} 0 \\ G_a \end{pmatrix} \text{ uma matriz } p \times (p-1), \text{ com } 0 \text{ vetor nulo } 1 \times (p-1) \text{ e}$$

$$G_a = 2\text{Diag}(\beta_2^{-1}\rho_2(1-\rho_2), \dots, \beta_p^{-1}\rho_p(1-\rho_p)) \text{ uma matriz diagonal de ordem } p-1;$$

$$G_3 = -\sigma_x^{-2}\text{Diag}(\rho_1^2, \beta_2^{-2}\rho_2^2, \dots, \beta_p^{-2}\rho_p^2) \text{ uma matriz diagonal de ordem } p.$$

Finalmente, as matrizes $\Omega_{\alpha_*}, \Omega_{\alpha_*\beta_*}, \Omega_{\alpha_*\phi}, \Omega_{\beta_*}, \Omega_{\beta_*\phi}, \Omega_{\phi\beta_*}$ e Ω_ϕ são como em (2.72).

Prova: (i) Notamos que $\theta^* = g(\theta) = (g_1(\theta), \dots, g_{3p}(\theta))$, onde g segue de (1.20) e (2.73).

Logo, usando o método delta temos que

$$\sqrt{n}(\hat{\theta}^* - \theta^*) \xrightarrow{\mathcal{D}} N_{3p}(0, G\Omega_\theta G') ,$$

onde Ω_θ é como em (2.71) e

$$G = \left[\frac{\partial(g_i(\theta))}{\partial\theta_j} \right] = \begin{bmatrix} I_p & 0 & 0 \\ 0 & I_p & 0 \\ 0 & G_1 & G_2 \end{bmatrix}$$

é uma matriz $3p \times 3p$. Daqui segue que $\Omega_{\theta^*} = G\Omega_\theta G'$.

A prova de (ii) é similar e será omitida. Ver Bolfarine e Galea (1995b).

2.3.3 Região de confiança assintótica para ρ

Usando os resultados da seção anterior, podemos obter intervalos de confiança assintóticos para qualquer parâmetro de interesse. Aqui apresentamos uma região de confiança assintótica para $\rho = (\rho_1, \rho_2, \dots, \rho_p)'$. Para isto utilizaremos o seguinte fato devido a Sidák (1968):

Lema 2.5 Se $W = (w_1, w_2, \dots, w_p)'$ tem uma distribuição $N_p(0, \Sigma)$, então temos que

$$P \left[\bigcap_{i=1}^p \{|W_i| \leq c_i\} \right] \geq \prod_{i=1}^p P[|W_i| \leq c_i] , \quad (2.76)$$

para todo $c = (c_1, c_2, \dots, c_p)'$, $c_i > 0$, $i = 1, 2, \dots, p$.

Notemos que a condição sobre a distribuição de W é a normalidade e média zero. Mas este resultado também é válido para a classe de distribuições elípticas, ver Das Gupta et al. (1972) e Tong (1990).

Por outro lado, do Corolário 2.2, temos que

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{\mathcal{D}} N_p(0, \Omega_\rho) , \quad (2.77)$$

onde Ω_p como em (2.74) ou (2.75). Sejam $\omega_{11}, \dots, \omega_{pp}$ os elementos da diagonal de Ω_p .

Logo, usando o Lema 2.5 e (2.77), temos

$$P \left[\bigcap_{i=1}^p \left\{ \frac{\sqrt{n}|\hat{\rho}_i - \rho_i|}{\sqrt{\omega_{ii}}} \leq c_i \right\} \right] \geq \prod_{i=1}^p P \left[\left| \frac{\sqrt{n}(\hat{\rho}_i - \rho_i)}{\sqrt{\omega_{ii}}} \right| \leq c_i \right]. \quad (2.78)$$

Seja $c_i = z_a$, $i = 1, 2, \dots, p$, onde $a = 1 - (1 - \alpha_a)^{1/p}$ e z_a é o percentil $100(1 - \frac{a}{2})$ da distribuição normal padrão. Assim, (2.78) fica

$$P \left[\bigcap_{i=1}^p \left\{ \frac{\sqrt{n}|\hat{\rho}_i - \rho_i|}{\sqrt{\omega_{ii}}} \leq z_a \right\} \right] \geq 1 - \alpha_a;$$

e a região de confiança simultânea de $100(1 - \alpha_a)\%$ é dada por

$$\hat{\rho}_i - z_a \sqrt{\frac{\hat{\omega}_{ii}}{n}} \leq \rho_i \leq \hat{\rho}_i + z_a \sqrt{\frac{\hat{\omega}_{ii}}{n}}, \quad (2.79)$$

onde $\hat{\omega}_{ii}$ é o EMV de ω_{ii} , $i = 1, 2, \dots, p$.

Notemos que a região (2.79) está baseada no método do módulo máximo para construir regiões de confiança simultâneas, ver Seber (1977). Shyr e Gleser (1986) usam este método para construir uma região de confiança simultânea para as precisões $\pi = (\pi_1, \pi_2, \dots, \pi_p)'$. Alternativamente, poderíamos utilizar o método de Scheffé ou Bonferroni para construir intervalos de confiança simultâneos para $\rho_1, \rho_2, \dots, \rho_p$. Mas, em geral, esses métodos produzem intervalos de maior comprimento, ver Seber (1977), Shyr e Gleser (1986) e Savin (1980).

O mesmo procedimento pode ser seguido para construir intervalos de confiança simultâneos para os elementos de $\pi = (\pi_1, \pi_2, \dots, \pi_p)'$ e $\tau = (\tau_1, \tau_2, \dots, \tau_p)'$; as precisões e precisões relativas, respectivamente.

Do Corolário 2.1 e (2.72), temos que:

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_* - \beta_* \\ \hat{\phi} - \phi \end{pmatrix} \xrightarrow{D} N_{2p} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{\beta_*} & \Omega_{\beta_*\phi} \\ \Omega_{\phi\beta_*} & \Omega_{\phi} \end{pmatrix} \right), \quad (2.80)$$

de onde segue que

$$\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{D} N_p(0, \Omega_{\pi}), \quad (2.81)$$

com

$$\begin{aligned}\Omega_\pi &= G_1 \Omega_{\beta_*} G_1 + G_2 \Omega_{\phi \beta_*} G_1 + G_1 \Omega_{\beta_* \phi} G_2 + G_2 \Omega_\phi G_2 ; \\ G_1 &= 2\text{Diag}(0, \beta_2/\phi_2, \dots, \beta_p/\phi_p) \text{ e} \\ G_2 &= -\text{Diag}(1/\phi_1^2, \beta_2^2/\phi_2^2, \dots, \beta_p^2/\phi_p^2) ,\end{aligned}$$

são matrizes diagonais de ordem p . Shyr e Gleser (1986) obtêm (2.81) usando resultados gerais para modelos multivariados com erros nas variáveis (Fuller, 1987). Contudo, a expressão dada por Shyr e Gleser (1986) para Ω_π depende, como eles mesmos reconhecem, da inversa de uma matriz que não tem forma explícita; o que não acontece em nosso caso.

De forma similar, usando a parametrização de análise de fatores; do Teorema 2.2 segue que

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{D} N_p(0, \Omega_\tau) , \quad (2.82)$$

com

$$\Omega_\tau = G_1 \Omega_\lambda G_1 + G_2 \Omega_{\phi \lambda} G_1 + G_1 \Omega_{\lambda \phi} G_2 + G_2 \Omega_\phi G_2 ,$$

onde, agora,

$$\begin{aligned}G_1 &= 2\text{Diag}(\lambda_1/\phi_1, \lambda_2/\phi_2, \dots, \lambda_p/\phi_p) \text{ e} \\ G_2 &= -\text{Diag}(\lambda_1^2/\phi_1^2, \dots, \lambda_p^2/\phi_p^2) .\end{aligned}$$

As matrizes Ω_λ , $\Omega_{\phi \lambda}$, $\Omega_{\lambda \phi}$ e Ω_ϕ são como em (2.71).

Podemos, então, usar estes resultados ((2.81) ou (2.82)) e o Lema 2.5 para construir intervalos de confiança simultâneos para os elementos de π ou τ .

2.3.4 Testes de hipóteses

Nesta seção apresentamos testes assintóticos para as hipóteses indicadas na Seção 1.3 do Capítulo 1.

Uma escolha natural para testar as hipóteses, H_{01} , H_{02} , H_{03} e H_{04} (ver Seção 1.3), em amostras grandes, é o teste da razão de verossimilhança (RV). Mas este teste requer das EMV dos parâmetros sob a hipótese nula, para o qual precisamos de um processo iterativo, já que, em geral, não é possível obter expressões explícitas para estas estimativas sob a hipótese nula. O teste de Wald é equivalente ao teste da RV em amostras grandes e tem a vantagem de não requerer as EMV sob a hipótese nula. Devido a isto, ilustraremos o uso do teste de Wald para provar as hipóteses H_{01} , H_{02} , H_{03} e H_{04} ; na versão de análise de fatores do modelo estrutural. Para algumas propriedades deste teste ver Cox e Hinkley (1974).

Para provar se os p instrumentos medem a característica sem vício e com a mesma confiabilidade, a hipótese é $H_{01} : \mu_1 = \mu_2 = \dots = \mu_p, \lambda_1 = \lambda_2 = \dots = \lambda_p, \rho_1 = \rho_2 = \dots = \rho_p$. Notamos que H_{01} pode ser escrita como $H_{01} : C\theta^* = q_0$, onde

$$C = \begin{bmatrix} C_1 & 0 & 0 \\ 0 & C_1 & 0 \\ 0 & 0 & C_1 \end{bmatrix}, \quad (2.83)$$

uma matriz $3(p-1) \times 3p$ de posto $3(p-1)$, com

$$C_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix}, \quad (2.84)$$

uma matriz $(p-1) \times p$, $q_0 = 0$ vetor unitário nulo de dimensão $3(p-1)$.

Logo, a estatística de Wald para testar H_{01} , denotada por W_{01} , é dada por

$$W_{01} = n(C\hat{\theta}^* - q_0)'(C\hat{\Omega}_{\theta^*}C')^{-1}(C\hat{\theta}^* - q_0); \quad (2.85)$$

que converge em distribuição a uma variável aleatória $\chi^2(3(p-1))$ sob H_{01} ; $\hat{\Omega}_{\theta^*}$ é a matriz Ω_{θ^*} dada em (2.74) avaliada no EMV de θ^* , $\hat{\theta}^*$.

Portanto, H_{01} é rejeitada ao nível α se

$$W_{01} > \chi^2_{1-\alpha}(3(p-1)), \quad (2.86)$$

onde $\chi^2_{1-\alpha}(3(p-1))$ denota o percentil $100(1-\alpha)$ da distribuição $\chi^2(3(p-1))$.

Para testar se os instrumentos medem sem vício, temos que provar a hipótese $H_{02} : \mu_1 = \mu_2 = \dots = \mu_p, \lambda_1 = \lambda_2 = \dots = \lambda_p$. Também, neste caso, H_{02} pode ser escrita como $H_{02} : C\theta^* = q_0$, onde, agora,

$$C = \begin{bmatrix} C_1 & 0 & 0 \\ 0 & C_1 & 0 \end{bmatrix},$$

uma matriz $2(p-1) \times 3p$ de posto $2(p-1)$; C_1 como em (2.84) e $q_0 = 0$ um vetor de dimensão $2(p-1)$.

Assim, a estatística de Wald para provar H_{02} , denotada por W_{02} , é:

$$W_{02} = n(C\hat{\theta}^*)'(C\hat{\Omega}_{\theta^*}C')^{-1}(C\hat{\theta}^*) = W_{02\mu} + W_{02\lambda} \xrightarrow{D} \chi^2(2(p-1)), \quad (2.87)$$

sob H_{02} , onde

$$W_{02\mu} = n\hat{\mu}'C_1'(C_1\hat{\Omega}_\mu C_1')^{-1}C_1\hat{\mu} \text{ e}$$

$$W_{02\lambda} = n\hat{\lambda}'C_1'(C_1\hat{\Omega}_\lambda C_1')^{-1}C_1\hat{\lambda},$$

onde $\hat{\Omega}_\mu$ e $\hat{\Omega}_\lambda$ são as matrizes Ω_μ e Ω_λ dadas em (2.74) avaliadas no EMV $\hat{\theta}$.

Para provar se os instrumentos são igualmente confiáveis a hipótese de interesse é $H_{03} : \rho_1 = \rho_2 = \dots = \rho_p$. A estatística de Wald, denotada por W_{03} , é

$$W_{03} = n\hat{\rho}'C_1'(C_1\hat{\Omega}_\rho C_1')^{-1}C_1\hat{\rho}, \quad (2.88)$$

a qual $\xrightarrow{D} \chi^2(p-1)$ sob H_{03} , onde $\hat{\Omega}_\rho$ é a matriz Ω_ρ dada em (2.74) avaliada em $\hat{\theta}$ e C_1 como definida em (2.84).

Finalmente, para testar $H_{04} : \lambda_1 = \lambda_2 = \dots = \lambda_p$, a estatística de Wald é $W_{04} = W_{02\lambda}$, dada em (2.87), a qual tem $p-1$ graus de liberdade.

De maneira similar, usando os resultados dados na Seção 2.3.3, podemos provar hipóteses relativas a igualdade de precisões ou precisões relativas. Também podemos usar a versão de regressão do modelo estrutural para testar hipóteses de interesse como as citadas acima.

2.4 O modelo estrutural quando o instrumento de referência mede sem erro a característica de interesse

Existem situações, especialmente na área da medicina, ver Kaaks et al. (1994), Behre et al. (1989) e Leurgans (1980), onde é possível considerar, numa primeira aproximação, que o instrumento de referência mede a característica de interesse sem erro de medição. Usando a notação da seção anterior, isto significa $\phi_1 = 0$ ou equivalentemente $\rho_1 = 1$; isto é, temos que $y_{ij} = x_j$, $j = 1, 2, \dots, n$. Esta suposição, se for razoável, é atrativa pois simplifica a obtenção dos EMV já que, neste caso, não precisamos de um procedimento iterativo. Alternativamente, alguns autores, ver Jaech (1985), recomendaram tomar $\phi_1 = 0$ se ϕ_1 for estimado negativamente. Por outro lado, no caso $p = 2$, a restrição $\phi_1 = 0$ (ou $\rho_1 = 1$) é usada para identificar o modelo estrutural, ver Kaaks et al. (1994).

Nesta seção apresentamos expressões explícitas para os EMV quando $\phi_1 = 0$, válidas para dois ou mais instrumentos de medição. Também damos a matriz de covariância assintótica dos EMV. Como nas seções anteriores, usaremos as duas parametrizações do modelo.

A versão de análise de fatores do modelo, neste caso, é

$$Y_j = \mu + Lv_j, \quad j = 1, 2, \dots, n; \quad (2.89)$$

$$v_j = (F_j, \varepsilon_{2j}, \dots, \varepsilon_{pj})', \quad 1 \leq j \leq n \text{ são iid } N_p \left(0, \begin{pmatrix} 1 & 0 \\ 0 & D(\phi) \end{pmatrix} \right)$$

onde $\mu = (\mu_1, \mu_2, \dots, \mu_p)'$, $L = [\lambda, \begin{pmatrix} 0 \\ I_{p-1} \end{pmatrix}]$ uma matriz $p \times p$, $D(\phi) = \text{Diag}(\phi_2, \phi_3, \dots, \phi_p)$ uma matriz diagonal $(p-1) \times (p-1)$ e $\phi = (\phi_2, \phi_3, \dots, \phi_p)'$ agora de dimensão $p-1$.

A versão de regressão é dada por:

$$Y_j = a + Bu_j, \quad 1 \leq j \leq n, \quad (2.90)$$

$$u_j = (x_j, \varepsilon_{2j}, \dots, \varepsilon_{pj})', \quad 1 \leq j \leq n, \text{ são iid } N_p \left(\begin{pmatrix} \mu_x \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & D(\phi) \end{pmatrix} \right),$$

com $\mathbf{a} = \begin{pmatrix} 0 \\ \alpha \end{pmatrix}$ um vetor $p \times 1$, $B = [\mathbf{b}; \begin{pmatrix} 0 \\ I_{p-1} \end{pmatrix}]$ uma matriz $p \times p$ e $\mathbf{b} = \begin{pmatrix} 1 \\ \beta \end{pmatrix}$ um vetor $p \times 1$.

Aqui, $\theta = (\mu', \lambda', \phi')'$, onde, agora, ϕ é um vetor $(p-1) \times 1$ dado em (2.89), e

$$\begin{aligned} \gamma &= (\mu_x, \alpha', \sigma_x, \beta', \phi')' = g(\theta) \\ &= (\mu_1, \mu_2 - (\lambda_2/\lambda_1)\mu_1, \dots, \mu_p - (\lambda_p/\lambda_1)\mu_1, \lambda_1, (\lambda_2/\lambda_1), \dots, (\lambda_p/\lambda_1), \phi_2, \dots, \phi_p)' \end{aligned} \quad (2.91)$$

De (2.89) é fácil ver que

$$Y_j \sim N_p(\mu, V), \quad j = 1, 2, \dots, n,$$

onde

$$V = \begin{pmatrix} \lambda_1^2 & \lambda_1 \lambda'_* \\ \lambda_1 \lambda_* & \lambda_* \lambda'_* + D(\phi) \end{pmatrix}, \quad (2.92)$$

com $\lambda_* = (\lambda_2, \lambda_3, \dots, \lambda_p)'$, de onde segue que o logaritmo da função de verossimilhança observada é

$$\ell(\theta/Y) = \text{cte.} - \frac{n}{2} \ln |V| - \frac{1}{2} \sum_{j=1}^n (Y_j - \mu)' V^{-1} (Y_j - \mu). \quad (2.93)$$

Usando o Teorema 8.9.3 de Graybill (1983), temos

$$|V| = \lambda_1^2 \prod_{i=2}^p \phi_i \quad \text{e} \quad V^{-1} = \begin{pmatrix} c/\lambda_1^2 & -(1/\lambda_1)\lambda'_* D^{-1}(\phi) \\ -(1/\lambda_1)D^{-1}(\phi)\lambda_* & D^{-1}(\phi) \end{pmatrix}, \quad (2.94)$$

com $c = 1 + \sum_{i=2}^p (\lambda_i^2/\phi_i)$. Logo, usando (2.94), temos

$$\ell(\theta/Y) = \text{cte.} - \frac{n}{2} \ln \lambda_1^2 - \frac{n}{2} \sum_{i=2}^p \ln \phi_i - \frac{1}{2} \sum_{j=1}^n f_j, \quad (2.95)$$

com

$$f_j = \frac{c}{\lambda_1^2} (y_{ij} - \mu_i)^2 - \frac{2}{\lambda_1} (y_{ij} - \mu_i) \sum_{i=2}^p (\lambda_i/\phi_i) (y_{ij} - \mu_i) + \sum_{i=2}^p (1/\phi_i) (y_{ij} - \mu_i)^2,$$

$j = 1, 2, \dots, n$. De (2.95) é fácil ver que os EMV são:

$$\begin{aligned} \hat{\mu}_i &= \bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij}, \quad i = 1, 2, \dots, p; \\ \hat{\lambda}_1 &= \sqrt{S_{11}}, \quad \hat{\lambda}_i = S_{1i}/\sqrt{S_{11}}, \quad i = 2, 3, \dots, p; \\ \hat{\phi}_i &= S_{ii} - S_{1i}^2/S_{11}, \quad i = 2, 3, \dots, p, \end{aligned} \quad (2.96)$$

com

$$S_{kl} = \frac{1}{n} \sum_{j=1}^n (y_{kj} - \bar{y}_k)(y_{lj} - \bar{y}_l) , \quad k, l = 1, 2, \dots, p.$$

Usando estes estimadores e (2.94), obtemos o EMV de γ : $\hat{\gamma} = g(\hat{\theta})$.

No resultado seguinte, apresentamos a matriz de informação.

Lema 2.6 *Sob o modelo (2.89), a matriz de informação esperada por elemento denotado por $I_1(\theta)$ é:*

$$I_1(\theta) = \begin{bmatrix} I_{\mu\mu} & 0 & 0 \\ 0 & I_{\lambda\lambda} & 0 \\ 0 & 0 & I_{\phi\phi} \end{bmatrix} , \quad (2.97)$$

com

$$\begin{aligned} I_{\mu\mu} &= V^{-1} , \\ I_{\lambda\lambda} &= \begin{bmatrix} (1+c)/\lambda_1^2 & (-\lambda_1)\lambda'_* D^{-1}(\phi) \\ (-\lambda_1)D^{-1}(\phi)\lambda_* & D^{-1}(\phi) \end{bmatrix} , \\ I_{\phi\phi} &= \text{Diag}(1/2\phi_2^2, \dots, 1/2\phi_p^2) , \end{aligned}$$

V como em (2.92) e c como definido em (2.94).

Prova: Diferenciando duas vezes (2.95) (ou (2.93)) em relação a θ e tomando a esperança, o resultado segue, depois de alguma álgebra. Ver Apêndice C.

Teorema 2.3 *Sob o modelo (2.89), temos que*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N_{3p-1}(0; \Omega_\theta) ,$$

onde

$$\Omega_\theta = \begin{bmatrix} V & 0 & 0 \\ 0 & \Omega_\lambda & 0 \\ 0 & 0 & \Omega_\phi \end{bmatrix} , \quad (2.98)$$

com

$$\begin{aligned} V &\text{ como em (2.92),} \\ \Omega_\lambda &= \begin{pmatrix} \frac{\lambda_1^2}{2} & -(\lambda_1/2)\lambda'_* \\ -(\lambda_1/2)\lambda_* & D(\phi) + \frac{1}{2}\lambda_*\lambda'_* \end{pmatrix} e \\ \Omega_\phi &= 2\text{Diag}(\phi_2^2, \dots, \phi_p^2) . \end{aligned}$$

Prova: Sabemos que $\Omega_\theta = I_1^{-1}(\theta)$ e usando o Teorema 8.2.1 de Graybill (1983) o resultado segue.

Notemos que $\hat{\mu}$, $\hat{\lambda}$ e $\hat{\phi}$ são, neste caso, assintoticamente independentes. Além disso, os EMV $\hat{\phi}_2, \dots, \hat{\phi}_p$ são também assintoticamente independentes, o que não acontece no caso geral.

Corolário 2.3 *Sob o modelo (2.90) temos que*

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{D} N_{3p-1}(0, \Omega_\gamma) ,$$

onde

$$\Omega_\gamma = \begin{bmatrix} \Omega_{\alpha_*} & \Omega_{\alpha_*\beta_*} & 0 \\ \Omega_{\beta_*\alpha_*} & \Omega_{\beta_*} & 0 \\ 0 & 0 & \Omega_\phi \end{bmatrix} , \quad (2.99)$$

com

$$\Omega_{\alpha_*} = G_{11}V'G'_{11} + G_{12}\Omega_\lambda G'_{12} ,$$

$$\Omega_{\alpha_*\beta_*} = G_{12}\Omega_\lambda G'_{22} = \Omega'_{\beta_*\alpha_*} ,$$

$$\Omega_{\beta_*} = G_{22}\Omega_\lambda G'_{22};$$

V , Ω_λ e Ω_ϕ como em (2.98), $\alpha_* = \begin{pmatrix} \mu_x \\ \alpha \end{pmatrix}$, $\beta_* = \begin{pmatrix} \sigma_x \\ \beta \end{pmatrix}$, G_{11} , G_{12} e G_{22} como definidas em (1.14).

Prova: De (2.91) temos que $\gamma = g(\theta)$, aplicando o método delta obtemos:

$$\Omega_\gamma = G\Omega_\theta G' , \quad (2.100)$$

onde

$$G = \begin{bmatrix} G_{11} & G_{12} & 0 \\ 0 & G_{22} & 0 \\ 0 & 0 & I_{p-1} \end{bmatrix}$$

é uma matriz $(3p-1) \times (3p-1)$, $\theta = a(\gamma) = (\mu_x, \alpha' + \beta'\mu_x, \sigma_x, \sigma_x\beta', \phi')'$ e Ω_θ como em (2.98). O resultado segue, pois é fácil ver que Ω_γ , dado em (2.100), é igual a Ω_γ dada em (2.99).

Seja

$$\theta^* = (\mu', \lambda', \rho')' , \quad (2.101)$$

onde, agora, $\rho = (\rho_2, \rho_3, \dots, \rho_p)'$.

Analogamente ao Corolário 2.2, é fácil provar o seguinte resultado:

Corolário 2.4 *Sob o modelo (2.89) temos que*

$$\sqrt{n}(\hat{\theta}^* - \theta^*) \xrightarrow{\mathcal{D}} N_{3p-1}(0, \Omega_{\theta^*}) ,$$

onde

$$\Omega_{\theta^*} = \begin{bmatrix} V & 0 & 0 \\ 0 & \Omega_\lambda & \Omega_{\lambda\rho} \\ 0 & \Omega_{\rho\lambda} & \Omega_\rho \end{bmatrix} , \quad (2.102)$$

com

$$V \text{ e } \Omega_\lambda \text{ como em (2.98),}$$

$$\Omega_{\lambda\rho} = \Omega_\lambda G'_1 = \Omega'_{\rho\lambda} \text{ e}$$

$$\Omega_\rho = G_1 \Omega_\lambda G'_1 + G_2 \Omega_\phi G'_2 ,$$

onde $G_1 = [0, 2\text{Diag}(\lambda_2^{-1}\rho_2(1-\rho_2), \dots, \lambda_p^{-1}\rho_p(1-\rho_p))]$ é uma matriz $(p-1) \times p$ e $G_2 = -\text{Diag}(\lambda_2^{-2}\rho_2^2, \dots, \lambda_p^{-2}\rho_p^2)$ é uma matriz diagonal $(p-1) \times (p-1)$.

Para a versão de regressão do modelo, seja

$$\gamma^* = (\alpha', \beta', \rho')' \text{ com } \rho = (\rho_2, \dots, \rho_p)' .$$

Temos, então, o seguinte resultado, análogo ao Corolário 2.2 (ii), a prova será omitida. Ver Bolfarine e Galea (1995b).

Lema 2.7 *Sob o modelo (2.90), temos que*

$$\sqrt{n}(\hat{\gamma}^* - \gamma^*) \xrightarrow{\mathcal{D}} N_{3p-3}(0, \Omega_{\gamma^*}) ,$$

onde

$$\Omega_{\gamma^*} = \begin{bmatrix} \Omega_{\alpha} & \Omega_{\alpha\beta} & \Omega_{\alpha\rho} \\ \text{(simetria)} & \Omega_{\beta} & \Omega_{\beta\rho} \\ & & \Omega_{\rho} \end{bmatrix}, \quad (2.103)$$

com

$$\Omega_{\alpha} = G_1 \Omega_{\alpha^*} G_1', \quad \Omega_{\alpha\beta} = G_1 \Omega_{\alpha^*\beta^*} G_1',$$

$$\Omega_{\alpha\rho} = G_1 \Omega_{\alpha^*\beta^*} G_2', \quad \Omega_{\beta} = G_1 \Omega_{\beta^*} G_1',$$

$$\Omega_{\beta\rho} = G_1 \Omega_{\beta^*} G_2' \quad e$$

$$\Omega_{\rho} = G_2 \Omega_{\beta^*} G_2' + G_3 \Omega_{\phi} G_3';$$

$$\Omega_{\alpha^*}, \quad \Omega_{\alpha^*\beta^*}, \quad \Omega_{\beta^*} \quad e \quad \Omega_{\phi} \text{ como em (2.99);}$$

$$G_1 = (0, I_{p-1}) \text{ uma matriz } (p-1) \times p;$$

$$G_2 = (g_{21}, G_{22}) \text{ uma matriz } (p-1) \times p \text{ com}$$

$$g_{21} = 2\sigma_x^{-1}(\rho_2(1-\rho_2), \dots, \rho_p(1-\rho_p))' \text{ é um vetor } (p-1) \times 1$$

$$G_{22} = 2\text{Diag}(\beta_2^{-1}\rho_2(1-\rho_2), \dots, \beta_p^{-1}\rho_p(1-\rho_p)),$$

$$\text{uma matriz diagonal de ordem } p-1; \quad e$$

$$G_3 = -\sigma_x^{-2}\text{Diag}(\beta_2^{-2}\rho_2^2, \dots, \beta_p^{-2}\rho_p^2)$$

$$\text{é uma matriz diagonal de ordem } p-1.$$

Finalmente, usando os resultados anteriores, podemos construir regiões de confiança assintótica para ρ (ver Seção 2.3.3) e testar hipóteses de interesse como as discutidas na Seção 2.3.4. Destacamos que, neste caso, temos expressões explícitas para os EMV e as matrizes de covariância assintótica são também mais simples que o caso geral discutido na Seção 2.3.

2.5 Ilustração numérica

Considere os seguintes dados, tomados de Barnett (1969), relativos a um estudo médico para avaliar a qualidade de dois instrumentos usados para medir a capacidade vital pulmo-

nar (x) num grupo de pacientes. Um instrumento é de tipo padrão e outro novo, portátil e mais fácil de operar. Ademais, o instrumento padrão requer um operador especializado. Os dois instrumentos foram comparados num grupo comum de $n = 72$ pacientes, quando operados por um operador especializado e por um não especializado. As quatro combinações instrumento-operador serão referidas como (Theobald e Mallison, 1978):

Instrumento 1: Instrumento estandar e operador especializado;

Instrumento 2: Instrumento estandar e operador não especializado;

Instrumento 3: Instrumento novo e operador especializado;

Instrumento 4: Instrumento novo e operador não especializado.

Seja y_{ij} a medida da capacidade vital fornecida pelo instrumento i para o paciente j , com $i = 1, 2, 3, 4$ e $j = 1, 2, \dots, 72$. O conjunto de observações $\{y_{ij}\}$ é apresentado na Tabela 2.1

Seguindo Barnett (1969) e Theobald e Mallison (1978), vamos supor um modelo estrutural normal, ver (2.31) - (2.32), com $p = 4$ e $n = 72$.

A Tabela 2.2 mostra algumas iterações do algoritmo EM, indicando que a convergência é atingida aproximadamente na iteração 400, dando estimativas para λ_i e $\sigma_i^2 = \phi_i$, $i = 1, 2, 3, 4$.

Tabela 2.1: Medidas da capacidade vital para 72 pacientes, feitas por 4 instrumentos

Obs.	y_1	y_2	y_3	y_4	Obs.	y_1	y_2	y_3	y_4
1	3450	3530	4030	3720	37	1060	1000	850	600
2	1310	1320	1610	1600	38	2000	1800	1270	1700
3	3820	3720	4150	3700	39	2280	2280	2380	2350
4	2110	2880	2740	2520	40	1940	1800	1670	1580
5	1860	1420	1540	1690	41	2580	2700	2850	2110
6	1940	1780	2020	1800	42	1400	1440	1680	1480
7	2360	2260	2430	2350	43	1260	1100	1000	1030
8	2880	2920	2650	2860	44	2320	2420	2360	2360
9	1980	1720	1800	1660	45	2000	1940	1980	1980
10	3120	3180	3250	3040	46	2400	1900	1470	1740
11	1760	1630	1390	1200	47	2880	2980	3240	3140
12	1480	1760	1700	1640	48	3420	3150	3200	3200
13	1840	1660	1400	1650	49	1000	1130	650	840
14	3580	3480	3680	3960	50	1400	1400	1350	1380
15	1880	2000	2090	2070	51	1880	1710	1600	1350
16	2400	2320	2550	2480	52	1280	1260	1160	1330
17	2220	2120	2290	2270	53	3120	3000	3110	3250
18	2540	2500	2620	1960	54	3770	3340	3900	3700
19	920	1200	640	1030	55	3420	3220	3120	3290
20	2240	2160	2300	2300	56	2740	2880	2850	2880
21	2240	2130	2030	2140	57	2840	2920	2710	2750
22	2260	2510	2400	2450	58	3800	3740	3440	3400
23	3860	4180	3980	3680	59	2100	1680	1650	1930
24	2780	2100	1890	2000	60	1820	1400	1060	1050
25	2220	1400	1840	1360	61	1400	1320	1350	1100
26	1880	1820	1900	1840	62	2200	1680	1640	1110
27	940	960	1060	1000	63	1940	1900	1820	1270
28	2480	2220	2150	2150	64	3260	3200	3250	3270
29	1660	1780	1760	1800	65	1960	1940	1890	1920
30	4040	4180	4000	3770	66	1320	1260	1140	1000
31	2540	2560	2080	2250	67	2840	3060	3650	3510
32	1780	1700	1390	1200	68	2060	1840	1720	1780
33	1280	1300	800	1130	69	2200	1970	1900	2270
34	1940	2060	2030	1880	70	1260	1150	860	1150
35	1760	2000	1860	1860	71	3040	2840	2850	2670
36	2040	1660	1470	1160	72	2140	2180	2560	2720

Fonte: Barnett (1969).

Tabela 2.2: Convergência do algoritmo EM

	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	$\hat{\sigma}_4^2$
10	386.2	409.2	460.4	436.7	50578	19741	29520	38995
50	618.9	655.8	737.7	699.7	50286	19229	29284	38874
100	687.6	739.2	831.4	788.7	50257	19167	29284	38874
200	727.6	771.0	867.2	822.6	50248	19152	29236	38843
300	730.5	774.1	870.7	825.9	50248	19151	29236	38843
350	730.7	774.3	870.9	826.1	50248	19151	29236	38843
400	730.8	774.4	871.0	826.2	50248	19151	29236	38843
450	730.8	774.4	871.0	826.2	50248	19151	29236	38843

Estas estimativas coincidem com as obtidas por Theobald e Mallison (1978). Ademais, os EMV de μ e ρ são

$$\hat{\mu} = (2246, 2176, 2149, 2102)';$$

$$\hat{\rho} = (0.9140, 0.9683, 0.9629, 0.9462)'.$$

Do Teorema 2.2 segue que a matriz de covariância assintótica estimada de $\hat{\theta} = (\hat{\mu}', \hat{\lambda}', \hat{\phi}')'$ é dada por

$$\frac{1}{n} \Omega_{\hat{\theta}} = \quad (2.104)$$

[	8115																			
	7860	8595																		
	8841	9368	10943																	
	8386	8886	9995	10021																
	0	0	0	0	4417															
	0	0	0	0	3928	4438														
	0	0	0	0	4419	4681	5684													
	0	0	0	0	4192	4441	4996	5290												
	0	0	0	0	-15167	2141	1638	1037	92302159											
	0	0	0	0	2269	-11278	5173	3273	-3315623	27655075										
	0	0	0	0	1953	5818	-16621	2818	-2854356	-9011148	52697795									
	0	0	0	0	1172	3492	2673	-16026	-1713169	-5408444	-4656025	68392407								

Notemos que os elementos da matriz de covariância assintótica estimada, (2.104), correspondentes a $\hat{\mu}$ e $\hat{\lambda}$ coincidem com os dados por Theobald e Mallison (1978). Por outro lado, os elementos correspondentes a $\hat{\phi}$ não coincidem já que trabalhamos com σ_i^2 , $i = 1, 2, 3, \dots, p$, na obtenção da matriz de informação; Theobald e Mallison usam σ_i , $i = 1, 2, \dots, p$. Ver Bolfarine e Galea (1995a).

Dos resultados do Capítulo 1, Seção 1.2, segue que o EMV de γ é dado por

$$\hat{\gamma} = (2246, -205, -529, -437, 730.8, 1.060, 1.192, 1.131,$$

$$50248, 19151, 29236, 38843)' \quad (2.105)$$

e do Corolário 2.1 temos que a matriz de covariância assintótica estimada de $\hat{\gamma}$ é dada por

$$\frac{1}{n}\Omega_{\hat{\gamma}} = \quad (2.106)$$

[illegible]

Adicionalmente, a partir do Corolário 2.2 (i), (2.74), temos que a matriz de covariância assintótica estimada de $\hat{\theta}^* = (\hat{\mu}', \hat{\lambda}', \hat{\rho}')'$ é dada por:

$$\frac{1}{n} \Omega_{\hat{\theta}^*} = \quad (2.107)$$

[illegible]

Usando estes resultados e a estatística de Wald, podemos testar hipóteses de interesse como as indicadas no Capítulo 1. Com efeito, dos resultados da Seção 2.3.4, com $p = 4$ e $n = 72$, temos $W_{01} = 45.36$, com nove graus de liberdade, para H_{01} , para H_{02} temos que $W_{02} = 34.82$, com seis graus de liberdade; para H_{03} temos $W_{03} = 7.88$, com três graus de liberdade; e correspondendo a H_{04} , temos $W_{04} = 19.35$, com três graus de liberdade, que são todas rejeitadas a um nível de 5%. A conclusão geral é que os instrumentos são viciados, i.e., medem com vício a capacidade vital e não são igualmente confiáveis. Theobald e Mallison (1978) obtêm as mesmas conclusões utilizando o teste da razão de verossimilhança.

Algumas hipóteses como $H_{05} : \beta_2 = \dots = \beta_p$ não têm a correspondente linear na versão de análise de fatores do modelo estrutural. Em tais casos é conveniente trabalhar

diretamente com a versão de regressão do modelo. Note que H_{05} pode ser escrita como $H_{05} : C\gamma = q_0$, ver Seção 2.3.4. Neste caso, temos $W_{05} = 13.39$, com dois graus de liberdade, indicando que as linhas de calibração não são paralelas.

Finalmente, como uma forma de avaliar o ajuste do modelo estrutural normal, construiremos o gráfico da distância transformada, ver Little (1988) e Lange et al. (1989).

Seja

$$d_j^2 = (Y_j - \mu)'V^{-1}(Y_j - \mu), \quad j = 1, 2, \dots, n. \quad (2.108)$$

a distância de Mahalanobis. Sob normalidade, sabemos que d_j^2 tem uma distribuição χ^2 com p graus de liberdade, $j = 1, 2, \dots, n$. Ademais, $\hat{d}_j^2$ tem assintoticamente a mesma distribuição χ^2 , onde $\hat{d}_j^2 = d_j^2(\hat{\theta})$, $j = 1, 2, \dots, n$. Usando a aproximação de Wilson-Hilferty (Johnson e Kotz, 1970) temos que

$$z_j = \frac{\{(\hat{d}_j^2/p)^{1/3} - (1 - \frac{2}{9p})\}}{\{2/9p\}^{1/2}}, \quad (2.109)$$

$j = 1, 2, \dots, n$, é aproximadamente distribuído normal padrão. Um $Q - Q$ plot das distâncias transformadas z_j , $j = 1, 2, \dots, n$, dadas por (2.109), pode ser utilizado para avaliar o ajuste do modelo normal.

A Figura 2.1 mostra o gráfico da distância transformada para o modelo normal. Pode-se observar um ajuste razoável, apesar de que algumas observações poderiam ser consideradas “outliers”.

Figura 2.1: Gráfico da distância transformada para o modelo normal.

Capítulo 3

Extensões do modelo estrutural normal

3.1 Introdução

Neste capítulo apresentamos duas extensões do modelo estrutural normal discutido no Capítulo 2.

Em primeiro lugar, para o caso de dois instrumentos de medição, analisamos o modelo estrutural repetido (Chan e Mak, 1979) e o modelo ultraestrutural repetido (Cox, 1976 e Dolby, 1976). A Seção 3.2 está dedicada a estes modelos. Na Seção 3.3, estendemos o modelo estrutural normal (1.9), incorporando uma covariável no modelo linearmente relacionada com a característica de interesse x .

Em ambas as extensões usamos o algoritmo EM para obter os EMV. Testes de hipóteses são também discutidos.

3.2 O caso de dois instrumentos

Nesta seção discutimos a comparação de dois instrumentos de medição sob esquemas amostrais diferentes quando observações repetidas estão disponíveis (caso repetido) e quando as observações provêm de vários estratos ou subpopulações (caso estratificado). O modelo estrutural repetido (caso repetido) é tratado por Chan e Mak (1979) e o caso

estratificado por Cox (1976) e Dolby (1976), simultaneamente. Dolby (1976) chama o modelo estatístico correspondente de modelo ultraestrutural repetido. Nestes trabalhos são propostos métodos para encontrar os EMV. A matriz de informação também é obtida em ambos os casos. O interesse principal dos autores é fazer inferências sobre o coeficiente angular da reta de calibração.

Nossa idéia é usar o algoritmo EM para encontrar os EMV e a estatística de Wald para testar hipóteses de interesse.

3.2.1 Os modelos

(i) O modelo estrutural repetido (MER)

Suponha que temos dois instrumentos para medir uma característica x nas unidades de uma população dada. Suponha também que temos uma amostra aleatória de tamanho n desta população.

Seja x_i o verdadeiro valor da característica x na unidade i , $i = 1, 2, \dots, n$. Para cada i dividamos a unidade usada para medir x_i em r partes para obter r medidas repetidas de x_i por instrumento.

Seja y_{hij} a medida fornecida pelo instrumento h para a unidade i , na j -ésima réplica; $h = 1, 2$, $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, r$.

O modelo estatístico proposto por Chan e Mak (1979) é:

$$y_{1ij} = x_i + \varepsilon_{1ij} \quad (3.1)$$

$$y_{2ij} = \alpha_2 + \beta_2 x_i + \varepsilon_{2ij} ; \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, r,$$

onde x_i , $i = 1, 2, \dots, n$, são v.a.i.i.d. $N(\mu_x, \phi_x)$; ε_{1ij} ; $i = 1, 2, \dots, n$, $j = 1, 2, \dots, r$ são v.a.i.i.d $N(0, \phi_1)$ independentes de $\{x_i\}$, e ε_{2ij} , $i = 1, 2, \dots, n$, $j = 1, 2, \dots, r$, são v.a.i.i.d $N(0, \phi_2)$ independentes de $\{\varepsilon_{1ij}\}$ e $\{x_i\}$.

