Testes para componentes de variância em modelos lineares mistos: uma abordagem com valor-s

Jônatas de Oliveira Alves

Dissertação apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Programa: Estatística Orientador: Prof^o. Dr^o. Alexandre Galvão Patriota

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro do CNPq

São Paulo Julho de 2021

Testes para componentes de variância em modelos lineares mistos: uma abordagem com valor-s

Jônatas de Oliveira Alves

Esta é a versão original da dissertação elaborada pelo candidato Jônatas de Oliveira Alves, tal como submetida à Comissão Julgadora. Autorizo a reprodução e divulgação total ou parcial deste trabalho por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Dedico este trabalho À Deus que nos momentos mais difíceis me restaurou o ânimo e me inspirou com boas ideias

Resumo

Jônatas de Oliveira Alves. **Testes para componentes de variância em modelos lineares mistos: uma abordagem com valor-s**. Dissertação (Mestrado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2021.

Neste trabalho, estudamos a medida de evidência 'valor-s', desenvolvida por Patriota (2013), em testes envolvendo algumas componentes de variância iguais a zero em modelos lineares mistos. O valor-s é construído por meio de regiões de confiança assintóticas baseadas na razão de verossimilhanças. Um dos objetivos desta dissertação consiste em apresentar a distribuição assintótica da estatística razão de verossimilhanças para hipóteses simples em que o parâmetro está na borda do espaço paramétrico e, a partir dela, obter a medida de evidência assintótica valor-s. Por fim comparamos por meio de simulações de Monte Carlo, esse valor-s com o valor-p assintótico, o qual é também obtido a partir da estatística da razão de verossimilhanças sob a hipótese de que a variância do efeito aleatório é nula e cuja distribuição assintótica é uma mistura de qui-quadrados (Giampaoli e Singer, 2009; Stram e Lee, 1994).

Palavras-chave: Distribuição assintótica; Valor-s; Modelos mistos; Razão de verossimilhanças; Simulações de Monte Carlo; Teste de hipóteses.

Abstract

Jônatas de Oliveira Alves. **Testes para componentes de variância em modelos lineares mistos: uma abordagem com valor-s**. Thesis (Masters). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2021.

In this work, we study the evidence measure knwon as 's-value', which was developed by Patriota (2013), in tests involving some variance components equal to zero in linear mixed models. The s-value is computed from asymptotic confidence regions based on the likelihood ratio function. One of the main goals of this dissertation is to find the asymptotic distribution of the likelihood ratio statistics under a simple null hypothesis where the parameter is on the border of the parameter space and, from this asymptotic distribution, to obtain the s-value. We also compare by means of Monte Carlo simulations, this asymptotic s-value with the usual asymptotic p-value, which is attained from the likelihood ratio statistics under the composite null hypothesis that the random-effects variance is zero and whose asymptotic distribution is a mixture of chisquared random variables (Giampaoli e Singer, 2009; Stram e Lee, 1994).

Keywords: Asymptotic distributions; Hypothesis testing; Likelihood ratio; Mixed models; Monte Carlo simulations; S-value.

v

Lista de Símbolos

- ^{ind} Variáveis aleatórias independentes
- $\stackrel{iid}{\sim}$ Variáveis aleatórias independentes e identicamente distribuídas
- $\stackrel{D}{\rightarrow} \quad \text{Convergência em distribuição}$
- $\boldsymbol{\theta}$ Vetor de parâmetros
- k Dimensão de θ
- β Vetor de parâmetros dos efeitos fixos
- p Dimensão de β
- D Matriz de covariâncias dos efeitos aleatórios
- r Número de parâmetros funcionalmente independentes de D
- σ^2 Variância do erro aleatório
- n Número de grupos
- *n*_i Número de observações do i-ésimo grupo
- **b** Vetor de efeitos aleatórios
- *q* Dimensão de **b**
- e Vetor de erros aleatórios
- λ_n Razão de verossimilhanças
- $L(\cdot; \mathbf{y})$ Função de verossimilhança
 - $l_n(\cdot)$ Função de log-verossimilhança
 - \mathcal{H}_0 Hipótese nula
 - \mathcal{H}_1 Hipótese alternativa
 - Θ Espaço paramétrico
 - Θ_0 Espaço paramétrico da hipótese nula
 - Θ_1 Espaço paramétrico da hipótese alternativa
 - \mathcal{C} Cone
 - $\tilde{\mathcal{C}}$ Cone tangente
 - *D*_d Derivada direcional com respeito à direção d
- $S_n(\theta)$ Função escore
- $G_n(\theta)$ Informação de Fisher
 - $\|\cdot\|$ Norma Euclidiana
 - o(g) Função com ordem de crescimento inferior a de g

Lista de Figuras

4.1	$\Theta_{a\epsilon}$ para diferentes valores de a , sendo \bar{y} um deles	19
4.2	$\Theta_{ ilde{y}\epsilon}$ e seu subconjunto Θ_0 , com suas respectivas estimativas da razão de	
	verossimilhanças ϵ e ϕ	20
6.1	Função densidades de probabilidade da $\lambda_n(\{\theta\})$, em que $\sigma_b^2 = 0$	24
A.1	Derivadas direcionais da função $h(\theta) = -\theta^3 - \theta^2 + 2$, definida em $\theta \ge 0$. As	
	derivadas direcionais são observadas nos pontos em que θ = 0, θ = 0,5 e	
	θ = 1,5	26

Sumário

1	Introdução	1					
2	Modelos lineares mistos						
	2.1 Definição do modelo	5					
	2.2 Estimação	6					
3	Distribuição assintótica da Razão de Verossimilhanças						
4	Inconsistência do valor-p						
5	Valor-s						
6	Estudo de simulação						
Aj	pêndices						
A	A Definições e Condições de Regularidade						
B	Teoremas do Capítulo 5	29					
	B.1 Prova do Teorema 3	29					
	B.2 Prova do Teorema 4	29					
	B.3 Prova do Teorema 5	30					
С	Artigo 31						

Capítulo 1

Introdução

Os estudos com medidas repetidas são empregados em diversas áreas tais como Agricultura, Biologia, Economia, Medicina, entre outras (Kowalski, 1993 e Pinheiro e Bates, 2006). Por exemplo, em alguns estudos deve-se avaliar a evolução das condições clínicas dos pacientes ao longo do tempo após a administração de algum medicamento. Nesses estudos, as variáveis de interesse são observadas mais de uma vez em cada unidade experimental, caracterizando um estudo de medidas repetidas. Tais réplicas podem ser obtidas em momentos diferentes, em lugares diferentes e/ou em circunstâncias diferentes. Os estudos de medidas repetidas podem induzir uma dependência intra-unidade experimental, a qual pode ser modelada por meio de modelos mistos, uma vez que estes incorporam naturalmente uma dependência intra-unidade experimental. A fim de ilustrar a modelagem da dependência intra-unidade experimental, a presentamos a seguir um modelo misto simples.

Considere *n* unidades experimentais: $U_1, ..., U_n$. Na unidade experimental *i*, temos o interesse em mensurar k_i vezes uma quantidade numérica *y* como descrito na tabela abaixo:

$$\begin{vmatrix} U_1 & \dots & U_1 & U_2 & \dots & U_2 \\ y_{11} & \dots & y_{1k_1} & y_{21} & \dots & y_{2k_2} & \dots & y_{n1} & \dots & y_{nk_n} \end{vmatrix}$$

ou seja, y_{ij} é a *j*-ésima observação da *i*-ésima unidade experimental. Considere que todas as variáveis tem a mesma média populacional e que as observações entre unidades experimentais são independentes. Como as variáveis y_{i1}, \ldots, y_{ik_i} pertencem à mesma unidade experimental, pode ser o caso de que elas carreguem alguma dependência entre si. Assim, nesse contexto, uma forma de se modelar essa dependência é por meio do seguinte modelo

$$y_{ij} = \alpha + b_i + u_{ij}, \tag{1.1}$$

em que α é a média populacional comum a todas as variáveis, $b_i \stackrel{iid}{\sim} N(0, \sigma_b^2)$ é o efeito aleatório do modelo cujo objetivo principal é induzir dependência entre as observações de mesma unidade experimental e $u_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$ é o erro aleatório. Além disso, no modelo (1.1), b_i é independente de u_{ij} e "id" denota "variáveis independentes e identicamente distribuídas". Um modelo mais geral é apresentado no Capítulo 2.

A dependência intra-unidade experimental induzida pelo modelo (1.1) pode ser identificada calculando a covariância e correlação entre y_{ij} e $y_{ij'}$, dadas respectivamente por $cov(y_{ij}, y_{ij'}) = \sigma_b^2 e corr(y_{ij}, y_{ij'}) = \sigma_b^2/(\sigma_b^2 + \sigma^2), j \neq j'$. Observe que uma limitação deste modelo é que a covariância sempre será não-negativa. Como estamos considerando um modelo normal, concluímos que a independência entre as variáveis intra-unidades experimentais ocorre quando a variância do efeito aleatório é igual a zero, ou seja, $\sigma_b^2 = 0$.

Há outras formas de modelar a dependência intra-unidade experimental, como modelos de regressão multivariados (ver Davis, 2002). Esses modelos permitem correlação negativa intra-unidade experimental, entretanto, esperamos uma correlação positiva entre observações de uma mesma unidade experimental, o que torna plausível a utilização de modelos mistos.

O teste de nulidade da variância, que testa se $\sigma_b^2 = 0$ no modelo (1.1) por exemplo, tem sido um tema bastante abordado, principalmente por sua importância em modelos lineares mistos (ver Verbeke e Molenberghs, 2000 e Demidenko, 2013). O objetivo do teste é verificar se o componente aleatório tem relevância no modelo em questão, ou seja, no modelo (1.1) o teste verifica se as observações de uma mesma unidade experimental são não correlacionadas e, portanto, se há independência intra-unidade experimental, pois o modelo em questão é normal. Podemos testar essas hipóteses utilizando: testes F exatos (Khuri et al., 2011); estatística U (Nobre et al., 2013); testes assintóticos (Vu e Zhou, 1997 e Giampaoli e Singer, 2009). Para fazermos esses testes utilizaremos a estatística da razão de verossimilhanças. Em modelos mais complexos, encontrar uma estatística de teste com distribuição conhecida é algo difícil e que, por vezes, faz-se necessário o uso de funções especiais. Uma alternativa é utilizar teoria assintótica ou ainda procedimentos computacionais, tais como: Bootstrap; Monte Carlo; Jackknife (Shao e Tu, 2012). As condições de regularidade para obtermos a distribuição assintótica podem ser vistas em Cox e Hinkley (1979), Sen e Singer (1994) e Silvapulle e Sen (2005). Uma das condições é que o ponto a ser testado seja ponto interior do espaço paramétrico, porém σ_b^2 = 0 está na borda desse espaço. Silvapulle e Sen (2005) discutem que nesses casos a estatística da razão de verossimilhanças não converge necessariamente em distribuição para uma variável aleatória qui-quadrado.

Stram e Lee (1994) encontram a distribuição assintótica da estatística da razão de verossimilhanças para testar a nulidade de componentes de variância em modelos lineares mistos, utilizando resultados de Self e Liang (1987), que supõem unidades experimentais independentes e identicamente distribuídas, e provam que essa estatística converge para uma mistura de qui-quadrados. Giampaoli e Singer (2009) fazem uso de resultados obtidos por Vu e Zhou (1997) e mostram, para os casos com até dois efeitos aleatórios, que a distribuição assintótica da estatística da razão de verossimilhanças também é uma mistura de qui-quadrados, quando permitimos às unidades experimentais não serem identicamente distribuídas.

Stram e Lee (1994) mostram que a distribuição assintótica da estatística da razão de verossimilhanças para testar a nulidade de componentes de variância é uma mistura de quiquadrados. No desenvolvimento, os autores utilizam os resultados de Self e Liang (1987), assumindo unidades experimentais independentes e identicamente distribuídas. Para o caso em que as unidade experimentais não são identicamente distribuídas, Giampaoli e Singer (2009) mostram que a distribuição assintótica dessa estatística também é uma mistura de qui-quadrados.

Os testes de Stram e Lee (1994) e Giampaoli e Singer (2009) têm como principal função verificar a necessidade do efeito aleatório no modelo (1.1). Ou seja, podemos verificar se o efeito aleatório b_i é realmente necessário, testando a hipótese \mathcal{H}_0 : $\sigma_b^2 = 0$.

Na prática, podemos ter também interesse em testar outros tipos de hipóteses, como por exemplo, \mathcal{H}'_0 : $\alpha = \alpha_0 \in \mathcal{H}''_0$: $(\sigma_b^2, \alpha) = (0, \alpha_0)$. Na primeira hipótese o vetor de parâmetros é interior ao espaço paramétrico, portanto a estatística da razão de verossimilhanças tem distribuição assintótica qui-quadrado com um grau de liberdade. Já na segunda hipótese o vetor de parâmetros está na borda do espaço paramétrico, assim uma teoria assintótica não-padrão deve ser aplicada, como apresentado por Vu e Zhou (1997).

Observe que, $\mathcal{H}_0'' \implies \mathcal{H}_0 \in \mathcal{H}_0'' \implies \mathcal{H}_0'$, portanto, esperamos encontrar mais evidência contra \mathcal{H}_0'' do que contra, respectivamente, $\mathcal{H}_0' \in \mathcal{H}_0$. De maneira geral, se temos duas hipóteses $\mathcal{A} \in \mathcal{B}$ tal que $\mathcal{A} \implies \mathcal{B}$ então esperamos encontrar mais evidências contra \mathcal{A} do que contra \mathcal{B} , pois \mathcal{A} é uma hipótese mais restrita.

Schervish (1996) e Patriota (2013) apresentam alguns exemplos em que $\mathcal{A} \implies \mathcal{B}$, contudo rejeita-se \mathcal{B} , mas não se rejeita \mathcal{A} , utilizando o valor-p. Ou seja, o valor-p não respeita, necessariamente, as conclusões esperadas para essas duas hipóteses. No exemplo 1 no Capítulo 4, apresentamos uma simulação que explicita o problema 'inconsistência' do valor-p, considerando as hipóteses \mathcal{H}_0 , \mathcal{H}'_0 e \mathcal{H}''_0 . Já no exemplo 2 no Capítulo 4, desenvolvemos um exemplo teórico em que $\mathcal{A} \implies \mathcal{B}$, porém 'rejeitamos \mathcal{B} ' e não 'rejeitamos \mathcal{A} '. Para resolver esse problema, Patriota (2013) propõe uma medida de evidência alternativa chamada valor-s, que satisfaz relações lógicas não satisfeitas pelo valor-p, a qual utilizaremos neste trabalho.

O valor-s é calculado com base nas regiões de confiança geradas a partir da distribuição assintótica da estatística da razão de verossimilhanças para uma hipótese simples. Isso é diferente da estatística para uma hipótese composta utilizada por Giampaoli e Singer (2009). Um objetivo deste trabalho é derivar a distribuição assintótica da primeira estatística. A razão de verossimilhanças é definida formalmente no Capítulo 3.

Um outro objetivo é, utilizando a distribuição assintótica obtida, determinar a medida valor-s que será utilizada para testar a hipótese de nulidade de componentes da variância de efeitos aleatórios. A contribuição deste trabalho está em apresentar um novo método para verificar a presença de um efeito aleatório em um modelo linear e comparar o desempenho do valor-s com o do valor-p.

No Capítulo 2, apresentamos uma revisão literária sobre os modelos lineares mistos: definição e estimação. No Capítulo 3, fazemos uma breve revisão sobre testes de hipóteses em modelos lineares mistos. Além disso, derivamos a distribuição assintótica da estatística de teste para uma hipótese simples. O Capítulo 4 discute os problemas de incoerência lógica observados quando a medida de evidência utilizada é o valor-p. Apresentamos dois exemplos em que a incoerência lógica é observada, sendo que o primeiro é uma simulação e o segundo um exemplo teórico. Já no Capítulo 5, definimos o valor-s e apresentamos algumas de suas propriedades. Essa definição estende a dada por Patriota (2013), permitindo os casos em que o espaço paramétrico é fechado e/ou particionado. Utilizando essa definição

estendida, obtivemos o valor-s como medida para testar nulidade de componentes da variância.

Por fim, algumas comparações entre o valor-p e o valor-s são feitas, desenvolvendo discussões mais conceituais.

Capítulo 2

Modelos lineares mistos

Os modelos lineares mistos estão entre os principais recursos na análise de dados com medidas repetidas. Laird e Ware (1982) introduzem esses modelos, como modelos com efeitos aleatórios ou modelos em dois estágios, apresentando como principais vantagens: os dados podem ser desbalanceados; a relação entre e intra-unidades experimentais é explicitada; e por vezes, a predição dos efeitos aleatórios trazem consigo uma interpretação exploratória útil.

Nesses modelos os efeitos aleatórios são responsáveis por modelar a correlação intraunidade experimental e sua estrutura pode ser dividida em dois estágios. No primeiro estágio é considerada a distribuição das unidades uma normal, condicionada aos efeitos aleatórios. Já no segundo estágio assumimos uma distribuição normal para os efeitos aleatórios. Mais detalhes podem ser vistos na seção 2.1.