Em notação matricial o modelo (3.1) pode ser escrito como

$$Y_i = a + Bu_i, \quad i = 1, 2, \dots, n; \quad (3.2)$$

onde

$$Y_i = \begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix}$$

são vetores de dimensão $2r \times 1$ e

$$Y_{1i} = (y_{1i1}, y_{1i2}, \dots, y_{1ir})', \quad Y_{2i} = (y_{2i1}, y_{2i2}, \dots, y_{2ir})'$$

são vetores de dimensão $r \times 1$;

$$a = \begin{bmatrix} 0 \\ \alpha_2 \mathbf{1}_r \end{bmatrix}, \quad B = [b, I_{2r}] \quad \text{e} \quad b = \begin{pmatrix} \mathbf{1}_r \\ \beta_2 \mathbf{1}_r \end{pmatrix},$$

com $\mathbf{1}_r$ um vetor de uns de dimensão $r \times 1$ e 0 o vetor nulo de dimensão $r \times 1$;

$$u_i = \begin{pmatrix} x_i \\ \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix}$$

são vetores de dimensão $(2r + 1) \times 1$ e

$$\varepsilon_{1i} = (\varepsilon_{1i1}, \dots, \varepsilon_{1ir})', \quad \varepsilon_{2i} = (\varepsilon_{2i1}, \dots, \varepsilon_{2ir})'$$

são vetores de dimensão $r \times 1$, onde

$$u_i, \quad i = 1, 2, \dots, n, \text{ são i.i.d. } N_{2r+1} \left(\begin{pmatrix} \mu_x \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \phi_x & 0 & 0 \\ 0 & \phi_1 I_r & 0 \\ 0 & 0 & \phi_2 I_r \end{pmatrix} \right).$$

De (3.2) é fácil ver que

$$Y_i, \quad i = 1, 2, \dots, n, \text{ são i.i.d. } N_{2r}(\mu_Y, V_R) \quad (3.3)$$

onde $\mu_Y = \mu_Y(\theta_R) = a + b\mu_x$ e $V_R = V_R(\theta_R) = \phi_x b b' + D$, com $D = \begin{pmatrix} \phi_1 I_r & 0 \\ 0 & \phi_2 I_r \end{pmatrix}$ uma matriz de dimensão $2r \times 2r$ e $\theta_R = (\mu_x, \alpha_2, \beta_2, \phi_x, \phi_1, \phi_2)'$.

Este modelo foi estudado por Chan e Mak (1979), onde é proposto um método para encontrar os EMV. Em particular, Chan e Mak (1979) mostram que o EMV de β_2 é a raiz de um polinômio de quarto grau. Estes autores também obtêm a matriz de informação de Fisher e a variância assintótica do EMV de β_2 .

Aqui utilizaremos o algoritmo EM para encontrar os EMV. A distribuição assintótica do EMV de $(\alpha_2, \beta_2, \rho_1, \rho_2)$ é também obtida.

Note que se $r = 1$ o modelo (3.2) ou (3.1) reduz-se ao modelo linear simples com erros nas variáveis; ver Seção 1.2 do Capítulo 1.

(ii) O modelo ultraestrutural repetido (MUR)

Suponha que temos dois instrumentos para medir uma característica x nas unidades de uma população dada, a qual estratificamos em k estratos ou subpopulações. Suponhamos, também, que de cada subpopulação temos uma amostra aleatória de tamanho r .

Seja x_{ij} o verdadeiro valor da característica x na j -ésima unidade do estrato i , $j = 1, 2, \dots, r$, $i = 1, 2, \dots, k$.

Seja y_{hij} a medida fornecida pelo instrumento h para a j -ésima unidade da subpopulação i , $h = 1, 2$, $i = 1, 2, \dots, k$ e $j = 1, 2, \dots, r$.

O modelo proposto por Cox (1976) e Dolby (1976) é:

$$\begin{aligned} y_{1ij} &= x_{ij} + \varepsilon_{1ij} \\ y_{2ij} &= \alpha_2 + \beta_2 x_{ij} + \varepsilon_{2ij}, \quad i = 1, 2, \dots, k \quad \text{e} \quad j = 1, 2, \dots, r. \end{aligned} \quad (3.4)$$

onde x_{ij} são v.a.i. $\sim N(\mu_i, \phi_x)$; ε_{1ij} são v.a.i.i.d. $N(0, \phi_1)$; ε_{2ij} são v.a.i.i.d. $N(0, \phi_2)$ e x_{ij} , ε_{1ij} e ε_{2ij} todas independentes.

Em notação matricial, o modelo (3.4) pode ser escrito como

$$Y_{ij} = \mathbf{a} + B u_{ij}, \quad (3.5)$$

onde

$$u_{ij} = \begin{pmatrix} x_{ij} \\ \varepsilon_{1ij} \\ \varepsilon_{2ij} \end{pmatrix} \text{ são i.d. } N_3 \left(\begin{pmatrix} \mu_i \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \phi_x & 0 & 0 \\ 0 & \phi_1 & 0 \\ 0 & 0 & \phi_2 \end{pmatrix} \right),$$

$i = 1, 2, \dots, k$ e $j = 1, \dots, r$, onde, agora,

$$\mathbf{a} = \begin{pmatrix} 0 \\ \alpha_2 \end{pmatrix}, \quad B = [b, I_2] \quad \text{e} \quad b = \begin{pmatrix} 1 \\ \beta_2 \end{pmatrix}.$$

De (3.5) segue que

$$Y_{ij}, \quad i = 1, 2, \dots, k \quad \text{e} \quad j = 1, 2, \dots, r, \text{ são i.d. } N_2(\mu_{y_i}, V_E), \quad (3.6)$$

onde

$$\begin{aligned} \mu_{y_i} &= \mu_{y_i}(\theta_E) = \mathbf{a} + \mathbf{b}\mu_i, \quad i = 1, 2, \dots, k; \\ V_E &= V_E(\theta_E) = \begin{bmatrix} \phi_x + \phi_1 & \beta_2 \phi_x \\ \beta_2 \phi_x & \beta_2^2 \phi_x + \phi_2 \end{bmatrix} = \phi_x \mathbf{b} \mathbf{b}' + D(\phi), \text{ com} \\ D(\phi) &= \text{Diag}(\phi_1, \phi_2); \quad \theta_E = (\mu', \alpha_2, \beta_2, \phi_x, \phi_1, \phi_2)' \quad \text{e} \\ \mu &= (\mu_1, \mu_2, \dots, \mu_k)'. \end{aligned}$$

Cox (1976) e Dolby (1976) propõem um método para encontrar os EMV e observam que em alguns casos o valor de θ_E que maximizar a função de verossimilhança pode ficar fora do espaço paramétrico e ter, então, estimativas negativas de alguma variância (ϕ_x, ϕ_1, ϕ_2) em θ_E . Gleser (1983) nota que os EMV dados por Cox (1976) e Dolby (1976) podem ser obtidos como caso especial dos estimadores propostos em Anderson (1951). Brown (1978) mostra que os estimadores de mínimos quadrados generalizados de $\mu_1, \dots, \mu_k, \alpha_2$ e β_2 coincidem com os EMV obtidos por Cox e Dolby. A matriz de informação de Fisher é obtida por Dolby. Cox apresenta as variâncias assintóticas dos EMV.

Nesta seção utilizamos o algoritmo EM para encontrar os EMV e apresentamos a matriz de covariância assintótica do EMV de $(\alpha_2, \beta_2, \rho_1, \rho_2)$. Testes de hipóteses são também discutidos.

Note que este modelo, chamado modelo ultraestrutural por Dolby (1976), reduz-se ao modelo estrutural se $\mu_1 = \mu_2 = \dots = \mu_k = \mu_x$, ver Seção 1.2, e ao modelo funcional se $\phi_x = 0$.

3.2.2 Estimação de máxima verossimilhança via algoritmo EM

Neste parágrafo mostramos como obter os EMV de θ_R e θ_E usando o algoritmo EM. Como será mostrado, este enfoque produz sempre estimativas não negativas das variâncias.

A seguir obtemos o logaritmo da função de verossimilhança completa para o modelo estrutural repetido.

Seja $Z_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}$, $i = 1, 2, \dots, n$.

De (3.2) segue que

$$Z_i, \quad 1 \leq i \leq n, \text{ são v.a.i.i.d. } N_{2r+1}(\mu_Z, \Sigma_R), \quad (3.7)$$

onde

$$\mu_Z = \begin{pmatrix} \mu_x \\ \mathbf{a} + \mathbf{b}\mu_x \end{pmatrix} \quad \text{e} \quad \Sigma_R = \begin{pmatrix} \phi_x & \phi_x \mathbf{b}' \\ \phi_x \mathbf{b} & V_R \end{pmatrix}.$$

Depois de alguma álgebra, temos (Graybill, 1983),

$$|\Sigma_R| = \phi_x \phi_1^r \phi_2^r \quad \text{e} \quad \Sigma_R^{-1} = \begin{pmatrix} c_R/\phi_x & -\mathbf{b}'D^{-1} \\ -D^{-1}\mathbf{b} & D^{-1} \end{pmatrix}, \quad (3.8)$$

onde $c_R = 1 + \phi_x \mathbf{b}' D^{-1} \mathbf{b} = 1 + r \phi_x \left(\frac{1}{\phi_1} + \frac{\beta_2^2}{\phi_2} \right)$ e D como em (3.3).

Portanto, o logaritmo da função de verossimilhança completa, para o modelo estrutural repetido, é dado por

$$\ell(\theta_R|Z) = \sum_{i=1}^n \ell_i(\theta_R), \quad (3.9)$$

onde

$$\begin{aligned} \ell_i(\theta_R) &= \text{cte.} - \frac{1}{2} \ln(\Sigma_R) - \frac{1}{2} f_i \quad \text{e} \\ f_i &= (Z_i - \mu_Z)' \Sigma_R^{-1} (Z_i - \mu_Z), \quad i = 1, 2, \dots, n. \end{aligned}$$

Usando (3.8) temos, depois de alguma álgebra,

$$f_i = f_{1i} + f_{2i} + f_{3i}, \quad i = 1, 2, \dots, n, \quad (3.10)$$

com

$$\begin{aligned} f_{1i} &= \frac{c_R}{\phi_x} (x_i - \mu_x)^2 \\ f_{2i} &= -2(x_i - \mu_x) \left[\frac{1}{\phi_1} (Y_{1i} - \mu_x \mathbf{1}_r)' \mathbf{1}_r + \frac{\beta_2}{\phi_2} (Y_{2i} - (\alpha_2 + \beta_2 \mu_x) \mathbf{1}_r)' \mathbf{1}_r \right] \\ f_{3i} &= \frac{1}{\phi_1} \|Y_{1i} - \mu_x \mathbf{1}_r\|^2 + \frac{1}{\phi_2^2} \|Y_{2i} - (\alpha_2 + \beta_2 \mu_x) \mathbf{1}_r\|^2, \end{aligned}$$

onde $\|a\|$ denota a norma do vetor a . Note que $\ln |\Sigma_R| = \ln \phi_x + r \ln \phi_1 + r \ln \phi_2$.

Para o caso estratificado, seja

$$Z_{ij} = \begin{pmatrix} x_{ij} \\ Y_{ij} \end{pmatrix}, \quad i = 1, 2, \dots, k \quad \text{e} \quad j = 1, 2, \dots, r.$$

De (3.5) é fácil ver que

$$Z_{ij} \sim N_3(\mu_{Z_i}, \Sigma_E), \quad (3.11)$$

onde

$$\mu_{Z_i} = \begin{pmatrix} \mu_i \\ \mathbf{a} + \mathbf{b}\mu_i \end{pmatrix}, \quad i = 1, 2, \dots, k;$$

$$\Sigma_E = \begin{pmatrix} \phi_x & \phi_x \mathbf{b}' \\ \phi_x \mathbf{b} & V_E \end{pmatrix} \quad \text{e}$$

$\mathbf{a}$, $\mathbf{b}$ e V_E como em (3.5). Notando que $\Sigma_R = \Sigma_E$ se $r = 1$, de (3.8) temos que

$$|\Sigma_E| = \phi_x \phi_1 \phi_2 \quad \text{e} \quad \Sigma_E^{-1} = \begin{pmatrix} c_E/\phi_x & -\mathbf{b}' D^{-1}(\phi) \\ -D^{-1}(\phi) \mathbf{b} & D^{-1}(\phi) \end{pmatrix}, \quad (3.12)$$

onde, agora, $c_E = 1 + \phi_x \mathbf{b}' D^{-1}(\phi) \mathbf{b} = 1 + \phi_x \left(\frac{1}{\phi_1} + \frac{\beta_2^2}{\phi_2} \right)$.

Assim, o logaritmo da função de verossimilhança completa, para o modelo ultraestrutural, é dado por

$$\ell(\theta_E|Z) = \sum_{i=1}^k \sum_{j=1}^r \ell_{ij}(\theta_E), \quad (3.13)$$

onde $\ell_{ij}(\theta_E) = \text{cte.} - \frac{1}{2} \ln(\Sigma_E) - \frac{1}{2} f_{ij}$, com

$$f_{ij} = (Z_{ij} - \mu_{Z_i})' \Sigma_E^{-1} (Z_{ij} - \mu_{Z_i}), \quad i = 1, 2, \dots, k \quad \text{e} \quad j = 1, 2, \dots, r.$$

Usando (3.12) é fácil ver que

$$f_{ij} = f_{1ij} + f_{2ij} + f_{3ij}, \quad (3.14)$$

com

$$f_{1ij} = \frac{c_E}{\phi_x} (x_{ij} - \mu_i)^2$$

$$f_{2ij} = -2(x_{ij} - \mu_i) \left[\frac{1}{\phi_1} (y_{1ij} - \mu_i) + \frac{\beta_2}{\phi_2} (y_{2ij} - \alpha_2 - \beta_2 \mu_i) \right]$$

$$f_{3ij} = \frac{1}{\phi_1} (y_{1ij} - \mu_i)^2 + \frac{1}{\phi_2} (y_{2ij} - \alpha_2 - \beta_2 \mu_i)^2;$$

$i = 1, 2, \dots, k, j = 1, 2, \dots, r$ e c_E como em (3.12).

Agora, como já visto nas seções anteriores, cada iteração (ciclo) do algoritmo EM consiste de duas etapas; a etapa E e a etapa M. Para implementar a etapa E, precisamos calcular $E(X_i|Y_i, \theta_R)$ e $E(X_i^2|Y_i, \theta_R)$, $i = 1, 2, \dots, n$, no modelo estrutural repetido e $E(X_{ij}|Y_{ij}, \theta_E)$ e $E(X_{ij}^2|Y_{ij}, \theta_E)$ para o modelo ultraestrutural.

Com efeito, usando propriedades da distribuição normal (Muirhead, 1982), temos (Bolfarine et al., 1993)

$$\hat{x}_i = E(X_i|Y_i, \theta_R) = \mu_x + \frac{\phi_x}{c_R} \mathbf{b}' D^{-1} (Y_i - \mu_Y) \quad (3.15)$$

e $\text{Var}(X_i|Y_i, \theta_R) = \phi_x / c_R$, de onde segue que

$$\hat{x}_i^2 = E(X_i^2|Y_i, \theta_R) = \frac{\phi_x}{c_R} + (\hat{x}_i)^2, \quad (3.16)$$

$i = 1, 2, \dots, n$, para o caso repetido.

Similarmente, para o modelo ultraestrutural, de (3.11) é fácil ver que

$$\hat{x}_{ij} = E(X_{ij}|Y_{ij}, \theta_E) = \mu_i + \frac{\phi_x}{c_E} \mathbf{b}' D^{-1} (\phi) (Y_{ij} - \mu_{y_i}) \quad (3.17)$$

e $\text{Var}(x_{ij}|Y_{ij}, \theta_E) = \phi_x / c_E$, de onde segue que

$$\hat{x}_{ij}^2 = E(x_{ij}^2|Y_{ij}, \theta_E) = \frac{\phi_x}{c_E} + (\hat{x}_{ij})^2, \quad (3.18)$$

$i = 1, 2, \dots, k$ e $j = 1, 2, \dots, r$.

A etapa M é implementada maximizando o logaritmo da função de verossimilhança completa (3.9) para o MER e (3.13) para o MUR.

Para o MER (3.9) é maximizada em (Bolfarine et al., 1993):

$$\begin{aligned} \hat{\mu}_x &= \bar{x}, \quad \hat{\alpha}_2 = \bar{y}_2 - \hat{\beta}_2 \bar{x}, \\ \hat{\beta}_2 &= \sum_{i=1}^n \sum_{j=1}^r (x_i - \bar{x})(y_{2ij} - \bar{y}_2) / \sum_{i=1}^n \sum_{j=1}^r (x_i - \bar{x})^2, \end{aligned}$$

$$\begin{aligned}
\hat{\phi}_x &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \\
\hat{\phi}_1 &= \frac{1}{nr} \sum_{i=1}^n \|Y_{1i} - x_i \mathbf{1}_r\|^2, \\
\hat{\phi}_2 &= \frac{1}{nr} \sum_{i=1}^n \|Y_{2i} - (\hat{\alpha}_2 + \hat{\beta}_2) \mathbf{1}_r\|^2,
\end{aligned} \tag{3.19}$$

onde $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y}_2 = \frac{1}{nr} \sum_{i=1}^n \sum_{j=1}^r y_{2ij}$ e $\mathbf{1}_r = (1, 1, \dots, 1)'$, um vetor de dimensão $r \times 1$.

Analogamente, para o MUR o logaritmo da função de verossimilhança completa (3.13) é maximizado em

$$\begin{aligned}
\hat{\mu}_i &= \bar{x}_i = \frac{1}{r} \sum_{j=1}^r x_{ij}, \quad i = 1, 2, \dots, k, \\
\hat{\alpha}_2 &= \bar{y}_2 - \hat{\beta}_2 \bar{x}, \\
\hat{\beta}_2 &= \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - \bar{x})(y_{2ij} - \bar{y}_2) / \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - \bar{x})^2, \\
\hat{\phi}_x &= \frac{1}{kr} \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - \bar{x}_i)^2, \\
\hat{\phi}_1 &= \frac{1}{kr} \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - y_{1ij})^2 \quad \text{e} \\
\hat{\phi}_2 &= \frac{1}{kr} \left[\sum_{i=1}^k \sum_{j=1}^r (y_{2ij} - \bar{y}_2)^2 - \hat{\beta}_2^2 \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - \bar{x})^2 \right],
\end{aligned} \tag{3.20}$$

onde $\bar{x} = \frac{1}{kr} \sum_{i=1}^k \sum_{j=1}^r x_{ij}$ e $\bar{y}_2 = \frac{1}{kr} \sum_{i=1}^k \sum_{j=1}^r y_{2ij}$.

Notemos que, em cada iteração do algoritmo, as estimativas das variâncias (ϕ_x, ϕ_1, ϕ_2) são não negativas.

Portanto, se $\theta_R^{(m)}$ e $\theta_E^{(m)}$ maximizam (3.9) e (3.13), respectivamente na m -ésima iteração, então na iteração $m + 1$ o algoritmo EM procede como segue:

Etapa E: Calcule $\hat{x}_i = \hat{x}_i(\theta_R^{(m)})$ e $\hat{x}_i^2 = \hat{x}_i^2(\theta_R^{(m)})$, $i = 1, 2, \dots, n$, dados por (3.15) e (3.16), respectivamente, para o MER; $\hat{x}_{ij} = \hat{x}_{ij}(\theta_E^{(m)})$ e $\hat{x}_{ij}^2 = \hat{x}_{ij}^2(\theta_E^{(m)})$, $i = 1, 2, \dots, k$ e $j = 1, 2, \dots, r$, dados por (3.17) e (3.18), respectivamente, para o MUR.

Etapa M: Calcule $\theta_R^{(m+1)}$ a partir de (3.19) para o MER e $\theta_E^{(m+1)}$ a partir de (3.20) para o MUR.

Para obter um valor inicial $(\theta_R^{(0)}, \theta_E^{(0)})$ podemos proceder como segue. Para o MER, seja

$$x_i^0 = \frac{1}{r} \sum_{j=1}^r y_{1ij}, \quad i = 1, 2, \dots, n.$$

Substituindo x_i por x_i^0 em (3.1) e usando mínimos quadrados, podemos obter estimativas iniciais $\alpha_2^{(0)}, \beta_2^{(0)}, \phi_1^{(0)}$ e $\phi_2^{(0)}$. Para μ_x e ϕ_x podemos tomar

$$\mu_x^{(0)} = \bar{x}_0 = \frac{1}{n} \sum_{i=1}^n x_i^0 \quad \text{e} \quad \phi_x^{(0)} = \frac{1}{n} \sum_{i=1}^n (x_i^0 - \bar{x}_0)^2.$$

Desta forma, temos $\theta_R^{(0)}$.

De um modo similar para o MUR, seja $x_i^0 = \frac{1}{r} \sum_{j=1}^r y_{1ij}$, $i = 1, 2, \dots, k$. Então, substituindo x_{ij} em (3.14) por x_i^0 e usando mínimos quadrados obtemos as estimativas iniciais $\alpha_2^{(0)}, \beta_2^{(0)}, \phi_1^{(0)}$ e $\phi_2^{(0)}$. Para μ_i , $i = 1, 2, \dots, k$, tomamos $\mu_i^{(0)} = x_i^{(0)}$ e para ϕ_x usamos como valor inicial a $\phi_x^{(0)} = \frac{1}{r} \sum_{j=1}^r (y_{1ij} - \bar{y}_{11.})^2$, onde $\bar{y}_{11.} = \frac{1}{r} \sum_{j=1}^r y_{11j}$, com o qual temos uma estimativa inicial $\theta_E^{(0)}$.

3.2.3 Matriz de covariância assintótica dos EMV

A matriz de covariância assintótica dos EMV é a inversa da matriz de informação (total) de Fisher. Para o caso repetido, a matriz de informação foi obtida por Chan e Mak (1979) e para o caso estratificado por Dolby (1976); ver também Cox (1976). Estes autores apresentam, além disso, as variâncias assintóticas de cada elemento de $\hat{\theta}_E$. No entanto, Chan e Mak (1979) obtêm a variância assintótica de $\hat{\beta}_2$. Nos dois lemas a seguir, resumimos alguns destes resultados.

Lema 3.1 *Sob o modelo (3.2), temos que*

$$\sqrt{n}(\hat{\theta}_R - \theta_R) \xrightarrow{D} N_6(0, \Omega_R),$$

onde

$$\Omega_R^{-1} = I(\theta_R) = \begin{bmatrix} i_{\mu_x} & i_{\mu_x\alpha_2} & i_{\mu_x\beta_2} & 0 \\ & i_{\alpha_2} & i_{\alpha_2\beta_2} & 0 \\ & & i_{\beta_2} & I_{\beta_2\phi} \\ \text{simetria} & & & I_{\phi\phi} \end{bmatrix}, \quad (3.21)$$

com

$$\begin{aligned} i_{\mu_x} &= \frac{1}{\phi_x}(1 - c_R^{-1}), \quad i_{\mu_x\alpha_2} = r\beta_2/c_R\phi_2, \\ i_{\mu_x\beta_2} &= r\mu_x\beta_2/c_R\phi_2, \quad i_{\alpha_2} = \frac{r}{\phi_2}(1 - r\beta_2^2\phi_x/c_R\phi_2), \\ i_{\alpha_2\beta_2} &= \frac{r\beta_2}{\phi_2}(1 - r\beta_2^2\phi_x/c_R\phi_2), \\ i_{\beta_2} &= \frac{r^2\phi_2^2}{c_R\phi_1\phi_2}(1 + 2\beta_2^2\phi_1/c_R\phi_2) + \frac{\mu_x^2r}{\phi_2}(1 - r\beta_2^2\phi_x/c_R\phi_2) \\ I_{\beta_2\phi} &= \left(\frac{r\beta_2\phi_x}{\phi_1\phi_2c_R^2} \right) \left[r(1 + \beta_2^2\phi_1/\phi_2), \quad -\frac{r\phi_x}{\phi_1}, \quad \frac{\phi_1c_R}{\phi_2}(1 - r\beta_2^2\phi_x/c_R\phi_2) \right] \\ I_{\phi\phi} &= c \begin{pmatrix} \frac{1}{2c\phi_k^2}(1 - c_R^{-1})^2 & \phi_2^2 & \phi_1^2\beta_1^2 \\ & \phi_2^2(c_R^2 - \frac{2c_R\phi_x}{\phi_1} + \frac{r\phi_x^2}{\phi_1^2}) & r\beta_2^2\phi_x^2 \\ \text{simetria} & & \phi_1^2(c_R^2 - \frac{2c_R\phi_x\beta_2^2}{\phi_2} + \frac{r^2\phi_x^2\beta_2^4}{\phi_2^2}) \end{pmatrix} \\ c &= \frac{r}{2\phi_1^2\phi_2^2c_R^2} \quad e \quad \phi = (\phi_x, \phi_1, \phi_2)'. \end{aligned}$$

Prova: A matriz de informação por elemento, $I(\theta_R)$, é obtida por Chan e Mak (1979) e o resultado segue pelas propriedades dos EMV (Sen e Singer, 1993).

Lema 3.2 *Sob o modelo (3.5), temos que*

$$\sqrt{r}(\hat{\theta}_E - \theta_E) \xrightarrow{\mathcal{D}} N_{k+5}(0, \Omega_E),$$

onde

$$\Omega_E^{-1} = I(\theta_E) = \begin{bmatrix} I_{\mu\mu} & I_{\mu\alpha_2} & I_{\mu\beta_2} & 0 \\ & i_{\alpha_2} & i_{\alpha_2\beta_2} & 0 \\ & & i_{\beta_2} & I_{\beta_2\phi} \\ \text{simetria} & & & I_{\phi\phi} \end{bmatrix}, \quad (3.22)$$

com

$$I_{\mu\mu} = \frac{1}{\phi_x}(1 - c_E^{-1})I_k, \quad I_{\mu\alpha_2} = \left(\frac{\beta_2}{\phi_2c_E} \right) \mathbf{1}_k,$$

$$I_{\mu\beta_2} = \left(\frac{\beta_2}{\phi_2 c_E} \right) \mu, \quad i_{\alpha_2} = \left(\frac{k}{c_E \phi_2} \right) (1 + \phi_x / \phi_1), \quad i_{\alpha_2 \beta_2} = \left(\frac{\sum_{i=1}^k \mu_i}{\phi_2 c_E} \right) (1 + \phi_x / \phi_1),$$

$$i_{\beta_2} = \left(\frac{\sum_{i=1}^k \mu_i^2}{c_E \phi_2} \right) (1 + \phi_x / \phi_1) + \left(\frac{k \phi_x^2}{c_E \phi_1 \phi_2} \right) (1 + 2\beta_2^2 \phi_1 / c_E \phi_2),$$

$$I_{\beta_2 \phi} = \left(\frac{k \beta_1 \phi_x}{\phi_1 \phi_2 c_E^2} \right) \left[\frac{c_E \phi_1}{\phi_x} (1 - c_E^{-1}), -\frac{\phi_x}{\phi_1}, \frac{\phi_1}{\phi_2} (1 + \phi_x / \phi_2) \right],$$

$$I_{\phi\phi} = c \begin{pmatrix} \frac{k}{2c\phi_k^2} (1 - c_E^{-1})^2 & \phi_2^2 & \phi_1^2 \beta_2^2 \\ \text{simetria} & \phi_2^2 (1 + \beta_2^2 \phi_x / \phi_2)^2 & \phi_x^2 \beta_2^2 \\ & & \phi_1^2 (1 + \phi_x / \phi_1)^2 \end{pmatrix}$$

$$c = \frac{k}{2\phi_1^2 \phi_2^2 c_E^2} \quad e \quad \phi = (\phi_x, \phi_1, \phi_2)'.$$

Prova: Ver Dolby (1976).

Seja

$$\theta_M^* = g(\theta_M) = (\alpha_2, \beta_2, \rho_1, \rho_2)', \quad (3.23)$$

onde $M = R$ para o modelo estrutural repetido (MER) e $M = E$ para o modelo ultraestrutural repetido (MUR); ρ_i é o coeficiente de confiabilidade do instrumento i , $i = 1, 2$, definido na Seção 1.3 do Capítulo 1.

Usando os Lemas 3.1 e 3.2 dados acima e o método delta (Sen e Singer, 1993), é fácil provar o seguinte resultado:

Teorema 3.1 *Sejam $\hat{\theta}_M^*$ o EMV de θ_M^* . Então, temos que*

$$\sqrt{n^*}(\hat{\theta}_M^* - \theta_M^*) \xrightarrow{D} N_4(0, \Omega_M^*),$$

onde

$$\Omega_M^* = (O_M, G) \Omega_M (O_M, G)', \quad (3.24)$$

com O_M uma matriz nula de dimensão 4×1 se $M = R$ e de dimensão $4 \times k$ se $M = E$; $n^* = n$ para o MER e $n^* = r$ para o MUR; G é uma matriz 4×5 dada por

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \phi_x^{-1} \rho_1 (1 - \rho_1) & -\phi_x^{-1} \rho_1^2 & 0 \\ 0 & 2\beta_2^{-1} \rho_2 (1 - \rho_2) & \phi_x^{-1} \rho_2 (1 - \rho_2) & 0 & -\beta_2^{-2} \phi_x^{-1} \rho_2^2 \end{pmatrix}$$

e Ω_M como definida no Lema 3.1 para $M = R$ e como no Lema 3.2 para $M = E$.

Usando este resultado podemos construir intervalos de confiança assintóticos para qualquer parâmetro de interesse ($\alpha_2, \beta_2, \rho_1$ ou ρ_1), como também construir regiões de confiança assintótica para (ρ_1, ρ_2) , em ambos os modelos. Para detalhes, ver Seção 2.3.3 do Capítulo 2.

3.2.4 Testes de hipóteses

Usando o Teorema 3.1 podemos testar hipóteses de interesse, tais como H_{01} , H_{02} , H_{03} e H_{04} dadas na Seção 1.3 do Capítulo 1, mediante a estatística de Wald, ver Seção 2.3.4. Também podemos utilizar o teste da razão de verossimilhança para testar estas hipóteses. Para estimar os parâmetros sob a hipótese nula é possível usar ainda o algoritmo EM. Por exemplo, consideremos a hipótese $H_{03} : \rho_1 = \rho_2$. Notemos que, neste caso, H_{03} é equivalente a $H_{03} : \phi_2 = \beta_2^2 \phi_1$. Sob H_{03} a etapa E do algoritmo está dada por (3.15) e (3.16) para o caso repetido e por (3.17) e (3.18) para o caso estratificado, mas com ϕ_2 substituído por $\beta_2^2 \phi_1$. Para implementar a etapa M temos que, no modelo estrutural repetido, o valor de θ_R que maximiza (3.9) sob H_{03} é dado por (Bolfarine et al. 1993) $\hat{\mu}_x$, $\hat{\alpha}_2$ e $\hat{\phi}_x$ como em (3.19) e por

$$\hat{\beta}_2 = \left[\frac{\sum_{i=1}^n \|Y_{2i} - \bar{y}_2 \mathbf{1}_r\|^2}{r \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n \|Y_{1i} - x_i \mathbf{1}_r\|^2} \right]^{1/2}, \quad (3.25)$$

e

$$\hat{\phi}_1 = \frac{1}{nr} \left[\frac{1}{\beta_2^2} \sum_{i=1}^n \|Y_{2i} - \bar{y}_2 \mathbf{1}_r\|^2 - \frac{1}{\beta_2} \sum_{i=1}^2 (x_i - \bar{x})(Y_{2i} - \bar{Y}_2 \mathbf{1}_r)' \mathbf{1}_r \right].$$

Para o modelo ultraestrutural repetido temos que o logaritmo da função de verossimilhança completa (3.13) é maximizado sob H_{03} em $\hat{\mu}_i$, $i = 1, 2, \dots, k$, $\hat{\alpha}_2$ e $\hat{\phi}_x$ dados em (3.20) e em

$$\hat{\beta}_2 = \left[\frac{\sum_{i=1}^k \sum_{j=1}^r (y_{2ij} - \bar{y}_2)^2}{\sum_{i=1}^k \sum_{j=1}^r (x_{ij} - y_{1ij})^2 + \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - \bar{x}_i)^2 - r \sum_{i=1}^k (\bar{x}_i - \bar{x})^2} \right] \quad (3.26)$$

e

$$\hat{\phi}_2 = \frac{1}{2kr} \left[\sum_{i=1}^k \sum_{j=1}^r (x_{ij} - y_{1ij})^2 + \sum_{i=1}^k \sum_{j=1}^r \left((x_{ij} - \bar{x}) - \frac{1}{\hat{\beta}_2} (y_{2ij} - \bar{y}_2) \right)^2 \right].$$

Desta forma, temos os EMV sob H_{03} e podemos implementar então o teste da razão de verossimilhança para provar H_{03} em qualquer modelo. De maneira similar podemos obter o teste da razão de verossimilhança para testar as hipóteses H_{01} , H_{02} e H_{04} ou outras hipóteses de interesse.

3.3 O modelo estrutural normal com uma covariável

Existem situações experimentais onde, além das medições da característica x feitas pelos p instrumentos, é observada em cada unidade experimental uma (ou outras) variável que está correlacionada com x . Nesta direção, estudamos o modelo estrutural normal proposto por Barnett (1969), incorporando no modelo uma covariável relacionada com x .

Por outro lado, no caso de dois instrumentos, uma forma de fazer com que o modelo seja identificável é incorporar uma covariável (ou variável instrumental) linearmente relacionada com x ; ver Fuller (1987) e Madansky (1959).

Usamos o algoritmo EM para encontrar os EMV. Como veremos o algoritmo é muito simples de ser implementado e a etapa M não requer qualquer processo iterativo extra. Obtemos a matriz de informação e testes de algumas hipóteses de interesse são considerados utilizando a estatística de Wald.

3.3.1 O modelo

O modelo estrutural normal proposto por Barnett (1969) é dado por (ver Seção 1.2)

$$\begin{aligned} Y_j &= \mathbf{a} + \mathbf{b}x_j + \varepsilon_j, \\ &= \mathbf{a} + Bu_j, \end{aligned} \quad (3.27)$$

e u_j , $1 \leq j \leq n$ são v.a.i.i.d. $N_{p+1} \left(\begin{pmatrix} \mu_x \\ 0 \end{pmatrix}; \begin{pmatrix} \phi_x & 0 \\ 0 & D(\phi) \end{pmatrix} \right)$, onde $\mathbf{a}$, $\mathbf{b}$ e B são definidos como em (1.9). Daqui segue que

$$Y_j, \quad 1 \leq j \leq n \text{ são v.a.i.i.d. } N_p(\mu_Y, V), \quad (3.28)$$

onde $\mu_Y = \mathbf{a} + \mathbf{b}\mu_x$ e $V = \phi_x \mathbf{b}\mathbf{b}' + D(\phi)$.

Segundo Madansky (1959) e Fuller (1987), vamos supor que uma covariável, digamos ω , está linearmente relacionada com a característica de interesse x .

Seja ω_j o valor da covariável na j -ésima unidade experimental, $j = 1, 2, \dots, n$. O modelo que consideraremos é dado por

$$\begin{aligned} Y_j &= \mathbf{a} + \mathbf{b}x_j + \varepsilon_j, \\ x_j &= \gamma_0 + \gamma_1\omega_j + e_j; \quad j = 1, 2, \dots, n, \end{aligned} \quad (3.29)$$

e $\begin{pmatrix} \varepsilon_j \\ e_j \\ \omega_j \end{pmatrix}$, $1 \leq j \leq n$ são v.a.i.i.d. $N_{p+2} \left(\begin{pmatrix} 0 \\ 0 \\ \mu_\omega \end{pmatrix}; \begin{pmatrix} D(\phi) & 0 & 0 \\ 0 & \phi_e & 0 \\ 0 & 0 & \phi_\omega \end{pmatrix} \right)$.

De (3.29) é fácil ver que

$$W_j = \begin{pmatrix} Y_j \\ \omega_j \end{pmatrix}, \quad 1 \leq j \leq n, \text{ são v.a.i.i.d. } N_{p+1}(\mu_w, V_w), \quad (3.30)$$

onde

$$\mu_w = \begin{pmatrix} \mathbf{a} + \mathbf{b}\mu_x \\ \mu_\omega \end{pmatrix} \quad \text{e} \quad V_w = \begin{pmatrix} \phi_x \mathbf{b}\mathbf{b}' + D(\phi) & \gamma_1 \phi_\omega \mathbf{b} \\ \gamma_1 \phi_\omega \mathbf{b}' & \phi_\omega \end{pmatrix},$$

com $\mu_x = \gamma_0 + \gamma_1\mu_\omega$ e $\phi_x = \gamma_1^2 \phi_\omega + \phi_e$.

Seja $\theta = (\mu_\omega, \gamma_0, \gamma_1, \alpha', \beta', \phi', \phi_e, \phi_\omega)'$ o vetor de parâmetros do modelo de dimensão $3(p+1) \times 1$.

3.3.2 Estimação via algoritmo EM

Neste parágrafo mostramos como obter o EMV de θ usando o algoritmo EM. Como será observado, a implementação do algoritmo é simples e produz estimativas das variâncias dos erros de medição não negativas.

Seja $Z_j = \begin{pmatrix} x_j \\ w_j \end{pmatrix}$, $j = 1, 2, \dots, n$.

De (3.29) temos que

$$Z_j, \quad 1 \leq j \leq n, \text{ são v.a.i.i.d. } N_{p+2}(\mu_Z, \Sigma_Z), \quad (3.31)$$

onde

$$\mu_Z = \begin{pmatrix} \mu_x \\ \mathbf{a} + \mathbf{b}\mu_x \\ \mu_\omega \end{pmatrix} \quad \text{e} \quad \Sigma_Z = \begin{pmatrix} \phi_x & \phi_x \mathbf{b}' & \gamma_1 \phi_\omega \\ \phi_x \mathbf{b} & \phi_x \mathbf{b} \mathbf{b}' + D(\phi) & \gamma_1 \phi_\omega \mathbf{b} \\ \gamma_1 \phi_\omega & \gamma_1 \phi_\omega \mathbf{b}' & \phi_\omega \end{pmatrix}.$$

Lema 3.3 *Seja Σ_Z como em (3.31). Então temos*

$$(i) \quad |\Sigma_Z| = \phi_\omega \phi_e \prod_{i=1}^p \phi_i \quad \text{e}$$

$$(ii) \quad \Sigma_Z^{-1} = \begin{pmatrix} c_e/\phi_e & -\mathbf{b}' D^{-1}(\phi) & -\gamma_1/\phi_e \\ -D^{-1}(\phi) \mathbf{b} & D^{-1}(\phi) & 0 \\ -\gamma_1/\phi_e & 0 & \frac{\gamma_1^2}{\phi_e} + \frac{1}{\phi_\omega} \end{pmatrix},$$

com $c_e = 1 + \phi_e \mathbf{b}' D^{-1}(\phi) \mathbf{b}$.

Prova: Subdividamos Σ_Z da seguinte forma:

$$\Sigma_Z = \begin{pmatrix} \phi_x & \phi_x \mathbf{b}' & \gamma_1 \phi_\omega \\ \phi_x \mathbf{b} & \phi_x \mathbf{b} \mathbf{b}' + D(\phi) & \gamma_1 \phi_\omega \mathbf{b} \\ \gamma_1 \phi_\omega & \gamma_1 \phi_\omega \mathbf{b}' & \phi_\omega \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \phi_\omega \end{pmatrix}. \quad (3.32)$$

Logo, temos (Graybill, 1983):

$$\begin{aligned} (i) \quad |\Sigma_Z| &= \phi_\omega |\Sigma_{11} - \Sigma_{12}(\phi_\omega)^{-1} \Sigma_{21}| = \phi_\omega \left| \begin{pmatrix} \phi_e & \phi_e \mathbf{b} \mathbf{b}' \\ \phi_e \mathbf{b} & \phi_e \mathbf{b} \mathbf{b}' + D(\phi) \end{pmatrix} \right| \\ &= \phi_\omega \phi_e |D(\phi)| = \phi_\omega \phi_e \prod_{i=1}^p \phi_i. \end{aligned}$$

(ii) Usando os teoremas 8.2.1 e 8.9.3 em Graybill (1983), temos depois de alguma álgebra:

$$\Sigma_{11}^{-1} = \begin{pmatrix} \frac{c_x}{\phi_x} & -b'D^{-1}(\phi) \\ -D^{-1}(\phi)b & D^{-1}(\phi) \end{pmatrix}, \quad (3.33)$$

com $c_x = 1 + \phi_x b'D^{-1}(\phi)b$.

Agora, notando que

$$\Sigma_{11} - \Sigma_{12}\phi_\omega^{-1}\Sigma_{21} = \begin{pmatrix} \phi_e & \phi_e b' \\ \phi_e b & \phi_e b b' + D(\phi) \end{pmatrix},$$

temos, usando novamente os resultados em Graybill (1983):

$$[\Sigma_{11} - \Sigma_{12}(\phi_\omega)^{-1}\Sigma_{21}]^{-1} = \begin{pmatrix} \frac{c_e}{\phi_e} & -b'D^{-1}(\phi) \\ -D^{-1}(\phi)b & D^{-1}(\phi) \end{pmatrix}.$$

De (3.33) é fácil ver que $\phi_\omega - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \phi_\omega\phi_e/\phi_x$ e $-\Sigma_{11}^{-1}\Sigma_{12}[\phi_\omega - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}]^{-1} = -(\gamma_1/\phi_e)$ e o resultado segue.

Agora, de (3.31) segue que o logaritmo da função de verossimilhança completa é dado por

$$\ell(\theta/Z) = \sum_{j=1}^n \ell_j(\theta), \quad (3.34)$$

onde

$$\begin{aligned} \ell_j(\theta) &= \text{cte.} - \frac{1}{2} \ln |\Sigma_Z| - \frac{1}{2} f_j \quad \text{e} \\ f_j &= (Z_j - \mu_Z)' \Sigma_Z^{-1} (Z_j - \mu_Z), \quad j = 1, 2, \dots, n. \end{aligned}$$

Usando a parte (ii) do Lema 3.3, temos

$$f_j = f_{1j} + f_{2j} + f_{3j}, \quad j = 1, 2, \dots, n, \quad (3.35)$$

onde

$$\begin{aligned} f_{1j} &= (x_j - \mu_x)^2 \frac{c_e}{\phi_e} - 2(x_j - \mu_x)b'D^{-1}(\phi)(Y_j - \mu_Y) + (Y_j - \mu_Y)'D^{-1}(\phi)(Y_j - \mu_Y), \\ f_{2j} &= -\frac{2\gamma_1}{\phi_e}(x_j - \mu_x)(\omega_j - \mu_\omega), \\ f_{3j} &= \left(\frac{\gamma_1^2}{\phi_e} + \frac{1}{\phi_\omega} \right) (\omega_j - \mu_\omega)^2 \quad \text{e} \quad \mu_x = \gamma_0 + \gamma_1\mu_\omega. \end{aligned}$$

Notemos também que $\ln |\Sigma_Z| = \ln \phi_\omega + \ln \phi_e + \sum_{i=1}^p \ln \phi_i$.

Usando propriedades da distribuição normal multivariada (Muirhead, 1982), de (3.31) temos que

$$\hat{x}_j = E(x_j|W_j, \theta) = \mu_x + (\phi_x \mathbf{b}', \gamma_1 \phi_\omega) V_w^{-1} (W_j - \mu_w) \quad (3.36)$$

e

$$\text{Var}(x_j|W_j, \theta) = \phi_x - (\phi_x \mathbf{b}', \gamma_1 \phi_\omega) V_w^{-1} \begin{pmatrix} \phi_x \mathbf{b} \\ \gamma_1 \phi_\omega \end{pmatrix}, \quad (3.37)$$

onde V_w é como em (3.30). De maneira similar ao Lema 3.3, podemos provar o seguinte resultado:

Lema 3.4 *Seja V_w como em (3.30). Então, temos que*

$$(i) |V_w| = c_e \phi_\omega \prod_{i=1}^p \phi_i.$$

$$(ii) V_w^{-1} = \begin{pmatrix} D^{-1}(\phi) - (\phi_e | c_e) D^{-1}(\phi) \mathbf{b} \mathbf{b}' D^{-1}(\phi) & -(\gamma_1 | c_e) D^{-1}(\phi) \mathbf{b} \\ -(\gamma_1 | c_e) \mathbf{b}' D^{-1}(\phi) & (c_x | c_e) \phi_\omega \end{pmatrix},$$

com c_e e c_x como no Lema 3.3 e (3.33), respectivamente.

Logo, usando a parte (ii) do Lema 3.4, temos

$$\hat{x}_j = E(x_j|W_j, \theta) = \mu_x + \frac{\phi_e}{c_e} \mathbf{b}' D^{-1}(\phi) (Y_j - \mu_Y) + \frac{\gamma_1}{c_e} (\omega_j - \mu_\omega) \quad (3.38)$$

e $\text{Var}(x_j|W_{j1}, \theta) = \phi_e / c_e$, de modo que

$$\hat{x}_j^2 = E(x_j^2|W_j, \theta) = \frac{\phi_e}{c_e} + (\hat{x}_j)^2. \quad (3.39)$$

A etapa M do algoritmo é implementada maximizando a função de verossimilhança completa (3.34). Derivando $\ell(\theta|Z)$ em relação aos elementos de θ e igualando a zero, temos, depois de alguma álgebra, as seguintes equações:

$$\hat{\mu}_\omega = \bar{\omega}, \quad \hat{\gamma}_0 = \bar{x} - \hat{\gamma}_1 \bar{\omega}$$

$$\hat{\gamma}_1 = \frac{\sum_{j=1}^n (\omega_j - \bar{\omega})(x_j - \bar{x})}{\sum_{j=1}^n (\omega_j - \bar{\omega})^2}$$

$$\begin{aligned}
\hat{\alpha}_i &= \bar{y}_i - \hat{\beta}_i \bar{x}, \quad i = 2, 3, \dots, p, \\
\hat{\beta}_i &= \sum_{j=1}^n (x_j - \bar{x})(y_{ij} - \bar{y}_i) / \sum_{j=1}^n (x_j - \bar{x})^2, \quad i = 2, 3, \dots, p, \\
\hat{\phi}_1 &= \frac{1}{n} \sum_{j=1}^n (x_j - y_{1j})^2 \\
\hat{\phi}_i &= \frac{1}{n} \sum_{j=1}^n [(y_{ij} - \bar{y}_i) - \hat{\beta}_i (x_j - \bar{x})]^2 \\
&= \frac{1}{n} \left[\sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 - \hat{\beta}_i^2 \sum_{j=1}^n (x_j - \bar{x})^2 \right], \quad i = 2, \dots, p, \\
\hat{\phi}_\omega &= \frac{1}{n} \sum_{j=1}^n (\omega_j - \bar{\omega})^2 \\
\hat{\phi}_\epsilon &= \frac{1}{n} \sum_{j=1}^n [(x_j - \bar{x}) - \hat{\gamma}_1 (\omega_j - \bar{\omega})]^2 = \frac{1}{n} \left[\sum_{j=1}^n (x_j - \bar{x})^2 - \hat{\gamma}_1^2 \sum_{j=1}^n (\omega_j - \bar{\omega})^2 \right],
\end{aligned} \tag{3.40}$$

onde $\bar{\omega} = \frac{1}{n} \sum_{j=1}^n \omega_j$, $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$ e $\bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij}$, $i = 2, \dots, p$.

Portanto, se $\theta^{(m)}$ maximiza (3.34) na m -ésima iteração, então na iteração $m + 1$ o algoritmo EM procede como segue:

Etapa E: Calcule $\hat{x}_j = \hat{x}_j(\theta^{(m)})$ e $\hat{x}_j^2 = \hat{x}_j^2(\theta^{(m)})$, $j = 1, 2, \dots, n$, dados por (3.38) e (3.39), respectivamente, com $\theta = \theta^{(m)}$.

Etapa M: Calcule $\theta^{(m+1)}$ a partir de (3.40).

Notemos que as estimativas das variâncias são sempre não negativas. Notemos, também, que a implementação do algoritmo EM é bastante simples e que as estimativas $\hat{\mu}_\omega$ e $\hat{\phi}_\omega$ não dependem de x_j , $j = 1, 2, \dots, n$; ou seja, não é requerido um processo iterativo.

Para obter um valor inicial $\theta^{(0)}$, podemos proceder como segue:

- (i) Com o método proposto por Barnett (1969), ou Seção 2.3.1, obtemos estimativas iniciais $\alpha^{(0)}$, $\beta^{(0)}$ e $\phi^{(0)}$.
- (ii) Seja $x_j^0 = \frac{1}{p} \sum_{i=1}^p y_{ij}$, $j = 1, 2, \dots, n$. Usando x_j^0 no lugar de x_j na segunda equação

de (3.29) e o método dos mínimos quadrados obtemos estimativas iniciais para γ_0 , γ_1 e ϕ_e .

3.3.3 A matriz de informação

Nesta seção obtemos a matriz de informação esperada.

De (3.30), segue que o logaritmo da função de verossimilhança observada, $\ell(\theta|W)$, é dado por

$$\ell(\theta|W) = \sum_{j=1}^n \ell_j(\theta) , \quad (3.41)$$

onde

$$\begin{aligned} \ell_j(\theta) &= \text{cte.} - \frac{1}{2} \ln |V_w| - \frac{1}{2} f_j \quad \text{e} \\ f_j &= (W_j - \mu_w)' V_w^{-1} (W_j - \mu_w) , \quad j = 1, 2, \dots, n. \end{aligned}$$

Usando o Lema 3.4 é fácil ver que

$$\ln |V_w| = \ln \phi_w + \ln c_e + \sum_{i=1}^p \phi_i \quad \text{e} \quad f_j = f_{1j} + f_{2j} + f_{3j} ,$$

onde

$$\begin{aligned} f_{1j} &= \frac{1}{\phi_1} (y_{1j} - \mu_x)^2 + (Y_{2j} - \alpha - \beta \mu_x)' D_2^{-1} (Y_{2j} - \alpha - \beta \mu_x) \\ &\quad - \frac{\phi_e}{c_e} \left[\frac{1}{\phi_1} (y_{1j} - \mu_x) + (Y_{2j} - \alpha - \beta \mu_x)' D_2^{-1} \beta \right]^2 \\ f_{2j} &= -\frac{2\gamma_1}{c_e} (\omega_j - \mu_\omega) \left[\frac{1}{\phi_1} (y_{1j} - \mu_x) + (Y_{2j} - \alpha - \beta \mu_x)' D_2^{-1} \beta \right] \\ f_{3j} &= (\omega_j - \mu_\omega)^2 \left[\frac{\gamma_1^2}{\phi_e} (1 - c_e^{-1}) + \frac{1}{\phi_\omega} \right] , \quad j = 1, 2, \dots, n, \end{aligned}$$

com

$$\begin{aligned} \mu_x &= \gamma_0 + \gamma_1 \mu_\omega , \quad Y_{2j} = (y_{2j}, \dots, y_{pj})' , \quad j = 1, 2, \dots, n, \quad \text{e} \\ D_2^{-1} &= \text{Diag}(\phi_2^{-1}, \phi_3^{-1}, \dots, \phi_p^{-1}) . \end{aligned}$$

Para obter a matriz de informação precisamos do seguinte fato:

Lema 3.5 *Seja $Y \sim N_p(0, V)$, onde $V = \phi_x \mathbf{b} \mathbf{b}' + D(\phi)$. Então,*

- (i) $E[(Y' D^{-1}(\phi) \mathbf{b})^2] = \phi_x \left(\frac{c_e - 1}{\phi_e} \right)^2 + \left(\frac{c_e - 1}{\phi_e} \right);$
- (ii) $E[D(Y) D(Y)] = \phi_x D(\mathbf{b}) D(\mathbf{b}) + D(\phi);$
- (iii) $E[(Y' D^{-1}(\phi) \mathbf{b}) D(Y)] = \frac{\phi_x}{\phi_e} (c_e - 1) D(\mathbf{b}) + D(\mathbf{b}).$

Prova: (i) Temos que

$$\begin{aligned}
 E[(Y' D^{-1}(\phi) \mathbf{b})^2] &= E(Y' D^{-1}(\phi) \mathbf{b} \mathbf{b}' D^{-1}(\phi) Y) \\
 &= \text{tr}[D^{-1}(\phi) \mathbf{b} \mathbf{b}' D^{-1}(\phi) V] \\
 &= \frac{\phi_x}{\phi_e} (c_e - 1) \text{tr}(D^{-1}(\phi) \mathbf{b} \mathbf{b}') + \text{tr}(D^{-1}(\phi) \mathbf{b} \mathbf{b}') \\
 &= \phi_x \left(\frac{c_e - 1}{\phi_e} \right)^2 + \left(\frac{c_e - 1}{\phi_e} \right).
 \end{aligned}$$

(ii) É fácil ver que

$$E[D(Y) D(Y)] = D(V) = D(\phi_x \mathbf{b} \mathbf{b}' + D(\phi)) = \phi_x D(\mathbf{b}) D(\mathbf{b}) + D(\phi).$$

(iii) Notemos que

$$\begin{aligned}
 E[(Y' D^{-1}(\phi) \mathbf{b}) D(Y)] &= E[D(Y (Y' D^{-1}(\phi) \mathbf{b}))] \\
 &= D(V D^{-1}(\phi) \mathbf{b}) = D\left(\frac{\phi_x}{\phi_e} (c_e - 1) \mathbf{b} + \mathbf{b}\right) \\
 &= \frac{\phi_x}{\phi_e} (c_e - 1) D(\mathbf{b}) + D(\mathbf{b}).
 \end{aligned}$$

Agora, derivando $\ell(\theta|W)$ em (3.41), duas vezes em relação a θ (ver Nel, 1980, e Apêndice C) obtemos a matriz de informação observada. Tomando a esperança e usando o Lema 3.5 e fatos conhecidos sobre a distribuição normal (Graybill, 1983), temos, depois de uma longa álgebra, o seguinte resultado:

Lema 3.6 *Sob o modelo (3.29) a matriz de informação esperada, por elemento, denotada por $I(\theta)$, é dada por*

$$I(\theta) = \begin{bmatrix} i_{\mu_x} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & i_{\gamma_0} & i_{\gamma_0\gamma_1} & I_{\gamma_0\alpha} & I_{\gamma_0\beta} & 0 & 0 & 0 \\ & & i_{\gamma_1} & I_{\gamma_1\alpha} & I_{\gamma_1\beta} & 0 & 0 & 0 \\ & & & I_{\alpha\alpha} & I_{\alpha\beta} & 0 & 0 & 0 \\ & & & & I_{\beta\beta} & I_{\beta\phi} & I_{\beta\phi_e} & 0 \\ & & & & & I_{\phi\phi} & I_{\phi\phi_e} & 0 \\ \text{(simetria)} & & & & & & i_{\phi_e} & 0 \\ & & & & & & & i_{\phi_\omega} \end{bmatrix}, \quad (3.42)$$

onde

$$\begin{aligned} i_{\mu_\omega} &= 1/\phi_\omega ; \quad i_{\gamma_0} = (1 - c_e^{-1})/\phi_e ; \quad i_{\gamma_0\gamma_1} = \mu_\omega(1 - c_e^{-1})/\phi_e ; \\ I_{\gamma_0\alpha} &= c_e^{-1}\beta'D_2^{-1} ; \quad I_{\gamma_0\beta} = \mu_x c_e^{-1}\beta'D_2^{-1} ; \quad i_{\gamma_1} = (1 - c_e^{-1})(\mu_\omega^2 + \phi_\omega)/\phi_e ; \\ I_{\gamma_1\alpha} &= \mu_\omega c_e^{-1}\beta'D_2^{-1} ; \quad I_{\gamma_1\beta} = c_e^{-1}(\gamma_1\phi_\omega + \mu_\omega\mu_x)\beta'D_2^{-1} ; \\ I_{\alpha\alpha} &= D_2^{-1} - \phi_e c_e^{-1}M_2 ; \quad I_{\alpha\beta} = \mu_x(D_2^{-1} - \phi_e c_e^{-1}M_2) ; \\ I_{\beta\beta} &= D_2^{-1}(\mu_x^2 + \gamma_1^2\phi_\omega + \phi_e(1 - c_e^{-1})) - c_e^{-1}\phi_e M_2(\mu_x^2 + \gamma_1^2\phi_\omega + \phi_e(1 - 2c_e^{-1})) ; \\ I_{\beta\phi} &= c_e^{-1}\phi_e[c_e^{-1}\phi_e D_2^{-1}\beta/\phi_1^2; D_2^{-1}D(\beta) - c_e^{-1}\phi_e M^*] \\ I_{\beta\phi_e} &= -2c_e^{-1}(1 - 2c_e^{-1})D_2^{-1}\beta ; \\ I_{\phi\phi} &= D^{-1}(\phi) - c_e^{-1}\phi_e D^{-3}(\phi)D(b)D(b) + \frac{c_e^{-2}}{2}\phi_e^2 M^{**} ; \\ I_{\phi\phi_e} &= \frac{c_e^{-2}}{2}D^{-2}(\phi)D(b)b ; \\ i_{\phi_e} &= c_e^{-2}(c_e - 1)^2/2\phi_e^2 ; \quad i_{\phi_\omega} = 1/2\phi_\omega^2 ; \end{aligned}$$

com

$$\begin{aligned} M_2 &= D_2^{-1}\beta\beta'D_2^{-1} , \quad M^* = M_2 D_2^{-1}D(\beta) , \\ M^{**} &= D^{-2}(\phi)D(b)bb'D(b)D^{-2}(\phi) \quad e \quad D_2^{-2} = \text{Diag}(\phi_2^{-2}, \phi_3^{-2}, \dots, \phi_p^{-2}) . \end{aligned}$$

A matriz de covariância assintótica do EMV $\hat{\theta}$ é dada, então, por $\frac{1}{n}I^{-1}(\theta)$.

Seja $\theta^* = (\gamma_1, \alpha', \beta', \rho')'$, onde $\rho = (\rho_1, \rho_2, \dots, \rho_p)'$. Logo, temos o seguinte resultado:

Teorema 3.2 *Seja $\hat{\theta}^*$ o EMV de θ^* . Então temos que*

$$\sqrt{n}(\hat{\theta}^* - \theta^*) \xrightarrow{D} N_{3p-1}(0; \Omega_{\theta^*}) ,$$

onde

$$\Omega_{\theta^*} = GI^{-1}(\theta)G' \quad (3.43)$$

com

$$G = \begin{pmatrix} O_{p \times 2} & I_p & O_{p \times (p-1)} & O_{p \times (p+2)} \\ O_{(p-1) \times 2} & O_{(p-1) \times p} & I_{p-1} & O_{(p-1) \times (p+2)} \\ O_{p \times 2} & G_1 & G_2 & G_3 \end{pmatrix} ,$$

uma matriz de dimensão $(3p-1) \times (3p+3)$, onde

$G_1 = 2\gamma_1^{-1}[d_1; O_{p \times (p-1)}]$ é uma matriz de dimensão $p \times p$ e

$$d_1 = \rho_x[\rho_1(1 - \rho_1), \dots, \rho_p(1 - \rho_p)]'$$

um vetor $p \times 1$;

$G_2 = \begin{pmatrix} O_{p \times (p-1)} \\ G_{22} \end{pmatrix}$ uma matriz $p \times (p-1)$ e

$$G_{22} = \text{Diag}(2\beta_2^{-1}\rho_2(1 - \rho_2), \dots, 2\beta_p^{-1}\rho_p(1 - \rho_p));$$

$G_3 = [G_{31}; \phi_\omega^{-1}d_1; \gamma_1^{-2}\phi_\omega^{-1}d_1]$ uma matriz $p \times (p+2)$ com

$$G_{31} = -\phi_x^{-1}\text{Diag}(\rho_1^2, \beta_2^{-2}\rho_2^2, \dots, \beta_p^{-2}\rho_p^2)$$

uma matriz diagonal $p \times p$ e $\rho_x = \frac{\gamma_1^2\phi_\omega}{\gamma_1^2\phi_\omega + \phi_\epsilon}$.

Notemos que ρ_x pode ser interpretado como a “proporção da variabilidade de x explicada pela covariável ω ”.

Prova: Observemos que $\theta^* = g(\theta)$, onde g segue de (1.19). Logo, usando o método delta, o resultado segue depois de alguma álgebra. Ver Corolário 2.2.

Utilizando este método, podemos construir intervalos de confiança assintóticos para qualquer parâmetro de interesse; também podemos construir regiões de confiança assintótica para $\rho = (\rho_1, \rho_2, \dots, \rho_p)'$, ver Seção 2.3.3. Além disso, podemos testar hipóteses

de interesse (H_{01}, H_{02}, H_{03} ou H_{04}) usando a estatística de Wald, ver Seção 2.3.4. Neste caso, uma outra hipótese, para testar a importância da covariável, é $H_{05} : \gamma_1 = 0$.

Alternativamente, podemos usar o teste da razão de verossimilhança para provar hipóteses como as mencionadas acima. Para estimar os parâmetros sob a hipótese nula usamos novamente o algoritmo EM, ver Seção 3.2.4 para uma ilustração.

3.3.4 Ilustração numérica: dados sobre terremotos na Alasca

Para ilustrar a metodologia desenvolvida nesta seção, consideremos os dados da Tabela 3.1, tomados de Fuller (1987), sobre a magnitude de terremotos no Alasca no período de 1969 a 1976. Aqui, x é a verdadeira intensidade do terremoto, que é medida por dois instrumentos, dos quais um é de referência. Além disso, temos uma covariável ω relacionada com x . Ver Exemplo 1.4.1 em Fuller (1987), para maiores detalhes sobre as variáveis.

Todos os pares de gráficos das variáveis observadas y_1 , y_2 e ω não mostrados aqui mostram evidência de linearidade. Logo, vamos supor o modelo (3.29) com $p = 2$ e $n = 62$.

Usando o algoritmo EM desenvolvido na Seção 3.3.2, temos os seguintes EMV:

$$\begin{aligned}\hat{\mu}_\omega &= 5,2435, & \hat{\gamma}_0 &= 2,2883, & \hat{\gamma}_1 &= 0,5581, \\ \hat{\alpha}_2 &= -4,2829, & \hat{\beta}_2 &= 1,7960, & \hat{\phi}_1 &= 0,0622, \\ \hat{\phi}_2 &= 0,1375, & \hat{\phi}_e &= 0,0223, & \hat{\phi}_\omega &= 0,3986, \\ \hat{\rho}_x &= 0,8477, & \hat{\rho}_1 &= 0,7018, & \hat{\rho}_2 &= 0,7745.\end{aligned}$$

Além disso, do Teorema 3.2, temos que a matriz de covariância assintótica estimada de $\hat{\theta}^* = (\hat{\gamma}_1, \hat{\alpha}_2, \hat{\beta}_2, \hat{\rho}_1, \hat{\rho}_2)'$ é dada por

$$\frac{1}{n}\hat{\Omega}_{\theta^*} = \begin{pmatrix} 0,0033 & 0,0393 & -0,0075 & 0,0019 & -0,0004 \\ 0,0393 & 1,1176 & -0,2133 & 0,0203 & -0,0392 \\ -0,0075 & -0,2133 & 0,0409 & -0,0039 & 0,0075 \\ 0,0019 & 0,0203 & -0,0039 & 0,0050 & 0,0015 \\ -0,0004 & -0,0392 & 0,0075 & 0,0015 & 0,0056 \end{pmatrix}.$$

Notemos que as variâncias assintóticas de $\hat{\alpha}_2$ e $\hat{\beta}_2$ são próximas das obtidas por Fuller

Tabela 3.1: Medidas de intensidade de 62 terremotos no Alasca

Observação j	y_2	y_1	ω	Observação j	y_2	y_1	ω
1	5,5	5,1	5,6	32	5,4	5,7	5,1
2	5,7	5,5	6,0	33	5,3	5,7	5,7
3	6,0	6,0	6,4	34	5,7	5,9	5,8
4	5,3	5,2	5,2	35	4,8	5,8	5,7
5	5,2	5,5	5,7	36	6,4	5,8	5,7
6	4,7	5,0	5,1	37	4,2	4,9	4,9
7	4,2	5,0	5,0	38	5,8	5,7	5,9
8	5,2	5,7	5,5	39	4,6	4,8	4,3
9	5,3	4,9	5,0	40	4,7	5,0	5,2
10	5,1	5,0	5,2	41	5,8	5,7	5,9
11	5,6	5,5	5,8	42	5,6	5,0	5,4
12	4,8	4,6	4,9	43	4,8	5,1	5,5
13	5,4	5,6	5,9	44	6,2	5,2	5,8
14	4,3	5,2	4,7	45	6,8	5,5	6,2
15	4,4	5,1	4,9	46	6,0	5,8	5,8
16	4,8	5,5	4,6	47	4,6	4,9	4,7
17	3,6	4,7	4,3	48	4,1	4,7	4,5
18	4,6	5,0	4,8	49	4,4	4,9	4,6
19	4,5	5,1	4,5	50	4,0	5,3	5,2
20	4,2	4,9	4,6	51	5,0	5,0	5,6
21	4,4	4,7	4,6	52	5,9	5,6	5,9
22	3,6	4,7	4,3	53	5,7	5,5	5,9
23	3,9	4,5	4,6	54	5,0	4,1	5,3
24	4,0	4,8	4,6	55	5,3	5,5	5,5
25	5,5	5,7	4,9	56	5,7	5,5	5,4
26	5,6	5,7	5,5	57	4,7	4,8	4,7
27	5,1	4,7	4,7	58	4,6	4,8	4,6
28	5,5	4,9	4,1	59	4,2	4,5	4,6
29	4,4	4,8	4,9	60	6,5	6,0	7,1
30	7,0	6,2	6,5	61	4,9	4,8	4,6
31	6,6	6,0	6,3	62	4,4	5,0	5,3

Fonte: Fuller (1987).

(1987). Vale salientar também que no Teorema 1.4.1 em Fuller contempla só a distribuição assintótica de $(\hat{\alpha}_2, \hat{\beta}_2)$.

Usando a estatística de Wald podemos testar hipóteses de interesse. A hipótese $H_{05} : \gamma_1 = 0$, por exemplo, pode ser escrita como $H_{05} : C\theta^* = q_0$, onde $C = (1, 0, 0, 0, 0)'$ e $q_0 = 0$.

Assim, a estatística de Wald é dada por

$$W_{05} = n(C\hat{\theta}^* - q_0)'(C\hat{\Omega}_{\theta^*}C')^{-1}(C\hat{\theta}^* - q_0) ,$$

com um grau de liberdade. Neste caso, $W_{05} = 93,88$, de modo que $H_{05} : \gamma_1 = 0$ é rejeitada, o que indica que a variável ω pode ser utilizada como uma covariável. Isto coincide com um resultado similar, dado em Fuller (1987). Analogamente, para a hipótese $H_{01} : \alpha_2 = 0$, $\beta_2 = 1$, $\rho_1 = \rho_2$, temos que $W_{01} = 23,60$, com três graus de liberdade, indicando que os dois instrumentos não são viciados e igualmente confiáveis simultaneamente. Com respeito a $H_{02} : \alpha_2 = 0$, $\beta_2 = 1$, temos $W_{02} = 18,70$, com dois graus de liberdade, o que indica que o instrumento 2 é viciado. Finalmente, para $H_{03} : \rho_1 = \rho_2$, temos $W_{03} = 0,69$, com um grau de liberdade, ou seja, os dois instrumentos são igualmente confiáveis.

Como conclusão podemos dizer que os instrumentos medem com vício a intensidade do terremoto e são igualmente confiáveis.

3.3.5 Um estudo de simulação

Nesta seção apresentamos um pequeno estudo de simulação, no qual comparamos a “eficiência” dos EMV das confiabilidades em presença e ausência de uma covariável.

500 amostras foram geradas de acordo com o modelo (3.29), com $p = 3$, $\alpha_2 = 2$, $\alpha_3 = -2$, $\beta_2 = 0,9$, $\beta_3 = 1,2$, $\phi_1 = 5,0$, $\phi_2 = 7,5$, $\phi_3 = 5,0$, $\phi_x = 50$, $\gamma_0 = 10$, $\mu_\omega = 0,0$, $\phi_\omega = 1,0$; para $n = 25, 50, 75, 100$ e $\rho_x = 0,50$ ($\gamma_1^2 = 25$, $\phi_e = 25$), $\rho_x = 0,80$ ($\gamma_1^2 = 40$, $\phi_e = 10$), $\rho_x = 0,95$ ($\gamma_1^2 = 47,5$, $\phi_e = 2,5$).

Sejam:

$\hat{\rho}_{ij}$: a estimativa de MV, ver Seção 3.3.2, de ρ_i na amostra j , $i = 1, 2, 3$, $j = 1, 2, \dots, 500$;

ρ_{iT} : valor verdadeiro de ρ_i , $i = 1, 2, 3$.

$EQ_{0i} = \frac{1}{500} \sum_{j=1}^{500} (\hat{\rho}_{ij} - \rho_{iT})^2$ o erro quadrático médio $\hat{\rho}_i$ no modelo sem covariável, $i = 1, 2, 3$.

EQ_{1i} idem, modelo com covariável e

$EFR_i = \frac{EQ_{1i}}{EQ_{0i}}$, $i = 1, 2, 3$, a eficiência relativa empírica de $\hat{\rho}_i$ no modelo sem covariável com respeito a $\hat{\rho}_i$ no modelo com covariável. Os resultados são mostrados na Tabela 3.2.

Pode-se observar que para n e ρ_x pequenos os EMV de ρ_i no modelo sem covariável são mais eficientes. No entanto, quando n e ρ_x crescem o EMV de ρ_i no modelo com covariável é muito mais eficiente. Note-se que, para n fixo, aumentando o grau de correlação entre a covariável (ω) e a característica de interesse (x) (medida aqui por ρ_x), aumenta também a eficiência dos EMV de ρ_i no modelo com covariável.

Tabela 3.2: Eficiência relativa dos EMV das confiabilidades no modelo sem covariável com respeito ao modelo com covariável

Tamanho da amostra (n)	$\rho_x = 0,50$			$\rho_x = 0,80$			$\rho_x = 0,95$		
	EFR_1	EFR_2	EFR_3	EFR_1	EFR_2	EFR_3	EFR_1	EFR_2	EFR_3
25	1,30	1,03	1,31	1,02	0,95	0,93	0,78	0,88	0,62
50	1,32	1,08	1,04	1,13	1,01	0,83	0,87	0,93	0,59
75	1,00	0,95	0,88	0,88	0,90	0,74	0,73	0,82	0,57
100	0,92	0,94	0,87	0,84	0,89	0,70	0,70	0,82	0,51

Capítulo 4

O modelo estrutural t -elíptico

4.1 Introdução

Nos capítulos anteriores consideramos modelos estruturais normais, i.e., onde tanto a característica de interesse x quanto os erros de medição dos instrumentos seguem uma distribuição normal. Como método de estimação foi usado máxima verossimilhança.

Certamente a distribuição normal é útil em muitos casos e grande parte da inferência estatística para variáveis contínuas tem sido desenvolvida sob a suposição de normalidade. Isto é particularmente o caso de análise multivariada e regressão linear.

No entanto, é sabido que a inferência baseada na distribuição normal é vulnerável à presença de “outliers” nos dados que provêm de distribuições com caudas mais pesadas. Ou seja, é um fato reconhecido que os “outliers” têm uma grande influência sobre os estimadores e testes baseados em máxima verossimilhança. Assim, a distribuição normal pode não ser apropriada em situações onde os dados provêm de distribuições com caudas mais pesadas.

Isto tem provocado inúmeros estudos sobre procedimentos ou métodos para a análise de dados que contém “outliers” ou observações aberrantes. E assim como na atualidade existem em diversas áreas da estatística uma variedade de métodos para analisar dados em presença de “outliers”.

Um enfoque é através da identificação e eliminação de “outliers” seguido de inferência

estatística (máxima verossimilhança) baseada na distribuição normal. De fato, uma ampla área de pesquisa é detecção de dados atípicos, especialmente em modelos lineares e análise multivariada. Algumas referências básicas são Barnett e Lewis (1994), Belsley et al. (1980), Chatterjee e Hadi (1988), Cook e Weisberg (1982), Hawkins (1980) e Peña e Yohai (1995). Na área de modelos de regressão com erros nas variáveis dois trabalhos pioneiros são Kelly (1984) e Wellman e Gunst (1991). Uma referência básica para influência local é Cook (1986).

Um outro enfoque é a acomodação de observações aberrantes usando procedimentos robustos clássicos. Nesta direção, métodos robustos têm sido desenvolvidos, modificando o método de máxima verossimilhança, de modo que os “outliers” tenham menos influência sobre as estimativas finais; ver Huber (1981), Hampel et al. (1986), Rousseeau e Leroy (1987) e Staudte e Sheather (1990). Na área de modelos de regressão com erros nas variáveis algumas referências são Kelly (1984), Zamar (1989) e Cheng e Van Ness (1990).

Uma outra possibilidade é usar distribuições simétricas, com caudas mais pesadas que a distribuição normal, as quais permitem reduzir a influência dos “outliers” sobre os EMV. Uma classe de distribuições que oferece várias alternativas neste sentido é a classe de distribuições elípticas. Esta classe de distribuições, que contém as distribuições normal, t e normal contaminada multivariadas, entre outras, tem tido um interesse crescente nos últimos anos; ver por exemplo, Fang e Zhang (1990), Fang et al. (1990), Fang e Anderson (1990), Kano et al. (1993), Berkane et al. (1994) e Arellano (1994). Dentro desta classe uma das distribuições mais utilizadas como alternativa à normal é a distribuição t .

Recentemente, vários autores têm sugerido a distribuição t multivariada como alternativa à normal multivariada por ter as caudas mais pesadas e, portanto, “acomoda” possíveis “outliers” presentes nos dados. Assim, o uso da distribuição t forneceria um procedimento robusto de análise de dados. Por exemplo, Rubin (1983) obteve os EMV para os parâmetros da distribuição t multivariada usando o algoritmo EM; Little (1988)

estende os resultados de Rubin (1983) a dados incompletos, Sutradhar e Ali (1986) usam máxima verossimilhança no modelo de regressão multivariado com erros distribuídos t . Lange et al. (1989) ilustram o uso da distribuição t em regressão e em algumas áreas de análise multivariada, ver também Taylor (1992). Mais recentemente, Sutradhar (1993) propôs um teste escore para testar que a matriz de covariância é igual a uma matriz dada, usando também a distribuição t ; Bolfarine e Arellano (1994) usam a distribuição t no modelo linear simples com erros nas variáveis.

Neste capítulo, devido à inferência estatística baseada na distribuição t combinar simplicidade conceitual e computacional com generalidade, já que pode ser aplicada numa grande variedade de situações, é que propomos como alternativa a distribuição normal, usada nos capítulos anteriores, a distribuição t multivariada, com segundo momento finito. Para estimar parâmetros usamos o algoritmo EM.

A distribuição t p -dimensional com média μ , matriz de covariância V e ν graus de liberdade, será referida como $T_p(\mu, V, \nu)$ ou simplesmente como $T(\mu, V, \nu)$. Se um vetor aleatório Y , $p \times 1$, $\sim T_p(\mu, V, \nu)$ sua densidade, ver Apêndice B, é dada por

$$f_Y(y) = C(\nu, p) |V|^{-\frac{1}{2}} \{(\nu - 2) + (y - \mu)' V^{-1} (y - \mu)\}^{-\frac{1}{2}(\nu + p)} \quad (4.1)$$

onde

$$C(\nu, p) = [(\nu - 2)^{\frac{p}{2}} \Gamma((\nu + p)/2) / \pi^{p/2} \Gamma(\nu/2)] , \quad \nu > 2.$$

Esta distribuição t é obtida da clássica distribuição t multivariada, ver Apêndice B, mediante uma reparametrização de modo que a matriz de covariância de Y não depende dos graus de liberdade (ν) da distribuição. Isto permite uma comparação direta com a distribuição normal p -dimensional. Ademais, quando $\nu \rightarrow \infty$ a distribuição t reduz-se à distribuição normal. Mais especificamente, no modelo estrutural discutido no Capítulo 2, propomos como alternativa à $N_p(\mu, V)$ a distribuição $T_p(\mu, V, \nu)$, $\nu > 2$. Esta distribuição incorpora um parâmetro adicional (ν) que permite ajustar a curtose da distribuição.

Este parâmetro pode ser fixado a priori; Lange et al. (1989) e Berkane et al. (1994) recomendam $\nu = 4$, ou estimado desde os dados. Ver, também, Taylor (1992).

Na Seção 4.2 analisamos o caso de dois instrumentos. Primeiro apresentamos a distribuição assintótica do EMV no modelo estrutural elíptico. Depois discutimos o modelo estrutural t -elíptico; o algoritmo EM é usado para encontrar os EMV. A matriz de covariância assintótica do EMV é obtida e discutimos implementação de teste de hipóteses.

A Seção 4.3 é dedicada ao caso de três ou mais instrumentos. Apresentamos a implementação do algoritmo EM e a obtenção da matriz de informação na parametrização de análise de fatores do modelo t -elíptico. Consideramos também testes de hipóteses usando a estatística de Wald. Finalmente, apresentamos a matriz de informação no modelo estrutural elíptico e uma ilustração numérica.

4.2 O caso de dois instrumentos

Nesta seção apresentamos alguns resultados utilizando a parametrização de análise de fatores do modelo estrutural. Resultados análogos, usando a versão de regressão do modelo, podem ser obtidos facilmente a partir das relações entre as parametrizações discutidas no Capítulo 1.

4.2.1 Matriz de informação no modelo estrutural elíptico

Consideremos o modelo (1.17):

$$Y_j = \mu + Lv_j, \quad j = 1, 2, \dots, n, \quad (4.2)$$

onde $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $L = (\lambda, I_2)$, $\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$ e $v_j = \begin{pmatrix} F_j \\ \epsilon_j \end{pmatrix}$ com $E(v_j) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, $\text{Var}(v_j) = \Sigma_v$,

$$j = 1, \dots, n \text{ e } \Sigma_v = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & \phi \end{pmatrix}.$$

Daqui segue que:

$$E(Y_j) = \mu \quad \text{e} \quad \text{Var}(Y_j) = V = \lambda\lambda' + \phi I_2, \quad j = 1, 2, \dots, n. \quad (4.3)$$

No Capítulo 2 assumimos que $v_1, v_2, \dots, v_n$ são v.a.i.i.d. $N_3(0, \Sigma_v)$. A idéia aqui é supor uma distribuição elíptica para v_j , $j = 1, 2, \dots, n$.

Seja Y um vetor aleatório p -dimensional. Diz-se que Y tem uma distribuição elíptica simétrica, com parâmetros $\mu(p \times 1)$ e $\Lambda(p \times p)$ se a densidade de Y é dada por, ver Apêndice A,

$$f_Y(y) = |\Lambda|^{-1/2} g[(y - \mu)' \Lambda^{-1} (y - \mu)] \quad (4.4)$$

para alguma função g e Λ definida positiva. Usaremos a notação $Y \sim El_p(\mu, \Lambda)$ para indicar que Y tem uma densidade dada por (4.4). A função característica é dada por $\exp(it'\mu)\psi(t'\Lambda t)$ para alguma função ψ , onde $i = \sqrt{-1}$, $t \in \mathbb{R}^p$. Supondo existência de segundo momento, temos que

$$E(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = c_g \Lambda \quad (4.5)$$

com $c_g = -2\psi'(0)$ uma constante. Algumas propriedades das distribuições elípticas são resumidas no Apêndice A.

Nesta seção vamos considerar o modelo (4.2):

$$Y_j = \mu + Lv_j, \quad j = 1, 2, \dots, n, \quad (4.6)$$

onde, agora, assumimos que v_j , $1 \leq j \leq n$, são v.a.i.i.d. $El_3(0, \Lambda_v)$ e matriz de covariância $\Sigma_v = c_g \Lambda_v$ dada por (4.2). Esta é a versão da análise de fatores do modelo estrutural elíptico.

Das propriedades das distribuições elípticas, ver Apêndice A, segue que:

$$Y_i, \quad 1 \leq j \leq n, \quad \text{são v.a.i.i.d. } El_2(\mu, \Lambda_Y), \quad (4.7)$$

onde $\Lambda_Y = L\Lambda_v L'$ e matriz de covariância $V = c_g \Lambda_Y$ dada em (4.3).

A seguir apresentamos a matriz de informação. Seja $\theta = (\mu_1, \mu_2, \lambda_1, \lambda_2, \phi)'$.

Lema 4.1 Sob o modelo (4.6) a matriz de informação $I_E(\theta)$ é dada por

$$I_E(\theta) = n \begin{bmatrix} I_{\mu\mu} & 0 & 0 \\ 0 & I_{\lambda\lambda} & I_{\lambda\phi} \\ 0 & I_{\phi\lambda} & i_{\phi\phi} \end{bmatrix} = n\Omega_E^{-1}, \quad (4.8)$$

onde

$$\begin{aligned} I_{\mu\mu} &= 2c_g a_1 V^{-1}, \\ I_{\lambda\lambda} &= \frac{1}{c^2 \phi^2} \left[1 + 2a_2 + \frac{a_3}{2} - \frac{a_3}{2} c(1 - 2c^{-1}) \right] \lambda \lambda' + \frac{a_3}{2\phi} c^{-1} (c - 1) I_2, \\ I_{\lambda\phi} &= \frac{1}{c^2 \phi^2} \left[(1 + c) \left(\frac{1}{2} + a_2 \right) + \frac{a_3}{4} (c + 3) \right] \lambda, \quad I_{\phi\lambda} = I'_{\lambda\phi}, \\ i_{\phi\phi} &= \frac{1}{2\phi^2} (1 + c^{-1})^2 \left(\frac{1}{2} + a_2 + \frac{a_3}{4} \right) + \frac{a_3}{4\phi^2} (1 + c^{-2}), \end{aligned}$$

com

$$\begin{aligned} a_1 &= E[\|z\|^2 (W_g(\|z\|^2))^2], \\ a_2 &= E[\|z\|^2 W_g(\|z\|^2)], \\ a_3 &= E[\|z\|^4 (W_g(\|z\|^2))^2], \\ c &= 1 + \lambda' \lambda / \phi, \quad z = \Lambda_Y^{-1/2} (Y_1 - \mu) \quad e \\ W_g(x) &= \frac{d}{dx} (\ln g(x)) = g'(x) / g(x). \end{aligned}$$

Prova: De (4.7) segue que o logaritmo da função de verossimilhança, $\ell_E(\theta)$, é dado por:

$$\ell_E(\theta) = \sum_{j=1}^n \ell_j, \quad (4.9)$$

onde

$$\ell_j = -\frac{1}{2} \ln |\Lambda_Y| + \ln g[(y_j - \mu)' \Lambda_Y^{-1} (y_j - \mu)],$$

$j = 1, 2, \dots, n$, com

$$|\Lambda_Y| = \phi^2 c / c_g^2 \quad e \quad \Lambda_Y^{-1} = c_g \left[\frac{1}{\phi} I_2 - \frac{c^{-1}}{\phi^2} \lambda \lambda' \right].$$

Agora, usando a seguinte fórmula dada por Lange et al. (1989) (ver também Arellano, 1994),

$$\begin{aligned}
 -E \left(\frac{\partial^2 \ell_E(\theta)}{\partial \theta_i \partial \theta_j} \right) &= \frac{4na_1}{p} \frac{\partial \mu'}{\partial \theta_i} \Lambda_Y^{-1} \frac{\partial \mu}{\partial \theta_j} \\
 &+ n \left(\frac{1}{4} + \frac{a_2}{p} + \frac{a_3}{p(p+2)} \right) \text{tr} \left(\Lambda_Y^{-1} \frac{\partial \Lambda_Y}{\partial \theta_i} \right) \text{tr} \left(\Lambda_Y^{-1} \frac{\partial \Lambda_Y}{\partial \theta_j} \right) \\
 &+ \frac{2na_3}{p(p+2)} \text{tr} \left(\Lambda_Y^{-1} \frac{\partial \Lambda_Y}{\partial \theta_i} \Lambda_Y^{-1} \frac{\partial \Lambda_Y}{\partial \theta_j} \right),
 \end{aligned} \tag{4.10}$$

onde $p = 2$ denota a dimensão de Y_j , $j = 1, 2, \dots, n$ e

$$\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (\mu_1, \mu_2, \lambda_1, \lambda_2, \phi)',$$

o resultado segue depois de uma longa álgebra.

Seja $\hat{\theta}_E$ o EMV de θ no modelo estrutural elíptico.

Teorema 4.1 *Sob o modelo estrutural elíptico (4.6), temos que*

$$\sqrt{n}(\hat{\theta}_E - \theta) \xrightarrow{D} N_5(0; \Omega_E),$$

onde

$$\Omega_E = \begin{bmatrix} \Omega_\mu & 0 & 0 \\ 0 & \Omega_\lambda & \Omega_{\lambda\phi} \\ 0 & \Omega_{\phi\lambda} & \omega_\phi \end{bmatrix}, \tag{4.11}$$

com

$$\begin{aligned}
 \Omega_\mu &= (2c_g a_1)^{-1} V, \\
 \Omega_\lambda &= \frac{1}{c_1} I_2 + \left[\frac{c_3^2}{h^2 c_1^2 i_{\phi\phi} - h c_1 c_3^2 (c-1)\phi} - \frac{c_2}{c_1^2 h} \right] \lambda \lambda', \\
 \Omega_{\lambda\phi} &= - \left(\frac{c_3}{\epsilon h c_1} \right) \lambda, \quad \omega_\phi = e^{-1}, \quad \Omega_{\phi\lambda} = \Omega'_{\lambda\phi} \quad e \\
 c_1 &= a_3(1 - c^{-1})/2\phi, \\
 c_2 &= (1 + 2a_2 + \frac{a_3}{2} - \frac{c a_3}{2}(1 - 2c^{-1}))/c^2 \phi^2, \\
 c_3 &= \left((1+c)(\frac{1}{2} + a_2) + \frac{a_3}{4}(c+3) \right) / c^2 \phi^2, \\
 h &= 1 + c_2(c-1)\phi/c, \quad e = i_{\phi\phi} - c_3^2(c-1)\phi/hc_1.
 \end{aligned}$$

Prova: O resultado segue das propriedades assintóticas do EMV (Sen e Singer, 1993), de inversa de matrizes subdivididas, do teorema 8.9.3 em Graybill (1983) e de bastante álgebra para simplificar.