2.1 Definição do modelo

Com o objetivo de modelar *n* unidades experimentais com n_i observações cada (i = 1, ..., n), considere o seguinte modelo linear com efeitos aleatórios para dados com medidas repetidas.

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i \\ \mathbf{e}_i &= \mathbf{Z}_i \mathbf{b}_i + \mathbf{u}_i \end{aligned}$$
 (2.1)

em que, $\mathbf{Y}_i = (Y_{i1}, ..., Y_{in_i})^{\top}$ é o vetor de observações da *i*-ésima unidade experimental, \mathbf{X}_i é a matriz $(n_i \times p)$ das variáveis regressoras (conhecida e de posto coluna completo), $\boldsymbol{\beta}$ é o vetor $(p \times 1)$ de parâmetros da média, $\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}_{i1} & \dots & \mathbf{z}_{in_i} \end{bmatrix}^{\top}$ é a matriz $(n_i \times q)$ de especificação (conhecida) dos efeitos aleatórios, \mathbf{b}_i é o vetor $(q \times 1)$ de efeitos aleatórios e \mathbf{u}_i é o vetor $(n_i \times 1)$ de erros. Além disso, $\mathbf{b}_i \stackrel{iid}{\sim} N_q(\mathbf{0}, \mathbf{D}) \perp \mathbf{u}_i \stackrel{ind}{\sim} N_{n_i} \left(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}\right)$, tal que \mathbf{D} é uma matriz não estruturada e não negativa definida (n.n.d.), e \mathbf{I}_{n_i} é matriz identidade de ordem n_i . Logo, $\mathbf{e}_i \stackrel{ind}{\sim} N_{n_i}(\mathbf{0}, \Sigma_i)$, com $\Sigma_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^{\top} + \sigma^2 \mathbf{I}_{n_i}$, por conseguinte, $\mathbf{Y}_i \stackrel{ind}{\sim} N_{n_i}(\mathbf{X}_i \boldsymbol{\beta}, \Sigma_i)$. Diremos que $\boldsymbol{\gamma}$ é o vetor de ordem $(q(q+1)/2 \times 1)$ contendo os elementos do triângulo inferior de \mathbf{D} . Por fim, o espaço paramétrico é denotado por $\Theta = \mathbb{R}^p \times \mathbb{R}_+ \times \Gamma$, em que $\mathbb{R}_+ = (0, +\infty)$, $\Gamma = \{ \boldsymbol{\gamma} \in \mathbb{R}^r : \boldsymbol{D} \text{ é uma matriz n.n.d.} \}$ e r = q(q+1)/2, e o vetor de parâmetros é denotado por $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\top}, \sigma^2, \boldsymbol{\gamma}^{\top})^{\top} \in \Theta$, em que a dimensão do espaço paramétrico é k = p + 1 + r.

Note que o modelo (2.1) pode ser visto em dois estágios, de modo que, no primeiro estágio consideramos a distribuição da unidade experimental condicionada ao efeito aleatório, como mostra a equação (2.2).

$$\boldsymbol{Y}_{i}|\boldsymbol{b}_{i} \stackrel{ind}{\sim} N_{n_{i}}\left(\boldsymbol{X}_{i}\boldsymbol{\beta} + \boldsymbol{Z}_{i}\boldsymbol{b}_{i}, \sigma^{2}\boldsymbol{I}_{n_{i}}\right).$$

$$(2.2)$$

No segundo estágio, temos a distribuição do efeito aleatório.

$$\boldsymbol{b}_i \stackrel{iid}{\sim} \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{D}) \,. \tag{2.3}$$

Os efeitos aleatórios permitem modelar a dependência intra-unidade experimental. Os resultados (2.4) e (2.5) implicam que a dependência intra-unidade experimental está diretamente relacionada à variância do efeito aleatório.

A covariância e a correlação intra-unidade experimental são dadas por:

$$\operatorname{Cov}_{\boldsymbol{\theta}}\left(Y_{ij}, Y_{il}\right) = \boldsymbol{z}_{ij}^{\top} \boldsymbol{D} \boldsymbol{z}_{il}$$
(2.4)

e

$$\operatorname{Corr}_{\theta}\left(Y_{ij}, Y_{il}\right) = \frac{\boldsymbol{z}_{ij}^{\top} \boldsymbol{D} \boldsymbol{z}_{il}}{\sqrt{\boldsymbol{z}_{ij}^{\top} \boldsymbol{D} \boldsymbol{z}_{ij} + \sigma^2} \sqrt{\boldsymbol{z}_{il}^{\top} \boldsymbol{D} \boldsymbol{z}_{il} + \sigma^2}}.$$
(2.5)

Perceba que, se D é matriz nula, então a covariância e a correlação serão ambas iguais a zero. Como a distribuição marginal de Y_i é normal, então correlação nula implica em independência. Dessa forma, quando verificamos a hipótese $\mathcal{H}_0 : D$ é matriz nula, na verdade estamos verificando a hipótese de independência intra-unidade experimental.

2.2 Estimação

Os parâmetros do modelo (2.1) são tipicamente estimados, sob o ponto de vista clássico, via máxima verossimilhança (MV) e/ou via máxima verossimilhança restrita (MVR). Neste trabalho abordaremos apenas MV, pois os estimadores obtidos por esse método são usados nos resultados do Capítulo 3.

Os parâmetros são estimados via máxima verossimilhança por meio da maximização da função de log-verossimilhança marginal, a qual é dada por

$$l_n(\boldsymbol{\theta}) = \sum_{i=1}^n \left\{ -\frac{n_i}{2} \log(2\pi) - \frac{1}{2} \log|\boldsymbol{\Sigma}_i| - \frac{1}{2} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \right\}$$
(2.6)

Os resultados (2.7, 2.8 e 2.11 - 2.14), que se seguem, foram extraídos de Laird e Ware (1982). Em um primeiro momento consideramos o modelo misto com estrutura de variância conhecida (Σ_i conhecida). Daí, temos que

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top} \boldsymbol{\Sigma}_{i}^{-1} \boldsymbol{X}_{i}\right)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top} \boldsymbol{\Sigma}_{i}^{-1} \boldsymbol{Y}_{i}$$
(2.7)

é o estimador de máxima verossimilhança de β quando Σ_i é conhecida, i = 1, 2, ..., n. Note que (2.7) também pode ser visto como estimador de mínimos quadrados ponderados de β .

Em um segundo momento, consideramos o modelo com estrutura de variância desconhecida, com Σ_i função de um vetor γ de ordem *r*. Nesse caso, desejamos estimar o vetor $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathsf{T}}, \boldsymbol{\gamma}^{\mathsf{T}})^{\mathsf{T}}$. O estimador de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ para $\boldsymbol{\theta}$ ocorre quando $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$ são conjuntamente pontos de máximo da função de log-verossimilhança. Portanto, obtemos $\hat{\boldsymbol{\beta}}$ $(\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}}))$, substituindo Σ_i por $\hat{\Sigma}_i$ $(\hat{\Sigma}_i(\hat{\boldsymbol{\gamma}}))$ na equação (2.7). Desse modo

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top} \hat{\boldsymbol{\Sigma}}_{i}^{-1} \boldsymbol{X}_{i}\right)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top} \hat{\boldsymbol{\Sigma}}_{i}^{-1} \boldsymbol{Y}_{i}.$$
(2.8)

Necessitamos usar algoritmos de estimação para estimar γ (componentes da variância). Laird e Ware (1982) propõe o algoritmo EM (*expectation-maximization*), desenvolvido por Dempster *et al.* (1977), para obter $\hat{\gamma}$. O processo de estimação dá-se em dois passos. No primeiro passo, também conhecido como passo E (*expectation*), calculamos a esperança da função de log-verossimilhança completa (log-verossimilhança da distribuição conjunta de Y_i e b_i), dado que $Y_i = y_i$ e temos $\hat{\theta}_{(r)}$ uma estimativa para θ no *r*-ésimo passo do algoritmo. Seja $l^*(\theta; y, b)$ a log-verossimilhança completa, expressa por

$$l^{*}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{b}) = \sum_{i=1}^{n} \left\{ -\frac{n_{i}}{2} \log(2\pi) - \frac{n_{i}}{2} \log(\sigma^{2}) - \frac{1}{2} (\boldsymbol{y}_{i} - \boldsymbol{X}_{i} \boldsymbol{\beta} - \boldsymbol{Z}_{i} \boldsymbol{b}_{i})^{\mathsf{T}} \boldsymbol{\Sigma}_{i}^{-1} (\boldsymbol{y}_{i} - \boldsymbol{X}_{i} \boldsymbol{\beta} - \boldsymbol{Z}_{i} \boldsymbol{b}_{i}) - \frac{1}{2} \log(2\pi) - \frac{1}{2} \log|\boldsymbol{D}| - \frac{1}{2} \boldsymbol{b}_{i}^{\mathsf{T}} \boldsymbol{D}^{-1} \boldsymbol{b}_{i} \right\}$$

então o passo E é dado por,

$$M(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{(r)}) = \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{b}_{i} | \boldsymbol{y}_{i}, \hat{\boldsymbol{\theta}}_{(r)}} \left[l^{*}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{b}) \right].$$
(2.9)

No segundo passo, também conhecido como passo M (*maximization*), atualizamos nossa estimativa pelo θ que maximiza a função M da equação (2.9), como vemos a seguir

$$\hat{\boldsymbol{\theta}}_{(r+1)} = \arg\max_{\boldsymbol{\theta}} M(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{(r)}).$$
(2.10)

Podemos reescrever o passo M, de modo que, $\hat{\beta}_{(r+1)}(\hat{\gamma}_{(r+1)})$ é conforme a equação (2.8) e $\hat{\gamma}_{(r+1)}$ é dado por

$$\hat{\sigma}_{(r+1)}^2 = \frac{\sum_{i=1}^n e_i^\top e_i}{\sum_{i=1}^n n_i} = \frac{t_1}{\sum_{i=1}^n n_i},$$
(2.11)

$$\hat{D}_{(r+1)} = \frac{\sum_{i=1}^{n} b_i b_i^{\top}}{n} = \frac{t_2}{n}.$$
(2.12)

Também podemos reescrever o passo E, obtendo os valores esperados de t_1 e t_2 , dado que já conhecemos y_i e dispomos de $\hat{\theta}_{(r)}$, como se segue

$$\hat{t}_{1} = \mathbb{E}_{\boldsymbol{b}_{i}|\boldsymbol{y}_{i},\hat{\boldsymbol{\theta}}_{(r)}} \left[\sum_{i=1}^{n} \boldsymbol{e}_{i}^{\top} \boldsymbol{e}_{i} \right]$$
(2.13)

$$\hat{\boldsymbol{t}}_2 = \mathbb{E}_{\boldsymbol{b}_i | \boldsymbol{y}_i, \hat{\boldsymbol{\theta}}_{(r)}} \left[\sum_{i=1}^n \boldsymbol{b}_i \boldsymbol{b}_i^\top \right]$$
(2.14)

Iniciamos o algoritmo escolhendo valores iniciais para $\hat{\theta}_{(0)}$, calculamos o passo E, dados nas equações (2.13) e (2.14), e, em seguida, executamos o passo M, ao substituirmos os valores de \hat{t}_1 e \hat{t}_2 nas equações (2.11) e (2.12). Repetimos os passos E e M até o algoritmo convergir computacionalmente.

Apesar de o algoritmo EM ser muito utilizado, ele vem com ressalvas, sendo uma delas a demora na convergência computacional. Demidenko (2013) destaca que o algoritmo, para estimar componentes de variância em modelos mistos, tem propriedades que desmotivam seu uso, quando a variância dos efeitos pode estar na borda do espaço paramétrico. Segundo Demidenko (2013), se o valor inicial da matriz de variância dos efeitos aleatórios for nula, então nas próximas interações permanecerá nula, mas se a matriz de variância dos efeitos aleatórios for positiva definida, ela permanecerá positiva definida.

Neste trabalho utilizaremos a função *lmer* do pacote *lme4*, proposto por Bates *et al.* (2015), para fazer a estimação por máxima verossimilhança dos parâmetros de um modelo linear misto. Bates *et al.* (2015) utilizam o método de mínimos quadrados penalizados e ponderados dos resíduos $(y_{ij} - \boldsymbol{x}_{ij}^{\top} \hat{\boldsymbol{\beta}})$ por repetidas vezes e fazem uso de métodos heurísticos para otimizar parte do processo, para encontrar a estimativa de máxima verossimilhança.

e

Capítulo 3

Distribuição assintótica da Razão de Verossimilhanças

Neste trabalho utilizaremos a distribuição assintótica de (-2 log) razão de verossimilhanças para fazermos todos os testes de hipóteses. Essa distribuição varia de acordo com a posição do seu vetor de parâmetros, apresentando comportamentos diferentes quando está na borda do espaço paramétrico e quando é ponto interior do espaço paramétrico. De forma geral, testaremos hipóteses na forma

$$\mathcal{H}_0: \boldsymbol{\theta}_0 \in \Theta_0 \quad \text{contra} \quad \mathcal{H}_1: \boldsymbol{\theta}_0 \in \Theta_1, \tag{3.1}$$

em que θ_0 é o verdadeiro valor do vetor de parâmetros e $\Theta_0 \cup \Theta_1 = \Theta$ e $\Theta_0 \cap \Theta_1 = \emptyset$, ou seja, Θ_0 e Θ_1 formam uma partição de Θ . A estatística de (-2 log) da razão de verossimilhanças é definida como

$$\lambda_n(\Theta_0,\Theta_1) = 2\left(\sup_{\boldsymbol{\theta}\in\Theta_1} l_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}\in\Theta_0} l_n(\boldsymbol{\theta})\right).$$
(3.2)

Quando a dimensão de Θ_0 é menor que a de Θ e l_n é contínua, a estatística da equação (3.2) pode ser reescrita como

$$\lambda_n(\Theta_0) = 2 \left(\sup_{\theta \in \Theta} l_n(\theta) - \sup_{\theta \in \Theta_0} l_n(\theta) \right).$$
(3.3)

Eventualmente escreveremos $\lambda_n(\mathcal{H}_0)$ em vez de $\lambda_n(\Theta_0)$ por questão de simplicidade.

Considere o modelo (2.1) e Θ_0 não contendo pontos singulares ou que estejam na borda do espaço paramétrico (Chernoff, 1954 e Drton, 2009). Nesse contexto, a estatística $\lambda_n(\Theta_0)$ tem como distribuição assintótica uma qui-quadrado com k – dim(Θ_0) graus de liberdade, pelo Teorema de Wilks (1938).

Vu e Zhou (1997) estudaram a distribuição assintótica de $\lambda_n(\Theta_0)$ em um caso mais

geral em que são utilizadas funções de estimação em um espaço paramétrico que contém pontos de borda. Nesse estudo, eles desenvolvem oito condições de regularidade que devem ser satisfeitas (ver condições (A1)-(A3) e (B1)-(B5) do Apêndice A). Giampaoli e Singer (2009) provaram que alguns casos particulares do modelo linear misto 2.1 satisfazem essas condições de regularidade.

A seguir, apresentamos os principais teoremas desenvolvidos por Self e Liang (1987) e por Vu e Zhou (1997). Esses teoremas utilizam a noção de cones, pois expressam a geometria local de subconjuntos de Θ no ponto $\theta_0 \in \Theta_0$ (Chernoff, 1954). Dizemos que $C \subset \mathbb{R}^k$ é um cone com vértice em $\theta_0 \in \Theta_0$ se

$$\boldsymbol{\theta} \in \mathcal{C} \Longrightarrow \boldsymbol{\theta}_0 + \alpha(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \in \mathcal{C} \quad \forall \alpha \ge 0$$

Teorema 1 (Vu e Zhou, 1997). Seja $\tilde{C}_{\Theta_0}^{\theta_0}$ e $\tilde{C}_{\Theta}^{\theta_0}$ cones que satisfazem as condições (A2) e (A3) do Apêndice A, em que dim(Θ_0) < k. Assim, sob as condições (A1) e (B1-B5) do Apêndice A, para cada $\theta_0 \in \Theta_0$,

$$\lambda_n(\Theta_0) \stackrel{D}{\longrightarrow} \inf_{oldsymbol{ heta}\in ilde{\mathcal{C}}_{\Theta_0}^{\Theta_0}} \|oldsymbol{N}-oldsymbol{ heta}\|^2 - \inf_{oldsymbol{ heta}\in ilde{\mathcal{C}}_{\Theta}^{\Theta_0}} \|oldsymbol{N}-oldsymbol{ heta}\|^2,$$

em que, " $\stackrel{D}{\rightarrow}$ "representa a convergência em distribuição, $\|\cdot\|$ é a distância Euclidiana e N é um vetor aleatório com distribuição normal padrão multivariada.

O teorema de Vu e Zhou (1997) pode ser visto como um extensão dos resultados de Self e Liang (1987), utilizando condições de regularidade menos restritivas.