No caso em que o modelo elíptico seja o modelo normal, temos que

$$g(x) = \text{cte.} e^{-x/2}, \quad c_g = 1, \quad W_g(x) = -\frac{1}{2}$$

e que $a_1 = 1/2$, $a_2 = -1$ e $a_3 = 2$. Neste caso a matriz de informação $I_E(\theta)$ reduz-se a $I(\theta)$ dada em (2.11) para o modelo estrutural normal. O mesmo acontece com a matriz de covariância assintótica Ω_E , que reduz-se a Ω_θ dada em (2.12).

A seguir discutimos estimação e testes de hipóteses no modelo estrutural usando a distribuição t .

4.2.2 Modelo estrutural t -elíptico

A versão de análise de fatores no modelo estrutural t -elíptico é dada por, ver (4.2),

$$\left. \begin{array}{l} Y_j = \mu + Lv_j, \quad j = 1, 2, \dots, n; \\ v_j, \quad 1 \leq j \leq n, \text{ são i.i.d. } T_3(0, \Sigma_v, \nu) \end{array} \right\} \quad (4.12)$$

Daqui segue que (ver Apêndice B):

$$Y_j, \quad j = 1, 2, \dots, n, \text{ são i.i.d. } T_2(\mu, V, \nu) \quad (4.13)$$

com μ e V como definidos em (4.3).

Em seguida mostramos como obter os EMV para θ e ν usando o algoritmo EM. Aqui, $\theta = (\mu', \lambda', \phi)$.

(i) Estimação via algoritmo EM

Sabemos que a distribuição t pode ser obtida como mistura de uma distribuição normal e uma χ^2 . Com efeito, seja

$$\left. \begin{array}{l} Y_j | q_j \sim N_2(\mu, q_j^{-1} \Lambda_Y) \quad \text{e} \\ q_j \sim \chi^2(\nu)/\nu, \quad j = 1, 2, \dots, n. \end{array} \right\} \quad (4.14)$$

Isto implica que $Y_j \sim T_2(\mu, V, \nu)$, $j = 1, 2, \dots, n$, como em (4.13).

Seja $f(y_j, q_j)$ a densidade conjunta de (Y_j, q_j) , $j = 1, 2, \dots, n$. Logo, podemos escrever

$$f(y_j, q_j) = f_1(y_j/q_j)f_2(q_j), \quad j = 1, 2, \dots, n.$$

Assim, o logaritmo da função de verossimilhança completa é:

$$\begin{aligned} \ell_z(\theta) &= \sum_{j=1}^n \ln f_1(y_j/q_j) + \sum_{j=1}^n \ln f_2(q_j) \\ &= \text{cte.} - n \ln[(\nu - 2)/\nu] - \frac{n}{2} \ln |V| \\ &\quad - \frac{\nu}{2(\nu - 2)} \sum_{j=1}^n q_j (y_j - \mu)' V^{-1} (y_j - \mu) + \sum_{j=1}^n \ln f_2(q_j). \end{aligned} \quad (4.15)$$

Etapa E: Para implementar a etapa E do algoritmo calculamos $E(q_j/Y_j, \theta)$, $j = 1, 2, \dots, n$.

Do Apêndice B, temos que

$$q_j/Y_j \sim \chi^2(\nu + 2)/(\nu + d_j^2), \quad (4.16)$$

onde

$$\begin{aligned} d_j^2 &= (Y_j - \mu)' \Lambda_Y^{-1} (Y_j - \mu) \\ &= \left(\frac{\nu}{\nu - 2} \right) (y_j - \mu)' V^{-1} (y_j - \mu) \\ &= \left(\frac{\nu}{\nu - 2} \right) (h_{1j} - h_{2j}), \end{aligned}$$

com

$$\begin{aligned} h_{1j} &= \frac{1}{\phi} [(y_{1j} - \mu_1)^2 + (y_{2j} - \mu_2)^2] \quad \text{e} \\ h_{2j} &= \frac{c^{-1}}{\phi^2} [\lambda_1 (y_{1j} - \mu_1) + \lambda_2 (y_{2j} - \mu_2)]^2, \quad j = 1, 2, \dots, n. \end{aligned}$$

De (4.16) segue, então, que

$$E(q_j|Y_j, \theta) = \frac{\nu + 2}{\nu + d_j^2}, \quad j = 1, 2, \dots, n. \quad (4.17)$$

Etapa M: Esta etapa do algoritmo é implementada maximizando $\ell_z(\theta)$ dada em (4.15).

Se ν for fixado a priori, o valor de θ que maximiza $\ell_z(\theta)$ é dado por

$$\left. \begin{aligned} \hat{\mu}_i &= \sum_{j=1}^n q_j y_{ij} / \sum_{j=1}^n q_j, \quad i = 1, 2, \\ \hat{\lambda}_1 &= \left\{ \left(\frac{\nu}{\nu-2} \right) (S_{q11} - S_{q22} + 2bS_{q12}) / (1+b^2) \right\}^{1/2}, \\ \hat{\lambda}_2 &= b\hat{\lambda}_1, \\ \hat{\phi} &= \left(\frac{\nu}{\nu-2} \right) (S_{q22} - bS_{q12}) \end{aligned} \right\} \quad (4.18)$$

onde

$$b = \left[\frac{S_{q22} - S_{q11}}{2S_{q12}} \right] + \left[\left(\frac{S_{q22} - S_{q11}}{2S_{q12}} \right)^2 + 1 \right]^{1/2} \text{ e}$$

$$S_{qik} = \frac{1}{n} \sum_{j=1}^n q_j (y_{ij} - \hat{\mu}_i)(y_{kj} - \hat{\mu}_k); \quad i, k = 1, 2.$$

Se ν não for fixado a priori, diferenciando $\ell_z(\theta)$ em (4.15) em relação a ν , temos a equação

$$2d_1/\nu(\nu-2) - 4/\nu(\nu-2) - (\psi(\nu/2) - \ln(\nu/2)) + d_2 = 0, \quad (4.19)$$

onde

$$d_1 \equiv \frac{\text{tr}(S_q)}{S_{q22} - S_{q12}b} - \frac{(S_{q11} - S_{q22} + 2bS_{q12})b'_* S_q b_*}{c(1+b^2)},$$

$$d_2 = 1 - \frac{1}{n} \sum_{j=1}^n q_j + \frac{1}{n} \sum_{j=1}^n \ln q_j,$$

com $b_* = \begin{pmatrix} 1 \\ b \end{pmatrix}$, $S_q = [(S_{qij})]$ (2×2) e $\psi(\nu/2) = \Gamma'(\nu/2)/\Gamma(\nu/2)$ é a função digama.

A equação anterior, (4.19), deve ser resolvida numericamente dentro da etapa M, fazendo com que a convergência do algoritmo seja mais lenta. Alternativamente, como sugerido por Lange et al. (1989), uma estimativa de ν pode ser obtida calculando o logaritmo da função de verossimilhança observada, $\ell(\theta)$, sobre um reticulado de valores de ν , e escolher o valor de ν que maximiza $\ell(\theta)$. Como estimativa inicial para θ podemos usar o EMV no caso normal, ver (2.9), para ν podemos começar com $\nu_0 = 4$.

(ii) Matriz de covariância assintótica do EMV

Aqui apresentamos a matriz de covariância assintótica do EMV, $\hat{\theta}_T$, de θ no modelo estrutural t .

Lema 4.2 *Sob o modelo estrutural t (4.12), temos:*

(i) *Se ν for fixado a priori, então:*

(i1) *A matriz de informação $I_T(\theta)$ é dada por*

$$I_T(\theta) = n \begin{bmatrix} I_{\mu\mu} & 0 & 0 \\ 0 & I_{\lambda\lambda} & I_{\lambda\phi} \\ 0 & I_{\phi\lambda} & i_{\phi\phi} \end{bmatrix}, \quad (4.20)$$

onde

$$\begin{aligned} I_{\mu\mu} &= \left(\frac{\nu}{\nu-2} \right) \left(\frac{\nu+2}{\nu+4} \right) V^{-1}, \\ I_{\lambda\lambda} &= \frac{1}{\phi^2} \left[\left(\frac{\nu+2}{\nu+4} \right) c^{-1} (3c^{-1} - 1) - c^{-2} \right] \lambda\lambda' + \frac{1}{\phi} (1 - c^{-1}) \left(\frac{\nu+2}{\nu+4} \right) I_2 \\ I_{\lambda\phi} &= \frac{1}{2c^2\phi^2} \left[-(1+c) + (c+3) \left(\frac{\nu+2}{\nu+4} \right) \right] \lambda, \quad I_{\phi\lambda} = I'_{\lambda\phi} \\ i_{\phi\phi} &= \frac{1}{4\phi^2} \left[-(1+c^{-1})^2 + (3+2c^{-1}+3c^{-2}) \left(\frac{\nu+2}{\nu+4} \right) \right]. \end{aligned}$$

(i2) $\sqrt{n}(\hat{\theta}_T - \theta) \xrightarrow{D} N_5(0, \Omega_T)$, onde

$$\Omega_T = \begin{bmatrix} \Omega_\mu & 0 & 0 \\ 0 & \Omega_\lambda & \Omega_{\lambda\phi} \\ 0 & \Omega_{\phi\lambda} & \omega_\phi \end{bmatrix}, \quad (4.21)$$

com

$$\begin{aligned} \Omega_\mu &= d \left(1 - \frac{2}{\nu} \right) V, \\ \Omega_\lambda &= \frac{1}{c_1} I_2 + \omega_1(d) \lambda\lambda', \quad \Omega_{\lambda\phi} = \omega_2(d) \lambda, \quad \Omega_{\phi\lambda} = \Omega'_{\lambda\phi}, \\ \omega_\phi &= \phi^2(3d-1)/d(2d-1) \quad e \\ d &= (\nu+2)/(\nu+4), \quad c_1 = d(c-1)/c\phi, \\ \omega_1(d) &= (9d^2 - 6d + 1)/4d(3d-1)(2d-1) \quad e \\ \omega_2(d) &= 2\phi((1+c) - d(c+3))/4d(2d-1)(c-1). \end{aligned}$$

(ii) Se ν não for fixado a priori, os elementos adicionais da matriz de informação $(I_T(\theta))$, $I_T(\theta, \nu)$, são:

$$\begin{aligned} I_{\mu\nu} &= 0, \quad I_{\lambda\nu} = \left[\frac{8c^{-1}\phi^{-1}}{(\nu-2)(\nu+2)(\nu+4)} \right] \lambda, \\ i_{\phi\nu} &= 4\phi^{-1}(1+c^{-1})/(\nu-2)(\nu+2)(\nu+4) \quad e \\ i_{\nu\nu} &= -(\nu-2)/2(\nu-2)^2 - \frac{1}{4}\psi' \left(\frac{\nu+2}{2} \right) + \frac{1}{4}\psi' \left(\frac{\nu}{2} \right) \\ &\quad + \frac{1}{2}\nu(\nu^2-20)/(\nu-2)^2(\nu+2)(\nu+4), \end{aligned} \quad (4.22)$$

onde $\psi(x) = \Gamma'(x)/\Gamma(x)$ é a função digama.

Prova: Na distribuição t p -variada temos que, ver Apêndice A,

$$\begin{aligned} g(x) &= \text{cte.} (1+x/\nu)^{-\frac{1}{2}(\nu+p)}, \quad x \geq 0; \\ c_g &= \nu/(\nu-2), \quad \nu > 2 \quad e \\ W_g(x) &= -\frac{(\nu+p)}{2\nu} (1+x/\nu)^{-1}. \end{aligned} \quad (4.23)$$

Para $p = 2$, temos que,

$$a_1 = (\nu+2)/2(\nu+4), \quad a_2 = -1 \quad e \quad a_3 = 2(\nu+2)/(\nu+4).$$

(i) Para provar (i1), basta usar os resultados anteriores e o Lema 4.1. Uma prova alternativa, sem usar o Lema 4.1, pode ser obtida utilizando os fatos dados no Apêndice D. Para provar (i2) usamos o Teorema 4.1 em conjunto com os fatos anteriores.

(ii) Ver Apêndice D.

Observe que quando $\nu \rightarrow \infty$ obtemos as matrizes de informação e de covariância assintótica para o modelo estrutural normal, ver Seção 2.2.

(iii) Testes de hipóteses

Usando os resultados anteriores, podemos implementar testes assintóticos, para provar hipóteses de interesse tais como H_{01} e H_{03} , utilizando por exemplo a estatística de Wald, ver Seção 2.2. Também é possível utilizar razão de verossimilhança e o teste escore.

A seguir, no entanto, discutiremos um teste assintótico alternativo para testar se os instrumentos são igualmente confiáveis: $H_{03} : \rho_1 = \rho_2$ ou equivalentemente $H_{04} : \lambda_1 = \lambda_2$; ver Seção 2.2.

Com efeito, temos que

$$Y_j, 1 \leq j \leq n, \text{ são i.i.d. } T_2 \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}; \nu \right), \quad (4.24)$$

onde

$$\sigma_{11} = \lambda_1^2 + \phi, \quad \sigma_{12} = \sigma_{21} = \lambda_1 \lambda_2 \quad \text{e} \quad \sigma_{22} = \lambda_2^2 + \phi.$$

Isto é, temos, ver Seção 2.2, que H_{03} é equivalente a $H : \sigma_{11} = \sigma_{22}$.

Consideremos novamente a transformação dada em (2.20),

$$X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \end{pmatrix} = \begin{pmatrix} y_{1j} + y_{2j} \\ y_{1j} - y_{2j} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} Y_j; \quad j = 1, 2, \dots, n. \quad (4.25)$$

Das propriedades da distribuição t , ver Apêndice B, segue que

$$X_j, \quad j = 1, \dots, n, \text{ são i.i.d. } T_2 \left(\begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1 - \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} + \sigma_{22} + 2\sigma_{12} & \sigma_{11} - \sigma_{22} \\ \sigma_{11} - \sigma_{22} & \sigma_{11} + \sigma_{22} - 2\sigma_{12} \end{pmatrix}, \nu \right) \quad (4.26)$$

Seja ρ_{12} o coeficiente de correlação entre x_1 e x_2 . Como discutido na Seção 2.2, temos que a hipótese $H : \sigma_{11} = \sigma_{22}$ é equivalente à hipótese $H : \rho_{12} = 0$. Ou seja, testar que os instrumentos são igualmente confiáveis é equivalente a provar que o coeficiente de correlação entre x_1 e x_2 é zero. Para isto, consideremos o seguinte fato, provado por Muirhead (1980, 1982).

Lema 4.3 *Seja $X_1, X_2, \dots, X_n$ uma amostra aleatória de tamanho n de uma distribuição elíptica bivariada com coeficiente de correlação ρ_{12} e parâmetro de curtose κ . Então tem-se que:*

$$\frac{(n-1)(r_{12} - \rho_{12})^2}{(1 - \rho_{12}^2)(1 + \kappa)} \xrightarrow{\mathcal{D}} \chi^2(1) , \quad (4.27)$$

onde $r_{12} = S_{12}/(S_{11}S_{22})^{1/2}$ é o coeficiente de correlação amostral.

Este fato pode ser usado para testar $H : \rho_{12} = 0$. Com efeito, se H é verdadeira, então,

$$\frac{(n-1)r_{12}^2}{(1 + \kappa)} \xrightarrow{\mathcal{D}} \chi^2(1) ,$$

de modo que um teste aproximado de tamanho α é rejeitar H se

$$\frac{(n-1)r_{12}^2}{(1 + \kappa)} > \chi_{1-\alpha}^2(1) , \quad (4.28)$$

onde $\chi_{1-\alpha}^2(1)$ denota o percentil $100(1 - \alpha)$ da distribuição $\chi^2(1)$.

Se κ não é conhecido, pode ser substituído por uma estimativa consistente.

Do Apêndice A temos que $\kappa = \beta_{2,p}/p(p+2) - 1$, onde $\beta_{2,p}$ é o coeficiente de curtose multivariada. Aqui, $p = 2$ é uma estimativa consistente de $\beta_{2,2}$ e é dada por (Mardia, 1970):

$$b_{2,2} = \frac{1}{n} \sum_{j=1}^n [(X_j - \bar{X})' S_x^{-1} (X_j - \bar{X})]^2 , \quad (4.29)$$

com

$$S_x = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})' \quad \text{e} \quad \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j .$$

Logo, uma estimativa consistente para κ é

$$\tilde{\kappa} = \frac{b_{2,2}}{8} - 1 . \quad (4.30)$$

Notemos que o teste (4.28) é válido em qualquer distribuição elíptica com quarto momento finito.

No caso da distribuição t p -variada sabemos que (ver Apêndice B) $\kappa = 2/(\nu - 4)$ se $\nu > 4$. Daqui segue que uma estimativa consistente para ν é dada por (ver também (4.30)):

$$\tilde{\nu} = 4(b_{2,2} - 4)/(b_{2,2} - 8) . \quad (4.31)$$

4.3 O caso de três ou mais instrumentos

Nesta seção, discutimos estimação e testes de hipóteses no modelo estrutural no caso $p \geq 3$, usando a distribuição t -elíptica. Em primeiro lugar, mostramos como obter os EMV, utilizando o algoritmo EM na parametrização de análise de fatores do modelo. Obtemos também a matriz de informação e discutimos testes de hipóteses baseados em amostras grandes. Finalmente, apresentamos a matriz de informação no modelo estrutural elíptico e uma ilustração numérica.

4.3.1 Modelo estrutural t -elíptico

Consideremos o modelo estrutural (1.11):

$$Y_j = \mu + Lv_j, \quad j = 1, 2, \dots, n. \quad (4.32)$$

com $\mu = (\mu_1, \dots, \mu_p)'$, $\lambda = (\lambda_1, \dots, \lambda_p)'$, $L = (\lambda, I_p)$ e $v_j = (F_j, \varepsilon_j')'$, $j = 1, 2, \dots, n$, vetores aleatórios não correlacionados, com

$$E(v_j) = 0 \quad \text{e} \quad \text{Var}(v_j) = \begin{pmatrix} 1 & 0 \\ 0 & D(\phi) \end{pmatrix} = \Sigma_v \quad (4.33)$$

$j = 1, 2, \dots, n$.

Daqui segue que:

$$E(Y_j) = \mu \quad \text{e} \quad \text{Var}(Y_j) = \lambda\lambda' + D(\phi); \quad j = 1, \dots, n. \quad (4.34)$$

No Capítulo 2 assumimos normalidade dos vetores v_j e, logo, normalidade dos Y_j , $j = 1, 2, \dots, n$. Nesta seção propomos substituir a normal pela distribuição t .

A versão de análise de fatores do modelo estrutural t -elíptico é dada por

$$\left. \begin{array}{l} Y_j = \mu + Lv_j, \quad j = 1, 2, \dots, n; \\ v_j, \quad 1 \leq j \leq n, \text{ são i.i.d. } T_{p+1}(0, \Sigma_v, \nu) \end{array} \right\} \quad (4.35)$$

Pelas propriedades da distribuição t , ver Apêndice B, temos que:

$$Y_j, \quad 1 \leq j \leq n, \text{ são i.i.d. } T_p(\mu, V, \nu) \quad (4.36)$$

com μ e V definidos em (4.34).

(i) Estimação usando o algoritmo EM

Sejam $\theta = (\mu', \lambda', \phi')'$. A seguir discutimos a implementação do algoritmo EM para encontrar os EMV.

Seja $Z_j = \begin{pmatrix} F_j \\ Y_j \end{pmatrix}$, $j = 1, 2, \dots, n$.

Notemos que

$$E(Z_j) = \mu_Z \quad \text{e} \quad \text{Var}(Z_j) = \begin{pmatrix} 1 & \lambda' \\ \lambda & V \end{pmatrix} = \Sigma_Z, \quad (4.37)$$

$j = 1, 2, \dots, n$, com

$$\mu_Z = \begin{pmatrix} 0 \\ \mu \end{pmatrix} \quad \text{e} \quad V = \lambda\lambda' + D(\phi).$$

Como na seção anterior, sejam

$$\left. \begin{array}{l} Z_j | q_j \sim N_{p+1}(\mu_Z; q_j^{-1} \Lambda_Z) \quad \text{e} \\ q_j \sim \chi^2(\nu)/\nu; \quad j = 1, 2, \dots, n. \end{array} \right\} \quad (4.38)$$

onde $\Lambda_Z = (\nu - 2)\Sigma_Z/\nu$.

Isto implica que $Z_j \sim T_{p+1}(\mu_Z, \Sigma_Z, \nu)$, de modo que

$$Y_j \sim T_p(\mu, V, \nu), \quad j = 1, 2, \dots, n.$$

Como considerado na Seção 4.2, seja $f(z_j, q_j)$ a densidade conjunta de (Z_j, q_j) , $j = 1, 2, \dots, n$. Logo, podemos escrever

$$f(z_j, q_j) = f_1(z_j | q_j) f_2(q_j), \quad j = 1, 2, \dots, n. \quad (4.39)$$

Daqui segue que o logaritmo da função de verossimilhança completa é dado por

$$\begin{aligned}\ell_Z(\theta, \nu) &= \sum_{j=1}^n \ln f_1(z_j|q_j) + \sum_{j=1}^n \ln f_2(q_j) \\ &= \text{cte.} - \frac{1}{2}n(p+1) \ln[(\nu-2)/\nu] - \frac{n}{2} \ln |\Sigma_Z| \\ &\quad - \frac{\nu}{2(\nu-2)} \sum_{j=1}^n q_j(z_j - \mu_Z)' \Sigma_Z^{-1} (z_j - \mu_Z) + \sum_{j=1}^n \ln f_2(q_j),\end{aligned}\tag{4.40}$$

onde $|\Sigma_Z|$ e Σ_Z^{-1} são dados em (2.35).

Etapa E: Para implementar a etapa E do algoritmo EM devemos calcular $E(q_j|Y_j, \theta)$, $E(F_j|Y_j, \theta)$ e $E(F_j^2|Y_j, \theta)$, $j = 1, 2, \dots, n$. Com efeito, usando as propriedades da distribuição t , ver Apêndice B, temos que

$$E(q_j|Y_j, \theta) = (\nu + p)/(\nu + d_j^2)\tag{4.41}$$

$$E(F_j|Y_j, \theta) = \lambda' V^{-1}(Y_j - \mu) = \frac{1}{c} \lambda' D^{-1}(\phi)(Y_j - \mu)\tag{4.42}$$

$$\begin{aligned}\text{Var}(F_j|Y_j, \theta) &= (\nu - 2 + (Y_j - \mu)' V^{-1}(Y_j - \mu))/c(\nu + p - 2) \\ &= (\nu - 2)(\nu + d_j^2)/c\nu(\nu + p - 2),\end{aligned}\tag{4.43}$$

onde

$$d_j^2 = \left(\frac{\nu}{\nu - 2} \right) (Y_j - \mu)' V^{-1}(Y_j - \mu), \quad j = 1, 2, \dots, n,\tag{4.44}$$

é a distância de Mahalanobis e c como em (2.35). De (4.42) e (4.43), segue que

$$E(F_j^2|Y_j, \theta) = \frac{(\nu - 2)(\nu + d_j^2)}{c\nu(\nu + p - 2)} + \left[\frac{1}{c} \lambda' D^{-1}(\phi)(Y_j - \mu) \right]^2,\tag{4.45}$$

$j = 1, 2, \dots, n$.

Notemos também que, como na Seção 4.2,

$$d_j^2 = \left(\frac{\nu}{\nu - 2} \right) (h_{1j} - h_{2j}), \quad j = 1, 2, \dots, n,\tag{4.46}$$

onde, agora,

$$\begin{aligned}h_{1j} &= \sum_{i=1}^p (y_{1j} - \mu_i)^2 / \phi_i \quad \text{e} \\ h_{2j} &= c^{-1} \left[\sum_{i=1}^p \lambda_i (y_{ij} - \mu_i) / \phi_i \right]^2, \quad j = 1, 2, \dots, n.\end{aligned}$$

Etapa M: Dado ν , derivando $\ell_Z(\theta, \nu)$ dada em (4.40), em relação a μ_i , λ_i e ϕ_i e igualando a zero, obtemos as seguintes equações:

$$\hat{\mu}_i = \bar{y}_{qi} - \bar{F}_q \hat{\lambda}_i, \quad (4.47)$$

$$\hat{\lambda}_i = \frac{\sum_{j=1}^n q_j F_j (y_{ij} - \bar{y}_{qi})}{\sum_{j=1}^n q_j (F_j - \bar{F}_q)^2} \quad (4.48)$$

e

$$\begin{aligned} \hat{\phi}_i &= \left(\frac{\nu}{\nu - 2} \right) \frac{1}{n} \sum_{j=1}^n q_j [(y_{ij} - \bar{y}_{qi}) - \hat{\lambda}_i (F_j - \bar{F}_q)]^2 \\ &= \left(\frac{\nu}{\nu - 2} \right) \frac{1}{n} \left\{ \sum_{j=1}^n q_j (y_{ij} - \bar{y}_{qi})^2 - \hat{\lambda}_i^2 \sum_{j=1}^n q_j (F_j - \bar{F}_q)^2 \right\}, \end{aligned} \quad (4.49)$$

onde

$$\bar{y}_{qi} = \sum_{j=1}^n q_j y_{ij} / \sum_{j=1}^n q_j, \quad i = 1, 2, \dots, p, \quad \text{e} \quad \bar{F}_q = \sum_{j=1}^n q_j F_j / \sum_{j=1}^n q_j.$$

Se ν for desconhecido, diferenciando $\ell_Z(\theta, \nu)$ em relação a ν e igualando a zero, obtemos a equação

$$h(\nu) = 0, \quad (4.50)$$

onde

$$h(\nu) = 2(p + 1 - c_2)(\nu - 2) - 2c_1\nu + \nu(\nu - 2)^2(\psi(\nu/2) - \ln(\nu/2)) - \nu(\nu - 2)^2c_3,$$

com

$$\begin{aligned} c_1 &= \frac{1}{n} \sum_{j=1}^n q_j F_j^2, \quad c_2 = \frac{1}{n} \sum_{j=1}^n q_j t_j, \\ c_3 &= \frac{1}{n} \sum_{j=1}^n \ln q_j - \frac{1}{n} \sum_{j=1}^n q_j + 1, \\ t_j &= \sum_{i=1}^p \left(\frac{1}{\omega_i} \right) (F_j \hat{\lambda}_i - (y_{ij} - \bar{y}_{qi}))^2, \quad j = 1, 2, \dots, n, \\ \omega_i &= \frac{1}{n} \sum_{j=1}^n q_j [(y_{ij} - \bar{y}_{qi}) - \hat{\lambda}_i (F_j - \bar{F}_q)]^2, \quad i = 1, 2, \dots, p, \quad \text{e} \\ \psi(\nu/2) &= \Gamma'(\nu/2)/\Gamma(\nu/2) \text{ é a função digama.} \end{aligned}$$

A equação (4.50) deve ser resolvida numericamente, dentro da etapa M, usando, por exemplo,

$$\nu^{(m+1)} = \nu^{(m)} - \frac{h(\nu^{(m)})}{h'(\nu^{(m)})}, \quad m = 0, 1, 2, \dots \quad (4.51)$$

fazendo com que a convergência do algoritmo EM seja mais lenta. Recentemente Liu e Rubin (1994) e Meng e Rubin (1993) propõem extensões do algoritmo, quando a etapa M requer um processo iterativo.

Alternativamente, como foi sugerido por Lange et al. (1989), o EMV de ν pode ser também obtido, calculando o logaritmo da função de verossimilhança observada, num reticulado de valores de ν e escolher o valor de ν que maximiza a verossimilhança observada. Dado ν , o logaritmo da função de verossimilhança observada pode ser escrito como (sem o termo cte.):

$$\ell(\theta|\nu) = \sum_{j=1}^n \ell_j, \quad (4.52)$$

onde

$$\begin{aligned} \ell_j = & \ln \left\{ \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \right\} - \frac{1}{2} \sum_{i=1}^p \ln \phi_i - \frac{1}{2} \ln \{c(\nu-2)^2\} \\ & - \frac{1}{2}(\nu+p) \ln(1 + d_j^2/\nu), \quad j = 1, 2, \dots, n; \end{aligned}$$

e d_j^2 como em (4.46), $j = 1, 2, \dots, n$.

Para obter os EMV sob a versão de regressão do modelo estrutural t -elíptico, basta utilizar as relações entre as duas parametrizações estabelecidas no Capítulo 1.

(ii) A matriz de informação esperada

Nesta seção obtemos a matriz de informação na parametrização de análise de fatores do modelo estrutural t -elíptico (4.35).

O logaritmo da função de verossimilhança observada é dado por

$$\ell(\theta, \nu) = \sum_{j=1}^n \ell_j(\theta, \nu), \quad (4.53)$$

onde

$$\ell_j(\theta, \nu) = \ln C(\nu, p) - \frac{1}{2}g_1(\lambda, \phi) - \frac{1}{2}(\nu + p) \ln r_j ,$$

$j = 1, 2, \dots, n$, com

$$g_1(\lambda, \phi) = \ln |V| = \ln c + \sum_{i=1}^p \ln \phi_i,$$

$$r_j = (\nu - 2) + W_j' V^{-1} W_j \quad \text{e}$$

$$W_j = Y_j - \mu \sim T_p(0, V, \nu), \quad j = 1, 2, \dots, n.$$

A seguir apresentamos a matriz de informação.

Teorema 4.2 *Sob o modelo estrutural t -elíptico (4.35) temos que a matriz de informação esperada é dada por:*

$$I_F(\theta, \nu) = n \begin{pmatrix} I_{\mu\mu} & 0 & 0 & 0 \\ 0 & I_{\lambda\lambda} & I_{\lambda\phi} & I_{\lambda\nu} \\ 0 & I_{\phi\lambda} & I_{\phi\phi} & I_{\phi\nu} \\ 0 & I_{\nu\lambda} & I_{\nu\phi} & i_{\nu\nu} \end{pmatrix}, \quad (4.54)$$

onde

$$I_{\mu\mu} = \left(\frac{\nu}{\nu - 2} \right) \left(\frac{\nu + p}{\nu + p + 2} \right) (\lambda\lambda' + D(\phi))^{-1},$$

$$I_{\lambda\lambda} = \left[(1 - c^{-1}) - \frac{2c^{-1}(c - 1)}{(\nu + p + 2)} \right] D^{-1}(\phi) + \left[-c^{-2} - \frac{c^{-1}(1 - 3c^{-1})(\nu + p)}{(\nu + p + 2)} \right] M,$$

$$I_{\lambda\phi} = c^{-1} [1 - 2/(\nu + p + 2)] D^{-2}(\phi) D(\lambda) + \frac{c^{-2}}{2} \left[1 - \frac{3(\nu + p)}{(\nu + p + 2)} \right] M^* \\ - [c^{-1}/(\nu + p + 2)] D^{-1}(\phi) \lambda \phi' D^{-2}(\phi),$$

$$I_{\lambda\nu} = [2c^{-1}(p + 2)/(\nu - 2)(\nu + p + 2)(\nu + p)] D^{-1}(\phi) \lambda,$$

$$I_{\phi\phi} = \left[\frac{1}{2} - (\nu + p + 2)^{-1} \right] D^{-2}(\phi) - [c^{-1} - 2c^{-1}(\nu + p + 2)^{-1}] D^{-3}(\phi) D(\lambda) D(\lambda)$$

$$\begin{aligned}
& + \frac{1}{4}[-c^{-2} + 3c^{-2}(\nu + p)(\nu + p + 2)^{-1}]\overline{M} \\
& + \frac{c^{-1}}{2}(\nu + p + 2)^{-1}[D^{-2}(\phi)(D(\lambda)\lambda\phi' + \phi\lambda'D(\lambda))D^{-2}(\phi)] \\
& - \frac{1}{2}(\nu + p + 2)^{-1}D^{-2}(\phi)\phi\phi'D^{-2}(\phi) ,
\end{aligned}$$

$$I_{\phi\nu} = \left[\frac{-(p+2)}{(\nu-2)(\nu+p+2)(\nu+p)} \right] [c^{-1}D^{-2}(\phi)D(\lambda)\lambda - D^{-2}(\phi)\phi] \quad e$$

$$\begin{aligned}
i_{\nu\nu} = & - \left[\frac{\nu-4}{2(\nu-2)^2} \right] - \frac{1}{4}\psi' \left(\frac{\nu+p}{2} \right) + \frac{1}{4}\psi'(\nu/2) \\
& + \frac{1}{2} \left[\frac{(\nu^2 + \nu p - 2\nu - 6p - 8)\nu}{(\nu-2)^2(\nu+p+2)(\nu+p)} \right] ,
\end{aligned}$$

com

$$M = D^{-1}(\phi)\lambda\lambda'D^{-1}(\phi) , \quad M^* = MD^{-1}(\phi)D(\lambda) ,$$

$$\overline{M} = D^{-1}(\phi)D(\lambda)MD(\lambda)D^{-1}(\phi) , \quad \psi(x) = \Gamma'(x)/\Gamma(x) ,$$

$$I_{\phi\lambda} = I'_{\lambda\phi} , \quad I_{\nu\phi} = I'_{\phi\nu} , \quad I_{\nu\lambda} = I'_{\lambda\nu} , \quad c = 1 + \lambda'D^{-1}(\phi)\lambda \quad e$$

$$D^{-i}(\phi) = \text{Diag}(\phi_1^{-i}, \phi_2^{-i}, \dots, \phi_p^{-i}), \quad i = 1, 2, 3.$$

Prova: Ver Apêndice D.

Se ν for fixado a priori, basta eliminar a última linha e a última coluna de $I_F(\theta, \nu)$ para obter a matriz de informação correspondente a θ ; $I_F(\theta)$, digamos. Também, se $\nu \rightarrow \infty$ obtemos a matriz de informação correspondente ao modelo estrutural normal dada no Lema 2.4. Ademais, usando os resultados dados no Capítulo 1, podemos obter a matriz de informação na versão de regressão do modelo estrutural t -elíptico.

Como no Capítulo 1, seja $\rho = (\rho_1, \dots, \rho_p)'$, onde $\rho_i = \lambda_i^2/(\lambda_i^2 + \phi_i)$, $i = 1, 2, \dots, p$, os coeficientes de confiabilidade, e $\theta^* = (\mu', \lambda', \rho')'$. De forma similar ao Corolário 2.2, podemos provar o seguinte resultado:

Corolário 4.1 *Sob o modelo (4.35), temos que*

$$\sqrt{n}(\hat{\theta}^* - \theta^*) \xrightarrow{\mathcal{D}} N_{3p}(0, \Omega_{\theta^*}) ,$$

onde

$$\Omega_{\theta^*} = nGI_F^{-1}(\theta, \nu)G' , \quad (4.55)$$

com G uma matriz $3p \times 3p + 1$, dada por

$$G = \begin{pmatrix} I_p & 0 & 0 \\ 0 & I_p & 0 \\ 0 & G_1 & G_2 \end{pmatrix} , \quad (4.56)$$

com

$$G_1 = 2\text{Diag}(\lambda_1^{-1}\rho_1(1-\rho_1), \dots, \lambda_p^{-1}\rho_p(1-\rho_p)), \quad p \times p, \quad e$$

$$G_2 = \{\text{Diag}(-\lambda_1^{-2}\rho_1^2, \dots, -\lambda_p^{-2}\rho_p^2), 0\} \text{ uma matriz } p \times p + 1.$$

Utilizando estes resultados, podemos usar a estatística de Wald para testar hipóteses de interesse, como no caso normal. Ver Seção 2.3.4. Também podemos construir regiões de confiança assintótica para ρ , como na Seção 2.3.3.

4.3.2 Modelo estrutural elíptico

Apresentamos a seguir o modelo estrutural elíptico e a matriz de informação esperada para este modelo.

A versão de análise de fatores do modelo estrutural elíptico pode ser definida como, ver Seção 4.2,

$$\left. \begin{aligned} Y_j &= \mu + Lv_j, \quad j = 1, 2, \dots, n; \\ v_j, \quad 1 \leq j \leq n, &\text{ são i.i.d. } E\ell_{p+1}(0, \Lambda_v) \end{aligned} \right\} \quad (4.57)$$

com matriz de covariância $\Sigma_v = c_g \Lambda_v$ dada em (4.33).

De (4.57) segue que, ver Apêndice A,

$$Y_j, \quad 1 \leq j \leq n, \text{ são i.i.d. } E\ell_p(\mu, \Lambda_Y) . \quad (4.58)$$

com matriz de covariância $V = c_g \Lambda_Y = \lambda \lambda' + D(\phi)$.

Seja $\theta = (\mu', \lambda', \phi')$. A seguir obtemos a matriz de informação esperada e a expressamos de forma simples em notação matricial.

Teorema 4.3 *Sob o modelo estrutural elíptico (4.57), a matriz de informação, $I_E(\theta)$, é dada por:*

$$I_E(\theta) = n \begin{pmatrix} I_{\mu\mu} & 0 & 0 \\ 0 & I_{\lambda\lambda} & I_{\lambda\phi} \\ 0 & I_{\phi\lambda} & I_{\phi\phi} \end{pmatrix} = n \Omega_E^{-1}, \quad (4.59)$$

onde

$$\begin{aligned} I_{\mu\mu} &= \left(\frac{4a_1 c_g}{p} \right) V^{-1}, \\ I_{\lambda\lambda} &= 4c_1 c^{-2} M + c_2 \{ c^{-1} (2c^{-1} - 1) M + 2(1 - c^{-1}) D^{-1}(\phi) \}, \\ I_{\lambda\phi} &= c_1 \{ 2c^{-1} D^{-1}(\phi) \lambda 1' D^{-1}(\phi) - 2c^{-2} M^* \} \\ &\quad + c_2 \{ 2c^{-1} D^{-2}(\phi) D(\lambda) - 2c^{-2} M^* \}, \\ I_{\phi\phi} &= c_1 \{ D^{-1}(\phi) 1 1' D^{-1}(\phi) - c^{-1} D^{-1}(\phi) 1 \lambda' D(\lambda) D^{-2}(\phi) \\ &\quad - c^{-1} D^{-2}(\phi) D(\lambda) \lambda 1' D^{-1}(\phi) + c^{-2} \overline{M} \} \\ &\quad + c_2 \{ D^{-2}(\phi) - 2c^{-1} D^{-3}(\phi) D(\lambda) D(\lambda) + c^{-2} \overline{M} \} \quad e \\ I_{\phi\lambda} &= I'_{\lambda\phi}, \quad c_1 = \frac{1}{4} + \frac{a_2}{p} + \frac{a_3}{p(p+2)}, \quad c_2 = \frac{2a_3}{p(p+2)}; \end{aligned}$$

a_1, a_2 e a_3 como no Lema 4.1; c M , M^* e $\overline{M}$ como no Teorema 4.2 e 1 é um vetor de uns de dimensão $p \times 1$.

Prova: Para provar este resultado usaremos a fórmula (4.10).

Primeiro notamos que:

$$\begin{aligned} \Lambda_Y^{-1} &= c_g V^{-1} = c_g \{ D^{-1}(\phi) - c^{-1} M \}, \\ \frac{\partial}{\partial \lambda_i} (\Lambda_Y) &= c_g^{-1} \{ e_i \lambda' + \lambda e_i' \}, \quad i = 1, 2, \dots, p; \\ \frac{\partial}{\partial \phi_i} (\Lambda_Y) &= c_g^{-1} \{ e_i e_i' \}, \quad i = 1, 2, \dots, p; \\ D^{-1}(\phi) e_i &= \left(\frac{1}{\phi_i} \right) e_i \quad e \quad \lambda' D^{-1}(\phi) e_i = \lambda_i / \phi_i, \quad i = 1, \dots, p, \end{aligned} \quad (4.60)$$

onde e_i é um vetor com um na i -ésima componente e zeros nas demais, $i = 1, 2, \dots, p$.

De (4.60) segue que

$$\begin{aligned} \text{tr} \left\{ \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \lambda_i} \right\} &= 2\lambda_i / c\phi_i \\ \text{e} \\ \text{tr} \left\{ \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \phi} \right\} &= (1/\phi_i)(1 - \lambda_i^2 / c\phi_i), \quad i = 1, 2, \dots, p. \end{aligned} \quad (4.61)$$

Também, utilizando (4.60), obtemos, depois de alguma álgebra para simplificar, o seguinte resultado:

$$\text{tr} \left\{ \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \lambda_i} \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \phi_j} \right\} = \begin{cases} \frac{2c^{-1}\lambda_i}{\phi_i^2} - \frac{2c^{-2}\lambda_i^2}{\phi_i^3} & \text{se } i = j, \\ -\frac{2c^{-2}\lambda_i\lambda_j^3}{\phi_i\phi_j^2} & \text{se } i \neq j. \end{cases} \quad (4.62)$$

$i, j = 1, 2, \dots, p$;

$$\text{tr} \left\{ \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \phi_i} \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \phi_j} \right\} = \begin{cases} \frac{1}{\phi_i^2}(1 - 2c^{-1}\lambda_i^2/\phi_i) + \frac{c^{-2}\lambda_i^4}{\phi_i^4} & \text{se } i = j, \\ \frac{c^{-2}\lambda_i^2\lambda_j^2}{\phi_i^2\phi_j^2} & \text{se } i \neq j. \end{cases} \quad (4.63)$$

$i, j = 1, 2, \dots, p$; e

$$\text{tr} \left\{ \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \lambda_i} \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \lambda_j} \right\} = \begin{cases} \frac{2c^{-1}(c-1)}{\phi_i} + \frac{2c^{-1}(2c^{-1}-1)\lambda_i^2}{\phi_i^2} & \text{se } i = j, \\ \frac{2c^{-1}(2c^{-1}-1)\lambda_i\lambda_j}{\phi_i\phi_j} & \text{se } i \neq j. \end{cases} \quad (4.64)$$

$i, j = 1, 2, \dots, p$;

Logo, usando (4.10) e escrevendo (4.61) - (4.64) em forma matricial, obtemos a matriz de informação dada em (4.59).

Este fato corresponde a uma generalização de alguns resultados anteriores. Com efeito, no caso normal é fácil ver que

$$a_1 = p/4, \quad a_2 = -p/2 \quad \text{e} \quad a_3 = p(p+2)/4, \quad (4.65)$$

e a matriz $I_E(\theta)$ reduz-se à matriz de informação dada no Lema 2.4 para o modelo estrutural normal. Analogamente, para a distribuição t , segue do Lema D2 do Apêndice D, que

$$a_1 = \frac{p(\nu+p)}{4(\nu+p+2)}, \quad a_2 = -\frac{p}{2} \quad \text{e} \quad a_3 = \frac{(\nu+p)p(p+2)}{4(\nu+p+2)}; \quad (4.66)$$

neste caso, a matriz $I_E(\theta)$ reduz-se à matriz de informação dada no Teorema 4.2 para o modelo estrutural t -elíptico, supondo ν conhecido.

Finalmente, do Teorema 4.3, temos o seguinte

Corolário 4.2 *Sob o modelo estrutural elíptico (4.57), temos que:*

$$\sqrt{n}(\hat{\theta}_E - \theta) \xrightarrow{\mathcal{D}} N_{3p}(0, \Omega_E) ,$$

onde $\hat{\theta}_E$ é o EMV de θ no modelo (4.57) e Ω_E como definida no teorema.

4.3.3 Ilustração numérica

Consideremos, novamente, os dados da Tabela 2.1, analisados no Capítulo 2 supondo um modelo estrutural normal. Agora, para ilustrar alguns resultados deste capítulo, vamos supor um modelo estrutural t -elíptico. Aqui temos $p = 4$ e $n = 72$.

Para estimar ν usaremos o método proposto por Lange et al. (1989). A Tabela 4.1 apresenta o máximo do logaritmo da função de verossimilhança observada (4.52) para vários valores de ν , mostrando que $\hat{\nu} = 17$ é um valor razoável a considerar.

Tabela 4.1: Logaritmo da função de verossimilhança observada maximizada

$\nu :$	4	10	15	16	17	18	20	100
$-\ell(\hat{\theta} \nu) :$	1903.75	1899.10	1898.75	1898.74	1898.72	1898.73	1898.74	1899.33

Para mostrar a habilidade do algoritmo em detectar observações extremas, as quatro medições correspondentes ao paciente número um foram substituídas pelo valor atípico 5000. A Tabela 4.2 mostra agora que uma estimativa dos graus de liberdade seria $\hat{\nu} = 14$ ou 15.

Tabela 4.2: Logaritmo da função de verossimilhança observada maximizada

$\nu :$	4	10	13	14	15	16	17	30
$-\ell(\hat{\theta} \nu) :$	1905.94	1901.61	1901.44	1901.43	1901.43	1901.44	1901.46	1901.78

Na Tabela 4.3 apresentamos várias iterações do algoritmo EM, mostrando que a convergência é atingida aproximadamente na 1000-ésima iteração, dando estimativas para λ_i e σ_i^2 , $i = 1, 2, 3, 4$, tomando $\hat{\nu} = 17$.

Tabela 4.3: Convergência do algoritmo EM

iteração	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	$\hat{\sigma}_4^2$
10	654.6	687.2	774.4	738.8	49262	19065	31039	39929
50	684.9	718.9	809.8	772.3	49380	19147	31405	40315
100	699.4	734.0	826.7	788.4	49483	19259	31503	40425
500	707.7	742.8	836.4	797.5	49531	19346	31554	40482
1000	707.7	742.8	836.4	797.6	49531	19346	31555	40482
1500	707.7	742.8	836.4	797.6	49531	19346	31555	40482

Também os EMV para μ e ρ são:

$$\hat{\mu} = (2208, 2140, 2110, 2073)'$$

e

$$\hat{\rho} = (0.9100, 0.9661, 0.9568, 0.9402)' .$$

Usando as estimativas de (λ, ϕ) apresentadas na última iteração da Tabela 4.3, temos desde (4.54) que a matriz de covariância assintótica estimada de $(\hat{\theta}, \hat{\nu})$ é dada por:

$$I_F^{-1}(\hat{\theta}, \hat{\nu}) =$$

[illegible]

Note que os elementos desta matriz de covariância assintótica correspondentes a $\hat{\mu}$, $\hat{\lambda}$ e $\hat{\phi}$ são próximos aos obtidos no caso normal, ver Seção 2.5. Isto deve-se a que o valor dos graus de liberdade é alto, (17).

do modelo estrutural t -elíptico, é dado por:

$$(\hat{\gamma}, \hat{\nu}) = (2208, -178, -500, -415, 707.7, 1.0496, 1.1819, \\ 1.1270, 49531, 19346, 31555, 40482, 17)'$$

e a matriz de covariância assintótica estimada de $(\hat{\gamma}, \hat{\nu})$ é dada por

$$I_R^{-1}(\hat{\gamma}, \hat{\nu}) =$$

[illegible]

Adicionalmente, do Corolário 4.1, segue que a matriz de covariância assintótica estimada de $\hat{\theta}^* = (\hat{\mu}', \hat{\lambda}', \hat{\rho}')'$ é dada por

$$\frac{1}{n}\Omega_{\hat{\theta}^*} =$$

[illegible]

Usando estes resultados e a estatística de Wald podemos testar hipóteses de interesse como as indicadas no Capítulo 1. A hipótese H_{01} , por exemplo, pode ser escrita como $H_{01} : C\theta^* = q_0$, onde, ver Seção 2.3.4,

$$C = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & C_1 & 0 \\ 0 & 0 & C_1 \end{pmatrix},$$

uma matriz 9×12 de posto 9,

$$C_1 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

e $q_0 = 0$ o vetor nulo de dimensão 9×1 . Assim, a estatística de Wald pode ser escrita como

$$W_{01} = n(C\hat{\theta}^* - q_0)'(C\Omega_{\hat{\theta}^*}C')^{-1}(C\hat{\theta}^* - q_0), \quad (4.67)$$

com 9 graus de liberdade. Neste caso, $W_{01} = 39.43$, i.e., rejeitamos a hipótese de que os instrumentos medem sem vício e são igualmente confiáveis. Analogamente, correspondendo à hipótese H_{02} , temos que $W_{02} = 29.82$, com seis graus de liberdade; correspondendo à hipótese H_{03} , temos que $W_{03} = 6.68$, com três graus de liberdade e correspondendo a H_{04} , temos que $W_{04} = 15.39$, com três graus de liberdade, as quais são todas rejeitadas ao nível de 5%. Ou seja, a principal conclusão é de que os instrumentos medem com vício a capacidade vital do paciente e não são igualmente confiáveis. Note que estes mesmos resultados foram obtidos usando o modelo estrutural normal, ver Seção 2.5.

Por outro lado, sabemos que (ver Apêndice B) d_j^2/p tem uma distribuição F com p e ν graus de liberdade, com d_j^2 dada por (4.44), $j = 1, 2, \dots, n$. Ademais, $F_j = \hat{d}_j^2/p$ tem assintoticamente a mesma distribuição F , onde $\hat{d}_j^2 = d_j^2(\hat{\theta}, \hat{\nu})$, $j = 1, 2, \dots, n$. Como sugerido por Lange et al. (1989), um método prático para avaliar o ajuste do modelo t é o gráfico da distância transformada. Utilizando a aproximação de Wilson-Hilferty (Johnson e Kotz, 1970), temos que

$$z_j = \frac{\left(1 - \frac{2}{9\nu}\right) F_j^{1/3} - \left(1 - \frac{2}{9p}\right)}{\left\{\left(\frac{2}{9\nu}\right) F_j^{2/3} + \left(\frac{2}{9p}\right)\right\}^{1/2}}, \quad (4.68)$$

$j = 1, 2, \dots, n$, tem aproximadamente uma distribuição normal padrão. Assim, um $Q-Q$ plot (estatísticos de ordem esperados versus observados) das distâncias transformadas z_j , $j = 1, 2, \dots, n$, pode ser usado para avaliar o ajuste da distribuição t multivariada.

A Figura 4.1 mostra o gráfico da distância transformada para a distribuição t para os dados em Barnett (1969). Como podemos observar o modelo estrutural t -elíptico apresenta um melhor ajuste do que o modelo estrutural normal, ver Seção 2.5.

Figura 4.1: Gráfico da distância transformada para o modelo t -elíptico

Capítulo 5

Modelo elíptico de Grubbs

5.1 Introdução

Um modelo que tem tido um desenvolvimento paralelo ao modelo estrutural proposto por Barnett (1969) e usado freqüentemente na área industrial para comparar instrumentos de medição é o modelo de Grubbs. Neste modelo, introduzido por Grubbs (1948) (ver também Grubbs, 1973, 1983) assume-se que os instrumentos medem a característica de interesse x na mesma escala, ou seja, $\beta_2 = \dots = \beta_p = 1$ na versão de regressão do modelo estrutural ou $\lambda_1 = \lambda_2 = \dots = \lambda_p = \lambda$ na versão de análise de fatores. Neste capítulo usaremos, como nos capítulos anteriores, a parametrização de análise de fatores do modelo.

Consideremos o modelo estrutural (1.11) com $\lambda_1 = \lambda_2 = \dots = \lambda_p = \lambda$, digamos

$$\begin{aligned} Y_j &= \mu + \mathbf{1}_p \lambda F_j + \varepsilon_j = \mu + \mathbf{1}_p z_j + \varepsilon_j \\ &= \mu + L v_j; \quad j = 1, 2, \dots, n; \end{aligned} \quad (5.1)$$

onde

$$L = (\mathbf{1}_p, I_p), \quad z_j = \lambda F_j \quad \text{e} \quad v_j = \begin{pmatrix} z_j \\ \varepsilon_j \end{pmatrix}, \quad j = 1, 2, \dots, n,$$

tais que

$$E(v_j) = 0 \quad \text{e} \quad \text{Var}(v_j) = \Sigma_v = \begin{pmatrix} \phi_x & 0 \\ 0 & D(\phi) \end{pmatrix} \quad (5.2)$$

$j = 1, 2, \dots, n$ e $\phi_x = \lambda^2$ que corresponde à variância de x .

De (5.1) - (5.2) segue que

$$E(Y_j) = \mu \quad \text{e} \quad \text{Var}(Y_j) = V = \phi_x \mathbf{1}_p \mathbf{1}_p + D(\phi), \quad (5.3)$$

$j = 1, 2, \dots, n$.

O modelo de Grubbs é dado por (5.1) - (5.2) mais a suposição de normalidade dos v_j , $j = 1, 2, \dots, n$. Grubbs (1948, 1973) trabalha com as diferenças pareadas $d_{ijk} = y_{ik} - y_{jk}$ e propõe estimadores tipo momentos. O uso dos estimadores propostos por Grubbs tem sido contestado, devido às dificuldades encontradas no desenvolvimento de procedimentos de inferência estatística. Alguns resultados, no entanto, são dados em Grubbs (1973), Brindley e Bradley (1985) e Jaech (1985). Neste capítulo utilizaremos o método de máxima verossimilhança.

No contexto do modelo de Grubbs a qualidade da medição é avaliada em termos do vício (aditivo) ou acurácia e da precisão (inverso da variância) dos diferentes instrumentos. Assim, uma hipótese inicial para avaliar a acurácia das medições feitas pelos diferentes instrumentos é $H_{g1} : \mu_1 = \mu_2 = \dots = \mu_p$. Para comparar as precisões dos instrumentos a hipótese é $H_{g2} : \pi_1 = \pi_2 = \dots = \pi_p$. Note que neste caso esta hipótese é equivalente a $H_{g2} : \phi_1 = \dots = \phi_p$. Lembremos também que H_{g2} é equivalente (ver Capítulo 1) a $\rho_1 = \rho_2 = \dots = \rho_p$; igualdade dos coeficientes de confiabilidade dos instrumentos. Dependendo da aplicação estas hipóteses podem ser testadas juntas ou separadamente. Ademais, dada a rejeição de uma delas, é freqüente considerar subhipóteses para testar igualdade de subconjuntos de variâncias e vícios.

Para $p = 2$, Maloney e Rastogi (1970) mostram que o teste de Pitman (Pitman, 1939) é equivalente a testar H_{g2} . Blackwood e Bradley (1991) usando regressão simples propõem um teste para provar H_{g1} e H_{g2} simultaneamente. Recentemente, Christensen e Blackwood (1993) propõem um modelo linear multivariado para testar H_{g1} e H_{g2} (ou subhipóteses) para qualquer número de instrumentos ($p \geq 2$).

Em todos estes trabalhos assume-se que os vetores aleatórios v_j , $j = 1, 2, \dots, n$, são i.i.d. $N_{p+1}(0, \Sigma_v)$. Neste capítulo vamos supor que v_j , $j = 1, 2, \dots, n$, são i.i.d. $El_{p+1}(0, \Lambda_v)$ com matriz de covariância $\Sigma_v = c_g \Lambda_v$ dada em (5.2). Mais especificamente, o modelo elíptico de Grubbs é dado por

$$\left. \begin{array}{l} Y_j = \mu + (\mathbf{1}_p, I_p)v_j, \quad j = 1, 2, \dots, n; \\ v_j, \quad 1 \leq j \leq n, \text{ são i.i.d. } El_{p+1}(0, \Lambda_v), \\ \text{com matriz de covariância } \Sigma_v \text{ dada por (5.2).} \end{array} \right\} \quad (5.4)$$

De (5.4) e das propriedades das distribuições elípticas, ver Apêndice A, segue que

$$Y_j, \quad 1 \leq j \leq n, \text{ são i.i.d. } El_p(\mu, \Lambda_Y), \quad (5.5)$$

com matriz de covariância $V = \phi_x \mathbf{1}_p \mathbf{1}_p' + D(\phi) = c_g \Lambda_v$.

A organização do capítulo é como segue. Na Seção 5.2 discutimos a implementação do algoritmo EM para os modelos normal e t . Na Seção 5.3 apresentamos a matriz de informação e a matriz de covariância assintótica dos EMV no modelo elíptico geral (5.4). Especificamos estes resultados nos modelos normal e t . Finalmente, na Seção 5.4 discutimos a implementação de testes de hipóteses.

5.2 Estimação de máxima verossimilhança

Nesta seção mostramos como obter os EMV usando o algoritmo EM, nos modelos normal e t .

5.2.1 Estimação no modelo normal

A seguir vamos supor que

$$v_j, \quad 1 \leq j \leq n, \text{ são i.i.d. } N_{p+1}(0, \Sigma_v); \quad (5.6)$$

de onde segue que

$$Y_j, \quad 1 \leq j \leq n, \text{ são i.i.d. } N_p(\mu, V). \quad (5.7)$$

Seja $Z_j = \begin{pmatrix} z_j \\ y_j \end{pmatrix}$, $j = 1, 2, \dots, n$.

Logo, temos que

$$Z_j, \ 1 \leq j \leq n, \text{ são i.i.d. } N_p(\mu_Z, \Sigma_Z), \quad (5.8)$$

onde

$$\mu_Z = \begin{pmatrix} 0 \\ \mu \end{pmatrix}, \quad \Sigma_Z = \begin{pmatrix} \phi_x & \phi_x \mathbf{1}'_p \\ \phi_x \mathbf{1}_p & V \end{pmatrix}.$$

Assim, o logaritmo da função de verossimilhança completa é dado por

$$\ell_Z(\theta) = \text{cte.} - \frac{n}{2} \ln |\Sigma_Z| - \frac{1}{2} \sum_{j=1}^n (Z_j - \mu_Z)' \Sigma_Z^{-1} (Z_j - \mu_Z), \quad (5.9)$$

onde

$$|\Sigma_Z| = \prod_{i=0}^p \phi_i \quad \text{e} \quad \Sigma_Z^{-1} = \begin{pmatrix} c/\phi_0 & -\mathbf{1}'_p D^{-1}(\phi) \\ -D^{-1}(\phi) \mathbf{1}_p & D^{-1}(\phi) \end{pmatrix}, \quad (5.10)$$

com $\phi_0 = \phi_x$, $c = 1 + \phi_x \sum_{i=1}^p (1/\phi_i)$ e $\theta = (\mu', \phi_0, \phi')'$.

De maneira similar à discutida na subseção 2.3.1, temos

Etapa E: Esta etapa é implementada calculando

$$\begin{aligned} E(z_j | Y_j, \theta) &= c^{-1} \phi_0 \mathbf{1}'_p D^{-1}(\phi) (Y_j - \mu) \quad \text{e} \\ E(z_j^2 | Y_j, \theta) &= \text{Var}(z_j | Y_j, \theta) + \{E(z_j | Y_j, \theta)\}^2 \\ &= c^{-1} \phi_0 \{1 + \mathbf{1}'_p D^{-1}(\phi) (Y_j - \mu)\} \end{aligned} \quad (5.11)$$

$j = 1, 2, \dots, n$.

Etapa M: Nesta etapa calculamos

$$\begin{aligned} \hat{\mu}_i &= \bar{y}_i - \bar{z}, \quad i = 1, 2, \dots, p, \\ \hat{\phi}_0 &= \frac{1}{n} \sum_{j=1}^n z_j \\ \hat{\phi}_i &= \frac{1}{n} \sum_{j=1}^n \{(z_j - \bar{z}) - (y_{ij} - \bar{y}_i)\}^2, \quad i = 1, 2, \dots, p, \end{aligned} \quad (5.12)$$

com

$$\bar{z} = \frac{1}{n} \sum_{j=1}^n z_j \quad \text{e} \quad \bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij}, \quad i = 1, 2, \dots, p.$$

Note que a implementação deste algoritmo é extremamente simples e que as estimativas das variâncias (ϕ_0, ϕ) são sempre não negativas, o que não acontece necessariamente com os estimadores propostos por Grubbs (1973). Como estimativas iniciais podemos utilizar estimadores de momentos, ver subseção 2.3.1.

5.2.2 Estimação no modelo t -elíptico

Nesta subseção vamos supor que

$$v_j, \quad 1 \leq j \leq n, \quad \text{são i.i.d. } T_{p+1}(0, \Sigma_v, \nu), \quad (5.13)$$

de onde segue que, ver Apêndice B,

$$Y_j, \quad 1 \leq j \leq n, \quad \text{são i.i.d. } T_p(\mu, V, \nu), \quad (5.14)$$

Sejam $Z_j = \begin{pmatrix} z_j \\ y_j \end{pmatrix}$, $j = 1, 2, \dots, n$.

Como considerado no capítulo anterior, sejam

$$\left. \begin{aligned} Z_j | q_j &\sim N_{p+1}(\mu_Z, q_j^{-1} \Lambda_Z) \\ q_j &\sim \chi^2(\nu)/\nu, \quad j = 1, 2, \dots, n, \end{aligned} \right\} \quad (5.15)$$

onde $\Lambda_Z = \{(\nu - 2)/\nu\} \Sigma_Z$, com μ_Z e Σ_Z como em (5.8).

De (5.15) segue que $Z_j \sim T_{p+1}(\mu_Z, \Sigma_Z, \nu)$, de modo que $Y_j \sim T_p(\mu, V, \nu)$, $j = 1, 2, \dots, n$.

Seja $f(Z_j, q_j)$ a densidade conjunta de (Z_j, q_j) , $j = 1, 2, \dots, n$, que pode ser escrita como

$$f(Z_j, q_j) = f_1(Z_j | q_j) f_2(q_j), \quad j = 1, 2, \dots, n.$$

Logo, o logaritmo da função de verossimilhança completa, $\ell_Z(\theta, \nu)$, é dado por

$$\begin{aligned} \ell_Z(\theta, \nu) = & \text{cte.} - \frac{1}{2} n(p+1) \ln\{(\nu-2)/\nu\} - \frac{1}{2} n \ln |\Sigma_Z| \\ & - \frac{\nu}{2(\nu-2)} \sum_{j=1}^n q_j (Z_j - \mu_Z)' \Sigma_Z^{-1} (Z_j - \mu_Z) + \sum_{j=1}^n \ln f_2(q_j). \end{aligned} \quad (5.16)$$

Agora, usando (5.10) e simplificando, temos que (5.16) pode ser escrita como

$$\begin{aligned}\ell_Z(\theta, \nu) = & \text{cte.} - \frac{n(p+1)}{2} \ln\{(\nu-2)/\nu\} - \frac{n}{2} \sum_{i=0}^p \ln \phi_i \\ & - \frac{\nu}{2(\nu-2)} \sum_{j=1}^n q_j \left[z_j^2 \frac{c}{\phi_0} - 2z_j \sum_{i=1}^p [(y_{ij} - \mu_i)/\phi_i] \right. \\ & \left. + \sum_{i=1}^p \{(y_{ij} - \mu_i)^2/\phi_i\} \right] + \sum_{j=1}^n \ln f_2(q_j)\end{aligned}\quad (5.17)$$

De forma similar à discutida na subseção 4.3.1, temos

Etapa E: Nesta etapa do algoritmo calculamos:

$$\begin{aligned}E(q_j|Y_j, \theta, \nu) &= (\nu + p)/(\nu + d_j^2) \\ E(z_j|Y_j, \theta, \nu) &= \phi_x \mathbf{1}_p' V^{-1} (Y_j - \mu) \\ E(z_j^2|Y_j, \theta, \nu) &= \frac{\phi_x}{c} \left(\frac{\nu-2}{2} \right) \left(\frac{\nu + d_j^2}{\nu + p - 2} \right) + (\phi_x \mathbf{1}_p' V^{-1} (Y_j - \mu))^2\end{aligned}\quad (5.18)$$

onde

$$d_j^2 = \left(\frac{\nu}{\nu-2} \right) (Y_j - \mu)' V^{-1} (Y_j - \mu), \quad j = 1, 2, \dots, n. \quad (5.19)$$

Etapa M: As equações, supondo ν fixado a priori, para implementar esta etapa, são:

$$\begin{aligned}\hat{\mu}_i &= \bar{y}_{iq} - \bar{z}_q, \quad i = 1, 2, \dots, p, \\ \hat{\phi}_0 &= (\nu/(\nu-2)) \frac{1}{n} \sum_{j=1}^n q_j z_j^2 \\ \hat{\phi}_i &= (\nu/(\nu-2)) \frac{1}{n} \sum_{j=1}^n q_j [(z_j - \bar{z}_q) - (y_{ij} - \bar{y}_{iq})]^2,\end{aligned}\quad (5.20)$$

$i = 1, 2, \dots, p$, com

$$\bar{z}_q = \sum_{j=1}^n q_j z_j / \sum_{j=1}^n q_j \quad \text{e} \quad \bar{y}_{iq} = \sum_{j=1}^n q_j y_{ij} / \sum_{j=1}^n q_j, \quad i = 1, 2, \dots, p.$$

Se ν não for fixado a priori, diferenciando $\ell_Z(\theta, \nu)$ com respeito a ν e igualando a zero, obtemos a equação

$$2(p+1-a) - (b+1)\nu(\nu-2) + \nu(\nu-2)(\psi(\nu/2) - \ln(\nu/2)) = 0, \quad (5.21)$$

onde

$$a = \frac{1}{n} \sum_{j=1}^n q_j b_j \quad \text{e} \quad b = \frac{1}{n} \sum_{j=1}^n (\ln q_j - q_j),$$

com

$$\begin{aligned} b_j &= z_j^2 \left(\frac{1}{a_0} + \sum_{i=1}^p (1/a_i) \right) - 2z_j \sum_{i=1}^p \left(\frac{1}{a_i} \right) (y_{ij} - \mu_i) \\ &\quad + \sum_{i=1}^p \left(\frac{1}{a_i} \right) (y_{ij} - \mu_i)^2, \quad j = 1, 2, \dots, n, \\ a_0 &= \frac{1}{n} \sum_{j=1}^n q_j z_j^2, \\ a_i &= \frac{1}{n} \sum_{j=1}^n q_j [(z_j - \bar{z}_q) - (y_{ij} - \bar{y}_{iq})]^2, \quad i = 1, 2, \dots, p, \end{aligned}$$

e $\psi(\nu/2)$ é a função digama.

Como no Capítulo anterior, esta equação, (5.21), deve ser resolvida numericamente dentro da etapa M, fazendo com que a convergência do algoritmo seja mais lenta. No entanto, neste caso, também podemos aplicar um método sugerido por Lange et al. (1989) para obter uma estimativa de ν , ver subseção 4.3.1.

5.3 A matriz de informação esperada

Nesta seção obtemos a matriz de informação no modelo elíptico de Grubbs, (5.4). Apresentamos também a matriz de covariância dos EMV.

Seja $\theta = (\mu', \phi_0, \phi')'$.

Teorema 5.1 *Sob o modelo elíptico de Grubbs (5.4) a matriz de informação esperada por elemento, $I_G(\theta)$, é dada por*

$$I_G(\theta) = \begin{pmatrix} I_{\mu\mu} & 0 \\ 0 & I_{\phi\phi} \end{pmatrix}, \quad (5.22)$$

onde

$$\begin{aligned} I_{\mu\mu} &= (4a_1 c_g / p) V^{-1} \quad \text{e} \\ I_{\phi\phi} &= c_1 h h' + c_2 \{b b' + D(d)\}. \end{aligned}$$

com

$$\begin{aligned} h &= [(h_i)], \quad h_i = \frac{1}{\phi_i} \left(1 - \frac{\phi_0}{c\phi_i} \right), \quad i = 0, 1, \dots, p, \\ b &= [(b_i)], \quad b_i = \frac{\phi_0}{c\phi_i^2}, \quad i = 0, 1, \dots, p, \quad e \\ d &= [(d_i)], \quad d_i = \frac{1}{\phi_i^2} \left(1 - \frac{2\phi_0}{c\phi_i} \right), \quad i = 0, 1, \dots, p, \end{aligned}$$

todos vetores $(p+1) \times 1$, $\phi = (\phi_0, \phi')'$ e a_1, c_1, c_2 como no Teorema 4.3.

Prova: Primeiro notemos que

$$V = c_g \Lambda_Y = \phi_0 J + D(\phi) = \sum_{i=0}^p \phi_i G_i, \quad (5.23)$$

onde $J = \mathbf{1}_p \mathbf{1}_p' = G_0$ e $G_i = e_i e_i'$, $i = 1, 2, \dots, p$, com e_i um vetor $p \times 1$ com uns na i -ésima componente e o resto zero, $i = 1, 2, \dots, p$,

$$\Lambda_Y^{-1} = c_g [D^{-1}(\phi) - c^{-1} \phi_0 D^{-1}(\phi) J D^{-1}(\phi)] \quad e \quad (5.24)$$

$$\frac{\partial(\Lambda_Y)}{\partial \phi_i} = c_g^{-1} G_i, \quad i = 0, 1, \dots, p. \quad (5.25)$$

De (5.24) e (5.25) segue

$$\text{tr} \left\{ \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \phi_i} \right\} = \frac{1}{\phi_i} \left(1 - \frac{\phi_0}{c\phi_i} \right), \quad i = 0, 1, \dots, p. \quad (5.26)$$

$$\begin{aligned} \text{tr} \left\{ \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \phi_i} \Lambda_Y^{-1} \frac{\partial(\Lambda_Y)}{\partial \phi_j} \right\} &= \begin{cases} \left(\frac{1}{\phi_i} - \frac{\phi_0}{c\phi_i^2} \right)^2 & \text{se } i = j \\ \frac{\phi_0^2}{c^2 \phi_i^2 \phi_j^2} & \text{se } i \neq j \end{cases} \\ i, j &= 0, 1, 2, \dots, p. \end{aligned} \quad (5.27)$$

Agora, usando a fórmula (4.10), o resultado segue, depois de alguma álgebra para simplificar.

Corolário 5.1 *Seja $\hat{\theta}_G$ o EMV de θ no modelo elíptico de Grubbs (5.4). Então temos que*

$$\sqrt{n}(\hat{\theta}_G - \theta) \xrightarrow{\mathcal{D}} N_{2p+1}(0, \Omega_G),$$

onde

$$\Omega_G = \begin{pmatrix} \Omega_\mu & 0 \\ 0 & \Omega_\phi \end{pmatrix}; \quad (5.28)$$

com

$$\begin{aligned} \Omega_\mu &= (p/4a_1c_g)V; \\ \Omega_\phi &= \left(\frac{1}{c_2}\right)Q - \left(\frac{c_1}{c_2^2}\right)\frac{Qh h' Q}{1 + \left(\frac{c_1}{c_2}\right)h' Q h} \quad e \\ Q &= (bb' + D(d))^{-1} = D^{-1}(d) - \frac{D^{-1}(d)bb'D^{-1}(d)}{1 + b'D^{-1}(d)b}. \end{aligned}$$

Desta forma temos expressões simples e fechadas para as matrizes de informação e de covariância assintótica do EMV para qualquer modelo elíptico. A seguir especificamos estes resultados para os modelos normal e t -elíptico.

5.3.1 Modelo normal

Para o modelo normal temos que $c_g = 1$ e (ver (4.65)) $a_1 = p/4$, $a_2 = -p/2$ e $a_3 = p(p+2)/4$, de onde segue que $c_1 = 0$ e $c_2 = 1/2$. Ou seja, a matriz de informação, por elemento, para este modelo é dada por

$$I_{GN}(\theta) = \begin{pmatrix} V^{-1} & 0 \\ 0 & \frac{1}{2}(bb' + D(d)) \end{pmatrix} \quad (5.29)$$

e a matriz de covariância assintótica do EMV é dada por (ver (5.28)):

$$\Omega_{GN}(\theta) = \begin{pmatrix} V & 0 \\ 0 & 2Q \end{pmatrix}. \quad (5.30)$$

Estes fatos coincidem com os dados por Anderson (1973), para o caso normal com matriz de covariância com estrutura linear, ver (5.23).

5.3.2 Modelo t -elíptico

Para a distribuição t p -variada temos que $c_g = (\nu/(\nu-2))$, $\nu > 2$ e (ver (4.66))

$$a_1 = \frac{p(\nu+p)}{4(\nu+p+2)}, \quad a_2 = -p/2 \quad e \quad a_3 = \frac{(\nu+p)p(p+2)}{4(\nu+p+2)},$$

de onde segue que

$$c_1 = -\frac{1}{2(\nu + p + 2)} \quad \text{e} \quad c_2 = \frac{(\nu + p)}{2(\nu + p + 2)} . \quad (5.31)$$

Assim, supondo ν conhecido, a matriz de informação é dada por

$$I_{GT}(\theta) = \begin{pmatrix} I_{\mu\mu} & 0 \\ 0 & I_{\phi\phi} \end{pmatrix} , \quad (5.32)$$

onde

$$\begin{aligned} I_{\mu\mu} &= \left(\frac{\nu}{\nu - 2} \right) \left(\frac{\nu + p}{\nu + p + 2} \right) V^{-1} \quad \text{e} \\ I_{\phi\phi} &= \left(\frac{-1}{2(\nu + p + 2)} \right) hh' + \left(\frac{(\nu + p)}{2(\nu + p + 2)} \right) \{bb' + D(d)\} \end{aligned}$$

e a matriz de covariância assintótica do EMV de θ é dada por:

$$\Omega_{GT} = \begin{pmatrix} \Omega_{\mu} & 0 \\ 0 & \Omega_{\phi} \end{pmatrix} . \quad (5.33)$$

onde

$$\begin{aligned} \Omega_{\mu} &= \left\{ \frac{(\nu + 2)(\nu + p + 2)}{\nu(\nu + p)} \right\} V \quad \text{e} \\ \Omega_{\phi} &= \left\{ \frac{2(\nu + p + 2)}{\nu + p} \right\} \left[Q + \frac{Qhh'Q}{(\nu + p) - h'Qh} \right] . \end{aligned}$$

Notemos que se $\nu \rightarrow \infty$ obtemos os resultados correspondentes ao modelo normal. Se ν não for fixado a priori, a matriz de informação, por elemento, é dada por:

$$I_{GT}(\theta, \nu) = \begin{pmatrix} I_{\mu\mu} & 0 & 0 \\ 0 & I_{\phi\phi} & I_{\phi\nu} \\ 0 & I_{\nu\phi} & i_{\nu\nu} \end{pmatrix} , \quad (5.34)$$

onde $I_{\mu\mu}$ e $I_{\phi\phi}$ são como em (5.32),

$$I_{\phi\nu} = \left(\frac{p + 2}{(\nu - 2)(\nu + p)(\nu + p + 2)} \right) h = I'_{\nu}\phi$$

e $i_{\nu\nu}$ é como definido em (4.54). Este fato segue dos resultados do Apêndice D.

5.4 Testes de hipóteses

Usando os resultados da Seção 5.3, podemos implementar testes assintóticos para provar hipóteses de interesse tais como H_{g1} e H_{g2} , utilizando, por exemplo, a estatística de Wald.

A seguir, no entanto, discutiremos um teste assintótico para testar H_{g2} em qualquer modelo elíptico com quarto momento finito.

Com efeito, temos que

$$Y_j, \ 1 \leq j \leq n, \text{ são i.i.d. } El_p(\mu, \Lambda_Y),$$

com matriz de covariância $V = c_g \Lambda_Y = D(\phi) + \phi_0 J$ com elementos

$$v_{ij} = \begin{cases} \sigma_i = \phi_0 + \phi_i & \text{se } i = j \\ \phi_0 & \text{se } i \neq j \end{cases} \quad i, j = 1, 2, \dots, p. \quad (5.35)$$

Note que H_{g2} é equivalente a $H : \sigma_1 = \sigma_2 = \dots = \sigma_p$. Para este caso e supondo normalidade, Choi e Wette (1972) propõem um teste para verificar H , que é uma extensão do teste de Pitman (1939).

Sejam

$$\begin{aligned} x_{0j} &= \bar{y}_j = \frac{1}{p} \sum_{i=1}^p y_{ij}; \\ x_{ij} &= y_{ij} - \bar{y}_j, \quad i = 1, 2, \dots, p-1, \quad j = 1, 2, \dots, n. \end{aligned} \quad (5.36)$$

Choi e Wette (1972) (ver também Han, 1968) mostram que H é equivalente a $\text{Cov}(x_{0j}, x_{ij}) = 0, \forall i = 1, 2, \dots, p-1$.

A transformação (5.36) pode ser escrita como

$$X_j = \begin{bmatrix} x_{0j} \\ x_{1j} \\ x_{2j} \\ \vdots \\ x_{p-1j} \end{bmatrix} = TY_j, \quad j = 1, 2, \dots, n, \quad (5.37)$$

onde T é uma matriz $p \times p$ dada por

$$T = \begin{pmatrix} \frac{1}{p} & \frac{1}{p} & \frac{1}{p} & \dots & \frac{1}{p} \\ (1 - \frac{1}{p}) & -\frac{1}{p} & -\frac{1}{p} & \dots & -\frac{1}{p} \\ -\frac{1}{p} & (1 - \frac{1}{p}) & -\frac{1}{p} & \dots & -\frac{1}{p} \\ \vdots & & & & \\ -\frac{1}{p} & -\frac{1}{p} & \dots & (1 - \frac{1}{p}) & -\frac{1}{p} \end{pmatrix}. \quad (5.38)$$

Das propriedades das distribuições elípticas, ver Apêndice A, temos que

$$X_j, j = 1, 2, \dots, n, \text{ são i.i.d. } El_p(\mu_X, \Lambda_X) \quad (5.39)$$

onde $\mu_X = T\mu$, $\Lambda_X = T\Lambda_T T'$ e matriz de covariância $V_X = TVT'$.

Sejam $\bar{\mu} = \frac{1}{p} \sum_{i=1}^p \mu_i$ e $\bar{\phi} = \frac{1}{p} \sum_{i=1}^p \phi_i$. Logo, é fácil ver que

$$\begin{aligned} \mu_X &= \begin{bmatrix} \bar{\mu} \\ \mu_1 - \bar{\mu} \\ \mu_2 - \bar{\mu} \\ \vdots \\ \mu_{p-1} - \bar{\mu} \end{bmatrix} \text{ e} \\ V_X &= \frac{1}{p} \begin{bmatrix} \bar{\phi} + p\phi_0 & \phi_1 - \bar{\phi} & \phi_2 - \bar{\phi} & \phi_3 - \bar{\phi} & \dots & \phi_{p-1} - \bar{\phi} \\ (p-2)\phi_1 + \bar{\phi} & -(\phi_1 + \phi_0) + \bar{\phi} & & & \dots & -(\phi_1 + \phi_{p-1}) + \bar{\phi} \\ & (p-2)\phi_2 + \bar{\phi} & -(\phi_2 + \phi_3) + \bar{\phi} & \dots & & -(\phi_2 + \phi_{p-1}) + \bar{\phi} \\ & & & & \vdots & \\ \text{(simetria)} & & & & & (p-2)\phi_{p-1} + \bar{\phi} \end{bmatrix} \\ &= \begin{pmatrix} v_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}. \end{aligned} \quad (5.40)$$

Ou seja, temos que (Choi e Wette, 1972) $H_{g2} : \phi_1 = \phi_2 = \dots = \phi_p$ se e somente se $V_{12} = 0$, mas isto é equivalente a $H : \bar{R} = 0$, onde

$$\bar{R} = \left\{ \frac{V_{12} V_{22}^{-1} V_{21}}{v_{11}} \right\}^{1/2}, \quad (5.41)$$

é o coeficiente de correlação múltipla entre x_0 e $x_1, x_2, \dots, x_{p-1}$. Isto é, testar que os instrumentos são igualmente confiáveis é equivalente a verificar se o coeficiente de correlação múltipla entre x_0 e $x_1, \dots, x_{p-1}$ é zero. Para isto, usaremos o seguinte resultado, provado por Muirhead (1980, 1982).

Lema 5.1 *Seja $X_1, X_2, \dots, X_n$ uma amostra aleatória de tamanho n de uma distribuição $El_p(\mu_X, \Lambda_X)$ com coeficiente de correlação múltipla $\bar{R} = 0$ e parâmetro de curtose κ . Então tem-se que*

$$\frac{(n-1)R^2}{1+\kappa} \xrightarrow{\mathcal{D}} \chi^2(p-1),$$

onde

$$R = \left\{ \frac{S_{12}S_{22}^{-1}S_{21}}{s_{11}} \right\}^{1/2}, \quad (5.42)$$

é o coeficiente de correlação múltipla amostral com

$$S_x = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})' = \begin{pmatrix} s_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \quad (5.43)$$

a matriz de covariância amostral.

Usando este lema temos que um teste aproximado de tamanho α para testar H é dado por

$$\text{Rejeitamos } H \text{ se } \frac{(n-1)R^2}{1+\kappa} > \chi_{1-\alpha}^2(p-1), \quad (5.44)$$

onde $\chi_{1-\alpha}^2(p-1)$ é o percentil $100(1-\alpha)$ da distribuição $\chi^2(p-1)$. Se κ for desconhecido pode ser substituído por uma estimativa consistente. Com efeito, como discutido na Seção 4.2, uma estimativa consistente para κ é

$$\tilde{\kappa} = \frac{b_{2,p}}{p(p+2)} - 1, \quad (5.45)$$

onde

$$b_{2,p} = \frac{1}{n} \sum_{j=1}^n [(X_j - \bar{X})' S_x^{-1} (X_j - \bar{X})]^2,$$

com S_x dada em (5.43).

Por outro lado, no caso normal temos um teste exato. Neste caso, sabemos que (Muirhead, 1982) o teste da razão de verossimilhança para testar $H : \bar{R} = 0$, é dado por

$$F = \frac{(n-p)}{(p-1)} \frac{R^2}{1-R^2} \sim F(p-1, n-p) \text{ sob } H.$$

Também supondo normalidade, Christensen e Blackwood (1993) usam um modelo de regressão multivariado para testar H_{g2} .

Finalmente, no caso do modelo t -elíptico, temos que $\kappa = 2/(\nu - 4)$ se $\nu > 4$, de onde segue que uma estimativa consistente para ν é dada por, ver Seção 4.2,

$$\tilde{\nu} = 2(1 - 2b_{2,p}/p(p+2))/(1 - b_{2,p}/p(p+2)) . \quad (5.46)$$

com $b_{2,p}$ como definido em (5.45).

Capítulo 6

Um modelo funcional normal

6.1 Introdução

Neste capítulo tratamos do problema de calibração comparativa funcional, onde desejamos comparar p (≥ 2) formas diferentes (instrumentos) de medir alguma quantidade desconhecida, x , num grupo comum de unidades experimentais.

Como considerado nos capítulos anteriores, sejam x_j o valor da característica x na unidade j e y_{ij} a medida fornecida pelo instrumento i para a característica x na unidade j , $i = 1, 2, \dots, p$ e $j = 1, 2, \dots, n$.

Um modelo freqüentemente adotado na literatura (Williams, 1969, Kimura, 1992) é:

$$\begin{aligned} y_{1j} &= x_j + \varepsilon_{1j} \\ y_{ij} &= \alpha_i + \beta_i x_j + \varepsilon_{ij}, \quad i \geq 2, \end{aligned} \tag{6.1}$$

onde ε_{ij} são v.a.i. com $\sim N(0, \phi \lambda_i)$, $\lambda_1, \dots, \lambda_p$ constantes conhecidas e $x_1, x_2, \dots, x_n$ constantes desconhecidas, parâmetros. Sem perda de generalidade, podemos supor que $\lambda_i = 1$, $i = 1, 2, \dots, p$. Ver Seção 6.6.

Este modelo em notação matricial pode ser escrito como

$$Y_j = a + b x_j + \varepsilon_j, \tag{6.2}$$

onde $Y_j = (y_{1j}, \dots, y_{pj})'$ e $\varepsilon_j = (\varepsilon_{1j}, \dots, \varepsilon_{pj})'$ são vetores aleatórios $p \times 1$ com

$$\begin{aligned} \varepsilon_j, \quad 1 \leq j \leq n, \text{ i.i.d. } N_p(0, \phi I_p); \\ \mathbf{a} = \begin{pmatrix} 0 \\ \alpha \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ \beta \end{pmatrix} \quad \text{com } \alpha = (\alpha_2, \dots, \alpha_p)' \\ \text{e } \beta = (\beta_2, \dots, \beta_p)'. \end{aligned}$$

Daqui segue que

$$Y_j, \quad 1 \leq j \leq n \quad \text{são v.a.i. tal que } Y_j \sim N_p(\mathbf{a} + \mathbf{b}x_j, \phi I_p) \quad (6.3)$$

O modelo (6.2) é denominado (Kendall e Stuart, 1979, e Fuller, 1987) “Modelo Linear Funcional com Erros nas Variáveis” ou simplesmente Modelo Funcional. Em geral o interesse é fazer inferências sobre os parâmetros α e β e a diferença com o modelo estrutural é que aqui os x_j 's são considerados fixos. O número de parâmetros neste modelo é $2p-1+n$ e cresce com o tamanho da amostra n .