Considere a seguinte hipótese de uma versão mais restrita do modelo 2.1

$$\mathcal{H}_{0}^{(1)}: \boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}^{*} & 0\\ 0 & 0 \end{pmatrix}, \text{ contra } \mathcal{H}_{1}^{(1)}: \boldsymbol{D} \text{ positiva definida,}$$
(3.4)

em que, D^* é positiva definida de dimensões $(q - 1) \times (q - 1)$, ou seja, apenas uma das componentes do efeito aleatório pode ter variância zero.

Stram e Lee (1994) derivaram a distribuição assintótica da estatística da razão de verossimilhanças para testar a hipótese $\mathcal{H}_0^{(1)}$, utilizando os resultados de Self e Liang (1987). Desse modo, sob $\mathcal{H}_0^{(1)}$,

$$\lambda_n(\mathcal{H}_0^{(1)}) \xrightarrow{D} 0, 5\chi_{q-1}^2 + 0, 5\chi_q^2.$$
(3.5)

em que, " \xrightarrow{D} " representa a convergência em distribuição, χ_q^2 é a distribuição qui-quadrado com q graus de liberdade e , por convenção, χ_0^2 é uma distribuição degenerada em 0. Assim, o valor-p é calculado da seguinte forma,

$$p(\mathcal{H}_{0}^{(1)}; \boldsymbol{y}) = \mathbb{P}\left(0, 5\chi_{q-1}^{2} + 0, 5\chi_{q}^{2} > \lambda_{n}\right),$$
(3.6)

de modo que, λ_n é o valor observado de $\lambda_n(\mathcal{H}_0^{(1)})$.

Os testes de Stram e Lee (1994) têm como base os resultados de Self e Liang (1987), e portanto, valem apenas quando os Y_i 's são independentes e identicamente distribuídos. Giampaoli e Singer (2009) estendem os resultados de Stram e Lee (1994) para o caso em que os Y_i 's são apenas independentes e o modelo linear misto tem no máximo dois efeitos aleatórios.

Giampaoli e Singer (2009) utilizam o Teorema 1 (ver também Vu e Zhou, 1997) e, sob algumas condições de regularidade mais flexíveis (ver Apêndice A), provam que o resultado de Stram e Lee (1994) para testar a hipótese $\mathcal{H}_0^{(1)}$, dado em (3.5), permanece válido nos casos em que q = 0 e q = 1.

Considere o modelo linear misto como definido no Capítulo 2, para o caso em que q = 1 (um único efeito aleatório). Utilizaremos os resultados de Vu e Zhou (1997) e Giampaoli e Singer (2009) para calcular a distribuição assintótica da razão de verossimilhanças para testar hipóteses pontuais, ou seja, estamos interessados na distribuição assintótica de $\lambda_n(\{\theta\})$ para cada $\theta \in \Theta$.

O Teorema 2 abaixo é um dos resultados desenvolvidos nesta dissertação.

Teorema 2. Sob as condições de regularidade dispostas no Apêndice A, a distribuição assintótica de $\lambda_n(\{\theta\})$ é uma mistura de qui-quadrados, tal que

$$\lambda_n(\{\boldsymbol{\theta}\}) \xrightarrow{D} \left\{ \begin{array}{cc} \chi_{p+2}^2 & , se \ \boldsymbol{\theta} \in \{\boldsymbol{\theta}^* \in \Theta : \boldsymbol{D} > 0\} \\ 0.5\chi_{p+1}^2 + 0.5\chi_{p+2}^2 & , se \ \boldsymbol{\theta} \in \{\boldsymbol{\theta}^* \in \Theta : \boldsymbol{D} = 0\}. \end{array} \right.$$
(3.7)

Demonstração. Giampaoli e Singer (2009) demonstraram que o modelo linear misto com um efeito aleatório, o qual estamos utilizando, satisfaz as condições (A1) e (B1–B5) das condições de regularidade de Vu e Zhou (1997) dispostas no Apêndice A. A seguir, mostraremos que, em nosso caso, as condições (A2) e (A3) também são válidas.

Seja $G_{n,\theta_0}^{1/2}$ conforme definido no Apêndice A. O cone $C_{\{\theta_0\}}^{\theta_0} = \{\theta_0\}$ satisfaz (A2), pois

$$\inf_{\boldsymbol{w}\in\mathcal{C}_{\Omega}^{\boldsymbol{\theta}_{0}}} \|\boldsymbol{G}_{\boldsymbol{n},\boldsymbol{\theta}_{0}}^{1/2}(\boldsymbol{w}-\boldsymbol{v})\| = \inf_{\boldsymbol{v}\in\Omega} \|\boldsymbol{G}_{\boldsymbol{n},\boldsymbol{\theta}_{0}}^{1/2}(\boldsymbol{w}-\boldsymbol{v})\| = 0 \quad \forall \ \boldsymbol{v}\in\Omega, \boldsymbol{u}\in\mathcal{C}_{\Omega}^{\boldsymbol{\theta}_{0}}$$

em que, $\Omega = \{\boldsymbol{\theta}_0\}$. A primeira e segunda parte da condição (A2) são garantidas ao fazer $u(\boldsymbol{v}) = ||\boldsymbol{v} - \boldsymbol{\theta}_0||$ e $v(\boldsymbol{w}) = ||\boldsymbol{w} - \boldsymbol{\theta}_0||$, então $C_{\{\boldsymbol{\theta}_0\}}^{\boldsymbol{\theta}_0}$ satisfaz (A2).

Para cada $\theta_0 \in \Theta = \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_0$, nós temos dois casos possíveis, a saber, (a) $\theta_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$ e (b) $\theta_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \{0\}$. Encontraremos cones associados a Θ distintos para o caso (a) e (b).

No caso (a), o vetor de parâmetros é um ponto interior de Θ , então, considere o cone $C_{\Theta}^{\theta_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}$ com vértice em $\theta_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$. A primeira parte da condição (A2) é válida ao fazermos $u(v) = ||v - \theta_0||$, uma vez que $\Theta \subseteq C_{\Theta}^{\theta_0}$ o que implica

$$\inf_{w\in \mathcal{C}_{\Theta}^{\theta_0}} \|\boldsymbol{G}_{n,\theta_0}^{1/2}(\boldsymbol{w}-\boldsymbol{v})\| = 0, \quad \boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+, \forall \ \boldsymbol{v} \in \Theta.$$

Para a segunda parte da condição (A2), nós precisamos analisar o (p + 1)-ésimo e o (p + 2)-ésimo elementos de $w \in C_{\Theta}^{\theta_0}$. Sejam g_{p+1} e g_{p+2} , respectivamente, as colunas de

ordem (p + 1) e (p + 2) de $G_{n,\theta_0}^{1/2}$.

- Se os (p + 1)-ésimo e (p + 2)-ésimo elementos são não negativos, então $\inf_{v \in \Theta} ||G_{n,\theta_0}^{1/2}(w v)|| = 0.$
- Do Apêndice A, nós concluímos que para ambos os casos, a saber, "se (p + 1)-ésimo elemento é negativo e (p + 2)-ésimo elemento é não negativo" ou "se (p + 1)-ésimo elemento é não negativo e (p + 2)-ésimo elemento é negativo", então inf_{v∈Θ} ||G^{1/2}_{n,θ₀}(w v)|| = 0.
- Se os (p + 1)-ésimo e (p + 2)-ésimo elementos são negativos, então $\inf_{v \in \Theta} ||G_{n,\theta_0}^{1/2}(w v)|| \le ||g_{p+1}||||w_{p+1}|| + ||g_{p+2}||||w_{p+2}||$. Além disso, note que $||G_{n,\theta_0}^{1/2}(w \theta_0)|| = c_2 + ||g_{p+1}||||w_{p+1} \sigma_0^2|| + ||g_{p+2}||||w_{p+2} \gamma_0||$, em que $c_2 \ge 0$. Como $w_{p+1} < 0$ e $w_{p+2} < 0$, então $||w_{p+1}|| < ||w_{p+1} \theta_0||$ e $||w_{p+2}|| < ||w_{p+2} \theta_0||$. Assim,

$$\inf_{\boldsymbol{v}\in\Theta} \|\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{v})\| < \|\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{\theta}_0)\|.$$

Seja

$$\upsilon(\boldsymbol{w}) = \begin{cases} ||\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w} - \boldsymbol{\theta}_0)|| & \text{, se } w_{p+1} > 0 \text{ ou } w_{p+2} > 0, \\ 1 & \text{, caso contrário.} \end{cases}$$

Quando $w \to \theta_0$, $v(w) \to 0$, pois quando w está próximo o suficiente de θ_0 , w_{p+1} e w_{p+2} são positivos. Portanto, $C_{\Theta}^{\theta_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}$ com vértice em $\theta_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$ satisfaz a condição (A2).

No caso (b), o vetor de parâmetros é um ponto na borda de Θ , então, considere o cone $C_{\Theta}^{\theta_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}_0$ com vértice em $\theta_0 \in \Theta'_0$. Note que $\Theta \subset C_{\Theta}^{\theta_0}$, então a primeira parte da condição (A2) é satisfeita ao fazer $u(v) = ||v - \theta_0||$, desde que

$$\inf_{w\in C_{\Theta}^{\theta_0}} \|\boldsymbol{G}_{n,\theta_0}^{1/2}(\boldsymbol{w}-\boldsymbol{v})\| = 0 \quad \forall \ \boldsymbol{v}\in \Theta.$$

Para a segunda parte da condição (A2), nós precisamos analisar o (p+1)-ésimo elemento de w para cada $w \in C_{\Theta}^{\theta_0}$. Se este (p + 1)-ésimo é não negativo, então $\inf_{v \in \Theta} ||G_{n,\theta_0}^{1/2}(w - v)|| = 0$. Portanto, a única mudança ocorre quando o (p + 1)-ésimo elemento de w é negativo. Nesse caso,

$$\inf_{v\in\Theta} ||G_{n,\theta_0}^{1/2}(w-v)|| = ||g_{p+1}||||w_{p+1}||,$$

em que g_{p+1} é o (p + 1)-ésima coluna de $G_{n,\theta_0}^{1/2}$. Como esse ínfimo não é zero, nós temos que fazer uma análise mais minuciosa.

Perceba que $||G_{n,\theta_0}^{1/2}(w - \theta_0)|| = c_1 + ||g_{p+1}||||w_{p+1} - \sigma_0^2||$, em que $c_1 \ge 0$. Por suposição, $w_{p+1} < 0$ and $\sigma_0^2 > 0$, isso implica que $||w_{p+1} - \sigma_0^2|| > ||w_{p+1}||$ e, portanto,

$$\inf_{\boldsymbol{v}\in\Theta} \|\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{v})\| = \|\boldsymbol{g}_{p+1}\|\|\boldsymbol{w}_{p+1}\| < c_1 + \|\boldsymbol{g}_{p+1}\|\|\boldsymbol{w}_{p+1} - \sigma_0^2\| = \|\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{\theta}_0)\|.$$

Seja

$$\upsilon(\boldsymbol{w}) = \begin{cases} \|\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w} - \boldsymbol{\theta}_0)\| & \text{, se } w_{p+1} > 0, \\ 1 & \text{, caso contrário.} \end{cases}$$

A medida que $w \to \theta_0$, $v(w) \to 0$, pois quando w é suficientemente próxima de θ_0 , w_{p+1} é positivo. Logo, $C_{\Theta}^{\theta_0}$ satisfaz a condição (A2).

Segundo Giampaoli e Singer (2009), sabemos que, para n suficientemente grande, G_{n,θ_0} é positivo definido para cada $\theta_0 \in \Theta$. Portanto, para T_n uma matriz $(p + 2) \times (p + 2)$ ortogonal e não estocástica, nós temos que $\tilde{C}_{\{\theta_0\}_n}^{\theta_0} = T_n G_{n,\theta_0}^{1/2} (C_{\{\theta_0\}}^{\theta_0} - \theta_0) = \{0\}$ e $\tilde{C}_{\Theta_n}^{\theta_0} = T_n G_{n,\theta_0}^{1/2} (C_{\Theta}^{\theta_0} - \theta_0) = \{0\}$ e $\tilde{C}_{\Theta_n}^{\theta_0} = T_n G_{n,\theta_0}^{1/2} (C_{\Theta}^{\theta_0} - \theta_0) = C_{\Theta}^{\theta_0}$ para cada $\theta_0 \in \Theta$. Logo, (A3) é válido. Nós concluímos que os cones $\tilde{C}_{\{\theta_0\}}^{\theta_0} = \{0\}$ e $\tilde{C}_{\Theta}^{\theta_0} = C_{\Theta}^{\theta_0}$ são os empregados no Teorema 1. Seja $N = (N_1, \dots, N_p, N_{p+1}, N_{p+2})$ um vetor aleatório com distribuição normal multivariada padrão. Considere os dois casos:

Caso (a) Se $\theta_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$, nós temos que $\tilde{C}^{\theta_0}_{\{\theta_0\}} = \{\mathbf{0}\}$ e $\tilde{C}^{\theta_0}_{\Theta} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}$, então

$$\inf_{\boldsymbol{\theta}\in\tilde{\mathcal{C}}_{\{\boldsymbol{\theta}_0\}}^{\boldsymbol{\theta}_0}} \|\boldsymbol{N}-\boldsymbol{\theta}\|^2 - \inf_{\boldsymbol{\theta}\in\tilde{\mathcal{C}}_{\Theta}^{\boldsymbol{\theta}_0}} \|\boldsymbol{N}-\boldsymbol{\theta}\|^2 = \sum_{j=1}^{p+2} N_j^2.$$

Ou seja, esta é uma qui-quadrado com p + 2 graus de liberdade.

Caso (b) Se $\boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_0$, nós temos que $\tilde{C}^{\boldsymbol{\theta}_0}_{\{\boldsymbol{\theta}_0\}} = \{\mathbf{0}\}$ e $\tilde{C}^{\boldsymbol{\theta}_0}_{\Theta} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}_0$, então

$$\inf_{\boldsymbol{\theta}\in \tilde{\mathcal{C}}_{\{\boldsymbol{\theta}_0\}}^{\boldsymbol{\theta}_0}} \|\boldsymbol{N}-\boldsymbol{\theta}\|^2 - \inf_{\boldsymbol{\theta}\in \tilde{\mathcal{C}}_{\Theta}^{\boldsymbol{\theta}_0}} \|\boldsymbol{N}-\boldsymbol{\theta}\|^2 = \sum_{j=1}^{p+1} N_j^2 + N_{p+2}^2 I(N_{p+2} \ge 0).$$

Ou seja, está é uma mistura 50-50% de qui-quadrados com g e g + 1 graus de liberdade.

Neste Capítulo, definimos a razão de verossimilhanças e fazemos uma pequena revisão literária sobre testes assintóticos, em especial, casos não-regulares. Apresentamos o Teorema 1 e, com base nele, derivamos o Teorema 2. Esse último nos mostra a distribuição assintótica da razão de verossimilhanças para testar uma hipótese simples, ou seja, a distribuição assintótica da razão de verossimilhanças num ponto. Isso é bastante útil, uma vez que nos permite construir regiões de confiança. Tais resultados assintóticos podem ser usados para calcular o valor-p. No entanto, veremos no Capítulo 4 que o valor-p dispõe de inconsistências lógicas em suas conclusões.

Capítulo 4

Inconsistência do valor-p

O conhecimento científico atual é o resultado de diversas pesquisas. Nesse processo, hipóteses são feitas e verificadas. Uma hipótese proposta só é considerada científica se puder ser testada. Nesse contexto, a estatística é um dos principais recursos utilizados. Uma das principais ferramentas estatísticas utilizadas no meio científico é o valor-p. De forma geral, o valor-p é a probabilidade de ocorrer em outros experimentos um evento tão extremo quanto o evento observado no experimento executado, sob as condições impostas pela hipótese nula que desejamos testar. De modo que, se o valor-p for menor que um nível crítico pré-determinado, rejeitamos a hipótese testada, porém, se o valor-p for maior que o nível crítico, então dizemos não haver evidências para rejeitar a hipótese nula.

A seguir apresentamos exemplos cujas conclusões, derivadas de testes de hipóteses, se contradizem. Esse tipo de problema ocorre quando se utiliza o valor-p em hipóteses nulas aninhadas. Autores como Schervish (1996) e Patriota (2013), apresentam exemplos em que o valor-p se mostra inconsistente. Como alternativa, Patriota (2013) desenvolveu o valor-s, uma medida de evidência construída com base em regiões de confiança e que é logicamente consistente. A seguir discutiremos os problemas de inconsistência do valor-p.

Considere o seguinte exemplo. Nele, exploraremos os testes de hipóteses abordados até então, como também os problemas envolvendo o uso do valor-p.