Este tipo de modelo tem sido tratado entre outros por Gleser (1981) e Fuller (1987). Embora o tratamento teórico dado por Gleser e Fuller seja bastante geral e completo, a teoria, baseada em valores e vetores próprios pode parecer um tanto complicada, especialmente para leitores interessados nas aplicações, como é notado também por Kimura (1992).

Devido a isto, Kimura (1992) propõe um simples algoritmo tipo EM para estimar os parâmetros α , β e ϕ . Para estimar a matriz de covariância assintótica dos estimadores, no entanto, Kimura propõe a inversa da matriz de informação observada, a qual nesta situação não necessariamente corresponde à matriz de covariância assintótica dos EMV, como foi notado por Patefield (1977, 1978) no modelo funcional simples ($p = 2$) e no modelo ultraestrutural, respectivamente.

Neste capítulo mostramos que os estimadores dados por Kimura são um caso especial dos estimadores propostos por Mak (1982) e apresentamos uma expressão explícita para a matriz de covariância assintótica dos EMV que não é baseada na matriz de informação.

Mostramos também que a inversa da matriz de informação esperada não corresponde à matriz de covariância assintótica dos EMV como foi provado por Patefield (1977) para o caso particular $p = 2$.

Além disso, propomos um estimador de momentos para β e estudamos suas propriedades assintóticas.

Na Seção 6.2 resumimos os principais resultados de Mak (1982). Estes fatos são usados na Seção 6.3 para encontrar os EMV e sua matriz de covariância assintótica. Um estimador consistente desta matriz é proposto. Também é mostrado que a inversa da matriz de informação esperada não é a matriz de covariância assintótica dos EMV.

Na Seção 6.4 propomos um estimador de momentos para β e estudamos suas propriedades assintóticas.

Na Seção 6.5 consideramos testes de hipóteses usando a estatística de Wald. Na Seção 6.6 apresentamos uma extensão do modelo funcional (6.1); na Seção 6.7 apresentamos uma ilustração numérica. Finalmente, na Seção 6.8 apresentamos um pequeno estudo de simulação.

6.2 Estimação em presença de parâmetros incidentais

Nesta seção apresentamos de uma forma resumida os principais resultados de Mak (1982), que serão utilizados posteriormente para obter os EMV dos parâmetros do modelo funcional (6.2).

No caso de observações iid e sob condições de regularidade é sabido (Sen e Singer, 1993) que os EMV são consistentes e assintoticamente normais (AN), a matriz de covariância assintótica é dada pela inversa da matriz de informação.

Em muitas situações, no entanto, a “suposição” de observações iid não é satisfeita, como por exemplo no modelo funcional (6.2).

Seja $Y_1, Y_2, \dots$ uma seqüência de vetores aleatórios independentes $p \times 1$. Para cada j , seja $f_j(Y_j, \theta, \tau_j)$ a função de densidade de probabilidade de Y_j , onde θ é um vetor de parâmetros $k \times 1$ e τ_j é de dimensão $r \times 1$. Aqui θ é o mesmo para cada j e é chamado parâmetro “estrutural” e $\tau_j, j = 1, 2, \dots$ são denominados parâmetros “incidentais”. O interesse principal é a estimação do parâmetro estrutural θ .

Na ausência de parâmetros incidentais, mas quando os Y_j não são iid, as propriedades assintóticas dos EMV, sob condições de regularidade, foram provadas por Hoadley (1971) e Philippou e Roussas (1975).

Em presença de parâmetros incidentais o número de parâmetros desconhecidos cresce quando o tamanho da amostra cresce e a justificativa para utilizar o EMV é menos óbvia, como no modelo funcional (6.2). Nesta situação, os EMV podem ser inconsistentes e sua matriz de covariância assintótica não é dada, necessariamente, pela inversa da matriz de informação de Fisher.

Mak (1982) propõe um método de estimação que inclui máxima verossimilhança e dá propriedades assintóticas do estimador, sob certas condições de regularidade. A seguir resumimos seus principais resultados.

Para uma amostra de tamanho n , o logaritmo da função de verossimilhança, $\ell(\theta, \tau)$, é dado por

$$\ell(\theta, \tau) = \sum_{j=1}^n \ln f_j(Y_j, \theta, \tau_j) , \quad (6.4)$$

onde $\theta \in \Theta$ é o espaço paramétrico de θ que contém o verdadeiro valor de θ , θ_0 . Seja τ_{0j} o verdadeiro valor de τ_j e $\tau = (\tau_1, \dots, \tau_n)'$.

Suponha que para cada j e dado θ , existe um estimador $\hat{\tau}_j = \hat{\tau}_j(Y_j, \theta)$ de τ_j que depende de θ . Em particular, $\hat{\tau}_j$ pode ser o EMV condicional de τ_j dado θ .

Substituindo τ_j por $\hat{\tau}_j$ em (6.4), temos:

$$\ell_M(\theta) = \sum_{j=1}^n h_j(Y_j, \theta) , \quad (6.5)$$

onde $h_j(Y_j, \theta) = \ln f_j(Y_j, \theta, \hat{\tau}_j)$, $j = 1, 2, \dots, n$.

Sejam

$$q_j(Y_j, \theta) = (q_{j1}(Y_j, \theta), \dots, q_{jk}(Y_j, \theta))' \quad (6.6)$$

um vetor aleatório de dimensão $k \times 1$, onde

$$q_{ji}(Y_j, \theta) = \frac{\partial}{\partial \theta_i}(h_j(Y_j, \theta)) \ , \ i = 1, 2, \dots, k,$$

e

$$A_n(\theta) = [(a_{il})] = \left[\left(\frac{1}{n} \sum_{j=1}^n q_{jil}(Y_j, \theta) \right) \right] \quad (6.7)$$

uma matriz aleatória simétrica $k \times k$, com

$$q_{jie}(Y_j, \theta) = \frac{\partial^2}{\partial \theta_e \partial \theta_i}(l_{1j}(Y_j, \theta)) \ , \ i, e = 1, 2, \dots, k;$$

supondo que as derivadas parciais existam, q.c., para todo j .

Suponha que, para n grande, existe $\theta_* \in \Theta$ que maximiza a função

$$\bar{h}(\theta) = \frac{1}{n} \sum_{j=1}^n E_0(h_j(Y_j, \theta)) \ , \quad (6.8)$$

onde E_0 indica que a esperança é tomada em θ_0 e τ_{0j} .

Mak (1982) mostra, sob certas condições de regularidade, o seguinte resultado:

Lema 6.1 (i) $\ell_M(\theta)$ dada em (6.5) tem um máximo $\hat{\theta}_n$ tal que

$$\hat{\theta}_n \xrightarrow{P} \theta_* \text{ quando } n \mapsto \infty \ .$$

$$(ii) \ \sqrt{n}V_n^{-1/2}E_0[A_n(\theta_*)](\hat{\theta}_n - \theta_*) \xrightarrow{\mathcal{D}} N_k(0, I_k)$$

ou, equivalentemente, $\hat{\theta}_n$ é $AN(\theta_*, \frac{1}{n}\Omega)$, onde

$$\Omega = [E_0(A_n(\theta_*))]^{-1}V_n[E_0(A_n(\theta_*))]^{-1} \quad (6.9)$$

com

$$V_n = \frac{1}{n} \sum_{j=1}^n \text{Var}_0(q_j(Y_j, \theta_*)) \ , \quad (6.10)$$

onde Var_0 indica que a variância é tomada sob os verdadeiros valores dos parâmetros

θ_0 e τ_{0j} .

Para detalhes sobre as condições de regularidade, comentários e algumas aplicações, ver Mak (1982).

6.3 Estimação de máxima verossimilhança

Nesta seção aplicamos os resultados de Mak (1982), resumidos no Lema 6.1, para obter os EMV, e suas propriedades assintóticas dos parâmetros do modelo funcional (6.2).

6.3.1 Obtenção dos EMV

Seja $\theta = (\alpha', \beta', \phi)'$ o vetor de parâmetros estruturais e $\tau = \mathbf{x} = (x_1, x_2, \dots, x_n)'$ o vetor de parâmetros incidentais.

De (6.3) segue que a densidade de Y_j é dada por

$$f_j(Y_j, \theta, \mathbf{x}) = (2\pi)^{-p/2} (\phi)^{-p/2} e^{-\frac{1}{2\phi} \|Y_j - \alpha + \beta x_j\|^2}, \quad (6.11)$$

$j = 1, 2, \dots, n$. Logo, o logaritmo da função de verossimilhança, $\ell(\theta, \mathbf{x})$, (6.4), é dado por

$$\begin{aligned} \ell(\theta, \mathbf{x}) &= \sum_{j=1}^n \ln f_j(Y_j, \theta, \mathbf{x}) \\ &= \sum_{j=1}^n \left\{ -\frac{p}{2} \ln(2\pi) - \frac{p}{2} \ln \phi - \frac{1}{2\phi} \|Y_j - \alpha + \beta x_j\|^2 \right\} \end{aligned} \quad (6.12)$$

É fácil ver que o EMV de x_j dado θ é

$$\hat{x}_j = \hat{x}_j(\theta) = (Y_j' \mathbf{b} - \alpha' \mathbf{b}) / \mathbf{b}' \mathbf{b} = \frac{1}{c} [y_{1j} + (Y_{2j} - \alpha)' \beta], \quad j = 1, 2, \dots, n, \quad (6.13)$$

onde $Y_{2j} = (y_{2j}, \dots, y_{pj})'$, $j = 1, 2, \dots, n$ e $c = \mathbf{b}' \mathbf{b} = 1 + \beta' \beta$.

Substituindo x_j em (6.12) por $\hat{x}_j$, dado acima, temos:

$$\ell_M(\theta) = \sum_{j=1}^n h_j(Y_j, \theta), \quad (6.14)$$

onde

$$\begin{aligned} h_j(Y_j, \theta) &= -\frac{p}{2} \ln(2\pi) - \frac{p}{2} \ln \phi - \frac{1}{2\phi} [(y_{1j} - \hat{x}_j)^2 + \|Y_{2j} - \alpha - \beta \hat{x}_j\|^2] \\ &= -\frac{p}{2} \ln(2\pi) - \frac{p}{2} \ln \phi - \frac{1}{2\phi} [y_{1j}^2 + \|Y_{2j} - \alpha\|^2 - \frac{1}{c} (y_{1j} + (Y_{2j} - \alpha)' \beta)^2], \end{aligned}$$

$j = 1, 2, \dots, n$.

Logo, os EMV de α , β e ϕ satisfazem:

$$\frac{\partial}{\partial \alpha}(\ell_M(\theta)) = 0; \quad (6.15)$$

$$\frac{\partial}{\partial \beta}(\ell_M(\theta)) = 0; \quad (6.16)$$

$$\frac{\partial}{\partial \phi}(\ell_M(\theta)) = 0. \quad (6.17)$$

Seja

$$B_1 = cI_{p-1} - \beta\beta'. \quad (6.18)$$

É fácil ver que (Graybill, 1983):

$$B_1^{-1} = \frac{1}{c}[I_{p-1} + \beta\beta'], \quad (6.19)$$

de onde segue que

$$B_1^{-1}\beta = \beta. \quad (6.20)$$

Usando (6.20), depois de alguma álgebra temos que (6.15) - (6.17) é equivalente a

$$\alpha = \bar{Y}_2 - \beta\bar{y}_1; \quad (6.21)$$

$$\beta = \frac{\sum_{j=1}^n (Y_{2j} - \bar{Y}_2)\hat{x}_j}{\sum_{j=1}^n (\hat{x}_j - \bar{y}_1)^2}; \quad (6.22)$$

$$\phi = \frac{1}{np} \sum_{j=1}^n \{(y_{1j} - \hat{x}_j)^2 + \|Y_{2j} - \alpha - \beta\hat{x}_j\|^2\}, \quad (6.23)$$

onde

$$\bar{Y}_2 = \frac{1}{n} \sum_{j=1}^n Y_{2j} \quad \text{e} \quad \bar{y}_1 = \frac{1}{n} \sum_{j=1}^n y_{1j} = \frac{1}{n} \sum_{j=1}^n \hat{x}_j.$$

A solução de (6.21) - (6.23) deve ser obtida usando um método iterativo e será denotado por $\hat{\theta}_n = (\hat{\alpha}'_n, \hat{\beta}'_n, \hat{\phi}_n)'$. Com efeito, seja $\alpha^{(m)}$ e $\beta^{(m)}$ a solução de (6.21) e (6.22), respectivamente, na m -ésima iteração, na iteração $m + 1$ o algoritmo procede como:

Etapas 1: Calcule $\hat{x}_j = \hat{x}_j(\alpha^{(m)}, \beta^{(m)})$ dada em (6.13).

Etapa 2: Calcule $\alpha^{(m+1)}$ e $\beta^{(m+1)}$ a partir de (6.21) e (6.22), respectivamente.

Como estimativa inicial, podemos tomar $\alpha^{(0)} = 0$ e $\beta^{(0)} = 1$.

Notemos que este algoritmo de estimação, proposto por Kimura (1992) é similar ao algoritmo EM. A estimativa de ϕ , que não precisa do processo iterativo, é obtida a partir de (6.23) e é dada por

$$\hat{\phi}_n = \frac{1}{np} \sum_{j=1}^n \{(y_{1j} - \hat{x}_j)^2 + \|Y_{2j} - \hat{\alpha}_n - \hat{\beta}_n \hat{x}_j\|^2\}. \quad (6.24)$$

Do anterior temos então que os estimadores propostos por Kimura (1992) são um caso especial dos dados em Mak (1982).

6.3.2 Propriedades assintóticas do EMV

Para obter as propriedades assintóticas do EMV $\hat{\theta}_n = (\hat{\alpha}'_n, \hat{\beta}'_n, \hat{\phi}_n)'$ usaremos os resultados de Mak (1982), resumidos no Lema 6.1.

De (6.14), é fácil ver que

$$h_j(Y_j, \theta) = -\frac{p}{2} \ln(2\pi) - \frac{p}{2} \ln \phi - \frac{1}{2\phi c} (Y_j - \mathbf{a})' B (Y_j - \mathbf{a}), \quad (6.25)$$

$j = 1, 2, \dots, n$, onde $B = cI_p - \mathbf{b}\mathbf{b}'$ e c como em (6.13).

Notemos que $BB = cB$.

Sejam $\theta_0 = (\alpha'_0, \beta'_0, \phi_0)'$ e $\mathbf{x}_0 = (x_{01}, \dots, x_{0n})'$ o verdadeiro valor de θ e de $\mathbf{x}$, respectivamente. Tomando a esperança de $h_j(Y_j, \theta)$ em $(\theta_0, \mathbf{x}_0)$, temos, usando propriedades da distribuição normal (Muirhead, 1982):

$$\begin{aligned} E_0(h_j(Y_j, \theta)) &= -\frac{p}{2} \ln(2\pi) - \frac{p}{2} \ln \phi - \frac{1}{2\phi} \{ \phi_0(p-1) \\ &\quad + \frac{1}{c} [x_{0j}^2 \mathbf{b}'_0 B \mathbf{b}_0 - 2x_{0j} \mathbf{b}'_0 B (\mathbf{a} - \mathbf{a}_0) + (\mathbf{a} - \mathbf{a}_0)' B (\mathbf{a} - \mathbf{a}_0)] \}, \end{aligned} \quad (6.26)$$

$j = 1, 2, \dots, n$; com $\mathbf{a}_0 = \begin{pmatrix} 0 \\ \alpha_0 \end{pmatrix}$ e $\mathbf{b}_0 = \begin{pmatrix} 1 \\ \beta_0 \end{pmatrix}$.

É fácil ver que (6.26) tem um máximo em $\theta_* = (\alpha'_0, \beta'_0, ((p-1)/p)\phi_0)'$, independente de x_{0j} . Logo, θ_* maximiza a função $\bar{h}(\theta)$ dada em (6.8), com $E_0(h_j(Y_j, \theta))$, $j = 1, 2, \dots, n$, dada por (6.26).

Para obter a matriz V_n , definida em (6.10), precisamos do seguinte resultado:

Lema 6.2 *Sejam $Y \sim N_p(\mu, \Sigma)$, $a \in \mathbb{R}^p$ constante e A uma matriz simétrica $p \times p$. Então:*

- (i) $E[(Y'AY)(YY')] = 2\Sigma A\Sigma + \Sigma \text{tr}(A\Sigma) + \text{tr}(A\Sigma)\mu\mu' + 2\Sigma A\mu\mu' + 2\mu\mu' A\Sigma + (\mu' A\mu)\Sigma + (\mu' A\mu)\mu\mu'$;
- (ii) $E[(a'Y)(YY')] = \Sigma a\mu' + \mu a'\Sigma + (a'\mu)\Sigma + (a'\mu)\mu\mu'$;
- (iii) $E[(Y'AY)Y] = \text{tr}(A\Sigma)\mu + 2\Sigma A\mu + (\mu' A\mu)\mu$.

Prova: Notemos que $Y - \mu \sim N_p(0, \Sigma)$ e, aplicando os resultados do Apêndice D, o resultado segue.

Lema 6.3 *Sob o modelo funcional (6.2), temos que*

$$(i) \quad V_n = \left(\frac{dp}{p-1} \right) \begin{bmatrix} K_2 \otimes B_{01} & 0 \\ 0 & p^2 c_0 / 2\phi_0(p-1) \end{bmatrix} ; \quad (6.27)$$

$$(ii) \quad E_0(A_n(\theta_*)) = -d \begin{bmatrix} K_1 \otimes B_{01} & 0 \\ 0 & p^2 c_0 / 2\phi_0(p-1) \end{bmatrix} , \quad (6.28)$$

onde

$$K_1 = \begin{bmatrix} 1 & \bar{x}_0 \\ \bar{x}_0 & \frac{1}{n} \sum_{j=1}^n x_{0j}^2 \end{bmatrix} , \quad K_2 = \begin{bmatrix} 1 & \bar{x}_0 \\ \bar{x}_0 & \frac{\phi_0}{c_0} + \frac{1}{n} \sum_{j=1}^n x_{0j}^2 \end{bmatrix} ,$$

$d = p/c_0\phi_0(p-1)$ e B_{01} como dada em (6.18) avaliada em $\theta = \theta_0$; $\otimes$ indica o produto Kronecker entre duas matrizes.

Prova: (i) De (6.6), temos, neste caso, que

$$q_j(Y_j, \theta) = \begin{bmatrix} \frac{\partial}{\partial \alpha} (h_j(Y_j, \theta)) \\ \frac{\partial}{\partial \beta} (h_j(Y_j, \theta)) \\ \frac{\partial}{\partial \phi} (h_j(Y_j, \theta)) \end{bmatrix} = \begin{bmatrix} \frac{1}{\phi} [(Y_{2j} - \alpha) - \beta \hat{x}_j] \\ \frac{1}{\phi} [(Y_{2j} - \alpha) \hat{x}_j - \beta \hat{x}_j^2] \\ -\frac{p}{2\phi} + \frac{1}{2\phi^2} Z_j Z_j \end{bmatrix} , \quad (6.29)$$

onde $Z_j = Y_j - a - b\hat{x}_j = \frac{1}{c}B(Y_j - a)$, $j = 1, 2, \dots, n$, com B definida em (6.25).

Logo, fazendo $\theta = \theta_*$ em (6.29), tomando variância e usando o Lema 6.2, temos, depois de alguma álgebra, a matriz V_n dada em (6.27).

(ii) A matriz $A_n(\theta)$, definida em (6.7), pode ser escrita como

$$A_n(\theta) = \frac{1}{n} \sum_{j=1}^n A_{nj}(\theta) ,$$

onde

$$A_{nj}(\theta) = \begin{pmatrix} A_{11}(\theta) & A_{12}(\theta) & A_{13}(\theta) \\ & A_{22}(\theta) & A_{23}(\theta) \\ \text{(simetria)} & & A_{33}(\theta) \end{pmatrix} , \quad (6.30)$$

onde, neste caso,

$$\begin{aligned} A_{11}(\theta) &= -(1/\phi c)B_1 , \\ A_{12}(\theta) &= -\frac{1}{\phi} \left[\hat{x}_j I_{p-1} + \frac{1}{c} \beta ((Y_{2j} - \alpha) - 2\hat{x}_j \beta)' \right] , \\ A_{13}(\theta) &= -\frac{1}{\phi^2} [(Y_{2j} - \alpha) - \beta \hat{x}_j] , \\ A_{22}(\theta) &= \frac{1}{\phi c} [(Y_{2j} - \alpha)(Y_{2j} - \alpha)' - 2\hat{x}_j (Y_{2j} - \alpha) \beta' - c \hat{x}_j^2 I_{p-1} \\ &\quad - 2\hat{x}_j \beta (Y_{2j} - \alpha)' + 4\hat{x}_j^2 \beta \beta'] , \\ A_{23}(\theta) &= -\frac{1}{\phi^2} [(Y_{2j} - \alpha) \hat{x}_j - \beta \hat{x}_j^2] \text{ e} \\ A_{33}(\theta) &= \frac{p}{2\phi^2} - \frac{1}{\phi^3 c} (Y_j - a)' B (Y_j - a) , \end{aligned}$$

$j = 1, 2, \dots, n$.

Fazendo $\theta = \theta_*$ em (6.30) e tomando a esperança, obtemos a matriz dada em (6.28).

Agora, como

$$[E_0(A_n(\theta_*))]^{-1} = -(1/d) \begin{bmatrix} K_1^{-1} \otimes B_{01}^{-1} & 0 \\ 0 & \frac{2\phi_0(p-1)}{c_0 p^2} \end{bmatrix} , \quad (6.31)$$

temos:

Teorema 6.1 Sob o modelo funcional (6.2), o EMV $\hat{\theta}_n = (\hat{\alpha}'_n, \hat{\beta}'_n, \hat{\phi}_n)'$ é $AN(\theta_*, \frac{1}{n}\Omega)$, onde

$$\Omega = c_0\phi_0 \begin{bmatrix} K \otimes B_{01}^{-1} & 0 \\ 0 & \frac{2\phi_0(p-1)}{c_0p^2} \end{bmatrix}, \quad (6.32)$$

com

$$K = \frac{1}{S_{xx}} \begin{bmatrix} S_{xx} + \bar{x}_0^2 d_0 & -\bar{x}_0 d_0 \\ -\bar{x}_0 d_0 & d_0 \end{bmatrix},$$

onde $d_0 = 1 + \phi_0/c_0 S_{xx}$ e $S_{xx} = \frac{1}{n} \sum_{j=1}^n (x_{0j} - \bar{x}_0)^2$.

Prova: É só aplicar os Lemas 6.1 e 6.3, usar (6.31) e alguma álgebra para simplificar.

Notemos que se $p = 2$, obtemos os resultados de Mak (1982) e Patefield (1977) como casos especiais. Note, também, que $\hat{\theta}_n$ não é consistente, pois $\hat{\theta}_n \xrightarrow{P} \theta_* \neq \theta_0$. Mas, $\tilde{\theta}_n = (\hat{\alpha}'_n, \hat{\beta}'_n, \tilde{\phi}_n)'$, com $\tilde{\phi}_n = (p/(p-1))\hat{\phi}_n$ é consistente e $AN(\theta_0, \frac{1}{n}\Omega_0)$, onde

$$\Omega_0 = c_0\phi_0 \begin{bmatrix} K \otimes B_{01}^{-1} & 0 \\ 0 & 2\phi_0/c_0(p-1) \end{bmatrix}. \quad (6.33)$$

Agora, para obter uma estimativa consistente de Ω_0 , só precisamos estimar consistentemente $\bar{x}_0$ e S_{xx} . Para isto, utilizaremos o seguinte resultado em Arellano et al. (1994b).

Lema 6.4 Seja $Y_1, Y_2, \dots, Y_j, \dots$ uma seqüência de vetores aleatórios independentes de dimensão $p \times 1$, com $E(Y_j) = \mu_j$ e $\text{Var}(Y_j) = \Sigma$, $j \geq 1$. Seja $W_j = Y_j - \mu_j$ de modo que $E(W_j) = 0$ e $\text{Var}(W_j) = \Sigma$, $j \geq 1$. Suponha que existem $\mu^* \in \mathbb{R}^p$ e Σ^* uma matriz simétrica $p \times p$ tais que:

$$\lim_{n \rightarrow \infty} \bar{\mu} = \mu^* \quad e \quad \lim_{n \rightarrow \infty} S^* = \Sigma^*, \quad (6.34)$$

onde

$$\bar{\mu} = \frac{1}{n} \sum_{j=1}^n \mu_j \quad e \quad S^* = \frac{1}{n} \sum_{j=1}^n (\mu_j - \bar{\mu})(\mu_j - \bar{\mu})'.$$

Então,

$$(i) \quad \bar{Y} \xrightarrow{q.c.} \mu^*;$$

(ii) $S \xrightarrow{q.c.} \Sigma + \Sigma^*$,

com

$$\bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j \quad e \quad S = \frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y})(Y_j - \bar{Y})' .$$

Usando este lema podemos provar o seguinte resultado:

Teorema 6.2 *Sob o modelo funcional (6.2) e se*

S1: $\lim_{n \rightarrow \infty} \bar{x}_0 = u^ < \infty$;*

S2: $\lim_{n \rightarrow \infty} S_{xx} = \sigma_{xx}^ < \infty$,*

então $\bar{y}_1$ e $\hat{S}_{xx} = \frac{\hat{\mathbf{b}}' S \hat{\mathbf{b}}}{\hat{c}^2} - \frac{\tilde{\phi}}{\hat{c}}$ são estimadores consistentes para $\bar{x}_0$ e S_{xx} , respectivamente, onde

$$\hat{\mathbf{b}} = (1, \hat{\beta}'_n)', \quad \hat{c} = \hat{\mathbf{b}}' \hat{\mathbf{b}} \quad e \quad \tilde{\phi} = \tilde{\phi}_n .$$

Prova: Aqui, neste caso, temos que $\mu_j = \mathbf{a} + \mathbf{b}x_{0j}$, $j = 1, \dots, n$; de modo que

$$\lim_{n \rightarrow \infty} \bar{\mu} = \lim_{n \rightarrow \infty} (\mathbf{a} + \mathbf{b}\bar{x}_0) = \mathbf{a} + \mathbf{b}u^* \quad e$$

$$\lim_{n \rightarrow \infty} S^* = \lim_{n \rightarrow \infty} S_{xx} \mathbf{b} \mathbf{b}' = \sigma_{xx}^* \mathbf{b} \mathbf{b}' .$$

Da parte (i) do lema anterior, temos que $\bar{Y} \xrightarrow{q.c.} \mathbf{a} + \mathbf{b}u^*$, de onde segue que $\bar{y}_1 \xrightarrow{q.c.} u^*$, implicando que $\bar{y}_1 - \bar{x}_0 \xrightarrow{P} 0$, ou seja, $\bar{y}_1$ é uma estimativa consistente de $\bar{x}_0$.

Da parte (ii) do Lema 6.4, segue que $S \xrightarrow{q.c.} \phi_0 I_p + \sigma_{xx}^* \mathbf{b} \mathbf{b}'$.

Notemos que $S_{xx} = \frac{1}{n} \sum_{j=1}^n (x_{0j} - \bar{x}_0)^2 = \mathbf{b}' S \mathbf{b} / c_0^2$.

Logo, é fácil ver que $\hat{S}_{xx} - S_{xx} \xrightarrow{P} 0$, de onde temos que $\hat{S}_{xx}$ é um estimador consistente de S_{xx} .

Finalmente, notemos que

$$\begin{pmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{pmatrix} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \quad e \quad AN \left(\begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} ; \frac{c_0 \phi_0}{n} K \otimes B_{01}^{-1} \right) \quad (6.35)$$

que é assintoticamente independente de $\tilde{\phi} = \tilde{\phi}_n$ que, por sua vez, é $AN(\phi_0, 2\phi_0^2/n(p-1))$. Kimura (1992) propõe como estimador de ϕ_0 a $\bar{\phi}_n = \left(\frac{n}{n-2}\right) \tilde{\phi}_n$.

Vale salientar também que sob as condições S1 e S2 do Teorema 6.2 dado acima, Gleser (1981) obtém a distribuição assintótica dos EMV de (α', β', ϕ) , a qual coincide com nosso resultado dado no Teorema 6.1. No entanto, os EMV propostos por Gleser dependem dos valores e vetores próprios da matriz nS .

6.3.3 A matriz de informação esperada

O objetivo desta seção é obter a matriz de informação esperada e mostrar que sua inversa não corresponde à matriz de covariância assintótica dos EMV obtidos na Seção 6.3.1.

De (6.12) temos que o logaritmo da função de verossimilhança é dado por

$$\ell(\theta, \mathbf{x}) = \text{cte.} - \frac{np}{2} \ln \phi - \frac{1}{2\phi} \sum_{j=1}^n [(y_{1j} - x_j)^2 + \|Y_{2j} - \alpha - \beta x_j\|^2]. \quad (6.36)$$

Derivando duas vezes em relação a $(\theta, \mathbf{x}) = (\alpha', \beta', \phi, x_1, \dots, x_n)'$ e tomando esperança, temos, depois de alguma simplificação algébrica, o seguinte resultado:

Teorema 6.3 *Sob o modelo funcional (6.2), a matriz de informação esperada é dada por:*

$$I(\theta, \mathbf{x}) = \left(\frac{1}{\phi}\right) \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}, \quad (6.37)$$

onde

$$\begin{aligned} I_{11} &= nK_1 \otimes I_{p-1} \text{ uma matriz } 2(p-1) \times 2(p-1), \\ I_{12} &= [0, X'] \otimes \beta \text{ uma matriz } 2(p-1) \times (n+1), \\ I_{22} &= \begin{bmatrix} \frac{np}{2\phi} & 0 \\ 0 & cI_n \end{bmatrix} \text{ uma matriz diagonal } (n+1) \times (n+1), \\ I_{21} &= I'_{12}, \quad X' = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & & \cdots & x_n \end{bmatrix} \text{ uma matriz } 2 \times n \\ e \quad K_1 &= \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \frac{1}{n} \sum_{j=1}^n x_j^2 \end{bmatrix}. \end{aligned}$$

Daqui segue que:

$$I^{-1}(\theta, \mathbf{x}) = \phi \begin{bmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{bmatrix}, \quad (6.38)$$

onde

$$\begin{aligned} I^{11} &= [I_{11} - I_{12}I_{22}^{-1}I_{21}]^{-1} \\ I^{12} &= -I_{11}I_{12}[I_{22} - I_{21}I_{11}^{-1}I_{12}]^{-1} \\ I^{22} &= [I_{22} - I_{21}I_{11}^{-1}I_{12}]^{-1} \quad \text{e} \quad I^{21} = I^{12'}. \end{aligned}$$

Notemos que:

$$I_{12}I_{22}^{-1}I_{21} = \frac{n}{c}K_1 \otimes \beta\beta', \quad (6.39)$$

de modo que

$$I_{11} - I_{12}I_{22}^{-1}I_{21} = \left(\frac{n}{c}\right) K_1 \otimes B_1, \quad (6.40)$$

ou seja,

$$\phi I^{11} = \frac{c\phi}{n} K_1^{-1} \otimes B_1^{-1} \quad (6.41)$$

com

$$K_1^{-1} = \left(\frac{1}{S_{xx}}\right) \begin{bmatrix} \frac{1}{n} \sum_{j=1}^n x_j^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}.$$

É claro que a matriz $\frac{c\phi}{n} K_1^{-1} \otimes B_1^{-1}$ não corresponde à matriz de covariância assintótica dos EMV $\hat{\alpha}$ e $\hat{\beta}$ dada em (6.35), como já foi notado por Patefield (1977) no caso $p = 2$.

Mais especificamente, notemos que a diferença entre as matrizes dadas em (6.35) e (6.41) é devida a que $K_1^{-1} \neq K$, pois é fácil ver que

$$K = K_1^{-1} + \Delta, \quad (6.42)$$

com

$$\Delta = (\phi/cS_{xx}^2) \begin{bmatrix} -\bar{x} \\ 1 \end{bmatrix} [-\bar{x} \ 1]$$

e onde escrevemos θ e x em vez de θ_0 e x_0 , respectivamente. Substituindo K_1^{-1} dado por (6.42) em (6.41), temos:

$$\phi I^{11} = \frac{\phi c}{n}[(K - \Delta) \otimes B_1^{-1}] = \frac{\phi c}{n}[(K \otimes B_1^{-1}) - (\Delta \otimes B_1^{-1})] \quad (6.43)$$

de onde segue que usando a inversa da matriz de informação esperada como a matriz de covariância assintótica dos EMV de (α, β) , estamos subestimando as variâncias destes estimadores. Notemos, no entanto, que se $\Delta = 0$, ambas as matrizes coincidem.

6.4 Estimadores de momentos

Nesta seção propomos uma estimativa de momentos para β e estudamos suas propriedades assintóticas. Antes, precisamos do seguinte fato provado em Arellano et al. (1994b).

Lema 6.5 *Seja $\{Y_j\}$ uma seqüência de vetores aleatórios independentes de dimensão $p \times 1$ como definida no Lema 6.4. Suponha que*

$$\text{Var}[\text{Vec}(W_1 W_1')] = \Lambda < \infty \quad (6.44)$$

com $\text{Vec}(A) = \begin{bmatrix} A_{.1} \\ \vdots \\ A_{.p} \end{bmatrix}$ e $A_{.j}$ denota a j -ésima coluna de A , uma matriz $q \times p$. Note que $\text{Vec}(A)$ é de dimensão $qp \times 1$.

Então, se vale (6.34) e (6.44), temos que

$$\sqrt{n}(S - \Sigma - \Sigma^*) \xrightarrow{D} N_{p^2}(0, \Lambda + \Lambda^*),$$

onde

$$\Lambda^* = \Sigma \otimes \Sigma^* + \Sigma^* \otimes \Sigma + \Sigma \odot \Sigma^* + \Sigma^* \odot \Sigma, \quad (6.45)$$

com

$$A \odot B = [A \otimes B_{.1}, A \otimes B_{.2}, \dots, A \otimes B_{.p}]$$

e $B_{.j}$ denota a j -ésima coluna de B e S é a matriz de covariância amostral definida no Lema 6.4.

Em nosso caso, temos que $\Sigma = \phi I_p$ e do Lema 6.4 temos $\Sigma^* = \sigma_{xx}^* \mathbf{b}\mathbf{b}'$ e que $S \xrightarrow{q.c.} \phi I_p + \sigma_{xx}^* \mathbf{b}\mathbf{b}'$, de onde segue que

$$\tilde{\beta}_2 = S_{23}/S_{13} \xrightarrow{q.c.} \beta_2 \quad (6.46)$$

$$\tilde{\beta}_j = \frac{S_{2j}}{S_{12}} \xrightarrow{q.c.} \beta_j, \quad j = 3, 4, \dots, p.$$

Ou seja, $\tilde{\beta} = (\tilde{\beta}_1, \dots, \tilde{\beta}_p)'$ é uma estimativa consistente para β . A seguir especificaremos as matrizes Λ e Λ^* em nosso caso.

Do Lema 6.4, temos que $W_j \sim N_p(0, \phi I_p)$, $j = 1, 2, \dots, n$. Por outro lado, sabemos que

$$\text{Vec}(W_1 W_1') = W_1 \otimes W_1. \quad (6.47)$$

Usando o Lema 4.1 em Magnus e Neudecker (1979), ver também Graybill (1983), temos

$$E(W_1 \otimes W_1) = \phi \text{Vec}(I_p) \quad (6.48)$$

$$\text{Var}(W_1 \otimes W_1) = \Lambda = \phi^2(I_{p^2} + \mathbf{K}_{pp}),$$

onde $\mathbf{K}_{pp}$ é uma matriz de comutação $p^2 \times p^2$, dada por

$$\mathbf{K}_{pp} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1p} \\ C_{21} & C_{22} & \cdots & C_{2p} \\ \vdots & & & \vdots \\ C_{p1} & & \cdots & C_{pp} \end{bmatrix}, \quad (6.49)$$

com C_{ij} uma matriz $p \times p$ com o elemento (j, i) igual a +1 e o resto zero, $i, j = 1, 2, \dots, p$.

Também, usando propriedades do produto Kronecker (Graybill, 1983), é fácil ver que

$$\Sigma \otimes \Sigma^* = \phi \sigma_{xx}^* I_p \otimes \mathbf{b}\mathbf{b}',$$

$$\Sigma^* \otimes \Sigma = \phi \sigma_{xx}^* \mathbf{b}\mathbf{b}' \otimes I_p,$$

$$\Sigma \odot \Sigma^* = \phi \sigma_{xx}^* \mathbf{b}' \otimes I_p \otimes \mathbf{b} , \quad (6.50)$$

$$\Sigma^* \odot \Sigma = \phi \sigma_{xx}^* \mathbf{b} \otimes I_p \otimes \mathbf{b}' = (\Sigma \odot \Sigma^*)' ,$$

de onde segue que

$$\Lambda^* = \phi \sigma_{xx}^* [I_p \otimes \mathbf{b} \mathbf{b}' + \mathbf{b} \mathbf{b}' \otimes I_p + \mathbf{b}' \otimes I_p \otimes \mathbf{b} + \mathbf{b} \otimes I_p \otimes \mathbf{b}'] . \quad (6.51)$$

Com estes resultados e o uso do método delta é possível encontrar a matriz de covariância assintótica de $\tilde{\beta}$. Ilustraremos isto para $p = 3$.

Sejam

$$\tilde{S} = (S_{11}, S_{12}, S_{13}, S_{22}, S_{23}, S_{33})' \text{ e}$$

$$\tilde{\sigma} = (\phi + \sigma_{xx}^*, \sigma_{xx}^* \beta_2, \sigma_{xx}^* \beta_3, \phi + \sigma_{xx}^* \beta_2^2, \sigma_{xx}^* \beta_2 \beta_3, \phi + \sigma_{xx}^* \beta_3^2)'$$

os elementos distintos de S e $\Sigma + \Sigma^*$, respectivamente.

Logo, temos que:

$$\sqrt{n}(\tilde{S} - \tilde{\sigma}) \xrightarrow{\mathcal{D}} N_6(0, \Sigma_{\tilde{S}}) ,$$

onde

$$\Sigma_{\tilde{S}} = \tilde{\Lambda} + \tilde{\Lambda}^* \quad (6.52)$$

com

$$\tilde{\Lambda} = \phi^2 \text{Diag}\{2, 1, 1, 2, 1, 2\} \text{ e}$$

$$\tilde{\Lambda}^* = \phi \sigma_{xx}^* \begin{bmatrix} 4 & 2\beta_2 & 2\beta_3 & 0 & 0 & 0 \\ & 1 + \beta_2^2 & \beta_3 \beta_2 & 2\beta_2 & \beta_3 & 0 \\ & & 1 + \beta_3^2 & 0 & \beta_2 & 2\beta_3 \\ & & & 4\beta_2^2 & 2\beta_2 \beta_3 & 0 \\ \text{(simetria)} & & & & \beta_2^2 + \beta_3^2 & 2\beta_2 \beta_3 \\ & & & & & 4\beta_3^2 \end{bmatrix}$$

Agora sabemos que $\tilde{\beta}_3 = S_{23}/S_{12}$. Usando estes resultados e o método delta, temos, depois de alguma álgebra:

$$\sqrt{n}(\tilde{\beta}_3 - \beta_3) \xrightarrow{\mathcal{D}} N(0, \sigma_{\beta_3}^2) \quad (6.53)$$

com

$$\sigma_{\hat{\beta}_3}^2 = \frac{\phi(1 + \beta_3^2)}{\sigma_{xx}^*} (1 + \phi/\sigma_{xx}^* \beta_2^2) .$$

Por outro lado, a partir de (6.35) segue que

$$\sqrt{n}(\hat{\beta}_3 - \beta_3) \xrightarrow{D} N(0, \sigma_{\hat{\beta}_3}^2) \quad (6.54)$$

com

$$\sigma_{\hat{\beta}_3}^2 = \frac{\phi(1 + \beta_3^2)}{S_{xx}} (1 + \phi/cS_{xx}) ,$$

onde, em vez de θ_0 anotamos θ . De (6.53) - (6.54), temos que a eficiência relativa assintótica do EMV $\hat{\beta}_3$ com respeito ao estimador de momentos $\tilde{\beta}_3$ é dada por

$$\begin{aligned} ERA(\hat{\beta}_3/\tilde{\beta}_3) &= \left[\frac{S_{xx}}{\sigma_{xx}^*} \right] \left\{ \frac{1 + \frac{\phi}{\sigma_{xx}^* \beta_2^2}}{1 + \frac{\phi}{cS_{xx}}} \right\} \\ &\rightarrow \left\{ \frac{1 + \phi/\sigma_{xx}^* \beta_2^2}{1 + \phi/c\sigma_{xx}^*} \right\} > 1 \end{aligned} \quad (6.55)$$

pois $c = 1 + \beta_2^2 + \beta_3^2 > \beta_2^2$. Ou seja, o EMV $\hat{\beta}_3$ é mais eficiente que o estimador de momentos $\tilde{\beta}_3$.