Exemplo 1: Simulamos Normais padrão independentes, dividindo em 30 grupos com 5 observações cada (ver Tabela 4.1). O seguinte modelo, baseado no modelo (2.1), foi proposto para descrever o comportamento dos dados (y_{ij}), de modo que, para cada *grupo i* = 1, ..., 30 e cada observação *j* = 1, ..., 5,

$$y_{ij} = \alpha + b_i + u_{ij}$$

em que, $b_i \sim N(0, \sigma_b^2)$ é independente de $u_{ij} \sim N(0, \sigma^2)$. Na simulação, consideramos $\alpha = 0$, $\sigma_b^2 = 0$ e $\sigma^2 = 1$. Desejamos verificar se todos os grupos têm média zero. A média de todos os $y'_{ij}s \in \alpha$, porém, para cada grupo *i* a média é $\alpha + b_i$. Portanto, precisamos não apenas testar que se $\alpha = 0$, mas também verificar se todos os b'_is são zeros, ou seja, se $\sigma_b^2 = 0$. Um procedimento possível é, primeiramente, testarmos a hipótese \mathcal{H}_0 : $\sigma_b^2 = 0$ contra \mathcal{H}_1 : $\sigma_b^2 > 0$ e em seguida testarmos se $\alpha = 0$. Utilizando o resultado (3.6), temos que, o valor-p associado a hipótese \mathcal{H}_0 é 0,0332. Consideramos o efeito aleatório significativo a 0,05 de nível de significância. Não necessitamos fazer mais testes, pois já verificamos que os grupos têm médias distintas.

Um segundo procedimento é verificarmos, diretamente, a hipótese \mathcal{H}'_0 : $\alpha = 0$ e $\sigma_b^2 = 0$ contra \mathcal{H}'_1 : $\alpha \neq 0$ ou $\sigma_b^2 > 0$. Self e Liang (1987) mostram que, sob uma hipótese tal como \mathcal{H}'_0 , $\lambda_n(\mathcal{H}'_0)$ converge em distribuição para uma mistura 50:50 de uma χ_1^2 e uma χ_2^2 . Assim, o valor-p associado a \mathcal{H}'_0 é 0, 1034. Portanto, não rejeitamos a hipótese de que os grupos têm médias zero.

Os procedimentos obtiveram conclusões distintas. Além disso, \mathcal{H}_0 é uma hipótese menos restritiva que \mathcal{H}'_0 , porém, encontramos mais evidência para rejeitar \mathcal{H}_0 do que para rejeitar \mathcal{H}'_0 . Isso se dá, pois o valor-p é uma medida inconsistente e pode gera conclusões controversas. No Capítulo 5 abordamos o valor-s, proposto por Patriota (2013), que serve como uma alternativa ao valor-p.

A seguir discutiremos um exemplo em que o valor-p apresenta problemas de coerência lógica em um modelo normal multivariado com média desconhecida. No exemplo podemos observar sinais de que as incoerências são bastante presentes e que podem resultar em conclusões equivocadas.

Exemplo 2: Seja $\boldsymbol{Y} = (\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n)^{\top}$ uma amostra aleatória tal que $\boldsymbol{Y}_i \stackrel{iid}{\sim} N_k(\boldsymbol{\theta}, \boldsymbol{I}_k)$, em que \boldsymbol{I}_k é matriz identidade de ordem k. Para toda hipótese do tipo $\mathcal{H}_0 : \boldsymbol{\theta} \in \Theta_0$ contra $\mathcal{H}_1 : \boldsymbol{\theta} \in \Theta_1 = \mathbb{R}^k - \Theta_0, \Theta_0$ fechado, consideraremos $\lambda_n(\Theta_0, \Theta_1)$ a razão de verossimilhanças como dada em (3.2). Podemos obter com alguma álgebra que $\lambda_n(\Theta_0, \Theta_1) = \inf_{\boldsymbol{\theta} \in \Theta_0} n || \boldsymbol{Y} - \boldsymbol{\theta} ||^2 - \inf_{\boldsymbol{\theta} \in \Theta_1} n || \boldsymbol{Y} - \boldsymbol{\theta} ||^2$, em que \boldsymbol{Y} é média de \boldsymbol{Y} e $|| \boldsymbol{x} - \boldsymbol{z} ||$ é distância euclidiana entre \boldsymbol{x} e \boldsymbol{z} .

Neste exemplo, o valor-p pode ser definido como

$$p(\Theta_0; \boldsymbol{y}) = \sup_{\boldsymbol{\theta}^{\star} \in \Theta_0} \mathbb{P}_{\boldsymbol{\theta}^{\star}} \left(\inf_{\boldsymbol{\theta} \in \Theta_0} n || \bar{\boldsymbol{Y}} - \boldsymbol{\theta} ||^2 - \inf_{\boldsymbol{\theta} \in \Theta_1} n || \bar{\boldsymbol{Y}} - \boldsymbol{\theta} ||^2 \geq \inf_{\boldsymbol{\theta} \in \Theta_0} n || \bar{\boldsymbol{y}} - \boldsymbol{\theta} ||^2 - \inf_{\boldsymbol{\theta} \in \Theta_1} n || \bar{\boldsymbol{y}} - \boldsymbol{\theta} ||^2 \right),$$

 $\operatorname{com} y$, observação de Y.

Examinaremos agora um tipo de inconsistência do valor-p em que há dois conjuntos $\Theta'_0 \in \Theta_0$ tais que $\Theta'_0 \supset \Theta_0 \in p(\Theta'_0; y) < p(\Theta_0; y)$. Esse fenômeno ocorre sempre que o valor observado está fora de Θ_0 , ou seja, $\bar{y} \notin \Theta_0$.

Seja $\Theta_{a\epsilon}$ um subconjunto de Θ tal que $\Theta_{a\epsilon} = \{ \boldsymbol{\theta} \in \Theta : n || \boldsymbol{a} - \boldsymbol{\theta} ||^2 \ge \epsilon \}$ para todo $\boldsymbol{a} \in \mathbb{R}^k$. Uma característica importante de cada $\Theta_{a\epsilon}$ é que $\mathbb{R}^k - \Theta_{a\epsilon}$ descreve uma hiperesfera aberta de raio igual a inf $_{\boldsymbol{\theta} \in \Theta_{a\epsilon}} n || \boldsymbol{a} - \boldsymbol{\theta} || = \epsilon$.

Note que, se $\bar{\boldsymbol{y}} \notin \Theta_0$ existe um $\phi > 0$ tal que $\inf_{\boldsymbol{\theta} \in \Theta_0} n ||\bar{\boldsymbol{y}} - \boldsymbol{\theta}||^2 = \phi$. Portanto, existe um $0 < \epsilon < \phi$ de modo que se $\boldsymbol{\theta} \in \Theta_0$, então $n ||\bar{\boldsymbol{y}} - \boldsymbol{\theta}||^2 \ge \phi > \epsilon$ implica que $\boldsymbol{\theta} \in \Theta_{\bar{\boldsymbol{y}}\epsilon}$ (ver Figura 4.2). Logo, $\Theta_{\bar{\boldsymbol{y}}\epsilon} \supset \Theta_0$. Assim,

grupo	1	2	3	4	5
1	-0,51	2,49	1,01	0,29	-0,21
2	1,86	-0,07	-0,16	-0,20	0,30
3	-0,76	0,08	0,74	-0,08	-0,79
4	-0,92	0,86	2,00	0,94	-1,62
5	-0,58	-0,00	-0,68	-1,05	-0,54
6	0,56	0,25	-0,90	0,82	-1,56
7	0,54	-1,27	0,64	-0,54	-1,02
8	-0,81	0,06	0,95	-1,13	-0,20
9	-0,23	0,56	0,75	1,92	1,01
10	3,25	-0,16	0,33	-1,18	1,08
11	-1,00	0,53	-1,49	0,44	1,43
12	-0,07	1,61	-0,85	1,04	-0,31
13	-0,79	1,42	-0,15	-0,13	-0,88
14	1,71	0,16	2,11	-0,29	-0,66
15	-0,50	-0,20	-0,43	-0,39	0,35
16	-0,49	-0,82	0,22	2,12	0,90
17	1,09	0,36	-0,63	0,65	1,68
18	-1,53	0,59	-1,24	-1,27	-1,22
19	0,62	-0,01	0,45	-0,21	0,41
20	-0,07	-0,33	-0,40	-0,50	-0,77
21	-1,71	-0,80	-1,41	0,53	-0,40
22	-0,28	0,12	0,78	-0,88	-1,16
23	-0,15	-1,68	-1,32	-0,56	-2,49
24	1,18	0,79	1,61	0,07	1,03
25	-0,67	-0,35	-0,44	-1,95	-0,84
26	-0,74	-1,42	1,14	0,16	-0,48
27	-0,53	-0,92	-2,17	-0,42	1,23
28	0,49	-0,78	0,78	-1,41	1,82
29	1,60	0,15	0,19	-1,30	1,08
30	-0,41	-0,04	-1,66	-0,58	-0,82

Tabela 4.1: Dados simulados de Normais padrão independentes, divididos aleatoriamente em 30 grupos com 5 observações cada.

$$p(\Theta_{\bar{y}\epsilon}; y) = \sup_{\theta^* \in \Theta_{\bar{y}\epsilon}} \mathbb{P}_{\theta^*} \left(\inf_{\theta \in \Theta_{\bar{y}\epsilon}} n ||\bar{Y} - \theta||^2 - \inf_{\theta \in \mathbb{R}^k - \Theta_{\bar{y}\epsilon}} n ||\bar{Y} - \theta||^2 \ge \inf_{\theta \in \Theta_{\bar{y}\epsilon}} n ||\bar{y} - \theta||^2 \right)$$

$$= \sup_{\theta^* \in \Theta_{\bar{y}\epsilon}} \mathbb{P}_{\theta^*} \left(\inf_{\theta \in \Theta_{\bar{y}\epsilon}} n \|Y - \theta\|^2 - \inf_{\theta \in \mathbb{R}^k - \Theta_{\bar{y}\epsilon}} n \|Y - \theta\|^2 \ge \epsilon \right)$$

- $= \sup_{\theta^{*} \in \Theta_{\bar{y}\epsilon}} \mathbb{P}_{\theta^{*}} \left(\inf_{\theta \in \Theta_{\bar{y}\epsilon}} n \|\bar{Y} \theta\|^{2} \ge \epsilon \right)$ $= \sup_{\theta^{*} \in \Theta_{\bar{y}\epsilon}} \left[\mathbb{P}_{\theta^{*}} \left(\inf_{\theta \in \Theta_{\bar{y}\epsilon}} n \|\bar{Y} \theta\|^{2} \ge \epsilon, \bar{Y} \in \Theta_{\bar{y}\epsilon} \right) + \mathbb{P}_{\theta^{*}} \left(\inf_{\theta \in \Theta_{\bar{y}\epsilon}} n \|\bar{Y} \theta\|^{2} \ge \epsilon, \bar{Y} \notin \Theta_{\bar{y}\epsilon} \right) \right]$ $= \sup_{\theta^{*} \in \Theta_{\bar{y}\epsilon}} \left[0 + \mathbb{P}_{\theta^{*}} \left(\inf_{\theta \in \Theta_{\bar{y}\epsilon}} n \|\bar{Y} \theta\|^{2} \ge \epsilon, \bar{Y} \notin \Theta_{\bar{y}\epsilon} \right) \right].$

Perceba que, o conjunto

$$\left\{\phi \in \mathbb{R}^k : \inf_{\theta \in \Theta_{\bar{y}\epsilon}} n \|\phi - \theta\|^2 \ge \epsilon, \phi \notin \Theta_{\bar{y}\epsilon}\right\} = \left\{\phi \in \mathbb{R}^k - \Theta_{\bar{y}\epsilon} : \inf_{\theta \in \Theta_{\bar{y}\epsilon}} n \|\phi - \theta\|^2 \ge \epsilon\right\}.$$

Mas $\mathbb{R}^k - \Theta_{\bar{y}\epsilon}$ é uma hiperesfera aberta de raio $\inf_{\theta \in \Theta_{\bar{y}\epsilon}} n ||\bar{y} - \theta||^2 = \epsilon$. Assim,

$$\left\{\phi \in \mathbb{R}^{k} - \Theta_{\tilde{y}\epsilon} : \inf_{\theta \in \Theta_{\tilde{y}\epsilon}} n \|\phi - \theta\|^{2} \ge \epsilon\right\} = \left\{\phi \in \mathbb{R}^{k} - \Theta_{\tilde{y}\epsilon} : \inf_{\theta \in \Theta_{\tilde{y}\epsilon}} n \|\phi - \theta\|^{2} = \epsilon\right\}$$

Portanto,

$$p(\Theta_{\bar{y}\epsilon}; \boldsymbol{y}) \leq \sup_{\boldsymbol{\theta}^{*} \in \Theta_{\bar{y}\epsilon}} \mathbb{P}_{\boldsymbol{\theta}^{*}} \left(\inf_{\boldsymbol{\theta} \in \Theta_{\bar{y}\epsilon}} n \| \bar{\boldsymbol{Y}} - \boldsymbol{\theta} \|^{2} \geq \epsilon \right)$$

$$= \sup_{\boldsymbol{\theta}^{*} \in \Theta_{\bar{y}\epsilon}} \mathbb{P}_{\boldsymbol{\theta}^{*}} \left(\inf_{\boldsymbol{\theta} \in \Theta_{\bar{y}\epsilon}} n \| \bar{\boldsymbol{Y}} - \boldsymbol{\theta} \|^{2} = \epsilon \right)$$

$$= 0.$$

Logo, rejeitamos a hipótese $\mathcal{H}'_0 : \boldsymbol{\theta} \in \Theta_{\bar{y}\epsilon}$ e, portanto, também deveríamos rejeitar $\mathcal{H}_0 : \boldsymbol{\theta} \in \Theta_0$, pois $\Theta_0 \subset \Theta_{\bar{y}\epsilon}$. Concluímos que, se $\bar{y} \notin \Theta_0$, então sempre podemos formular uma hipótese mais geral com rejeição certa, o que nos leva a também rejeitar $\mathcal{H}_0 : \boldsymbol{\theta} \in \Theta_0$. Porém, o valor-p associado a \mathcal{H}_0 assume, na maioria dos casos, um valor maior que 0, gerando assim um problema de inconsistência, uma vez que o valor-p associado a uma hipótese mais geral \mathcal{H}'_0 foi rejeitada com valor-p igual a 0.

Na Figura 4.1 vemos exemplos de conjuntos $\Theta_{a\epsilon}$, para quando a dimensão de Θ é 1 (k = 1). Já na Figura 4.2 vemos um exemplo de $\Theta_{\tilde{y}\epsilon} \supset \Theta_0$, quando a dimensão de Θ é 2.

Com base nesses argumentos, deve-se tomar cuidado ao utilizar o valor-p para testar hipóteses nulas aninhadas. Não podemos, por exemplo, usar o valor-p de uma hipótese para testar outra hipótese encaixada. Como alternativa, utilizamos o valor-s proposto por Patriota (2013) visto que o valor-s não apresenta os problemas citados nesta seção.



Figura 4.1: $\Theta_{a\epsilon}$ para diferentes valores de a, sendo \bar{y} um deles.



Figura 4.2: $\Theta_{\bar{y}\epsilon}$ e seu subconjunto Θ_0 , com suas respectivas estimativas da razão de verossimilhanças $\epsilon e \phi$.
Capítulo 5

Valor-s

No Capítulo 4 abordamos os problemas de consistência lógica ao usarmos o valor-p como medida de evidência contra uma hipótese. De modo prático, se temos duas hipóteses $\mathcal{H}'_0: \boldsymbol{\theta} \in \Theta'_0 \in \mathcal{H}''_0: \boldsymbol{\theta} \in \Theta''_0$, tal que $\Theta'_0 \subseteq \Theta''_0$, não podemos afirmar que $p(\Theta'_0) \leq p(\Theta''_0)$. Friedman e Halpern (2001) definem o que seria uma medida de plausibilidade. Patriota (2013) definiu uma medida de evidência chamada valor-s, a qual satisfaz as propriedades descritas por Friedman e Halpern (2001). O valor-s é calculado com base na distância de uma hipótese proposta do evento observado, de modo que utilizamos a razão de verossimilhanças para medir tal distância.

A primeira definição do valor-s (ver Patriota, 2013) utiliza a razão de verossimilhanças assumindo que sua distribuição assintótica não depende do vetor de parâmetros. No entanto, no Capítulo 3, vimos que em modelos lineares mistos a distribuição da razão de verossimilhanças depende se o vetor de parâmetros é um ponto interior ou um ponto de borda do espaço paramétrico.

A seguir, apresentamos uma definição de valor-s considerando que a distribuição assintótica da estatística da razão de verossimilhanças pode depender do vetor de parâmetros. A distribuição assintótica será utilizada como uma aproximação.