O mesmo acontece para $\hat{\beta}_2$ e $\tilde{\beta}_2$. Com efeito, neste caso, temos que

$$\sigma_{\hat{\beta}_2}^2 = \frac{\phi(1 + \beta_2^2)}{\sigma_{xx}^*} (1 + \phi/\sigma_{xx}^* \beta_3^2) \quad (6.56)$$

e

$$\sigma_{\hat{\beta}_2}^2 = \frac{\phi(1 + \beta_2^2)}{S_{xx}} (1 + \phi/cS_{xx}) \quad (6.57)$$

de onde segue que

$$\begin{aligned} ERA(\hat{\beta}_2/\tilde{\beta}_2) &= \left[\frac{S_{xx}}{\sigma_{xx}^*} \right] \left\{ \frac{1 + \frac{\phi}{\sigma_{xx}^* \beta_3^2}}{1 + \frac{\phi}{cS_{xx}}} \right\} \\ &\rightarrow \left\{ \frac{1 + \phi/\sigma_{xx}^* \beta_3^2}{1 + \phi/c\sigma_{xx}^*} \right\} > 1 . \end{aligned} \quad (6.58)$$

6.5 Teste da hipótese da igualdade das linhas de calibração

Para testar se os instrumentos medem sem vício a característica de interesse x , podemos provar a hipótese

$$H : \begin{matrix} \alpha_2 = \alpha_3 = \dots = \alpha_p = 0 \\ \beta_2 = \beta_3 = \dots = \beta_p = 1 \end{matrix} \quad \text{ou} \quad H : \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} .$$

Logo, para testar H , podemos usar a estatística de Wald, dada por (ver (6.35)).

$$W_a = (n/\hat{c}\tilde{\phi})[\hat{\alpha}', (\hat{\beta} - \mathbf{1})'](\hat{K}^{-1} \otimes \hat{B}_{01}) \begin{bmatrix} \hat{\alpha} \\ (\hat{\beta} - \mathbf{1}) \end{bmatrix} \quad (6.59)$$

$$\xrightarrow{D} \chi^2(2(p-1)) \text{ sob } H \quad \text{e}$$

$$\hat{K}^{-1} = \begin{bmatrix} 1 & \bar{y}_1 \\ \bar{y}_1 & \bar{y}_1^2 + \frac{\hat{S}_{xx}}{\hat{d}_0} \end{bmatrix}, \quad \hat{d}_0 = 1 + \tilde{\phi}/\hat{c}\hat{S}_{xx} .$$

É fácil ver que W_a pode ser escrito como

$$W_a = (n/\hat{c}\tilde{\phi}) \left\{ \hat{\alpha}' \hat{B}_{01} \hat{\alpha} + 2\bar{y}_1 \hat{\alpha}' \hat{B}_{01} (\hat{\beta} - \mathbf{1}) + (\bar{y}_1^2 + \frac{\hat{S}_{xx}}{\hat{d}_0}) (\hat{\beta} - \mathbf{1})' \hat{B}_{01} (\hat{\beta} - \mathbf{1}) \right\} \quad (6.60)$$

Notemos que se $p = 2$,

$$W_a = (n/\hat{c}\tilde{\phi}) \left\{ \hat{\alpha}_2^2 + 2\bar{y}_1 \hat{\alpha}_2 (\hat{\beta}_2 - 1) + (\bar{y}_1^2 + \frac{\hat{S}_{xx}}{\hat{d}_0}) (\hat{\beta}_2 - 1)^2 \right\}; \quad (6.61)$$

e se $p = 3$

$$W_a = (n/\hat{c}\tilde{\phi}) \{w_1 + w_2 + w_3\}, \quad (6.62)$$

onde

$$\begin{aligned} w_1 &= \hat{c}(\hat{\alpha}_2^2 + \hat{\alpha}_3^2) - (\hat{\alpha}_2 \hat{\beta}_2 + \hat{\alpha}_3 \hat{\beta}_3)^2; \\ w_2 &= 2\bar{y}_1 [(\hat{\alpha}_2 \hat{\beta}_2 + \hat{\alpha}_3 \hat{\beta}_3)(1 + \hat{\beta}_2 + \hat{\beta}_3) - \hat{c}(\hat{\alpha}_2 + \hat{\alpha}_3)]; \\ w_3 &= \left(\bar{y}_1^2 + \frac{\hat{S}_{xx}}{\hat{d}_0} \right) (3\hat{c} - 1 - (\hat{\beta}_2 + \hat{\beta}_3)(2 + \hat{\beta}_2 + \hat{\beta}_3)). \end{aligned}$$

6.6 Extensão do modelo funcional

No modelo funcional (6.2) assumimos que $\text{Var}(\varepsilon_j) = \phi I_p$, $j = 1, 2, \dots, n$. É possível, no entanto, modificar esta suposição e considerar, Gleser (1981),

$$\text{Var}(\varepsilon_j) = \phi \Sigma_0, \quad j = 1, 2, \dots, n, \quad (6.63)$$

onde Σ_0 é uma matriz conhecida definida positiva. Nesta situação, podemos transformar os dados originais Y_j em

$$Y_j^* = \Sigma_0^{-1/2} Y_j, \quad j = 1, 2, \dots, n; \quad (6.64)$$

onde

$$\Sigma_0^{-1/2} = \begin{pmatrix} t_{11} & 0 \\ T_{21} & T_{22} \end{pmatrix}, \quad \text{com } t_{11} \text{ escalar,}$$

resultando um novo modelo da forma (6.2);

$$Y_j^* = a^* + b^* x_j^* + \varepsilon_j^*, \quad j = 1, 2, \dots, n; \quad (6.65)$$

com

$$a^* = \begin{pmatrix} 0 \\ \alpha^* \end{pmatrix} = \begin{pmatrix} 0 \\ T_{22}\alpha \end{pmatrix}, \quad b^* = \begin{pmatrix} 1 \\ \beta^* \end{pmatrix} = \begin{pmatrix} 1 \\ T_{21}t_{11}^{-1} + T_{22}t_{11}^{-1}\beta \end{pmatrix};$$

$$x_j^* = t_{11}x_j; \quad \varepsilon_j^* = \Sigma_0^{-1/2}\varepsilon_j \quad \text{e} \quad \text{Var}(\varepsilon_j^*) = \phi I_p;$$

$j = 1, 2, \dots, n$.

Um caso interessante, frequentemente utilizado na prática (Kimura, 1992) é quando

$$\Sigma_0 = D(\lambda) = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_p), \quad (6.66)$$

com $\lambda_1, \lambda_2, \dots, \lambda_p$ constantes conhecidas. Aqui

$$\Sigma_0^{-1/2} = \text{Diag} \left\{ \frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \dots, \frac{1}{\sqrt{\lambda_p}} \right\}$$

i.e.

$$t_{11} = \frac{1}{\sqrt{\lambda_1}}, T_{22} = \text{Diag} \left\{ \frac{1}{\sqrt{\lambda_2}}, \dots, \frac{1}{\sqrt{\lambda_p}} \right\} \quad \text{e} \quad T_{21} = 0,$$

de modo que

$$\begin{aligned} x_j^* &= x_j / \sqrt{\lambda_1}, \quad j = 1, 2, \dots, n, \\ \alpha_i^* &= \alpha_i / \sqrt{\lambda_i}, \quad i = 1, 2, \dots, p, \\ \beta_i^* &= \sqrt{\lambda_1} \beta_i / \sqrt{\lambda_i}, \quad i = 2, \dots, n, \end{aligned}$$

e os dados transformados são

$$y_{ij}^* = y_{ij} / \sqrt{\lambda_i}, \quad i = 1, 2, \dots, p, \quad j = 1, 2, \dots, n.$$

6.7 Ilustração numérica

Consideremos os seguintes dados, tomados de Jaech (1985), relativos a um experimento, no qual a densidade de 43 partes cilíndricas de urânio sintetizado (combustível para reatores nucleares) foram medidas por vários métodos. Utilizaremos os dados correspondentes aos métodos 1, 2 e 3.

Vamos supor um modelo funcional normal (6.2) com $n = 43$ e $p = 3$. O vetor de médias e a matriz de covariâncias amostrais são

$$\bar{Y} = \begin{pmatrix} 4.397 \\ 4.370 \\ 4.436 \end{pmatrix} \quad \text{e} \quad S = \begin{pmatrix} 0.04391 & 0.03449 & 0.03933 \\ & 0.04013 & 0.03483 \\ & & 0.06442 \end{pmatrix}$$

Usando o algoritmo descrito na Seção 6.3, obtemos os seguintes EMV

$$\begin{aligned} \hat{\alpha} &= (0.31786, -0.89276) \\ \hat{\beta} &= (0.92157, 1.21191) \quad \text{e} \\ \hat{\phi} &= 0.0083656 \end{aligned}$$

Estas estimativas coincidem com as propostas por Gleser (1981). Com efeito, sabemos que $nS = GDG'$, onde G é uma matriz ortogonal cuja i -ésima coluna é o vetor próprio de

Tabela 6.1: Medidas de densidade de urânio sintetizado, fornecidas por três métodos

Método 1	Método 2	Método 3	Método 1	Método 2	Método 3
4.30	4.06	4.14	4.11	4.23	4.07
4.51	4.35	4.53	3.97	3.98	4.19
4.42	4.52	4.57	4.40	4.36	4.33
4.90	4.67	4.75	4.24	4.30	4.36
4.60	4.70	4.84	4.51	4.52	4.57
4.29	4.30	4.21	4.49	4.33	4.61
4.30	4.47	4.01	4.42	4.14	4.22
4.69	4.59	4.47	4.49	4.59	4.69
4.02	4.05	4.09	4.14	4.32	4.45
4.61	4.78	4.60	4.38	4.25	4.44
4.35	4.25	4.60	4.45	4.60	4.28
4.46	4.55	4.70	4.39	4.38	4.64
4.08	4.12	3.85	4.56	4.35	4.83
3.95	3.95	4.13	4.38	4.40	4.60
4.21	4.27	3.94	4.44	4.34	4.66
4.62	4.47	4.50	4.21	4.24	4.07
4.81	4.81	4.98	4.51	4.52	4.54
4.47	4.34	4.33	4.23	4.12	4.35
4.38	4.24	4.47	4.33	4.47	4.68
4.43	4.37	4.57	4.27	4.21	4.45
4.74	4.52	4.58	4.67	4.52	4.60
4.35	4.37	4.45			

Fonte: Jaech (1985).

nS correspondente ao i -ésimo valor próprio de nS , d_i , $i = 1, 2, 3$, e $D = \text{Diag}\{d_1, d_2, d_3\}$.

Neste caso, temos

$$d_1 = 5.30445, \quad d_2 = 0.76529 \quad \text{e} \quad d_3 = 0.31387;$$

$$G = \begin{pmatrix} g_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} = \begin{pmatrix} 0.548985 & -0.394097 & 0.737091 \\ 0.505928 & -0.545282 & -0.668359 \\ 0.665321 & 0.739834 & -0.099967 \end{pmatrix}$$

Os EMV, propostos por Gleser (1981), são

$$\hat{\beta} = G_{21}/g_{11}, \quad \hat{\alpha} = \bar{Y}_2 - \hat{\beta}\bar{y}_1 \quad \text{e} \quad \hat{\phi} = \frac{1}{3n}(d_2 + d_3),$$

que coincidem com os dados acima.

Também temos que $\tilde{\phi} = 0.0125484$, $\hat{c} = 3.318017$ e $\hat{S}_{xx} = 0.03339675$. Finalmente, usando (6.62), temos que o valor da estatística de Wald, $W_a = 11.38$, com quatro graus de liberdade. Ou seja, ao nível de 5%, concluímos que os métodos 2 e 3 medem com vício a densidade do urânio sintetizado.

6.8 Um estudo de simulação

Nesta seção apresentamos um pequeno estudo de simulação, no qual comparamos os níveis de significância empíricos do teste F , proposto por Kimura (1992), e do teste de Wald dado em (6.59).

1000 amostras foram geradas de acordo com o modelo (6.2), com $p = 3$, $\alpha_2 = \alpha_3 = 0$; $\beta_2 = \beta_3 = 1$ e $\phi = (7.5)^2$ para $n = 50, 100$ e 150 . Usamos três seqüências de parâmetros $x_1, x_2, \dots, x_n$. Uma gerada de uma distribuição $U(0, 10)$, outra de uma $U(0, 50)$ e uma terceira gerada de uma distribuição $U(0, 100)$. Os resultados são mostrados na Tabela 6.2.

Podemos observar que ambos os testes superestimam o nível de significância, especialmente para n pequeno. Para n grande, o teste de Wald é superior na estimação do nível de significância. Note, no entanto, que ambos os testes superestimam bastante o

Tabela 6.2: Níveis de significância para os testes F e de Wald com nível nominal de 5%

Tamanho da amostra (n)	$x_j \sim U(0, 10)$		$x_j \sim U(0, 50)$		$x_j \sim U(0, 100)$	
	Teste de Wald	Teste F	Teste de Wald	Teste F	Teste de Wald	Teste F
50	19.7	43.4	7.1	6.6	6.7	6.0
100	14.8	41.7	5.8	6.3	5.5	6.3
150	14.0	35.2	5.5	5.8	5.0	4.2

nível de significância para a sequência de parâmetros incidentais, $x_1, x_2, \dots, x_n$, gerados da $U(0, 10)$, especialmente o teste F proposto por Kimura (1992). Isto pode ser devido a esta sequência ter variância pequena.

Apêndice A

Distribuições elípticas simétricas

Neste apêndice apresentamos algumas propriedades das distribuições elípticas simétricas multivariadas. A classe de distribuições elípticas tem sido estudada por vários autores, ver Kelker (1970), Muirhead (1980, 1982), Fang et al. (1990), Fang e Zhang (1990) e Fang e Anderson (1990). Ver também Arellano (1994).

As distribuições elípticas podem ser definidas de várias formas. Assumiremos existência de uma função de densidade.

Definição A.1 *Diz-se que um vetor aleatório Y , $p \times 1$, tem uma distribuição elíptica (simétrica) com parâmetro de localização μ , $p \times 1$, e matriz de escala definida positiva Λ , $p \times p$, se sua densidade é da forma*

$$f_Y(y) = |\Lambda|^{-1/2} g[(y - \mu)' \Lambda^{-1} (y - \mu)] , \quad y \in \mathbb{R}^p, \quad (\text{A.1})$$

para alguma função $g : \mathbb{R} \mapsto [0, \infty)$, tal que $\int_0^\infty x^{p-1} g(x^2) dx < \infty$. A função g é chamada função geradora de densidade.

Se Y tem distribuição elíptica com densidade dada por (A.1), escreveremos $Y \sim El_p(\mu, \Lambda, g)$ ou simplesmente $Y \sim El_p(\mu, \Lambda)$. Note que isto não significa que Y tenha uma particular distribuição elíptica, só indica que sua distribuição pertence à classe de distribuições elípticas multivariadas. Se $\mu = 0$ e $\Lambda = I_p$, a distribuição de Y é denominada de esférica.

Exemplos de distribuições elípticas são (Muirhead, 1980):

Exemplo A.1 A distribuição $N_p(\mu, \Lambda)$:

$$f_Y(y) = (2\pi)^{-p/2} |\Lambda|^{-1/2} \exp \left\{ -\frac{1}{2} (y - \mu)' \Lambda^{-1} (y - \mu) \right\} , \quad (\text{A.2})$$

com

$$g(u) = (2\pi)^{-p/2} e^{-u/2} , \quad u \geq 0.$$

Exemplo A.2 A distribuição t p -variada com ν graus de liberdade, denotada por $t_p(\mu, \Lambda, \nu)$:

$$f_Y(y) \equiv \frac{\Gamma \left[\frac{1}{2}(\nu + p) \right]}{\Gamma \left(\frac{\nu}{2} \right) (\nu\pi)^{p/2}} |\Lambda|^{-1/2} \left\{ 1 + \frac{1}{\nu} (y - \mu)' \Lambda^{-1} (y - \mu) \right\}^{-\frac{1}{2}(\nu+p)} , \quad (\text{A.3})$$

com $\nu > 0$. Note que, neste caso,

$$g(u) = \frac{\Gamma \left[\frac{1}{2}(\nu + p) \right]}{\Gamma \left(\frac{\nu}{2} \right) (\nu\pi)^{p/2}} \left(1 + \frac{u}{\nu} \right)^{-\frac{1}{2}(\nu+p)} , \quad u \geq 0.$$

Exemplo A.3 A distribuição normal ε -contaminada p -variada, denotada por $NC_p(\mu; \Lambda, \varepsilon, \phi)$:

$$\begin{aligned} f_Y(y) = & (1 - \varepsilon)(2\pi)^{-p/2} |\Lambda|^{-1/2} \exp \left\{ -\frac{1}{2} (y - \mu)' \Lambda^{-1} (y - \mu) \right\} \\ & + \varepsilon (2\pi)^{-p/2} |\phi \Lambda|^{-1/2} \exp \left\{ -\frac{1}{2\phi} (y - \mu)' \Lambda^{-1} (y - \mu) \right\} , \end{aligned} \quad (\text{A.4})$$

com $\phi > 0$ e $0 \leq \varepsilon \leq 1$. É fácil ver que

$$g(u) = (2\pi)^{-p/2} \left\{ (1 - \varepsilon) e^{-\frac{u}{2}} + \varepsilon \phi^{-\frac{p}{2}} e^{-\frac{u}{2\phi}} \right\} , \quad u \geq 0.$$

Alguns gráficos destas distribuições são mostradas a seguir

$t_2(0, I, 30)$

$N_2(0, I)$

$t_2(0, I, 1)$

$t_2(0, I, 4)$

$NC_2(0, I, 0.05, 0.1)$

$NC_2(0, I, 0.05, 5)$

Agora apresentamos algumas propriedades das distribuições elípticas. A seguir vamos supor que $Y \sim E\ell_p(\mu, \Lambda)$.

Propriedade A.1 A função característica de Y , $\phi_Y(t) = E(e^{it'Y})$, é dada por

$$\phi_Y(t) = e^{it'\mu} \psi(t'\Lambda t), \quad t \in \mathbb{R}^p, \quad (\text{A.5})$$

para alguma função ψ e $i = \sqrt{-1}$. Ver Kelker (1970) e Fang et al. (1990).

Propriedade A.2 Se existem

$$E(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = c_g \Lambda, \quad (\text{A.6})$$

onde c_g é uma constante que depende de g . Em termos da função característica esta constante é $c_g = -2\psi'(0)$.

Kelker (1970) nota que se $u^{\frac{1}{2}(p-k)-1}g(u)$ é integrável, então os k -ésimos momentos de Y existem. Também a variável aleatória $U = (Y - \mu)'\Lambda^{-1}(Y - \mu)$ tem densidade dada por

$$f_U(u) = \left\{ \pi^{p/2} / \Gamma(p/2) \right\} u^{\frac{p}{2}-1} g(u), \quad u \geq 0; \quad (\text{A.7})$$

e em termos de U , $c_g = E(U)/p$.

Exemplo A.4 (i) Se $Y \sim N_p(\mu, \Lambda)$, temos que $c_g = 1$, $E(Y) = \mu$, $\text{Var}(Y) = \Lambda$ e $\psi(x) = e^{-x/2}$, $x \geq 0$.

(ii) Se $Y \sim t_p(\mu, \Lambda, \nu)$, tem-se que $c_g = \nu(\nu - 2)$, ver Apêndice B,

$$E(Y) = \mu \text{ se } \nu > 1,$$

$$\text{Var}(Y) = (\nu/(\nu - 2))\Lambda \text{ se } \nu > 2.$$

A função característica da distribuição t tem forma complicada e foi derivada por Sutradhar (1986).

(iii) Se $Y \sim NC_p(\mu, \Lambda; \varepsilon, \phi)$, temos que $c_g = 1 + \varepsilon(\phi - 1)$, $E(Y) = \mu$, $\text{Var}(Y) = (1 + \varepsilon(\phi - 1))\Lambda$ e $\psi(x) = (1 - \varepsilon)e^{-x/2} + \varepsilon e^{-x\phi/2}$, $x \geq 0$.

Da propriedade A2 segue que todas as distribuições na classe $El_p(\mu, \Lambda)$ tem (supondo que existam) a mesma média μ e a mesma matriz de correlação $\rho = [(\rho_{ij})]$, onde

$$\rho_{ij} = c_g \delta_{ij} / (c_g \delta_{ii} c_g \delta_{jj})^{1/2} = \delta_{ij} / (\delta_{ii} \delta_{jj})^{1/2} ;$$

$i = 1, 2, \dots, p$.

Na prática, no entanto, é comum supor existência de segundo momento e trabalhar com a sub-classe de distribuições com a mesma matriz de covariância. Ver Brown (1984), Muirhead (1982) e Iwashita e Siotani (1994).

Propriedade A.3 Seja $a \in \mathbb{R}^q$, $q \leq p$, A uma matriz $q \times p$ de posto q e $X = a + AY$.

Então, $X \sim El_q(a + A\mu, A\Lambda A')$. Esta propriedade segue diretamente de A1.

Agora partimos Y , μ e Λ como:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{e} \quad \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix},$$

onde Y_1 e μ_1 são $q \times 1$, $q \leq p$ e Λ_{11} é $q \times q$.

Usando novamente a propriedade A1 (ou A3), temos:

Propriedade A.4 $Y_1 \sim El_q(\mu_1, \Lambda_{11}, g)$, i.e., todas as distribuições marginais são também elípticas.

Propriedade A.5 A distribuição condicional de Y_1 dado Y_2 é elíptica q -variada com média e variância dadas por, desde que existam,

$$E(Y_1|Y_2) = \mu_1 + \Lambda_{12}\Lambda_{22}^{-1}(Y_2 - \mu_2) \quad \text{e}$$

$$\text{Var}(Y_1|Y_2) = h(Y_2)(\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}),$$

para alguma função h . Ver Fang et al. (1990).

Kelker (1970) mostra que se a matriz de covariância condicional, de Y_1 dado Y_2 não depende de Y_2 , então Y deve ter uma distribuição normal. Esta propriedade caracteriza a distribuição normal multivariada na classe das distribuições elípticas.

Propriedade A.6 Seja $Y \sim El_p(\mu, \Lambda)$, onde Λ é diagonal. Se os componentes de Y , $y_1, \dots, y_p$ são todos independentes, então Y tem distribuição normal. Ver Muirhead (1982).

Ou seja, na classe de distribuições elípticas, se os componentes do vetor Y são independentes, isto implica que a distribuição de Y é normal.

Sejam γ_1^i e γ_2^i a assimetria e curtose da distribuição de y_i , o i -ésimo componente de Y , $i = 1, 2, \dots, p$. Sabemos que

$$\gamma_1^i = \kappa_3^i / (\kappa_2^i)^{3/2} \quad \text{e} \quad \gamma_2^i = \kappa_4^i / (\kappa_2^i)^2, \quad i = 1, 2, \dots, p, \quad (\text{A.8})$$

onde κ_r^i é o cumulante de ordem r de y_i , $i = 1, \dots, p$. Se y_i tem distribuição normal, todos os cumulantes κ_r^i de ordem $r > 2$ são zero, $i = 1, 2, \dots, p$.

Propriedade A.7 Seja $Y \sim El_p(\mu, \Lambda)$ com matriz de covariância Σ e quartos momentos finitos. Então a distribuição de y_i , $i = 1, 2, \dots, p$, tem assimetria zero ($\gamma_1^i = 0$, $i = 1, 2, \dots, p$) e a mesma curtose

$$\gamma_2^i = 3\kappa, \quad i = 1, 2, \dots, p, \quad (\text{A.9})$$

onde

$$\kappa = \left\{ \frac{\psi''(0)}{\psi'(0)^2} - 1 \right\} = \left\{ \frac{E[(y_i - \mu_i)^2]}{3\sigma_{ii}^2} - 1 \right\}, \quad i = 1, 2, \dots, p.$$

e κ é denominado parâmetro de curtose. Ver Muirhead (1982).

Mardia (1970) define o coeficiente de curtose multivariado, $\beta_{2,p}$, por

$$\beta_{2,p} = E\{(Y - \mu)' \Sigma^{-1} (Y - \mu)\}^2, \quad (\text{A.10})$$

onde Σ é a matriz de covariância de Y . No caso elíptico tem-se, ver Berkane e Bentler (1987), que

$$\kappa + 1 = \beta_{2,p}/p(p+2) . \quad (\text{A.11})$$

Exemplo A.5 (i) Se $Y \sim N_p(\mu, \Lambda)$ tem-se que $\kappa = 0$.

(ii) Se $Y \sim t_p(\mu, \Lambda, \nu)$, então temos que, ver Apêndice B,

$$\kappa = 2/(\nu - 4) \text{ se } \nu > 4.$$

(iii) Se $Y \sim NC_p(\mu, \Lambda; \varepsilon, \phi)$, temos que, ver Berkane e Bentler (1987),

$$\kappa = \frac{1 + \varepsilon(\phi^2 - 1)}{[1 + \varepsilon(\phi - 1)]^2} - 1 .$$

Apêndice B

A distribuição t multivariada

Neste apêndice apresentamos algumas propriedades da distribuição t multivariada usadas neste trabalho.

Definição B.1 Um vetor aleatório p -dimensional $Y = (Y_1, \dots, Y_p)'$ é dito ter uma distribuição t multivariada, com locação $\mu \in \mathbb{R}^p$, matriz de escala Λ , $p \times p$, definida positiva e ν graus de liberdade, se sua densidade é dada por:

$$f_Y(y) = K(\nu, p)(1 + (y - \mu)' \Lambda^{-1}(y - \mu)/\nu)^{-\frac{1}{2}(\nu+p)}, \quad y \in \mathbb{R}^p; \quad (\text{B.1})$$

onde

$$K(\nu, p) = |\Lambda|^{-\frac{1}{2}} \Gamma\left(\frac{\nu + p}{2}\right) / \Gamma\left(\frac{1}{2}\right)^p \Gamma\left(\frac{\nu}{2}\right) \nu^{\frac{p}{2}}, \quad \nu > 0.$$

Se um vetor aleatório Y tem uma densidade dada por (B1) o denotaremos por $Y \sim t_p(\mu, \Lambda, \nu)$.

Recentemente, esta distribuição t , introduzida por Dunnett e Sobel (1954) e Bennett (1961), tem sido bastante utilizada, para ν pequeno, em análise multivariada devido às suas caudas pesadas e, conseqüentemente, os EMV de μ e Λ são resistentes a outliers. Ver Maronna (1976), Rubin (1983), Little (1988), Lange et al. (1989), Berkane e Bentler (1994), Kent et al. (1994). Outros trabalhos interessantes baseados na distribuição t são Zellner (1976), Sutradhar e Ali (1986), Singh (1991), Sutradhar (1990, 1993, 1994) e Arellano (1994).

Como foi mencionado no Apêndice A, a distribuição t pertence à classe das distribuições elípticas, de modo que suas propriedades podem ser obtidas a partir dessa classe. Contudo, devido à importância adquirida pela distribuição t multivariada, neste apêndice apresentamos suas principais propriedades.

Propriedade B.1 Seja $Y = \mu + Z/\sqrt{U}$, onde $Z \sim N_p(0, \Lambda)$ independente de $U \sim \chi^2(\nu)/\nu$. Então, tem-se que $Y \sim t_p(\mu, \Lambda, \nu)$.

Note que dado $U = u$, temos que

$$(Y|U = u) \sim N_p(\mu, u^{-1}\Lambda) . \quad (\text{B.2})$$

Daqui é fácil ver que a densidade de Y é dada por (B.1). Isto é,

$$f_Y(y) = \int_0^\infty |\Lambda|^{-1/2} (2\pi u^{-1})^{-p/2} e^{-\frac{u}{2}(y-\mu)'\Lambda^{-1}(y-\mu)} f_U(u) du ,$$

onde $f_U(u)$ é densidade de U . Ou seja, a distribuição t multivariada pode ser obtida como uma mistura de uma distribuição normal multivariada e uma distribuição χ^2 .

Propriedade B.2 Seja $(Y|U = u) \sim N_p(\mu, u^{-1}\Lambda)$, onde $U \sim \chi^2(\nu)/\nu$, $\nu > 0$. Então, temos que

- (i) $Y \sim t_p(\mu, \Lambda, \nu)$;
- (ii) $(U|Y = y) \sim \chi^2(\nu + p)/(\nu + d^2)$, onde $d^2 = (Y - \mu)'\Lambda^{-1}(Y - \mu)$;
- (iii) $\frac{d^2}{p} \sim F(p, \nu)$.

Note que esta propriedade segue diretamente da propriedade B1.

Propriedade B.3 Seja $Y \sim t_p(\mu, \Lambda, \nu)$. Então,

$$\begin{aligned} E(Y) &= \mu && \text{se } \nu > 1 \quad \text{e} \\ \text{Var}(Y) &= (\nu/(\nu - 2))\Lambda && \text{se } \nu > 2. \end{aligned}$$

Usando a propriedade B1, temos que

$$E(Y) = \mu + E(1/\sqrt{U})E(Z) \quad \text{e} \quad \text{Var}(Y) = E(U^{-1})E(ZZ')$$

e o resultado segue, pois $E(1/\sqrt{U}) = (\frac{\nu}{2})^{1/2}\Gamma(\frac{\nu-1}{2})/\Gamma(\frac{\nu}{2})$ se $\nu > 1$ e $E(U^{-1}) = (\nu/(\nu-2))$ se $\nu > 2$.

Também, usando a propriedade B1, segue a seguinte propriedade:

Propriedade B.4 Seja $Y \sim t_p(\mu, \Lambda, \nu)$ e $a \in \mathbb{R}^q$, $q \leq p$, A uma matriz $q \times p$ de posto q e $X = a + AY$. Então tem-se que

$$X \sim t_q(a + A\mu, A\Lambda A', \nu) .$$

Partamos Y , μ e Λ como

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{e} \quad \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix},$$

onde Y_1 e μ_1 são $q \times 1$, $q \leq p$ e Λ_{11} é $q \times q$.

Propriedade B.5 Seja $Y \sim t_p(\mu, \Lambda, \nu)$. Então, temos que

$$Y_1 \sim t_q(\mu_1, \Lambda_{11}, \nu).$$

Note que esta propriedade segue da propriedade B4.

Propriedade B.6 Seja $Y \sim t_p(\mu, \Lambda, \nu)$. Então tem-se que

$$(Y_1|Y_2 = y_2) \sim t_q(\mu_1 + \Lambda_{12}\Lambda_{22}^{-1}(y_2 - \mu_2), \Lambda_{1.2}, \nu + p - q) ,$$

onde $\Lambda_{1.2} = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}$. Ver Arellano (1994).

Daqui segue que

$$E(Y_1|Y_2 = y_2) = \mu_1 + \Lambda_{12}\Lambda_{22}^{-1}(y_2 - \mu_2) \quad \text{se } \nu + p - q > 1 \quad (\text{B.3})$$

e

$$\text{Var}(Y_1|Y_2 = y_2) = \left[\frac{\nu + (y_2 - \mu_2)'\Lambda_{22}^{-1}(y_2 - \mu_2)}{\nu + p - q - 2} \right] \Lambda_{1.2}, \quad (\text{B.4})$$

se $\nu + p - q > 2$.

Propriedade B.7 Seja $Y \sim t_p(\mu, \Lambda, \nu)$. Então o coeficiente de curtose multivariado é dado por

$$\beta_{2,p} = p(p+2)(\nu-2)/(\nu-4) \text{ para } \nu > 4.$$

Usando (A.10) e a propriedade B1, temos que

$$\beta_{2,p} = ((\nu-2)/\nu)^2 E(U^{-2}) E[(Z'\Lambda^{-1}Z)^2]$$

e o resultado segue, já que $E(U^{-2}) = \nu^2/(\nu-2)(\nu-4)$, $\nu > 4$ e como $Z \sim N_p(0, \Lambda)$, $E[(Z'\Lambda^{-1}Z)^2] = p(p+2)$, ver Graybill (1983).

Agora, como $\kappa + 1 = \beta_{2,p}/p(p+2)$, ver (A.11), segue que

$$\kappa = 2/(\nu-4) \text{ se } \nu > 4. \quad (\text{B.5})$$

Propriedade B.8 Seja $Y \sim t_p(\mu, \Lambda, \nu)$ com densidade $f_Y(y, \nu)$ dada por (B.1) e seja $f_N(y)$ a função de densidade de um vetor aleatório com $\sim N_p(\mu, \Lambda)$. Então,

$$\lim_{\nu \rightarrow \infty} f_Y(y, \nu) = f_N(y), \quad \forall y \in \mathbb{R}^p.$$

Esta propriedade segue diretamente do fato que (Tong, 1990):

$$\lim_{\nu \rightarrow \infty} (1 + \frac{1}{\nu}(y - \mu)'\Lambda^{-1}(y - \mu))^{-\frac{1}{2}(\nu+p)} = \exp \left\{ -\frac{1}{2}(y - \mu)'\Lambda^{-1}(y - \mu) \right\}$$

e que

$$\lim_{\nu \rightarrow \infty} K(\nu, p) = 1/(2\pi)^{p/2} |\Lambda|^{1/2}.$$

Combinando esta propriedade com o teorema de convergência de Scheffé, temos o seguinte fato (Tong, 1990):

Propriedade B.9 Seja $Y \sim t_p(\mu, \Lambda, \nu)$ e $X \sim N_p(\mu, \Lambda)$. Então tem-se que

$$\lim_{\nu \rightarrow \infty} P\{Y \in B\} = P\{X \in B\}$$

para todo subconjunto B Borel-mensurável de $\mathbb{R}^p$. Como caso especial, temos que $Y \xrightarrow{\mathcal{D}} X$, quando $\nu \mapsto \infty$.

Estes fatos mostram que a distribuição normal multivariada pode ser obtida a partir da distribuição t multivariada, fazendo $\nu \mapsto \infty$.

Por outro lado, na prática, é comum supor existência de segundo momento (Sutradhar, 1993) e trabalhar com a distribuição $t_p(\mu, \Lambda, \nu)$ com matriz de covariância dada por, digamos, $\Sigma = \text{Var}(Y)$. Neste caso, $\Lambda = ((\nu - 2)/\nu)\Sigma$ e substituindo Λ em (B.1), temos que

$$f_Y(y) = C(\nu, p) \{(\nu - 2) + (y - \mu)' \Sigma^{-1} (y - \mu)\}^{-\frac{1}{2}(\nu+p)}, \quad (\text{B.6})$$

onde

$$C(\nu, p) = \left[(\nu - 2)^{1/2} \Gamma\left(\frac{\nu + p}{2}\right) \right] / \pi^{p/2} \Gamma\left(\frac{\nu}{2}\right), \quad \nu > 2.$$

Se um vetor Y tem uma função de densidade dada por (B.6), escreveremos $Y \sim T_p(\mu, \Sigma, \nu)$. Ou seja, com esta versão reparametrizada da distribuição t multivariada temos que $E(Y) = \mu$ e $\text{Var}(Y) = \Sigma$, que também permite uma comparação direta com a distribuição $N_p(\mu, \Sigma)$.

Alguns resultados usando a $\sim T_p(\mu, \Sigma, \nu)$ são dados no Apêndice D.

Para mais detalhes e outras propriedades da $\sim t_p(\mu, \Lambda, \nu)$, ver Arellano (1994). Ver também Arellano et al. (1994a).

Apêndice C

Obtenção da matriz de informação no modelo estrutural normal

Neste apêndice obtemos a matriz de informação na parametrização de análise de fatores do modelo estrutural normal. Apresentamos também a matriz de informação perdida. Primeiro, apresentamos dois resultados prévios necessários para obter as matrizes de informação.

Lema C.1 *Sejam $a, b \in \mathbb{R}^p$, vetores de constantes e*

$$g_1 = g_1(\phi) = \ln |D(\phi)|,$$

$$g_2 = g_2(\phi) = a' D^{-1}(\phi) b,$$

$$g_3 = g_3(\phi) = D^{-1}(\phi) a,$$

$$g_4 = g_4(\phi) = D^{-2}(\phi) D(a) b \quad e$$

$$g_5 = g_5(\phi) = c^{-2} D^{-2}(\phi) D(\lambda) \lambda \text{ com } c = 1 + \lambda' D^{-1}(\phi) \lambda.$$

Então,

$$(i) \quad \frac{\partial}{\partial \phi}(g_1) = D^{-1}(\phi) \mathbf{1},$$

$$(ii) \quad \frac{\partial}{\partial \phi}(g_2) = -D^{-2}(\phi) D(a) b,$$

$$(iii) \quad \frac{\partial}{\partial \phi'}(g_3) = -D^{-2}(\phi) D(a),$$

$$(iv) \frac{\partial}{\partial \phi'}(g_4) = -2D^{-3}(\phi)D(a)D(b) \text{ e}$$

$$(v) \frac{\partial}{\partial \phi'}(g_5) = -2c^{-2}D^{-3}(\phi)D(\lambda)D(\lambda) + 2c^{-3}D^{-2}(\phi)D(\lambda)\lambda\lambda'D(\lambda)D^{-2}(\phi).$$

Prova: (i) Note que $g_1(\phi) = \sum_{k=1}^p \ln \phi_k$. Logo

$$\frac{\partial}{\partial \phi}(g_1) = \left[\left(\frac{\partial(g_1)}{\partial \phi_i} \right) \right] = [(1/\phi_i)] = D^{-1}(\phi)\mathbf{1}.$$

(ii) Temos que $g_2(\phi) = \sum_{k=1}^p a_k b_k / \phi_k$, de onde segue que $\frac{\partial}{\partial \phi_i}(g_2) = -a_i b_i \phi_i^{-2}$, $i = 1, 2, \dots, p$, implicando que

$$\frac{\partial}{\partial \phi}(g_2) = \left[\left(\frac{\partial(g_2)}{\partial \phi_i} \right) \right] = -D^{-2}(\phi)D(a)b.$$

(iii) Sabemos que $\frac{\partial}{\partial \phi'}(g_3) = \left[\left(\frac{\partial(g_{3i})}{\partial \phi_j} \right) \right]$. Mas

$$\frac{\partial}{\partial \phi_j}(g_{3i}) = \begin{cases} 0 & \text{se } i \neq j, \\ -a_i \phi_i^{-2} & \text{se } i = j. \end{cases}$$

Assim, temos

$$\frac{\partial}{\partial \phi'}(g_3) = \text{Diag}\{-a_1 \phi_1^{-2}, \dots, -a_p \phi_p^{-2}\} = -D^{-2}(\phi)D(a).$$

(iv) Similar a (iii).

(v) Use a regra do produto, (iv) e o fato que $\frac{\partial}{\partial \phi'}(c) = -D^{-2}(\phi)D(\lambda)\lambda$, que segue de (ii).

Lema C.2 Seja $W \sim N_p(0, V)$, onde $V = \lambda\lambda' + D(\phi)$. Então:

$$(i) E(W'D^{-1}(\phi)\lambda\lambda'D^{-1}(\phi)W) = c(c-1);$$

$$(ii) E(D(W)) = 0;$$

$$(iii) E(W'D^{-1}(\phi)\lambda D(W)) = cD(\lambda);$$

$$(iv) E(D(W)D(W)) = D(\lambda)D(\lambda) + D(\phi), \text{ com } c \text{ como no Lema C1.}$$

Prova: (i) Sabemos que (Graybill, 1983),

$$\begin{aligned} E(W'D^{-1}(\phi)\lambda\lambda'D^{-1}(\phi)W) &= \text{tr}(D^{-1}(\phi)\lambda\lambda'D^{-1}(\phi)V) \\ &= \text{tr}((c-1)D^{-1}(\phi)\lambda\lambda' + D^{-1}(\phi)\lambda\lambda') = \text{ctr}(D^{-1}(\phi)\lambda\lambda') \\ &= c(c-1) . \end{aligned}$$

$$(ii) E(D(W)) = D(E(W)) = D(0) = 0.$$

$$\begin{aligned} (iii) E(W'D^{-1}(\phi)\lambda D(W)) &= E(D(W(W'D^{-1}(\phi)\lambda))) = D(VD^{-1}(\phi)\lambda) \\ &= D(c\lambda) = cD(\lambda). \end{aligned}$$

$$(iv) E(D(W)D(W)) = D(E(W*W)) = D(V) = D(\lambda)D(\lambda) + D(\phi), \text{ onde } W*W = [(w_i^2)].$$

Por outro lado, de (2.50) e (2.53), temos que

$$q_j(\theta|Y_j) = \ell_{0j} + h_j(\theta|Y_j) , \quad j = 1, 2, \dots, n,$$

onde, ver Lema 2.3,

$$\begin{aligned} h_j(\theta|Y_j) &= c_h + \frac{1}{2} \ln c \quad e \\ \ell_{0j} &= -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \{f_{1j} + f_{2j}\}. \end{aligned}$$

com

$$f_{1j} = f_{1j}(\lambda, \phi) = \ln |V| = \ln c + \ln |D(\phi)| \quad (C.1)$$

$$f_{2j} = f_{2j}(\mu, \lambda, \phi) = (Y_j - \mu)' V^{-1} (Y_j - \mu) \quad (C.2)$$

$$= W_j' D^{-1}(\phi) W_j - c^{-1} [\lambda' D^{-1}(\phi) W_j]^2, \quad j = 1, 2, \dots, n, \quad e$$

$$W_j = Y_j - \mu, \quad j = 1, 2, \dots, n.$$

Agora, usando resultados de derivadas de vetores (Nel, 1980) e as partes (i) e (ii) do Lema C1, temos que

$$\frac{\partial}{\partial \mu}(f_{1j}) = 0 ,$$

$$\frac{\partial}{\partial \mu}(f_{2j}) = -2V^{-1}W_j, \quad (C.3)$$

$$\begin{aligned} \frac{\partial}{\partial \lambda}(f_{1j}) &= 2c^{-1}D^{-1}(\phi)\lambda, \\ \frac{\partial}{\partial \lambda}(f_{2j}) &= -2c^{-1}D^{-1}(\phi)W_jW_j'D^{-1}(\phi)\lambda + 2c^{-2}W_j'MW_jD^{-1}(\phi)\lambda, \end{aligned} \quad (C.4)$$

$$\begin{aligned} \frac{\partial}{\partial \phi}(f_{1j}) &= D^{-1}(\phi)\mathbf{1} - c^{-1}D^{-2}(\phi)D(\lambda)\lambda, \\ \frac{\partial}{\partial \phi}(f_{2j}) &= -D^{-2}(\phi)D(W_j)W_j + 2c^{-1}D^{-2}(\phi)D(\lambda)W_jW_j'D^{-1}(\phi)\lambda \\ &\quad - c^{-2}W_j'MW_jD^{-2}(\phi)D(\lambda)\lambda, \end{aligned} \quad (C.5)$$

$j = 1, 2, \dots, n$.

Usando (C.3) - (C.5) podemos obter a função escore dada por

$$S_{cj} = S_{cj}(\theta|Y_j) = \frac{\partial}{\partial \theta}(\ell_{0j}) = \begin{bmatrix} \frac{\partial}{\partial \mu}(\ell_{0j}) \\ \frac{\partial}{\partial \lambda}(\ell_{0j}) \\ \frac{\partial}{\partial \phi}(\ell_{0j}) \end{bmatrix}, \quad (C.6)$$

$j = 1, 2, \dots, n$, já que

$$\frac{\partial}{\partial \mu}(\ell_{0j}) = -\frac{1}{2} \frac{\partial}{\partial \mu}(f_{2j});$$

$$\frac{\partial}{\partial \lambda}(\ell_{0j}) = -\frac{1}{2} \left\{ \frac{\partial}{\partial \lambda}(f_{1j}) + \frac{\partial}{\partial \lambda}(f_{2j}) \right\};$$

$$\frac{\partial}{\partial \phi}(\ell_{0j}) = -\frac{1}{2} \left\{ \frac{\partial}{\partial \phi}(f_{1j}) + \frac{\partial}{\partial \phi}(f_{2j}) \right\}; \quad j = 1, 2, \dots, n.$$

Agora, usando as partes (iii), (iv) e (v) do Lema C1, temos depois de alguma álgebra, a seguinte matriz de informação, observada para a j -ésima observação:

$$J_{0j}(\theta|Y_j) = -\frac{\partial^2}{\partial \theta \partial \theta'}(\ell_{0j}) = - \begin{bmatrix} \frac{\partial^2}{\partial \mu \partial \mu'}(\ell_{0j}) & \frac{\partial^2}{\partial \mu \partial \lambda'}(\ell_{0j}) & \frac{\partial^2}{\partial \mu \partial \phi'}(\ell_{0j}) \\ \text{(simetria)} & \frac{\partial^2}{\partial \lambda \partial \lambda'}(\ell_{0j}) & \frac{\partial^2}{\partial \lambda \partial \phi'}(\ell_{0j}) \\ & & \frac{\partial^2}{\partial \phi \partial \phi'}(\ell_{0j}) \end{bmatrix}, \quad (C.7)$$

$j = 1, 2, \dots, n$, onde

$$\begin{aligned}
\frac{\partial^2}{\partial \mu \partial \mu'}(\ell_{0j}) &= -\frac{1}{2} \left\{ \frac{\partial^2}{\partial \mu \partial \mu'}(f_{1j}) + \frac{\partial^2}{\partial \mu \partial \mu'}(f_{2j}) \right\} \\
\frac{\partial^2}{\partial \mu \partial \lambda'}(\ell_{0j}) &= -\frac{1}{2} \left\{ \frac{\partial^2}{\partial \mu \partial \lambda'}(f_{1j}) + \frac{\partial^2}{\partial \mu \partial \lambda'}(f_{2j}) \right\} \\
\frac{\partial^2}{\partial \mu \partial \phi'}(\ell_{0j}) &= -\frac{1}{2} \left\{ \frac{\partial^2}{\partial \mu \partial \phi'}(f_{1j}) + \frac{\partial^2}{\partial \mu \partial \phi'}(f_{2j}) \right\} \\
\frac{\partial^2}{\partial \lambda \partial \lambda'}(\ell_{0j}) &= -\frac{1}{2} \left\{ \frac{\partial^2}{\partial \lambda \partial \lambda'}(f_{1j}) + \frac{\partial^2}{\partial \lambda \partial \lambda'}(f_{2j}) \right\} \\
\frac{\partial^2}{\partial \lambda \partial \phi'}(\ell_{0j}) &= -\frac{1}{2} \left\{ \frac{\partial^2}{\partial \lambda \partial \phi'}(f_{1j}) + \frac{\partial^2}{\partial \lambda \partial \phi'}(f_{2j}) \right\} \\
\frac{\partial^2}{\partial \phi \partial \phi'}(\ell_{0j}) &= -\frac{1}{2} \left\{ \frac{\partial^2}{\partial \phi \partial \phi'}(f_{1j}) + \frac{\partial^2}{\partial \phi \partial \phi'}(f_{2j}) \right\}, \quad j = 1, 2, \dots, n,
\end{aligned}$$

com

$$\frac{\partial^2}{\partial \mu \partial \mu'}(f_{1j}) = \frac{\partial^2}{\partial \mu \partial \lambda'}(f_{1j}) = \frac{\partial^2}{\partial \mu \partial \phi'}(f_{2j}) = 0 \quad (\text{C.8})$$

$$\frac{\partial^2}{\partial \lambda \partial \lambda'}(f_{1j}) = 2c^{-1} \{V^{-1} - c^{-1}M\} \quad (\text{C.9})$$

$$\frac{\partial^2}{\partial \lambda \partial \phi'}(f_{1j}) = -2c^{-1}V^{-1}D^{-1}(\phi)D(\lambda) \quad (\text{C.10})$$

$$\begin{aligned}
\frac{\partial^2}{\partial \phi \partial \phi'}(f_{1j}) &= -D^{-2}(\phi) + 2c^{-1}D^{-3}(\phi)D(\lambda)D(\lambda) \\
&\quad - c^{-2}D^{-1}(\phi)D(\lambda)MD(\lambda)D^{-1}(\phi)
\end{aligned} \quad (\text{C.11})$$

$$\frac{\partial^2}{\partial \mu \partial \mu'}(f_{2j}) = 2V^{-1} \quad (\text{C.12})$$

$$\frac{\partial^2}{\partial \mu \partial \lambda'}(f_{2j}) = 2c^{-1} \{D^{-1}(\phi)\lambda W'_j D^{-1}(\phi) + (\lambda' D^{-1}(\phi)W_j)[V^{-1} - c^{-1}M]\} \quad (\text{C.13})$$

$$\begin{aligned} \frac{\partial^2}{\partial \mu \partial \phi'}(f_{2j}) &= 2D^{-2}(\phi)D(W_j) - 2c^{-1}D^{-1}(\phi)\lambda W_j' D(\lambda)D^{-2}(\phi) \\ &\quad - 2c^{-1}(\lambda' D^{-1}(\phi)W_j)D^{-2}(\phi)D(\lambda) + 2c^{-2}MW_j\lambda' D(\lambda)D^{-2}(\phi) \end{aligned} \quad (C.14)$$

$$\begin{aligned} \frac{\partial^2}{\partial \lambda \partial \lambda'}(f_{2j}) &= -2c^{-1}D^{-1}(\phi)W_jW_j'D^{-1}(\phi) + 4c^{-2}D^{-1}(\phi)WW_j'M \\ &\quad + 2c^{-2}(W_j'MW_j)(V^{-1} - 3c^{-1}M) + 4c^{-1}MW_jW_j'D^{-1}(\phi) \end{aligned} \quad (C.15)$$

$$\begin{aligned} \frac{\partial^2}{\partial \lambda \partial \phi'}(f_{2j}) &= 2c^{-1}D^{-1}(\phi)W_jW_j'D(\lambda)D^{-2}(\phi) \\ &\quad - 2c^{-2}D^{-1}(\phi)W_jW_j'MD^{-1}(\phi) + 2c^{-1}(W_j'D^{-1}(\phi)\lambda)D^{-2}(\phi)D(W_j) \\ &\quad - 2c^{-2}(W_j'MW_j)D^{-2}(\phi)D(\lambda) + 4c^{-3}(W_j'MW_j)MD^{-1}(\phi)D(\lambda) \\ &\quad - 4c^{-2}MW_jW_j'D(\lambda)D^{-2}(\phi) \end{aligned} \quad (C.16)$$

$$\begin{aligned} \frac{\partial^2}{\partial \phi \partial \phi'}(f_{2j}) &= 2D^{-2}(\phi)D(W_j)D(W_j) \\ &\quad - 4c^{-1}D^{-3}(\phi)D(\lambda)(W_j'D^{-1}(\phi)\lambda)D(W_j) \\ &\quad - 2c^{-1}D^{-2}(\phi)D(\lambda)W_jW_j'D(\lambda)D^{-2}(\phi) \\ &\quad + 2c^{-2}D^{-2}(\phi)D(\lambda)W_jW_j'MD^{-1}(\phi)D(\lambda) \\ &\quad + 2c^{-2}(W_j'MW_j)D^{-3}(\phi)D(\lambda)D(\lambda) \\ &\quad - 2c^{-3}(W_j'MW_j)D^{-1}(\phi)D(\lambda)MD(\lambda)D^{-1}(\phi) \\ &\quad + 2c^{-2}D^{-1}(\phi)D(\lambda)MW_jW_j'D(\lambda)D^{-2}(\phi) \end{aligned} \quad (C.17)$$

$j = 1, 2, \dots, n.$

Assim, a matriz de informação observada total será

$$J_0(\theta|Y) = \sum_{j=1}^n J_{0j}(\theta|Y_j) . \quad (C.18)$$

De forma similar obtemos a matriz de informação observada perdida, por elemento, que é dada por

$$J_{mj}(\theta|Y_j) = -\frac{\partial^2}{\partial\theta\partial\theta'}(h_j(\theta|Y_j)) = J_m(\theta) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & H_{\lambda\lambda} & H_{\lambda\phi} \\ 0 & H_{\phi\lambda} & H_{\phi\phi} \end{bmatrix} \quad (C.19)$$

$j = 1, 2, \dots, n$, onde

$$\begin{aligned} H_{\lambda\lambda} &= -c^{-1}\{V^{-1} - c^{-1}M\} \\ H_{\lambda\phi} &= -c^{-1}V^{-1}D^{-1}(\phi)D(\lambda) \\ H_{\phi\phi} &= -c^{-1}D^{-3}(\phi)D(\lambda)D(\lambda) + \frac{c^{-2}}{2}D(\lambda)D^{-1}(\phi)MD^{-1}(\phi)D(\lambda) \\ H_{\phi\lambda} &= H'_{\lambda\phi} \quad \text{e} \quad V = \lambda\lambda' + D(\phi) . \end{aligned}$$

A matriz de informação observada perdida total é

$$J_m(\theta|Y) = nJ_m(\theta) \quad (C.20)$$

que não depende dos dados $Y = (Y_1, Y_2, \dots, Y_n)$.

Utilizando (C.18) e (C.20), obtemos a matriz de informação completa

$$J_c(\theta|Y) = J_0(\theta|Y) + J_m(\theta|Y) . \quad (C.21)$$

Notemos que a matriz de informação observada perdida por elemento coincide com a matriz de informação esperada perdida por elemento. Isto é, $J_m(\theta) = I_{Fm}(\theta)$, dada em (2.70).

Finalmente, tomando a esperança em (C.7), usando o Lema C2 e resultados conhecidos sobre a distribuição normal multivariada (Muirhead, 1982), temos, depois de algumas simplificações algébricas, a matriz de informação esperada por elemento, dada em (2.69).

Apêndice D

Obtenção da matriz de informação no modelo estrutural t -elíptico

Neste apêndice mostramos como obter a matriz de informação esperada no modelo estrutural t -elíptico, na versão de análise de fatores do modelo. Nos dois lemas a seguir, apresentamos alguns resultados prévios.

Lema D.1 *Seja $W \sim N_p(0, \Sigma)$, com $\Sigma = [(\sigma_{ij})]$, $a \in \mathbb{R}^p$, A uma matriz simétrica $p \times p$, $\mathbf{1}$ um vetor de uns $p \times 1$, J uma matriz de uns $p \times p$, $D(b) = \text{Diag}(b_1, \dots, b_1)$, se b é um vetor $p \times 1$ e $D(B) = \text{Diag}(b_{11}, \dots, b_{pp})$, se B é uma matriz $p \times p$. Então, temos que*

- (i) $E[(a'W)D(W)] = D(\Sigma a),$
- (ii) $E[D(W)D(W)] = D(\Sigma),$
- (iii) $E[D(W)W] = D(\Sigma)\mathbf{1},$
- (iv) $E[(a'W)WW'] = O_{p \times p},$
- (v) $E[(WW')D(W)] = O_{p \times p},$
- (vi) $E[(W'AW)] = \text{tr}(A\Sigma),$
- (vii) $E[(W'AW)(WW')] = 2\Sigma A\Sigma + \Sigma \text{tr}(A\Sigma),$
- (viii) $E[(W'AW)(W'W)] = 2\text{tr}(\Sigma A\Sigma) + \text{tr}(\Sigma)\text{tr}(A\Sigma),$
- (ix) $E[(a'W)(WW')D(W)] = \Sigma D(a)JD(\Sigma) + 2\Sigma D(\Sigma D(a)J),$

(x) $E[D(W)(WW')D(W)] = D(\Sigma)JD(\Sigma) + 2\Sigma * \Sigma$, com $\Sigma * \Sigma = [(\sigma_{ij}^2)]$ o produto de Hadamard entre duas matrizes.

Prova: (i) Notemos que

$$E[(a'W)D(W)] = E[D(WW'a)] = D(E(WW')a) = D(\Sigma a)$$

(ii) e (iii) são similares a (i).

(iv) Seja $Z \sim N_p(0, I)$. Usando o Teorema 10.9.1 em Graybill (1983), temos que $E[(a'Z)(ZZ')]$ $= O_{p \times p}$. Seja $W = T'Z$, onde $\Sigma = T'T$ e T é uma matriz triangular superior. Logo, $W \sim N_p(0, \Sigma)$. Assim,

$$\begin{aligned} E[(a'W)(WW')] &= E[(a'T'Z(T'ZZ'T))] \\ &= E[b'Z(T'ZZ'T)], \text{ com } b = Ta, \\ &= T'E[(b'Z)(ZZ')]T \\ &= O_{p \times p}. \end{aligned}$$

(v) Similar a (iv).

(vi) Ver Teorema 10.4.4 em Graybill (1983).

(vii) Seja $Z \sim N_p(0, I)$. Novamente, usando o Teorema 10.9.1 em Graybill (1983), podemos provar que

$$E(Z'AZ)(ZZ') = 2A + \text{tr}(A)I_p.$$

Com em (iv), seja $W = T'Z$. Logo,

$$\begin{aligned} E[(W'AW)(WW')] &= T\{E[(Z'BZ)(ZZ')]\}T; \quad B = TAT' \\ &= 2\Sigma A\Sigma + \text{tr}(A\Sigma)\Sigma. \end{aligned}$$

(viii) Notemos que

$$\begin{aligned}
E[(W'AW)(W'W)] &= E[\text{tr}\{(W'AW)(W'W)\}] \\
&= E[\text{tr}\{(W'AW)(WW')\}] \\
&= \text{tr}\{E[(W'AW)(WW')]\} \\
&= \text{tr}\{2\Sigma A\Sigma + \Sigma \text{tr}(A\Sigma)\}, \text{ por (vii),} \\
&= 2\text{tr}(\Sigma A\Sigma) + \text{tr}(\Sigma)\text{tr}(A\Sigma).
\end{aligned}$$

Note que se $A = I_p$, então $E[\|W\|^4] = 2\text{tr}(\Sigma\Sigma) + \text{tr}(\Sigma)\text{tr}(\Sigma)$.

(ix) Seja $H = H(W) = (a'W)(WW')D(W) = [(h_{ij})]$, onde

$$h_{ij} = \sum_{k=1}^p a_k W_k W_i W_j^2; \quad i, j = 1, 2, \dots, p.$$

Logo,

$$E(h_{ij}) = \sum_{k=1}^p a_k E(W_k W_i W_j^2), \quad i, j = 1, 2, \dots, p.$$

Notando que se

$$\begin{aligned}
i = j \neq k, \quad E[W_k W_i W_j^2] &= E[W_k W_i^3] = 3\sigma_{ii}\sigma_{ik}; \\
i = j = k, \quad E[W_k W_i W_j^2] &= E[W_i^4] = 3\sigma_{ii}^2; \\
i \neq j, \quad k = i, \quad E[W_k W_i W_j^2] &= E[W_i^2 W_j^2] = 2\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj}; \\
i \neq j, \quad k = j, \quad E[W_k W_i W_j^2] &= E[W_i W_j^3] = 3\sigma_{jj}\sigma_{ij}; \\
i \neq j \neq k, \quad E[W_k W_i W_j^2] &= 2\sigma_{ij}\sigma_{kj} + \sigma_{ki}\sigma_{jj},
\end{aligned}$$

temos que $E(h_{ij}) = 2\sigma_{ij} \sum_{k=1}^p a_k \sigma_{kj} + \sigma_{jj} \sum_{k=1}^p a_k \sigma_{ik}$. Logo,

$$\begin{aligned}
E[H] &= [(E(h_{ij}))] \\
&= 2 \left[\left(\sigma_{ij} \sum_{k=1}^p a_k \sigma_{kj} \right) \right] + \left[\left(\sigma_{jj} \sum_{k=1}^p a_k \sigma_{ik} \right) \right] \\
&= 2\Sigma D(\Sigma D(a)J) + \Sigma D(a)JD(\Sigma) .
\end{aligned}$$

(x) Notemos que $D(W)WW'D(W) = [(W_i^2 W_j^2)]$. Agora, como

$$E(W_i^2 W_j^2) = 2\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj}, \quad i, j = 1, 2, \dots, p,$$

temos que

$$\begin{aligned} E[(D(W)WW'D(W))] &= [E(W_i^2 W_j^2)] \\ &= [(\sigma_{ii}\sigma_{jj}) + 2(\sigma_{ij}^2)] \\ &= [(\sigma_{ii}\sigma_{jj})] + 2[(\sigma_{ij}^2)] \\ &= D(\Sigma)JD(\Sigma) + 2\Sigma * \Sigma. \end{aligned}$$

Lema D.2 *Seja $W \sim T_p(0, \Sigma, \nu)$, $r = (\nu - 2) + W'\Sigma^{-1}W$, a , A , 1 , $D(b)$ e $D(B)$ como no Lema D1. Então, tem-se que*

- 1 (i) $E(r^{-1}) = \nu/(\nu - 2)(\nu + p)$
- (ii) $E(r^{-1}W) = O_{p \times 1}$
- (iii) $E[r^{-1}WW'] = (\nu + p)^{-1}\Sigma$
- (iv) $E[r^{-1}(a'W)D(W)] = (\nu + p)^{-1}D(\Sigma a)$
- (v) $E[r^{-1}D(W)D(W)] = (\nu + p)^{-1}D(\Sigma)$
- (vi) $E[r^{-1}D(W)W] = (\nu + p)^{-1}D(\Sigma)1$
- (vii) $E[r^{-1}(W'AW)] = (\nu + p)^{-1}\text{tr}(A\Sigma)$
- (viii) $E[r^{-1}(W'AW)(WW')] = [(\nu - 2)/(\nu + 2)(\nu + p)]\{2\Sigma A\Sigma + \Sigma\text{tr}(A\Sigma)\}$
- (ix) $E[r^{-1}(W'AW)(W'W)] = [(\nu - 2)/(\nu + 2)(\nu + p)]\{2\text{tr}(\Sigma A\Sigma) + \text{tr}(\Sigma)\text{tr}(A\Sigma)\}$

- 2 (i) $E[r^{-2}] = \nu(\nu + 2)/(\nu - 2)^2(\nu + p)(\nu + p + 2)$
- (ii) $E[r^{-2}W] = O_{p \times 1}$
- (iii) $E[r^{-2}WW'] = [\nu/(\nu - 2)(\nu + p)(\nu + p + 2)]\Sigma$
- (iv) $E[r^{-2}(a'W)(WW')] = O_{p \times p}$
- (v) $E[r^{-2}(WW')D(W)] = O_{p \times p}$
- (vi) $E[r^{-2}(W'AW)] = [\nu/(\nu - 2)(\nu + p)(\nu + p + 2)]\text{tr}(A\Sigma)$
- (vii) $E[r^{-2}D(W)W] = [\nu/(\nu - 2)(\nu + p)(\nu + p + 2)]D(\Sigma)\mathbf{1}$
- (viii) $E[r^{-2}(W'AW)(WW')] = [(\nu + p)(\nu + p + 2)]^{-1}\{2\Sigma A\Sigma + \Sigma \text{tr}(A\Sigma)\}$
- (ix) $E[r^{-2}(W'AW)(W'W)] = [(\nu + p)(\nu + p + 2)]^{-1}\{2\text{tr}(\Sigma A\Sigma) + \text{tr}(\Sigma)\text{tr}(A\Sigma)\}$
- (x) $E[r^{-2}(a'W)(WW')D(W)] = [(\nu + p)(\nu + p + 2)]^{-1}\{\Sigma D(a)JD(\Sigma) + 2\Sigma D(\Sigma D(a)J)\}$
- (xi) $E[r^{-2}D(W)WW'D(W)] = [(\nu + p)(\nu + p + 2)]^{-1}\{D(\Sigma)J(\Sigma) + 2\Sigma * \Sigma\}$

Prova: Seja $H = H(W)$ uma função de W .

1. Um cálculo direto, como em Sutradhar (1993), produz

$$E[r^{-1}H(W)] = [\nu/(\nu - 2)(\nu + p)]E[H(W^*)] ,$$

onde W^* tem uma distribuição t com locação 0, matriz de escala $\Sigma^* = [(\nu - 2)/(\nu + 2)]\Sigma$ e $\nu + 2$ graus de liberdade, isto é, $W^* \sim t_p(0, \Sigma^*, \nu + 2)$.

Por outro lado, sabemos que se $W^* = W/\sqrt{u}$, onde $W \sim N_p(0, \Sigma^*)$, $u \sim \chi^2(\nu + 2)/(\nu + 2)$ independentes, então $W^* \sim t_p(0, \Sigma^*, \nu + 2)$. Assim, se consideramos, por exemplo, (v), $H(W) = D(W)D(W)$,

$$\begin{aligned} E[r^{-1}D(W)D(W)] &= [\nu/(\nu - 2)(\nu + p)]E[D(W^*)D(W^*)] \\ &= (\nu/(\nu - 2)(\nu + p))E(1/u)E[D(W)D(W^*)] , \end{aligned}$$

onde, agora, $W \sim N(0, \Sigma^*)$. Pelo Lema D1, temos que

$$E[D(W)D(W)] = D(\Sigma^*) = [(\nu - 2)/(\nu + 2)]D(\Sigma) .$$

Além disso, temos que $E(1/u) = (\nu + 2)/\nu$. Logo,

$$\begin{aligned} E[r^{-1}D(W)D(W)] &= [\nu/(\nu - 2)(\nu + p)][(\nu + 2)/\nu][(\nu - 2)/(\nu + 2)]D(\Sigma) \\ &= (\nu + p)^{-1}D(\Sigma) . \end{aligned}$$

Os outros itens são provados da mesma forma.

2. Neste caso, temos que (Sutradhar, 1993):

$$E[r^{-2}H(W)] = [\nu(\nu + 2)/(\nu - 2)^2(\nu + p)(\nu + p + 2)]E(H(W^*)) ,$$

onde, agora, $W^* \sim t_p(0, \Sigma^*, \nu + 4)$, com $\Sigma^* = [(\nu - 2)/(\nu + 4)]\Sigma$.

Assim, escrevendo W^* como mistura de uma distribuição normal com uma χ^2 , como na parte 1, e usando o Lema D1, o resultado segue.

Agora, o logaritmo da função de verossimilhança observada para o modelo estrutural t -elíptico (4.35) é dado por

$$\ell(\theta, \nu) = \sum_{j=1}^n \ell_j(\theta, \nu) , \quad (D.1)$$

onde

$$\ell_j(\theta, \nu) = \ell_j = \ln C(\nu, p) - \frac{1}{2} \{f_{1j} + (\nu + p) \ln(\nu - 2 + f_{2j})\}$$

com $C(\nu, p)$ como em (4.1), f_{1j} e f_{2j} como em (C.1) e (C.2), respectivamente, $j = 1, 2, \dots, n$.

Seja

$$\begin{aligned} r_j &= (\nu - 2) + f_{2j} \\ &= (\nu - 2) + W_j' V^{-1} W_j , \quad 1, 2, \dots, n, \end{aligned} \quad (D.2)$$

onde $W_j = Y_j - \mu \sim T_p(0, V, \nu)$, $j = 1, 2, \dots, n$.

Usando resultados de derivadas de vetores (Nel, 1980) e o Lema C1, obtemos a seguinte matriz de informação observada, por elemento:

$$J_{0j}(\theta, \nu/Y_j) = - \begin{bmatrix} \frac{\partial^2}{\partial \mu \partial \mu'}(\ell_j) & \frac{\partial^2}{\partial \mu \partial \lambda'}(\ell_j) & \frac{\partial^2}{\partial \mu \partial \phi'}(\ell_j) & \frac{\partial^2}{\partial \mu \partial \nu}(\ell_j) \\ & \frac{\partial^2}{\partial \lambda \partial \lambda'}(\ell_j) & \frac{\partial^2}{\partial \lambda \partial \phi'}(\ell_j) & \frac{\partial^2}{\partial \lambda \partial \nu}(\ell_j) \\ & & \frac{\partial^2}{\partial \phi \partial \phi'}(\ell_j) & \frac{\partial^2}{\partial \phi \partial \nu}(\ell_j) \\ \text{(simetria)} & & & \frac{\partial^2}{\partial \nu^2}(\ell_j) \end{bmatrix} \quad (\text{D.3})$$

$j = 1, 2, \dots, n$, onde

$$\begin{aligned} \frac{\partial^2}{\partial \mu \partial \tau'}(\ell_j) &= -\frac{1}{2}(\nu + p) \left\{ r_j^{-1} \frac{\partial^2}{\partial \mu \partial \tau'}(f_{2j}) - r_j^{-2} \frac{\partial}{\partial \mu}(f_{2j}) \frac{\partial}{\partial \tau'}(f_{2j}) \right\} \\ \tau &= \mu, \lambda \quad \text{e} \quad \phi; \end{aligned}$$

$$\frac{\partial^2}{\partial \mu \partial \nu}(\ell_j) = [r_j^{-1} - (\nu + p)r_j^{-2}]V^{-1}W_j;$$

$$\begin{aligned} \frac{\partial^2}{\partial \lambda \partial \tau'}(\ell_j) &= -\frac{1}{2} \frac{\partial^2}{\partial \lambda \partial \tau'}(f_{1j}) - \frac{1}{2}(\nu + p) \left\{ r_j^{-1} \frac{\partial^2}{\partial \mu \partial \tau'}(f_{2j}) - r_j^{-2} \frac{\partial}{\partial \lambda}(f_{2j}) \frac{\partial}{\partial \tau'}(f_{2j}) \right\} \\ \tau &= \lambda \quad \text{e} \quad \phi; \end{aligned}$$

$$\frac{\partial^2}{\partial \lambda \partial \nu}(\ell_j) = -\frac{1}{2}[r_j^{-1} - (\nu + p)r_j^{-2}] \frac{\partial}{\partial \lambda}(f_{2j});$$

$$\frac{\partial^2}{\partial \phi \partial \phi'}(\ell_j) = -\frac{1}{2} \frac{\partial^2}{\partial \phi \partial \phi'}(f_{1j}) - \frac{1}{2}(\nu + p) \left\{ r_j^{-1} \frac{\partial^2}{\partial \phi \partial \phi'}(f_{2j}) - r_j^{-2} \frac{\partial}{\partial \phi}(f_{2j}) \frac{\partial}{\partial \phi'}(f_{2j}) \right\};$$

$$\frac{\partial^2}{\partial \phi \partial \nu}(\ell_j) = -\frac{1}{2}[r_j^{-1} - (\nu + p)r_j^{-2}] \frac{\partial}{\partial \phi}(f_{2j}) \quad \text{e}$$

$$\begin{aligned} \frac{\partial^2}{\partial \nu^2}(\ell_j) &= [(\nu - 4)/2(\nu - 2)^2] + \frac{1}{4}\psi' \left(\frac{\nu + p}{2} \right) - \frac{1}{4}\psi' \left(\frac{\nu}{2} \right) \\ &\quad - \frac{1}{2}[2r_j^{-1} - (\nu + p)r_j^2], \quad j = 1, 2, \dots, n, \end{aligned}$$

com $\frac{\partial}{\partial \tau}(f_{2j})$, $\tau = \mu, \lambda$ e ϕ dada em (C.3), (C.4) e (C.5), respectivamente, $\frac{\partial^2}{\partial \lambda \partial \tau'}(f_{1j})$, $\tau = \lambda$ e ϕ , $\frac{\partial^2}{\partial \phi \partial \phi'}(f_{1j})$ como em (C.9), (C.10) e (C.11), $\frac{\partial^2}{\partial \mu \partial \tau'}(f_{2j})$, $\tau = \mu, \lambda$ e ϕ , $\frac{\partial^2}{\partial \lambda \partial \tau'}(f_{2j})$, $\tau = \lambda$, ϕ e $\frac{\partial^2}{\partial \phi \partial \phi'}(f_{2j})$ como definidas em (C.12) a (C.17), $j = 1, 2, \dots, n$.

A matriz de informação observada total é dada por

$$J_0(\theta, \nu|Y) = \sum_{j=1}^n J_{0j}(\theta, \nu|Y_j) . \quad (D.4)$$

Tomando a esperança em (D.4) e aplicando o Lema D2, temos, depois de bastante álgebra para simplificar, a matriz de informação esperada, $I_F(\theta, \nu)$, dada em (4.54).

REFERÊNCIAS

- Anderson, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical and Statistics*, 22, 327-351.
- Anderson, T.W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *The Annals of Statistics*, 1, 135-141.
- Arellano, R. (1994). *Distribuições Elípticas: Propriedades, Inferência e Aplicações a Modelos de Regressão*, Tese, IME-USP.
- Arellano, R. and Bolfarine, H. (1995). A note on the simple structural regression model. *Annals of the Institute of Statistical Mathematics*, to appear.
- Arellano, R. and Bolfarine, H. (1994). On score tests in structural regression models. *Relatório Técnico*, RT-MAE 9425, IME-USP.
- Arellano, R., Bolfarine, H. and Iglesias, P. (1994a). A predictivistic interpretation of the multivariate t distribution. *Test*, 3, 221-236.
- Arellano, R., Bolfarine, H. and Vilca, F. (1994b). Ultrastructural elliptical models. *Relatório Técnico*, RT-MAE 9424, IME-USP.
- Barnett, V. D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics*, 25, 129-142.
- Barnett, V. and Lewis, T. (1994). *Outliers in statistical data*. 3rd Edition. John Wiley.
- Behre, H.M., Nashan, D. and Nieschlag, M. (1989). Objective measurement of testicular volume by ultrasonography: Evaluation of the technique and comparison with orchidometer estimates. *International Journal of Andrology*, 12, 395-403.
- Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley. New York.
- Bennett, B.M. (1961). On a certain multivariate non-normal distribution. *Proceedings of the Cambridge Philosophical Society*, 57, 434-436.

- Berkane, M. and Bentler, P.M. (1987). Characterizing parameters of multivariate elliptical distributions. *Communications in Statistics- Simulation and Computation*, 16, 193-198.
- Berkane, M., Kano, Y. and Bentler, P.M. (1994). Pseudo maximum likelihood estimation in elliptical theory: Effects of misspecification. *Computational Statistics & Data Analysis*, 18, 255-267.
- Bickel, P.J. and Doksum, K.A. (1977). *Mathematical Statistics*, Holden-Day, Inc. San Francisco.
- Blackwood, L.G. and Bradley, E.L. (1991) An omnibus test for comparing two measuring devices. *Journal of Quality Technology*, 23, 12-16.
- Bolfarine, H. and Arellano, R. (1994). Robust modelling in measurement error models using the t distribution. *Revista Brasileira de Probabilidade e Estatística*, 1, 67-84.
- Bolfarine, H. and Galea, M. (1995a). Structural comparative calibration using the EM algorithm. *Journal of Applied Statistics*, 22, 277-292.
- Bolfarine, H. and Galea, M. (1995b). Maximum likelihood estimation of simultaneous pairwise linear structural relationships. *Biometrical Journal*, 37, 673-689.
- Bolfarine, H., Nascimento, J. and Galea, M. (1993). Maximum likelihood estimation of linear structural relationships via the EM algorithm. *Relatório Técnico*, RT-MAE 9319, IME-USP.
- Bradley, R.A. and Blackwood, L.G. (1989). Comparing paired data: a simultaneous test for means and variances. *American Statistician*, 43, 234-235.
- Brindley, D.A. and Bradley, R.A. (1985). Some new results on Grubb's estimators. *Journal of the American Statistical Association*, 80, 711-714.
- Brown, G.H. (1978). Generalized least squares applied to the linear ultra-structural model. *Biometrika*, 65, 441-444.
- Browne, M.W. (1984). Asymptotical distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83

- Carter, R. (1981). Restricted maximum likelihood estimation of bias and reliability in the comparison of several measuring methods. *Biometrics*, 37, 733-741.
- Chan, L.K. and Mak, T.K. (1979). Maximum likelihood estimation of a linear structural relationship with replications. *Journal of the Royal Statistical Society*, B, 41, 263-268.
- Chatterjee, S. Hadi, A.S. (1988). *Sensitivity Analysis in Linear Regression*, John Wiley. New York.
- Cheng, C.L. and Van Ness, J.W. (1990). Bounded influence errors- in- variables regression. *Contemporary Mathematics*, 112, 227-241.
- Choi, S.C. and Wette, R. (1972). A test for the homogeneity of variances among correlated variables. *Biometrics*, 28, 589-592.
- Christensen, R. and Blackwood, L. (1993). Tests for precision and accuracy of multiple measuring devices. *Technometrics*, 35, 411-420.
- Cochran, W.G. (1968). Errors of measurement in statistics. *Technometrics*, 10, 637-666.
- Cook, R.D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society*, B, 48, 133-169.
- Cook, R.D. and Weisberg, S. (1982). *Residuals and Infuence in Regression*, Chapman and Hall. London.
- Cox, N.R. (1976). The linear structural relation for several groups of data. *Biometrika*, 63, 231-237.
- Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statitics*, Chapman and Hall. London.
- Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society*, B, 49, 1-39.
- Das Gupta, S., Eaton, M.L, Olkin, I., Perlman, M.D., Savage, L.J. and Sobel, M. (1972). Inequalities on the probability content of convex regions for elliptical contoured distributions. *Proceeding of the 6th Berkeley Symposium*

on *Mathematical Statistics and Probability*, 2, 241-265.

Dempster, A., Laird, N., and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, B*, 39, 1-38.

Devlin, S.J., Gnanadesikan, R. and Kettenring, J.R. (1976). Some multivariate applications of elliptical distributions. *Essays in probability and Statistics*, Chapter 24, 365-393. Shinko Tsusho Co. Tokyo.

Dolby, G.R. (1976). The ultrastructural relation: A synthesis of the functional and structural relations. *Biometrika*, 63, 39-50.

Draper, N. and Guttman, I. (1975). Two simultaneous measurement procedures: A bayesian approach. *Journal of the American Statistical Association*, 70, 43-46.

Draper, N. and Guttman, I. (1976). K simultaneous measurement procedures: A bayesian approach. *Journal of the American Statistical Association*, 71, 605-607.

Dunn, G. (1992). *Design and Analysis of Reliability: The statistical evaluation of measurement errors*. Edward Arnold. New York.

Dunnett, C. W. and Sobel, M. (1954). A bivariate generalization of student's t-distribution, with tables for certain special cases. *Biometrika*, 41, 153-169.

Fang, K.T. and Anderson, T.W. (1990). *Statistical Inference in Elliptical Contoured and Related Distributions*, Allerton Press Inc.. New York.

Fang, K.T., Kotz, S. and Ng, K.W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall. London.

Fang, K.T. and Zhang, Y.T. (1990). *Generalized Multivariate Analysis*, Springer-Verlag. London.

Fuller, W.A. (1987). *Measurement error models*. Wiley, New York.

Gabrielsen, A. (1978). Consistency and identifiability. *Journal of Econometrics*, 8, 261-263.

Gleser, L.J. (1981). Estimation in a multivariate "error in variables" regres-

- sion model: Large sample results. *The Annals of Statistics*, 9, 24-44.
- Gleser, L.J. (1983). Functional, structural and ultrastructural errors-in-variables models. *Proceedings of the Business and Economic Statistic Section*, American Statistical Association, 57-66.
- Graybill, F.A. (1983). *Matrices with applications in statistics*. Second Edition, Wadsworth.
- Grubbs, F. E. (1948). On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association*, 43, 243-264.
- Grubbs, F. E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics* 15, 53-66.
- Grubbs, F.E. (1983). Grubbs' estimators. *Encyclopedia of Statistical Sciences*, 3, 542-549.
- Han, C. (1968). Testing the homogeneity of a set of correlated variances. *Biometrika*, 55, 317-326.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics*, John Wiley. New York.
- Hawkins, D.M. (1980). *Identification of Outliers*, Chapman and Hall. London.
- Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *The Annals of Mathematics and Statistics*, 42, 1977-1991.
- Huber, P.J. (1981). *Robust Statistics*, John Wiley. New York.
- Iwashita, T. and Siotani, M. (1994). Asymptotic distributions of functions of a sample covariance matrix under the the elliptical distribution. *The Canadian Journal of Statistics*, 22, 273-283.
- Jaech, J.L. (1964). A program to estimate measurement errors in nondestructive avaluation of fuel element quality. *Technometrics*, 6, 293-300.
- Jaech, J. L. (1985). *Statistical analysis of measurement errors*. Exxon Mono-

graphs. John Wiley, N.Y.

Johnson, N. and Kotz, S. (1970). *Distributions in statistics: continuous univariate distributions - 2*. New York. Houghton Mifflin.

Kaaks, R., Riboli, E., Estive, J., van Kappel and van Staveren (1994). Estimating the accuracy of questionnaire assessments: Validation in terms of structural equation models. *Statistics in Medicine*, 13, 127-142.

Kano, Y., Berkane, M. and Bentler, P.M. (1993). Statistical inference based on pseudo maximum likelihood in elliptical population. *Journal of the American Statistical Association*, 84, 881-896.

Kelker, D. (1970). Distribution theory of spherical distributions and a location scale parameter generalization. *Sankhya A*, 32, 419-430.

Kelly, G. (1984). The influence function in the errors in variables problem. *The Annals of Statistics*, 12, 87-100.

Kelly, G. (1985). Use of the structural equations model in assessing the reliability of a new measurement technique. *Applied Statistics*, 34, 258-263.

Kendall, M.G. and Stuart, A. (1979). *The Advanced Theory of Statistics*, 2, 4th ed. Hafner. New York.

Kent, J.T., Tyler, D.E. and Vardi, Y. (1994). A curious likelihood identity for the multivariate t distribution. *Communications in Statistics- Simulation and Computation*, 23, 441-453.

Kimura, D.K. (1992). Functional comparative calibration using an EM algorithm. *Biometrics*, 48, 1263-1271.

Lange, K.L., Little, R.J. and Taylor, J. (1989). Robust statistical modelling using the t -distribution. *Journal of the American Statistical Association*, 84, 881-896.

Lawley, D. N. and Maxwell, A. E. (1971). *Factor analysis as a statistical method*, 2nd ed. Butterworths, London.

Leurgans, S. (1980). Evaluating laboratory measurement techniques. In *Biostatistics Casebook*. Edited by Miller, R., Efron, B., Brown, B. and Moses,

L. John Wiley.

Little, R.J. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics*, 37, 23-38.

Little, R.J. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*, John Wiley. New York.

Liu, C. and Rubin, D.B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, 81, 633-648.

Louis, T.A. (1982). Finding the observed information when using the EM algorithm. *Journal of the Royal Statistical Society, B*, 44, 226-233.

Madansky, A. (1959). The fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association*, 54, 173-205.

Magnus, J.R. and Neudecker, H. (1979). The commutation matrix: Some properties and applications. *The Annals of Statistics*, 7, 381-394.

Mak, T.K. (1982). Estimation in the presence of incidental parameters. *The Canadian Journal of Statistics*, 10, 121-132.

Maloney, C.J. and Rastogi, S.C. (1970). Significance test for grubbs' estimators. *Biometrics*, 26, 671-676.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519-530.

Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics*, 4, 51-67.

Meng, X.L. and Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80, 267-278.

Muirhead, R. (1980). The effects of elliptical distributions on some standard procedures involving correlation coefficients, in *Multivariate Statistical Analysis* (ed. R.P. Gupta), North-Holland, 143-159.

Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley. New York.

- Nel, D.G. (1980). On matrix differentiation in Statistics. *South African Statistical Journal*, 14, 137-193.
- Orchad, T. and Woodbury, M.A. (1972). A missing information principle: Theory and applications. *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 697-715.
- Patefield, W.M. (1977). On the information matrix in the linear functional relationship problem. *Applied Statistics*, 26, 69-70.
- Patefield, W.M. (1978). The unreplicated ultrastructural relation: Large sample properties. *Biometrika*, 65, 535-540.
- Peña, D. and Yohai, V. (1995). The detection of influential subsets in linear regression by using an influence matrix. *Journal of the Royal Statistical Society, B*, 57, 145-156.
- Philippou, A.N. and Roussas, G.G. (1975). Asymptotic normality of the maximum likelihood estimate in the independent not identically case. *Annals of the Institute of Statistical Mathematics*, 27, 45-55.
- Pitman, E.G. (1939). A note on normal correlation. *Biometrika*, 31, 9-12.
- Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*, John Wiley. New York.
- Rubin, D.B. (1983). Iteratively reweighted least squares. *Encyclopedia of the Statistical Sciences*, 4, 272-275.
- Savin, N.E. (1980). The bonferroni and the scheffé multiple procedures. *Review of Economic Studies*, XLVII, 255-273.
- Seber, G.A. (1977). *Linear Regression Analysis*, John Wiley. New York.
- Sen, P.K. and Singer, J.M. (1993). *Large Sample Methods in Statistics: An Introduction With Applications*, Chapman & Hall. New York.
- Shyr, J.Y. and Gleser, L.J. (1986). Inference about comparative precision in linear structural relationships. *Journal of the Statistical Planning and Inference*, 14, 339-358.
- Sidák, Z. (1968). On multivariate normal probabilities of rectangles : Thier

dependence on the correlations. *The Annals of Mathematics and Statistics*, 39, 1425-1434.

Singh, R.S. (1991). James-Stein rule estimators in linear regression models with multivariate-t distributed error. *The Australian Journal of Statistics*, 33, 145-158.

Staudte, R.G. and Sheather, S.J. (1990). *Robust Estimation and Testing*, John Wiley. New York.

Sutradhar, B.C. (1986). On the characteristic function of multivariate student t distribution. *The Canadian Journal of Statistics*, 14, 329-337.

Sutradhar, B.C. (1990). Discrimination of observations into of two t populations. *Biometrics*, 46, 827-835.

Sutradhar, B.C. (1993). Score tests for the covariance matrix of the Elliptical t-distribution. *Journal of Multivariate Analysis*, 46, 1-12.

Sutradhar, B.C. (1994). On cluster regression and factor analysis models with elliptical t errors. *Lecture notes - Monograph series, Multivariate Analysis and Its Applications*, 24, 369-383.

Sutradhar, B.C. and Ali, M.M. (1986). Estimation of the parameter of a regression model with a multivariate t error variable. *Communications in Statistics - Theory and Methods*, 15, 429-450.

Taylor, J. (1992). Properties of modelling the error distribution with an extra shape parameter. *Computational Statistics & Data Analysis*, 13, 33-46.

Theobald, C. M. and Mallison, J. R. (1978). Comparative calibration, linear structural relationship and congeneric measurements. *Biometrics*, 34, 39-45.

Tong, Y.L. (1990). *The multivariate Normal Distribution*, Springer-Verlag. New York.

Wellman, J.M. and Gunst, R.F. (1991). Influence diagnostics for linear measurement error models. *Biometrika*, 78, 373-380.

Williams, E.J. (1969). Regression methods in calibration problems. *Bulletin of the International Statistical Institute*, 43, 17-28.

Wong, M.Y. (1989). Likelihood estimation of a simple linear regression model when both variables have error. *Biometrika*, 76, 141-148.

Zamar, R.H. (1989). Robust estimation in the errors-in-variables model. *Biometrika*, 76, 149-160.

Zellner, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student t-error term. *Journal of the American Statistics Association*, 71, 400-405.