Definição 1. Dizemos que $s(\cdot; y)$ é uma função de subconjuntos de Θ tal que

$$s(\Theta_0; \boldsymbol{y}) = \begin{cases} \sup\{(0, 1) : \Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_0 \neq \emptyset\} & \text{, se } \Theta_0 \neq \emptyset, \\ 0 & \text{, caso contrário,} \end{cases}$$
(5.1)

 $\begin{array}{l} tal \; que, \Lambda_{\alpha}(\boldsymbol{y}) = \left\{ \boldsymbol{\theta} \in \Theta : \: \lambda_{n}^{(obs)}(\{\boldsymbol{\theta}\}) \leq F_{\boldsymbol{\theta},1-\alpha} \: \right\}, \: F_{\boldsymbol{\theta},1-\alpha} \; \acute{e} \; o \; quantil \; de \; ordem \; 1-\alpha \; de \: \lambda_{n}(\{\boldsymbol{\theta}\}) \\ e \: \lambda_{n}^{(obs)}(\{\boldsymbol{\theta}\}) \: \acute{e} \; o \; valor \; observado \; de \: \lambda_{n}(\{\boldsymbol{\theta}\}). \end{array}$

Dizemos que um conjunto tem menos suporte, ou que é menos plausível, quando o valor-s está mais próximo de zero, e que um conjunto é mais plausível, ou que tem mais suporte, quando o valor-s está mais próximo de um. Pela definição 5.1, afirmamos que o θ_0 (verdadeiro θ) certamente não pertence (s = 0) ao conjunto vazio e certamente pertence (s = 1) ao Θ .

No modelo linear misto, conforme o modelo 2.1, a distribuição assintótica da razão de

verossimilhanças depende de θ_0 , podendo ser uma χ^2_{p+2} ou uma 0, $5\chi^2_{p+1}$ + 0, $5\chi^2_{p+2}$. Assim, a região de confiança assintótica é dada por

$$\Lambda_{\alpha}(\boldsymbol{y}) = \left\{\boldsymbol{\theta} \in \Theta^{(1)} : F_1(\lambda_n^{(obs)}(\{\boldsymbol{\theta}\})) \le 1 - \alpha\right\} \cup \left\{\boldsymbol{\theta} \in \Theta^{(2)} : F_2(\lambda_n^{(obs)}(\{\boldsymbol{\theta}\})) \le 1 - \alpha\right\},$$

em que, $\Theta^{(1)} = \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$ e $\Theta^{(2)} = \mathbb{R}^p \times \mathbb{R}_+ \times \{0\}$, F_1 é a função de distribuição acumulada de uma χ^2_{p+2} e F_2 é a função de distribuição acumulada de 0, $5\chi^2_{p+1} + 0$, $5\chi^2_{p+2}$.

Os Teoremas 3, 4 e 5 são resultados dessa dissertação. No Apêndice C, apresentamos os resultados deste trabalho compilados em um artigo escrito em parceria com Alexandre Galvão Patriota (orientador). Nesse artigo, outros resultados foram derivados, como, por exemplo, resultados sobre o nível de significância do valor-s.

Teorema 3. Considere o modelo 2.1 e assuma válidas as condições de regularidade (ver Apêndice A). Considere $\Theta^{(1)} e \Theta^{(2)}$ os conjuntos definidos na equação 5. Então, se $\Theta_0 \subset \Theta$

$$s(\Theta_0; \boldsymbol{y}) = 1 - F_i(\lambda_n^{(obs)}(\Theta_0)),$$

para $j = 1, 2, \lambda_n^{(obs)}(\Theta_0)$ é o valor observado de $\lambda_n(\Theta_0)$.

Demonstração. Ver Apêndice B.

Teorema 4. Considere o modelo 2.1 com q = 1 e condições de regularidade satisfeitas (ver Apêndice A). Seja $\Theta^{(1)} e \Theta^{(2)}$ os conjuntos definidos na equação 5. Se $\Theta_0 \cap \Theta^{(1)} \neq \emptyset e \Theta_0 \cap \Theta^{(2)} \neq \emptyset$, então

$$s(\Theta_0; \boldsymbol{y}) = \max\left\{s(\Theta_0 \cap \Theta^{(1)}; \boldsymbol{y}), s(\Theta_0 \cap \Theta^{(2)}; \boldsymbol{y})\right\}.$$

Demonstração. Ver Apêndice B.

Teorema 5. Considere o modelo 2.1 com q = 1 e condições de regularidade satisfeitas (ver Apêndice A). Então, $s(\emptyset; y) = 0$, $s(\Theta; y) = 1$ e $\Theta_1 \subset \Theta_2 \implies s(\Theta_1; y) \leq s(\Theta_2; y)$.

Demonstração. Ver Apêndice B.

Dessa forma, temos que o valor-s é uma medida de evidência monotônica, ou seja, se temos uma hipótese mais gerais que outra, o valor-s da hipótese menos geral será no máximo o valor-s da hipótese mais geral.

Capítulo 6

Estudo de simulação

Foram feitas simulações de Monte-Carlo para verificar de forma empírica o resultado dado em (3.7). Simulamos 3.000 réplicas de Monte-Carlo sob o modelo (2.1), com $\sigma_b^2 = 0$, n = 20, $n_i = 10$, i = 1, ..., n considerando oito formulações para $\boldsymbol{\theta}^{\top} = (\beta_0, \beta_1, \sigma_b^2, \sigma^2)$. Em cada uma das formulações $\sigma_b^2 = 0$, $\sigma^2 = 10$ ou $\sigma^2 = 4$, variando o $\boldsymbol{\beta}^{\top}$ entre (1, 10), (-1, -10), (-5, 5) e (5, -5). Os valores de x_{ij} foram gerados de uma distribuição U(0, 100). Obtivemos $\lambda_n(\{\boldsymbol{\theta}\})$ para cada réplica de Monte-Carlo e comparamos sua distribuição empírica com a mistura de qui-quadrados dada em (3.7) e a χ_4^2 (distribuição assintótica de $\lambda_n(\{\boldsymbol{\theta}\})$ quando $\boldsymbol{\theta}$ é ponto interior de Θ). No gráfico, chamaremos de $RV(\beta_0, \beta_1, \sigma_b^2, \sigma^2; \boldsymbol{y})$ a razão de verossimilhanças. Os resultados estão dispostos na Figura 6.1.



Figura 6.1: Função densidades de probabilidade da $\lambda_n(\{\theta\})$, em que $\sigma_b^2 = 0$.

Apêndice A

Definições e Condições de Regularidade

Logo a seguir, definimos derivadas direcionais conforme Vu e Zhou (1997), porém, com uma notação mais próxima de Giampaoli e Singer (2009).

Definição 2. Dizemos que C é um cone com vértice em θ_0 se para todo $x, y \in C$ então $\theta_0 + a(x - \theta_0) + b(y - \theta_0) \in C$, a, b > 0.

Perceba que Θ é um cone da forma

$$\Theta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^p \times \mathbb{R}_+ \times \Gamma : \boldsymbol{\theta} = \boldsymbol{\theta}_0 + \tilde{\theta}_1 \boldsymbol{d}_1 + \dots + \tilde{\theta}_k \boldsymbol{d}_k, \tilde{\theta}_j \in I_j \right\},\$$

em que $d_j \in \mathbb{R}^k$, j = 1, ..., k, são linearmente independentes e $I_j \subset \mathbb{R}$ são intervalos contendo zero.

Definição 3. Para cada direção d_j , j = 1, ..., k, definimos

$$D_{d_j} l_n(\boldsymbol{\theta}) = \lim_{t \to 0} \frac{l_n(\boldsymbol{\theta} + td_j) - l_n(\boldsymbol{\theta})}{t},$$
(A.1)

se $\tilde{\theta}_j$ é um ponto interior de I_j , ou

$$D_{d_j} l_n(\boldsymbol{\theta}) = \lim_{h \to 0, h + \tilde{\theta}_j \in I_j} D_{d_j} l_n(\boldsymbol{\theta} + h d_j), \tag{A.2}$$

se $\tilde{\theta}_i$ é um ponto de borda de I_i .

Para uma melhor compreensão das derivadas direcionais, observe a Figura (A.1).



Figura A.1: Derivadas direcionais da função $h(\theta) = -\theta^3 - \theta^2 + 2$, definida em $\theta \ge 0$. As derivadas direcionais são observadas nos pontos em que $\theta = 0$, $\theta = 0, 5 e \theta = 1,5$.

Usaremos neste trabalho um teorema apresentado por Vu e Zhou (1997). Reproduziremos algumas suposições e definições antes de utilizarmos os resultados de Vu e Zhou (1997) e Giampaoli e Singer (2009).

Seja $Y_1, ..., Y_n$ uma amostra aleatória do modelo (2.1). Considere $\boldsymbol{\theta}_0 = (\theta_{10}, ..., \theta_{k0})^{\top}$ o verdadeiro valor do parâmetro e a função de estimação de estimação

$$l_n(\boldsymbol{\theta}; \boldsymbol{Y}) = \sum_{i=1}^n l_n(\boldsymbol{\theta}; \boldsymbol{Y}_i)$$

Utilizando o conceito de derivadas direcionais dado na Definição 3, defina

$$\boldsymbol{S}_{n}(\boldsymbol{\theta}) = [\boldsymbol{D}_{d_{1}} \boldsymbol{l}_{n}(\boldsymbol{\theta}), \dots, \boldsymbol{D}_{d_{1}} \boldsymbol{l}_{n}(\boldsymbol{\theta})]^{\top}$$
(A.3)

e

$$\boldsymbol{F}_{n}(\boldsymbol{\theta}) = -[\boldsymbol{D}_{d_{i}}\boldsymbol{D}_{d_{i}}\boldsymbol{l}_{n}(\boldsymbol{\theta})]. \tag{A.4}$$

em que $[e_{il}]$ é o elemento da j-ésima linha e l-ésima coluna.

Seja $\boldsymbol{D}_n = E[\boldsymbol{S}_n(\boldsymbol{\theta}_0)\boldsymbol{S}_n^{\top}(\boldsymbol{\theta}_0)]$ e $\boldsymbol{G}_n = E[\boldsymbol{F}_n(\boldsymbol{\theta}_0)].$

Diremos que $[\langle \rangle_a]$ é um vetor cujo a-ésimo elemento é dado entre $\langle \rangle$ e $[\langle \rangle_{ab}]$ uma matriz tal que o valor entre $\langle \rangle$ ocupa a a-ésima linha e b-ésima coluna.

Considere as seguintes suposições:

- (A1) Existe uma vizinhança \mathcal{V} de θ_0 , em que $l_n(\theta)$ é contínua em $\mathcal{V} \cap \Theta$ e, além disso, as primeiras e segundas derivadas direcionais são finitas e contínuas em $\mathcal{V} \cap \Theta$.
- (A2) Dizemos que $\Omega \subset \Theta$ satisfaz (A2) se existe um cone fechado C_{Ω} com vértice em θ_0 tal que $\forall v \in \Omega$,

$$\inf_{\boldsymbol{w}\in\mathcal{C}_{\Omega}}\left|\boldsymbol{G}_{n}^{t/2}(\boldsymbol{w}-\boldsymbol{v})\right| \leq u(\boldsymbol{v})\left|\boldsymbol{G}_{n}^{t/2}(\boldsymbol{v}-\boldsymbol{\theta}_{0})\right| \tag{A.5}$$

e $\forall w \in C_{\Omega}$,

$$\inf_{\boldsymbol{v}\in\Omega} \left| \boldsymbol{G}_n^{t/2}(\boldsymbol{w}-\boldsymbol{v}) \right| \le \boldsymbol{v}(\boldsymbol{w}) \left| \boldsymbol{G}_n^{t/2}(\boldsymbol{w}-\boldsymbol{\theta}_0) \right|, \tag{A.6}$$

em que $G^{t/2} = (G^{1/2})^{\top}$, tal que $G^{1/2}G^{t/2} = G$. As funções reais u(v) e v(w) são definidas em Ω e C_{Ω} respectivamente, de modo que $u(v) \rightarrow 0$ quando $v \rightarrow \theta_0$ e $v(w) \rightarrow 0$ quando $w \rightarrow \theta_0$. Determine de forma arbitrária uma matriz ortogonal e não estocástica T_n . Assim, definimos

$$\tilde{C}_{\Omega_n}^{\theta_0} = \left\{ \tilde{\boldsymbol{\theta}} : \tilde{\boldsymbol{\theta}} = \boldsymbol{T}_n \boldsymbol{G}_n^{t/2} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0 \right), \boldsymbol{\theta} \in \boldsymbol{C}_{\Omega} \right\}.$$
(A.7)

Considere $\tau \subset \Theta$ outro conjunto fixado com seu respectivo cone fechado C_{τ} . C_{τ_n} é definido de forma análoga a C_{Ω_n} .

(A3) Para um conjunto $\Omega \subset \Theta$, existe um cone fechado $\tilde{C}_{\Omega}^{\theta_0}$ com vértice em 0 e não dependente de *n*, de modo que $\tilde{C}_{\Omega_n}^{\theta_0}$ coincide assintoticamente com $\tilde{C}_{\Omega}^{\theta_0}$ tal que, se $n \to \infty$,

$$\sup_{|\beta|=1} \left| \inf_{\theta \in \tilde{\mathcal{C}}_{\Omega_n}^{\theta_0}} |\beta - \theta| - \inf_{\theta \in \tilde{\mathcal{C}}_{\Omega}^{\theta_0}} |\beta - \theta| \right| \to 0.$$
(A.8)

- (B1) $E[S_n(\theta_0)] = 0$ e as matrizes D_n e G_n são finitas.
- (B2) $\lambda_{\min}(G_n) \to \infty$. A suposição (B2) garante que G_n é positivo definido para *n* suficientemente grande.
- (B3) $\sup_{\boldsymbol{\theta}^* \in N_n(A)} \left\| \boldsymbol{G}_n^{-1/2} \boldsymbol{F}_n(\boldsymbol{\theta}) \boldsymbol{G}_n^{-t/2} \boldsymbol{I}_p \right\| \stackrel{P}{\to} 0 \text{ em que}$

$$N_n(A) = \left\{ \boldsymbol{\theta} \in \Theta : (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \boldsymbol{G}_n(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \le A^2 \right\}$$
(A.9)

para algum A > 0 fixado.

- (B4) Existe uma matriz V positiva definida, tal que $\|G_n^{-1/2}D_nG_n^{-t/2} V\| \xrightarrow{P} 0$. Além disso, se (B2) também for satisfeito então D_n é positivo definido para n suficientemente grande.
- (B5) $D_n^{-1/2} S_n(\theta_0) \xrightarrow{\mathcal{D}} N(0, I_p).$

Apêndice B

Teoremas do Capítulo 5

B.1 Prova do Teorema 3

Demonstração. Para todo $\boldsymbol{\theta} \in \Theta_0 \subset \Theta^{(j)}$, a distribuição acumulada assintótica de $\lambda_n(\{\boldsymbol{\theta}\})$ é F_j . Portanto, quando $\Theta_0 \neq \emptyset$, temos que

$$\begin{split} \Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_0 &= \left\{ \boldsymbol{\theta} \in \Theta_0 \, : \, F_j(\lambda_n^{\mathrm{obs}}(\{\boldsymbol{\theta}\})) \leq 1 - \alpha \right\} \\ &= \left\{ \boldsymbol{\theta} \in \Theta_0 \, : \, 1 - F_j(\lambda_n^{\mathrm{obs}}(\{\boldsymbol{\theta}\})) \geq \alpha \right\}. \end{split}$$

Se houver um θ^* tal que $\inf_{\theta \in \Theta_0} \lambda_n(\{\theta\}) = \lambda_n(\{\theta^*\})$, então

$$\left\{ \alpha \in (0,1) : \left\{ \boldsymbol{\theta} \in \Theta_0 : 1 - F_j(\lambda_n^{\text{obs}}(\{\boldsymbol{\theta}\})) \ge \alpha \right\} \neq \emptyset \right\} = \left(0, \sup_{\boldsymbol{\theta} \in \Theta_0} \left(1 - F_j(\lambda_n^{\text{obs}}(\{\boldsymbol{\theta}\})) \right) \right]$$
$$= \left(0, \left(1 - \inf_{\boldsymbol{\theta} \in \Theta_0} F_j(\lambda_n^{\text{obs}}(\{\boldsymbol{\theta}\})) \right) \right]$$
$$= \left(0, \left(1 - F_j(\inf_{\boldsymbol{\theta} \in \Theta_0} \lambda_n^{\text{obs}}(\{\boldsymbol{\theta}\})) \right) \right]$$
$$= \left(0, \left(1 - F_j(\lambda_n^{\text{obs}}(\Theta_0)) \right) \right],$$

caso contrário, o intervalo é aberto. Finalizamos a prova, mostrando que

$$\sup\left\{\alpha\in(0,1): \left\{\boldsymbol{\theta}\in\Theta_0: 1-F_j(\lambda_n^{\mathrm{obs}}(\{\boldsymbol{\theta}\}))\geq\alpha\right\}\neq\emptyset\right\}=1-F_j(\lambda_n^{\mathrm{obs}}(\Theta_0)).$$

B.2 Prova do Teorema 4

Demonstração. Seja $\Theta_{0j} = \Theta_0 \cap \Theta^{(j)}$, com j = 1, 2. Então,

$$\begin{split} \Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_{0} &= (\Lambda_{\alpha}(\boldsymbol{y}) = \Theta_{01}) \cup (\Lambda_{\alpha}(\boldsymbol{y}) = \Theta_{02}) \\ &= \left\{ \boldsymbol{\theta} \in \Theta_{01} : F_{1}(\lambda_{n}^{\text{obs}}(\{\boldsymbol{\theta}\})) \leq 1 - \alpha \right\} \cup \left\{ \boldsymbol{\theta} \in \Theta_{02} : F_{2}(\lambda_{n}^{\text{obs}}(\{\boldsymbol{\theta}\})) \leq 1 - \alpha \right\} \\ &= \left\{ \boldsymbol{\theta} \in \Theta_{01} : 1 - F_{1}(\lambda_{n}^{\text{obs}}(\{\boldsymbol{\theta}\})) \leq \alpha \right\} \cup \left\{ \boldsymbol{\theta} \in \Theta_{02} : 1 - F_{2}(\lambda_{n}^{\text{obs}}(\{\boldsymbol{\theta}\})) \leq \alpha \right\}. \end{split}$$

Seja $A_0 = \{ \alpha \in (0,1) : \Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_0 \neq \emptyset \}$ e $A_j = \{ \alpha \in (0,1) : \Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_{0j} \neq \emptyset \}, j = 1, 2.$ Assim,

$$A_0 \subset A_1 \cup A_2 \implies \sup(A_0) \le \sup(A_1 \cup A_2) \le \max\{\sup(A_1), \sup(A_2)\}.$$

Perceba que, $A_1 \subset A_0$ e $A_2 \subset A_0$, então

$$\begin{cases} \sup(A_1) \le \sup(A_0) \\ \sup(A_2) \le \sup(A_0) \end{cases} \implies \sup(A_0) \ge \max\{\sup(A_1), \sup(A_2)\} \end{cases}$$

Como sup(A_0) = $s(\Theta_0; y)$ e sup(A_j) = $s(\Theta_0 \cap \Theta^{(j)}; y)$, j = 1, 2, então

$$s(\Theta_0; \boldsymbol{y}) = \max\{s(\Theta_0 \cap \Theta^{(1)}; \boldsymbol{y}), s(\Theta_0 \cap \Theta^{(2)}; \boldsymbol{y})\}.$$

B.3 Prova do Teorema 5

Demonstração. Por definição $s(\emptyset; y) = 0$. Seja $\Theta_1 \subset \Theta_2 \subset \Theta$ tal que $\Theta_{1j} = \Theta_1 \cap \Theta^{(1)}$ e $\Theta_{2j} = \Theta_2 \cap \Theta^{(j)}, j = 1, 2$. Pelo Teorema 4, temos que

$$\begin{cases} s(\Theta_1; \boldsymbol{y}) = \max\{s(\Theta_{11}; \boldsymbol{y}), s(\Theta_{12}; \boldsymbol{y})\} \\ s(\Theta_{2j}; \boldsymbol{y}) = \max\{s(\Theta_{21}; \boldsymbol{y}), s(\Theta_{22}; \boldsymbol{y})\}. \end{cases}$$

Para mostrar que $s(\Theta_1; y) \le s(\Theta_2; y)$ basta mostrar que $s(\Theta_{1j}; y) \le s(\Theta_{2j}; y)$, j = 1, 2. Pelo Teorema 3, sabemos que

$$\begin{cases} s(\Theta_{1j}; \boldsymbol{y}) = 1 - F_j(\lambda_n^{\text{obs}}(\Theta_{1j})) \\ s(\Theta_{2j}; \boldsymbol{y}) = 1 - F_j(\lambda_n^{\text{obs}}(\Theta_{2j})) \end{cases}.$$

Como F_j é uma função estritamente contínua, então $\Theta_{1j} \subset \Theta_{2j}$, implica que $\lambda_n^{\text{obs}}(\Theta_{2j}) \leq \lambda_n^{\text{obs}}(\Theta_{1j})$ e, portanto, $1 - F_j(\lambda_n^{\text{obs}}(\Theta_{2j})) \geq 1 - F_j(\lambda_n^{\text{obs}}(\Theta_{1j}))$, j = 1, 2. Assim, $s(\Theta_{1j}; \boldsymbol{y}) \leq s(\Theta_{2j}; \boldsymbol{y})$, j = 1, 2. Logo, $s(\Theta_1; \boldsymbol{y}) \leq s(\Theta_2; \boldsymbol{y})$.

Apêndice C

Artigo

A monotone frequentist measure of evidence for testing variance components in linear mixed models

Alexandre Galvão Patriota^{*} Jônatas de Oliveira Alves

Departamento de Estatística, IME, Universidade de São Paulo Rua do Matão, 1010, São Paulo/SP, 05508-090, Brazil

Abstract

This paper studies the s-value, a monotone frequentist measure of evidence, in a linear mixed model with a general parameter space that includes null variance components for the random effects. Under this nonregular parameter space, an asymptotic version of the s-value based on likelihood ratio statistics is defined and discussed under null variance components. Three sufficient regularity conditions to obtain the asymptotic distributions of the likelihood statistics are provided and some properties of the asymptotic s-value discussed. It is shown that the s-value satisfies logical consequences which are not satisfied by typical p-values and also that the test based on the s-value has significance level α , i.e., the maximum probability of the s-value not exceeding a fixed threshold α does not surpass α for each fixed $\alpha \in (0, 1)$. Finally, the theoretical results are applied to a famous Orthodontic growth data.

Key-words: Boundary of parameter space, Frequentist measure, Mixed

models, Monotonicity, Variance components, $S\mbox{-values}.$

^{*}Corresponding author; email: patriota@ime.usp.br

1 Introduction

Mixed models are typically applied in the analysis of repeated measures and, in particular, in longitudinal studies, since they naturally incorporate the positive dependence among the repeated observations of the same subject (sample or experimental unit). This dependence is modeled by means of random effects which have zero means and non-negative variances (Laird and Ware, 1982). It is common to test whether some of these variances are zero in order to simplify the model by removing the random effects with zero variances (at a fixed significance level). Since this is a nonregular problem (the parameter is on the boundary of the parameter space), the usual chisquared asymptotic distribution of the test statistics (e.g., likelihood ratio statistics) is not valid and a special treatment is required (Chernoff, 1954; Self and Liang, 1987; Drton, 2009).

Vu and Zhou (1997) derived a general asymptotic theory of likelihood ratio tests under nonstandard conditions where the parameters can be on the boundary of the parameter space. Their asymptotic theory requires eight regularity conditions on model quantities and on the null parameter space. The authors applied the proposed theory to a two-way nested random variance components model with covariates. Giampaoli and Singer (2009) applied this general asymptotic theory to a general linear mixed model which included the illustration presented in Vu and Zhou (1997) as a particular instance. From those results, asymptotic p-values can be computed to test whether some variance components are zero.

For the sake of illustration, consider the following random effect model

$$Y_{ij} = \beta + b_i + e_{ij},\tag{1.1}$$

where Y_{ij} is the *j*th observation from the *i*th experimental unity, β is the

populational mean, $b_i \stackrel{iid}{\sim} N(0, \gamma)$ is the random effect independent of the error term $e_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$, where γ is the random effect variance, σ^2 is the random error variance and $\stackrel{iid}{\sim}$ denotes "independent and identically distributed". For this model, the parameter space is $\Theta = \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_0$ and the parameter vector is $\theta = (\beta, \sigma^2, \gamma)$, where $\mathbb{R} = (-\infty, +\infty)$ is the set of real numbers, $\mathbb{R}_0 = [0, +\infty)$ and $\mathbb{R}_+ = (0, +\infty)$. Notice that the parameter space allows the randomeffect variance to be zero, namely, $\gamma = 0$. That is, the null random-effect variance hypothesis $\mathcal{H}_0 : \gamma = 0$, under which the model (1.1) reduces to the one mean model with independent response variables, lists parameter values on the boundary of the parameter space. For this reason, the standard asymptotic theory cannot be straightforwardly employed. The asymptotic distribution of the likelihood ratio statistics under nonstandard conditions derived by Vu and Zhou (1997), and its specialization studied by Giampaoli and Singer (2009), can be employed to test this hypothesis by means of (asymptotic) *p*-values based on likelihood ratio statistics.

In practice, however, practitioners may not be interested only in testing null random-effect variances but also in other types of hypotheses such as $\mathcal{H}'_0: (\beta, \gamma) = (\beta_0, 0)$ and $\mathcal{H}''_0: \beta = \beta_0$. Both hypotheses list elements on the boundary of Θ and, for this reason, a nonstandard asymptotic theory should be applied such as the one derived by Vu and Zhou (1997). Observe that there is a logical relationship among the aforementioned three hypotheses, namely, $\mathcal{H}'_0 \implies \mathcal{H}_0$ and $\mathcal{H}'_0 \implies \mathcal{H}''_0$. That is, the null hypothesis \mathcal{H}'_0 is the most restricted one among the threes and, therefore, we would expect more evidence against this null than against either \mathcal{H}_0 and \mathcal{H}''_0 from the same observed data. In general, let \mathcal{A} and \mathcal{B} be two hypotheses on Θ such that $\mathcal{A} \implies \mathcal{B}$ (i.e., if \mathcal{A} is true, then \mathcal{B} must be true), then we expect finding more evidence against the more restricted one, namely, \mathcal{A} , than \mathcal{B} . A measure of evidence (or support) that satisfies this feature is called to be *monotonic*. Schervish (1996) and Patriota (2013) presented examples for which *p*-values violate this monotonicity. The authors considered models with normal distributions and defined two (or more) hypotheses \mathcal{A} and \mathcal{B} on Θ such that $\mathcal{A} \implies \mathcal{B}$ but the corresponding *p*-values satisfy a non-expected relation

$$p$$
-value $(\mathcal{A}; \boldsymbol{y}) > p$ -value $(\mathcal{B}; \boldsymbol{y})$

for the same observed data \boldsymbol{y} rather than the one expected by the logical consequence

$$p$$
-value($\mathcal{A}; \boldsymbol{y}$) $\leq p$ -value($\mathcal{B}; \boldsymbol{y}$).

We present some numerical examples of this non-expected relation in a simple random effect model (1.1) by considering the hypotheses \mathcal{H}_0 , \mathcal{H}'_0 and \mathcal{H}''_0 , see example 4.1 in Section 4.

In order to overcome this problem, Patriota (2013) defined an alternative measure called (asymptotic) s-value that is based on (asymptotic) confidence regions built with likelihood ratio statistics. The smaller is the s-value, the larger is the distance of the null hypothesis from the observed data. Also, the s-value is a monotonic measure. In this paper, we extend the standard s-value definition to be also employed in cases where the parameter vector is on the boundary of the parameter space under a linear mixed model. In this context, the challenge is to find the asymptotic distribution of the likelihood ratio statistics for a simple hypothesis (not for the composite one as was developed by Giampaoli and Singer, 2009). Then, based on this nonstandard asymptotic distribution, we build a s-value version which is valid for testing also null variances for the random effects.

The rest of this paper is organized as follows. Section 2 introduces the linear mixed model. Section 3 presents the asymptotic distribution of the likelihood statistic for some very general hypotheses where the parameter vector is on the boundary of parameter space. Section 4 offers a numerical example to show that the *p*-value based on the asymptotic distribution is not monotonic. Section 5 defines the asymptotic *s*-value under model defined in Section 2 and presents some of its properties. Section 6 applies the theory in the famous Orthodontic growth data (Pinheiros and Bates, 2000) considering only male subjects. Finally, Section 7 ends the paper with some concluding remarks. For the sake of completeness, the reader can consult the eight nonstandard regularity conditions (Vu and Zhou, 1997) in Appendix A. The proofs and regularity conditions for the general linear mixed model are deferred to Appendix B.

2 Model

In this paper, we consider the following mixed model (Laird and Ware, 1982). Let $\mathbf{Y}_i = (Y_{1i}, \dots, Y_{1n_i})^{\top}$, $i = 1, \dots, n$, be random vectors such that

$$Y_i = X_i \beta + Z_i b_i + e_i, \qquad (2.2)$$

where n_i is the number of measurements in the *i*th experimental unit, X_i is a known design matrix $(n_i \times p)$ of the *i*th experimental unit with full column rank (its columns are linear independent), $\boldsymbol{\beta}$ is the vector of parameters $(p \times 1)$, Z_i is a known matrix $(n_i \times q)$, \boldsymbol{b}_i is the random effects vector $(q \times 1)$ and \boldsymbol{e}_i is the random error vector. We assume that $\boldsymbol{b}_i \stackrel{iid}{\sim} N_q(0, \boldsymbol{D})$ is independent of $\boldsymbol{e}_i \stackrel{ind}{\sim} N_{n_i} (0, \sigma^2 \boldsymbol{I}_{n_i})$, where \boldsymbol{D} is a nonnegative definite (n.n.d) matrix and \boldsymbol{I}_{n_i} is the identity matrix $(n_i \times n_i)$. Therefore, by the properties of the multivariate normal distribution,

$$\boldsymbol{Y}_{i} \stackrel{ind}{\sim} \mathrm{N}_{n_{i}}(\boldsymbol{X}_{i}\boldsymbol{\beta},\boldsymbol{\Sigma}_{i}), \quad i = 1,\ldots,n,$$

where $\Sigma_i = Z_i D Z_i^{\top} + \sigma^2 I_{n_i}$. Denote by $\gamma = \operatorname{vech}(D)^{\top} = (\gamma_{11}, \ldots, \gamma_{q1}, \gamma_{22}, \gamma_{22}$

 $\dots, \gamma_{q2}, \dots, \gamma_{q(q-1)}, \gamma_{qq})^{\top}$ the vector $(q(q+1)/2 \times 1)$ containing only elements of the lower triangular part of \boldsymbol{D} . The vector $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are not functionally related. The parameter space is $\Theta = \mathbb{R}^p \times \mathbb{R}_+ \times \Gamma$, where $\Gamma = \{\boldsymbol{\gamma} \in \mathbb{R}^r :$ \boldsymbol{D} is n.n.d matrix} with r = q(q+1)/2, and the parameter vector is $\boldsymbol{\theta} =$ $(\boldsymbol{\beta}^{\top}, \sigma^2, \boldsymbol{\gamma}^{\top})^{\top} \in \Theta$. We denote the dimension of the parameter space by k =p+r+1.

From this specification, the log-likelihood function is

$$\ell_n(\boldsymbol{\theta}) = c - \sum_{i=1}^n \left\{ \frac{1}{2} \log \det(\boldsymbol{\Sigma}_i) + \frac{1}{2} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \right\}, \quad (2.3)$$

where c is a constant term.

The maximum likelihood (ML) estimation of the parameter vector can be attained directly from this log-likelihood function or from a complete version considering the joint distribution of $(\mathbf{Y}_i, \mathbf{b}_i)$, $i = 1, \ldots, n$. Both are well established procedures and the reader can see more details in Laird and Ware (1982), Self and Liang (1987) and Demidenko (2004). In our numerical computations, we use the *lmer* function of the package *lme4* (R Core Team, 2020) written by Bates *et al.* (2015) to estimate the vector of parameters *via* ML approach.

3 Asymptotic distribution of the LR statistic

In this section, we present the asymptotic distribution of the likelihood ratio statistic under nonstandard regular conditions for the linear mixed model (2.2). The practitioner can test null hypotheses on Θ , under large samples, by computing the *p*-values based on this asymptotic distribution.

Typically, the following two general hypotheses are of interest

$$\mathcal{H}_0: \boldsymbol{\theta} \in \Theta_0 \text{ against } \mathcal{H}_1: \boldsymbol{\theta} \in \Theta_1, \tag{3.4}$$

where $\Theta_0 \cup \Theta_1 = \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$, that is, the sets Θ_0 and Θ_1 form a partition of Θ . The (-2 log) likelihood ratio statistic is defined by

$$\lambda_n(\Theta_0, \Theta_1) = 2 \left(\sup_{\boldsymbol{\theta} \in \Theta_1} \ell_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta} \in \Theta_0} \ell_n(\boldsymbol{\theta}) \right).$$
(3.5)

When ℓ_n is continuous and the dimension of Θ_0 is smaller than the dimension of Θ , the statistic (3.5) can be equivalently written as

$$\lambda_n(\Theta_0) = 2 \left(\sup_{\boldsymbol{\theta} \in \Theta} \ell_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta} \in \Theta_0} \ell_n(\boldsymbol{\theta}) \right).$$
(3.6)

Furthermore, if Θ_0 is an algebraic or a semi-algebraic subset of the parameter space, that is, if it is defined by polynomial equalities and inequalities in $\boldsymbol{\theta}$, then the limiting distribution of $\lambda_n(\Theta_0)$ always exists under some mild regularity conditions (Drton, 2009). In some special cases, e.g., when the set Θ_0 does not contain singularities or boundaries, the limiting distribution of $\lambda_n(\Theta_0)$ is a chi-squared with $s = k - \dim(\Theta_0)$ degrees-of-freedom, where $k = \dim(\Theta)$ (see Chernoff, 1954).

Vu and Zhou (1997) studied the asymptotic distribution of $\lambda_n(\Theta_0)$ considering a general estimating function for the case where the parameter space has elements on its boundary. As aforementioned, their development requires 8 regularity conditions on the restricted parameter spaces Θ_0 and Θ_1 , and on the first two directional derivatives of the estimating function (in our case, ℓ_n) to control the convergence of some quantities. For more details, see conditions (A1)–(A3) and (B1)–(B5) in Appendix A, see also Vu and Zhou (1997). Giampaoli and Singer (2009) showed that those conditions are valid for some special cases of the linear mixed model (2.2) and under some specific null hypotheses.

In what follows, we present the general theorem derived by Vu and Zhou (1997), which requires the notion of a cone. Recall that a subset $\mathcal{C} \subset \mathbb{R}^k$ is a

cone with vertex at $\boldsymbol{\theta}_0 \in \Theta$ if

$$\boldsymbol{\theta} \in \mathcal{C} \implies \{\boldsymbol{\theta}_0 + \alpha(\boldsymbol{\theta} - \boldsymbol{\theta}_0) : \alpha \in \mathbb{R}_0\} \subseteq \mathcal{C}.$$

Cones are very important to derive nonstandard asymptotics, since they express the local geometric of a subset of Θ at some point $\boldsymbol{\theta}_0 \in \Theta$ (Chernoff, 1954; Drton, 2009).

Theorem 3.1. Consider that $dim(\Theta_0) < k$ and assume valid the regularity conditions (A1)-(A3) and (B1)-(B5) described in the Appendix A. Then, for each fixed $\boldsymbol{\theta}_0 \in \Theta_0$,

$$\lambda_n(\Theta_0) \stackrel{d}{\longrightarrow} \inf_{oldsymbol{ heta}\in ilde{\mathcal{C}}_{\Theta_0}^{oldsymbol{ heta}_0}} |oldsymbol{N} - oldsymbol{ heta}|^2 - \inf_{oldsymbol{ heta}\in ilde{\mathcal{C}}_{\Theta}^{oldsymbol{ heta}_0}} |oldsymbol{N} - oldsymbol{ heta}|^2,$$

where $|\cdot|$ is the absolute value of a vector, " $\stackrel{d}{\longrightarrow}$ " stands for "converges in distribution to", **N** is a random vector following the standard multivariate normal distribution, and $\tilde{C}^{\theta_0}_{\Theta_0}$ and $\tilde{C}^{\theta_0}_{\Theta}$ are closed cones with vertex at $\mathbf{0} \in \mathbb{R}^k$ satisfying conditions (A2) and (A3).

Giampaoli and Singer (2009) showed that this asymptotic result is valid under model (2.2) with $q \leq 2$. In one of the cases, the authors considered that q = 1 and the null and alternative hypotheses to be $\mathcal{H}_0 : \gamma = 0$ and $\mathcal{H}_1 : \gamma > 0$, respectively. Under this setting, the null parameter space is $\Theta_0 = \mathbb{R}^p \times \mathbb{R}_+ \times \{0\}$. The authors showed that the required cones are $\tilde{C}_{\Theta_0}^{\theta_0} =$ $\mathbb{R}^p \times \mathbb{R} \times \{0\}$ and $\tilde{C}_{\Theta}^{\theta_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}_0$ for each $\theta_0 \in \Theta_0$. Under this specific setting, Giampaoli and Singer (2009) concluded that

$$\inf_{\boldsymbol{\theta}\in\tilde{c}_{\Theta_{0}}^{\boldsymbol{\theta}_{0}}}|\boldsymbol{N}-\boldsymbol{\theta}|^{2}-\inf_{\boldsymbol{\theta}\in\tilde{c}_{\Theta}^{\boldsymbol{\theta}_{0}}}|\boldsymbol{N}-\boldsymbol{\theta}|^{2}=N_{p+2}^{2}I(N_{p+2}\geq 0),$$

where $\mathbf{N} = (N_1, \ldots, N_{p+2})$ is a random vector following the standard mul-

tivariate normal distribution. That is, this asymptotic distribution is a 50-50% mixture of chisquared distributions with zero and one degrees-of-freedom: $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. The asymptotic *p*-value for testing this null hypothesis \mathcal{H}_0 is

$$p-value(\mathcal{H}_0, \boldsymbol{y}) = P\left(\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2 \ge \lambda_n\right), \qquad (3.7)$$

where λ_n is the observed value of $\lambda_n(\Theta_0)$.

We consider a small generalization of the above which regards also componentes of β in the null hypothesis. The following theorem is proved by considering three regularity conditions (C1)–(C3) which are deferred to Appendix B for the sake of readability. Roughly speaking, condition (C1) guarantees that some matrices related to X_i and Z_i have full rank, condition (C2) guarantees that the expectation of the second derivative of the log-likelihood function is somewhat limited and, finally, (C3) guarantees the convergence in probability of the second derivative of the log-likelihood function. See Appendix B for more details.

Theorem 3.2. Consider model (2.2) with q = 1 and conditions (C1)–(C3) in the Appendix B. Let the null hypothesis be $\mathcal{H}'_0 : \boldsymbol{\theta} \in \Theta'_0$ and the alternative one be $\mathcal{H}'_1 : \boldsymbol{\theta} \notin \Theta'_0$, where $\Theta'_0 = \{\boldsymbol{\theta} \in \Theta : \beta_j = 0, \forall j \in J, \gamma = 0\}$ and $J \subseteq \{1, \ldots, p\}$ a known set (which can be empty) of indexes with #(J) = g. Then, for each $\boldsymbol{\theta}_0 \in \Theta'_0$,

$$\lambda_n(\Theta'_0) \xrightarrow{d} \frac{1}{2}\chi_g^2 + \frac{1}{2}\chi_{g+1}^2.$$

Proof. See Appendix B.1.

Theorem 3.2 can be easily generalized in many ways, since we proved in Appendix B that, under (C1)–(C3), the conditions (A1), (A2), (B1)–(B5), stated in Vu and Zhou (1997), are valid for the general linear model (2.2)

when Θ_0 is (a union of sets) formed by polynomial equatilities or inequalities on $\boldsymbol{\theta}$. The only sensible condition that should be verified for each hypothesis is (A3), which involves cones. These generalizations are beyond the scope of this paper and will be further explored in details in a second paper.

The *p*-value for testing $\mathcal{H}'_0: \boldsymbol{\theta} \in \Theta'_0$ based on the asymptotic distribution of the likelihood ratio statistics, where Θ'_0 is defined in Theorem 3.2, is

$$p-value(\mathcal{H}'_0, \boldsymbol{y}) = P\left(\frac{1}{2}\chi_g^2 + \frac{1}{2}\chi_{g+1}^2 \ge \lambda'_n\right), \qquad (3.8)$$

where λ'_n is the observed value of $\lambda_n(\Theta'_0)$. In section 4, we present a numerical example in which the *p*-value based on the LR statistic violates monotonicity, i.e., for the very same data, we show that

$$\operatorname{p-value}(\mathcal{H}_0, \boldsymbol{y}) < \operatorname{p-value}(\mathcal{H}'_0, \boldsymbol{y}),$$

for null hypotheses such that $\mathcal{H}'_0 \implies \mathcal{H}_0$, where the opposite inequality would be expected by logical consequence. More specifically, we show that, by using *p*-values, the practitioner can find evidence to reject $\mathcal{H}_0 : \gamma = 0$ but at the same time cannot find evidence to reject $\mathcal{H}'_0 : (\beta, \gamma) = (\beta_0, 0)$ from the very same data. This feature would not be expected by the practitioner, since the logical consequence suggests the exact opposite. We shall see on Section 5 that an alternative classical measure called *s*-value satisfies the logical consequence for any null hypotheses the practitioner can be interested in testing.

4 *P*-values are not monotonic measures

It is not a trivial task to explain why a statistical conclusion is not in consonance with the expected logical consequences, specially when the statistical procedure is one of the most employed in scientific research such as the *p*-value. It is not rare to run a regression by using *lmer*, *glm* or *lm* functions and, after some statistical tests, come across to conclusions that contradict each other. Numerical examples of this feature where given by Schervish (1996) and Patriota (2013).

In this section, we present a numerical dataset that produces a 'contradictory' conclusion based on the asymptotic p-values (3.7) and (3.8). The same feature happens also for exact p-values, since the induced probability measure of any statistics varies with the null hypothesis. P-values computed under different null hypotheses are governed by different probabilities and, therefore, different 'metrics' are being applied to verify how discrepant the null hypotheses are from the observed data.

Definition 4.1. Let $Q(\cdot, \boldsymbol{y})$ be a function of subsets of Θ , where \boldsymbol{y} is the observed data. We say that Q is monotonic if, and only if, for all $\Theta_1 \subseteq \Theta_2 \subset \Theta$, we have $Q(\Theta_1, \boldsymbol{y}) \leq Q(\Theta_2, \boldsymbol{y})$.

One can also express the function Q in terms of null hypotheses \mathcal{H}_{01} : $\boldsymbol{\theta} \in \Theta_1$ and $\mathcal{H}_{02} : \boldsymbol{\theta} \in \Theta_2$ rather than their respective subsets of Θ , namely, $Q(\mathcal{H}_{01}, \boldsymbol{y})$ and $Q(\mathcal{H}_{02}, \boldsymbol{y})$, respectively. Notice that, if $\Theta_1 \subseteq \Theta_2$, then $\mathcal{H}_{01} \Longrightarrow$ \mathcal{H}_{02} and, therefore, Q would be monotonic if, and only if, $Q(\mathcal{H}_{01}, \boldsymbol{y}) \leq Q(\mathcal{H}_{02}, \boldsymbol{y})$.

The asymptotic *p*-value for testing a general null hypothesis $\tilde{\mathcal{H}}_0 : \boldsymbol{\theta} \in \tilde{\Theta}_0$ is typically defined as

$$\operatorname{p-value}(\tilde{\mathcal{H}}_0, \boldsymbol{y}) = \sup_{\boldsymbol{\theta}_0 \in \tilde{\Theta}_0} P_{\boldsymbol{\theta}_0}(\lambda_{\infty}(\tilde{\Theta}_0) \ge \lambda_n),$$

where $\lambda_{\infty}(\tilde{\Theta}_0)$ is the asymptotic random variable to which the likelihood ratio statistic $\lambda_n(\tilde{\Theta}_0)$ converges in distribution and λ_n is the observed likelihood ratio statistic. In Theorems 3.1 and 3.2 presented in Section 3, the distribution of $\lambda_{\infty}(\tilde{\Theta}_0)$ is the same for all $\theta_0 \in \tilde{\Theta}_0$, therefore, under those cases, the above p-value reduces simply to

p-value
$$(\tilde{\mathcal{H}}_0, \boldsymbol{y}) = P(\lambda_{\infty}(\tilde{\Theta}_0) \geq \lambda_n),$$

where P is the distribution of $\lambda_{\infty}(\tilde{\Theta}_0)$ which does not depend on specific values of $\boldsymbol{\theta}_0 \in \tilde{\Theta}_0$.

In the following example, we consider model (2.2) where p = q = 1 to illustrate the non-monotonicity of the above *p*-value.

Example 4.1. Let n = 35 and for each j = 1, ..., 5,

$$Y_{ij} = \beta + b_i + e_{ij},$$

where $b_i \stackrel{iid}{\sim} N(0, \gamma)$ is independent of $e_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$. The simulated data were generated by considering $\beta = 0$, $\gamma = 0$ and $\sigma^2 = 1$. We consider two null hypotheses, namely, $\mathcal{H}_0 : \gamma = 0$ and $\mathcal{H}'_0 : (\beta, \gamma) = (0, 0)$ under which the data are being simulated. Table 1 presents a simulated data from which we compute the observed likelihood statistics $\lambda_n = 3.07$ and $\lambda'_n = 3.13$ for \mathcal{H}_0 and \mathcal{H}'_0 , respectively, and their respective p-values: p-value($\mathcal{H}_0, \mathbf{y}$) = 0.040 and p-value($\mathcal{H}'_0, \mathbf{y}$) = 0.143. That is, we reject \mathcal{H}_0 at 5% significance level, however, we do not reject the more restricted one \mathcal{H}'_0 even at 10% significance level. This is not a problem with small samples.

We also consider $n \in \{35, 50, 150, 500, 1000\}$, $\sigma^2 \in \{0.05, 0.1, 0.5, 1, 2\}$ and the same null hypotheses \mathcal{H}_0 and \mathcal{H}'_0 to compute the empirical rate the event "p-value($\mathcal{H}_0, \mathbf{y}$) ≤ 0.04 and p-value($\mathcal{H}'_0, \mathbf{y}$) ≥ 0.1 " occurs from 50,000 Monte Carlo simulations. The results are presented in Table 2. As we can see, the event is somewhat rare, but not impossible: in one thousand MC samples we observe, in average, 5-6 events.

Table 1: Simulated data under the linear mixed model (2.2) and the nu	11
hypothesis \mathcal{H}'_0 : $(\beta, \gamma) = (0, 0)$ which generate conflicted conclusions on the	ıe
null hypotheses $\mathcal{H}_0: \gamma = 0$ and $\mathcal{H}'_0: (\beta, \gamma) = (0, 0)$.	

	Replications						
Ind	1	2	3	4	5		
1	-0.43	-0.45	-0.48	0.42	-0.42		
2	-1.19	-1.07	1.00	-1.11	1.26		
3	0.29	0.55	-0.87	-0.21	-0.53		
4	-1.62	0.83	0.78	-0.11	-0.08		
5	-0.31	-1.79	-0.24	0.12	-1.76		
6	-1.98	0.26	1.67	-0.06	0.41		
7	0.40	-1.50	-0.44	0.07	-0.01		
8	-0.28	1.49	1.44	1.19	0.21		
9	-0.53	0.87	-2.20	-0.99	-0.22		
10	-1.23	0.12	-0.26	-1.01	-0.03		
11	-1.24	2.09	1.24	-0.23	-0.45		
12	1.38	0.93	1.34	0.22	-1.21		
13	2.09	0.17	1.24	3.13	1.95		
14	-0.23	-0.61	-0.15	-0.20	1.14		
15	0.39	0.97	-0.53	-0.88	0.59		
16	-1.39	-0.41	1.08	-0.66	0.01		
17	-0.70	1.18	0.04	1.01	-0.33		
18	1.29	0.28	0.52	-0.22	1.58		
19	-0.39	-1.66	-2.09	-0.14	0.97		
20	-1.61	-0.49	-2.09	-0.90	-0.03		
21	-0.48	-1.16	0.97	0.51	0.54		
22	1.04	1.12	-0.75	1.87	2.27		
23	-1.09	-0.40	-0.18	0.55	-0.40		
24	0.32	-0.21	0.51	-0.77	0.18		
25	-0.33	0.77	0.37	-0.30	-0.41		
26	-0.21	1.56	-0.45	1.18	-0.66		
27	-1.50	-1.36	-0.12	-0.32	1.32		
28	0.66	-1.30	-1.25	1.05	0.80		
29	-0.51	1.08	-1.51	0.08	1.15		
30	0.36	1.68	-1.91	0.76	-0.56		
31	0.21	1.52	0.34	0.00	0.30		
32	-1.52	0.86	0.85	-0.14	-1.05		
33	-0.09	-0.41	0.51	1.40	-0.26		
34	0.52	0.95	-0.82	-0.36	-0.59		
35	-0.61	-0.36	0.35	-0.43	-0.56		

Notice that, under monotonicity, we should always observe "p-value($\mathcal{H}'_0, \boldsymbol{y}$) \leq p-value($\mathcal{H}_0, \boldsymbol{y}$)" whenever $\mathcal{H}'_0 \implies \mathcal{H}_0$. Therefore, it is sufficient to observe the event "p-value($\mathcal{H}_0, \boldsymbol{y}$) < p-value($\mathcal{H}'_0, \boldsymbol{y}$)" to build a contradictory conclu-

Table 2: Estimated probability (in percentage) from 50,000 Monte Carlo simulations of observing "p-value($\mathcal{H}_0, \mathbf{Y}$) ≤ 0.04 and p-value($\mathcal{H}'_0, \mathbf{Y}$) ≥ 0.1 ", where $\mathcal{H}_0: \gamma = 0$ and $\mathcal{H}'_0: (\beta, \gamma) = (0, 0)$ for $n_i = 5$ and $n \in \{35, 50, 150, 500, 1000\}$ and $\sigma^2 \in \{0.1, 0.5, 5.0, 10.0\}$.

				n		
		35	50	150	500	1000
	0.1	0.50%	0.53%	0.55%	0.50%	0.63%
	0.5	0.50%	0.57%	0.59%	0.62%	0.61%
σ^2	1.0	0.46%	0.44%	0.48%	0.58%	0.63%
	5.0	0.52%	0.49%	0.62%	0.62%	0.69%
	10.0	0.50%	0.51%	0.55%	0.59%	0.59%

sion for some significance level $\alpha \in (0, 1)$. The rate of the latter event is much more frequent than former event (it happens around 70% of the cases), as we can see in Table 3.

Table 3: Estimated probability (in percentage) from 50,000 Monte Carlo simulations of observing "p-value($\mathcal{H}_0, \mathbf{Y}$) < p-value($\mathcal{H}'_0, \mathbf{Y}$)", where $\mathcal{H}_0 : \gamma = 0$ and $\mathcal{H}'_0 : (\beta, \gamma) = (0, 0)$ for $n_i = 5$ and $n \in \{35, 50, 150, 500, 1000\}$ and $\sigma^2 \in \{0.1, 0.5, 5.0, 10.0\}.$

				n		
		35	50	150	500	1000
	0.1	70.69%	70.64%	71.62%	71.68%	71.61%
	0.5	70.60%	71.03%	71.60%	71.35%	71.56%
σ^2	1.0	70.48%	70.85%	71.32%	71.78%	71.82%
	5.0	70.96%	71.05%	71.22%	71.85%	72.35%
	10.0	70.68%	70.53%	71.65%	71.70%	71.36%

5 The *s*-value

In order to overcome this logical shortcoming of *p*-values, we resort to the *s*-value developed by Patriota (2013). In this section, we present its definition as a function of subsets of Θ and also discuss one of its fundamental properties: monotonicity. The main reason we advocate in favor of this measure is because it maintains the logical reasoning among the subsets of Θ without requiring any (subjective) Bayesian prior probabilities and it also preserves some desirable

frequentist features. In what follows, we define the s-value measure.

Definition 5.1. Let $s(\cdot, \boldsymbol{y})$ be a function of subsets of Θ such that

$$s(\Theta_0, \boldsymbol{y}) = \begin{cases} \sup\{\alpha \in (0, 1) : \Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_0 \neq \emptyset\}, & \text{if } \Theta_0 \neq \emptyset, \\ 0, & \text{if } \Theta_0 = \emptyset, \end{cases}$$

where $\Lambda_{\alpha}(\boldsymbol{y}) = \left\{ \boldsymbol{\theta} \in \Theta : \lambda_n^{obs}(\{\boldsymbol{\theta}\}) \leq F_{\boldsymbol{\theta},1-\alpha} \right\}$, $F_{\boldsymbol{\theta},1-\alpha}$ is the $1-\alpha$ quantile of the asymptotic distribution of $\lambda_n(\{\boldsymbol{\theta}\})$ and $\lambda_n^{obs}(\{\boldsymbol{\theta}\})$ is the value of $\lambda_n(\{\boldsymbol{\theta}\})$ for the observed data \boldsymbol{y} . The function $s(\cdot, \boldsymbol{y})$ is called 's-value measure' and $s(\Theta_0, \boldsymbol{y})$ is the s-value at Θ_0 .

We need to find the asymptotic distribution of $\lambda_n(\{\theta\})$, under model (2.2), for each fixed $\theta \in \Theta = \mathbb{R}^p \times \mathbb{R}_+ \times \Gamma$, where $\Gamma = \{\gamma \in \mathbb{R}^{q(q+1)/2} : \boldsymbol{D} \text{ is n.n.d. matrix}\}$. It is noteworthy that this asymptotic distribution is not a particular case of the ones already studied in Giampaoli and Singer (2009), but it is a particular case of the general Theorem 3.1. In this section, we specialize Theorem 3.1 to find the asymptotic distribution of $\lambda_n(\{\theta\})$ for each fixed $\boldsymbol{\theta} \in \Theta$ and then we provide a general expression for the *s*-value under the linear mixed model (2.2).

Theorem 5.1. Consider model (2.2) with q = 1 and conditions (C1)-(C3) in the Appendix B. Then, for each $\theta_0 \in \Theta$,

$$\lambda_n(\{\boldsymbol{\theta}_0\}) \stackrel{d}{\longrightarrow} W_{\boldsymbol{\theta}_0},$$

where

$$W_{\boldsymbol{\theta}_0} = \begin{cases} \chi_{p+2}^2, & \text{if } \boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+, \\ \frac{1}{2}\chi_{p+1}^2 + \frac{1}{2}\chi_{p+2}^2, & \text{if } \boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \{0\}. \end{cases}$$

Proof. See Appendix B.2.

For the general case, where q > 1, the analyst should find a partition

 $\Theta^{(1)}, \ldots, \Theta^{(m)}$ of Θ such that $\lambda_n(\{\boldsymbol{\theta}_0\}) \xrightarrow{d} W^{(j)}_{\boldsymbol{\theta}_0}$ for all $\boldsymbol{\theta}_0 \in \Theta^{(j)}, j = 1, \ldots, m$. This analysis is beyond the scope of this work and will be carefully explored in further papers. The theorem is also valid for the case where there is no covariate or intercept, that is, p = 1 or p = 0.

The asymptotic distribution of $\lambda_n(\{\theta_0\})$ is continuous but it is not the same for each $\theta_0 \in \Theta$. This feature affects the asymptotic confidence region, since the likelihood ratio statistic converges in distribution to either χ^2_{p+2} or $\frac{1}{2}\chi^2_{p+1} + \frac{1}{2}\chi^2_{p+2}$ depending on θ_0 . Therefore, the asymptotic confidence region is the union of two regions, namely,

$$\Lambda_{\alpha}(\boldsymbol{y}) = \left\{ \boldsymbol{\theta} \in \Theta^{(1)} : F_1(\lambda_n^{obs}(\{\boldsymbol{\theta}\})) \le 1 - \alpha \right\} \cup \left\{ \boldsymbol{\theta} \in \Theta^{(2)} : F_2(\lambda_n^{obs}(\{\boldsymbol{\theta}\})) \le 1 - \alpha \right\},$$
(5.9)

where $\Theta^{(1)} = \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$ and $\Theta^{(2)} = \mathbb{R}^p \times \mathbb{R}_+ \times \{0\}$, F_1 is the cumulative distribution of χ^2_{p+2} and F_2 is the cumulative distribution of $\frac{1}{2}\chi^2_{p+1} + \frac{1}{2}\chi^2_{p+2}$. From this confidence region (5.9), we derive the properties of the *s*-value measure under the linear mixed model (2.2) with q = 1. For the sake of readability, proofs of the Theorems below are deferred to the Appendix C.

Theorem 5.2. Consider model (2.2) with q = 1 and conditions (C1)-(C3) in the Appendix B. Let $\Theta^{(1)}$ and $\Theta^{(2)}$ be the sets defined in Equation (5.9). If $\Theta_0 \subseteq \Theta^{(j)}$ is nonempty, then

$$s(\Theta_0, \boldsymbol{y}) = 1 - F_j(\lambda_n^{obs}(\Theta_0))$$

for j = 1, 2, where $\lambda_n^{obs}(\Theta_0)$ is the observed (-2 log) likelihood ratio statistics. *Proof.* See Appendix C.1

Theorem 5.2 says how to compute the *s*-value at Θ_0 if all elements of Θ_0 are in one and only one of either $\Theta^{(1)}$ or $\Theta^{(2)}$. For instance, this result can be used to test the null hypothesis about the random effect variance $\mathcal{H}_0: \gamma = 0$, since the respective null parameter space is $\Theta_0 = \Theta^{(1)}$. If the *s*-value is smaller than a fixed threshold $\alpha \times 100\%$ (e.g., 5%), then it is required at least $(1-\alpha) \times 100\%$ confidence level to have at least one element of $\Theta^{(1)}$ in the confidence region. Next theorem guarantees that the limiting test size of the asymptotic *s*-value is not larger than the respective significance level. Based on this feature, the analyst can claim rejection of \mathcal{H}_0 at (an asymptotic) $\alpha \times 100\%$ significance level.

Theorem 5.3. Consider model (2.2) with q = 1 and conditions (C1)-(C3) in the Appendix B. Let $\Theta^{(1)}$ and $\Theta^{(2)}$ be the sets defined in Equation (5.9). If $\Theta_0 \subseteq \Theta^{(j)}$ is nonempty, then

$$\sup_{\boldsymbol{\theta}\in\Theta_0}\lim_{n\to\infty}P_{\boldsymbol{\theta}}\left(s(\Theta_0,\boldsymbol{Y})\leq\alpha\right)\leq\alpha$$

for each $\alpha \in (0, 1)$.

Proof. See Appendix C.2

Theorem 5.2 cannot be used to compute the *s*-value at $\Theta_0 = \{ \boldsymbol{\theta} \in \Theta : \gamma \leq 1 \}$, since some elements of Θ_0 are in $\Theta^{(1)}$ and others are in $\Theta^{(2)}$, i.e., $\Theta_0 = \Theta^{(1)} \cup \Theta_{02}$, where $\Theta_{02} = \{ \boldsymbol{\theta} \in \Theta : 0 < \gamma \leq 1 \} \subset \Theta^{(2)}$. Next, we show how to compute the *s*-value for the general case.

Theorem 5.4. Consider model (2.2) with q = 1 and conditions (C1)-(C3) in Apendix B. Let $\Theta^{(1)}$ and $\Theta^{(2)}$ be the sets defined in Equation (5.9). If $\Theta_0 \subseteq \Theta$, then

$$s(\Theta_0, \boldsymbol{y}) = \max\{s(\Theta_0 \cap \Theta^{(1)}, \boldsymbol{y}), s(\Theta_0 \cap \Theta^{(2)}, \boldsymbol{y})\}.$$

Proof. See Appendix C.3.

Theorem 5.5. Consider model (2.2) with q = 1 and conditions (C1)-(C3) in the Appendix B. Then, $s(\Theta, \boldsymbol{y}) = 1$, $s(\emptyset, \boldsymbol{y}) = 0$ and for any sets $\Theta_1 \subseteq \Theta_2 \subseteq$ Θ , we have $s(\Theta_1, \boldsymbol{y}) \leq s(\Theta_2, \boldsymbol{y})$.

Proof. See Appendix C.4.

Theorems 5.2–5.5 provide the main tools to compute the *s*-value for any subset of $\Theta = \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_0$. Theorem 5.5 shows that this *s*-value is a monotonic measure over the subsets of Θ and, unlike the *p*-values, the *s*-value can prevent conflicting conclusions. Next theorem generalizes Theorem 5.3 to be applied to any subset of Θ , that is, under the linear mixed model (2.2) with q = 1, the maximum probability of the *s*-value not exceeding a fixed threshold α cannot surpass α for all $\alpha \in (0, 1)$.

Theorem 5.6. Consider model (2.2) with q = 1 and conditions (C1)-(C3) in the Appendix B. If $\Theta_0 \subseteq \Theta$ is nonempty, then

$$\sup_{\boldsymbol{\theta}\in\Theta_0}\lim_{n\to\infty}P_{\boldsymbol{\theta}}\left(s(\Theta_0,\boldsymbol{Y})\leq\alpha\right)\leq\alpha$$

for each $\alpha \in (0, 1)$.

Proof. See Appendix C.5

Therefore, the test of a general null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta} \in \Theta_0$ with the rejection rule: "reject \mathcal{H}_0 if $s(\Theta_0, \boldsymbol{y}) \leq \alpha$ " will have significance level α . This means that the *s*-value respects the frequentist paradigm of controlling the maximum type I error.

6 Orthodontic grouth data

In this section, we consider the Orthodontic grouth data (Pinheiros and Bates, 2000) also available in the *nlme* library R Core Team (2020). The details provided by the R documentation follows "Investigators at the University of North Carolina Dental School followed the growth of 27 children (16 males, 11 females) from age 8 until age 14. Every two years they measured the distance

between the pituitary and the pterygomaxillary fissure, two points that are easily identified on x-ray exposures of the side of the head.". In this section, we consider only males patients and the following model

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + b_i + e_{ij},$$

where Y_{ij} is the *j*th measurement (j = 1, 2, 3, 4) of the distance from the pituitary to the pterygomaxillary fissure (mm) for the *i*th male individual (i = 1, ..., 16), and x_{1ij} is the age of the *i*th male individual in the *j*th measurement. We assume that the random effect $b_i \stackrel{iid}{\sim} N(0, \gamma)$ is independent of the error term $e_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$.

The following two null hypotheses will be tested, namely,

- $\mathcal{H}_0: \gamma = 0$ against $\mathcal{H}_1: \gamma > 0$; and
- $\mathcal{H}'_0: (\beta_0, \gamma) = (16, 0) \text{ against } \mathcal{H}'_1: (\beta_0, \gamma) \neq (16, 0).$

By using the asymptotic distributions derived in Theorem 3.2, we obtain the asymptotic p-values for both cases:

• for \mathcal{H}_0 , then g = 0 and

p-value
$$(\mathcal{H}_0, \boldsymbol{y}) = P\left(\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2 \ge 16.405\right) = 0.0000256$$

• for \mathcal{H}'_0 , then g = 1 and

p-value
$$(\mathcal{H}'_0, \boldsymbol{y}) = P\left(\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2 \ge 16.462\right) = 0.0001579.$$

Although those *p*-values are very small, p-value($\mathcal{H}'_0, \boldsymbol{y}$) is 6 times larger than p-value($\mathcal{H}_0, \boldsymbol{y}$). However, as $\mathcal{H}'_0 \implies \mathcal{H}_0$, we expect to have observed p-value($\mathcal{H}'_0, \boldsymbol{y}$) \leq p-value($\mathcal{H}_0, \boldsymbol{y}$). How can we explain to a practitioner that we have observed evidence to reject \mathcal{H}_0 at 0.01% significance level but they are not enough to reject \mathcal{H}'_0 at the same significance level?

From Theorem 5.1, we can compute the respective s-values

$$s(\mathcal{H}_0, \boldsymbol{y}) = P\left(\frac{1}{2}\chi_3^2 + \frac{1}{2}\chi_4^2 \ge 16.405\right) = 0.00172$$

and

$$s(\mathcal{H}'_0, \boldsymbol{y}) = P\left(\frac{1}{2}\chi_3^2 + \frac{1}{2}\chi_4^2 \ge 16.462\right) = 0.00168,$$

which satisfy the expected relation (monotonicity) as predicted by the theory.

7 Concluding remarks

This paper introduced the s-value measure under a linear mixed model with one random effect. This measure satisfies the expected logical consequences not met by typical p-values, namely, for two null hypotheses \mathcal{A} and \mathcal{B} such that $\mathcal{A} \implies \mathcal{B}$, we should find more evidence against \mathcal{A} than against \mathcal{B} with the same observed data. We provide three regularity conditions sufficient to compute the s-value through cumulative distributions of mixtures of chisquared distributions. From those conditions, we also showed that the proposed svalue also respects the frequentist criterion of controlling the maximum type I error. We recommend the use of the s-value as a measure of evidence to avoid eventual contradictory conclusions in testing null hypotheses.

8 Acknowledgements

This work was partially supported by CNPq and FAPESP (2018/21934-5).

References

- Bates, D. Mächler, M., Bolker B., Walker, S. (2015). Fitting linear mixedeffects models using lme4, *Journal of Statistical Software*, **67**, 1–48.
- Chernoff, H. (1954). On the distribution of the likelihood ratio, The Annals of Mathematical Statistics, 25, 573–578.
- Demidenko, E. (2004). Mixed Models Theory and Applications. John Wiley & Sons, Hoboken, NJ. ISBN 0-471-60161-6.
- Drton, M. (2009). Likelihood ratio tests and singularities, The Annals of Statistics, 37, 979–1012.
- Geyer, C.J. (1994). On the asymptotics of constrained M-estimation, The Annals of Statistics, 22, 1993–2010.
- Giampaoli, V., Singer, J. M. (2009). Likelihood ratio tests for variance components in linear mixed models, *Journal of Statistical Planning and Inference*, 139, 1435–1448.
- Laird, N.M., Ware, J.H. (1982). Random-effects models for longitudinal data, Biometrics, 38, 963–974.
- Patriota, A.G. (2011). A note on influence diagnostics in nonlinear mixedeffects elliptical models, *Computational Statistics & Data Analysis*, 55, 218– 225
- Patriota, A.G. (2013). A classical measure of evidence for general null hypotheses, *Fuzzy Sets and Systems*, 233, 74–88.
- Pinheiro, J.C., Bates, D.M. (2000). Mixed-Effects Models in S and S-PLUS, Springer, New York. (Appendix A.17)

- R Core Team (2020). R: A language and environment for statistical computing.R Foundation for Statistical Computing, Vienna, Austria.
- Self, S.G., Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal* of the American Statistical Association, 82, 605–610.
- Schervish, M.J. (1996). P values: what they are and what they are not, The American Statistician, 50, 203–206.
- Stram, D.O., Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model, *Biometrics*, 50, 1171–1177
- Vu, H.T.V, Zhou, S. (1997). Generalization of likelihood ratio tests under nonstandard conditions, *The Annals of Statistics*, 25, 897–916.

A Regularity Conditions

In this appendix, we replicate the regularity conditions presented in Vu and Zhou (1997) required in Theorem 3.1. For more details, we refer the reader to Vu and Zhou (1997) and Giampaoli and Singer (2009). In the following, consider that ℓ_n is the log-likelihood function (2.3) and Θ is the parameter space of model (2.2).

Before starting the main definitions, notice that for each $\theta_0 \in \Theta$ there exist k real (half closed or open) intervals containing 0, namely, I_1, \ldots, I_k and k linearly independent unit vectors d_1, \ldots, d_k such that

$$\Theta = \{ \boldsymbol{\theta} \in \mathbb{R}^k : \boldsymbol{\theta} = \boldsymbol{\theta}_0 + \tilde{\theta}_1 \boldsymbol{d}_1 + \ldots + \tilde{\theta}_k \boldsymbol{d}_k, \ \tilde{\theta}_j \in I_j, \ j = 1, \ldots, k \} = \mathbb{R}^p \times \mathbb{R}_+ \times \Gamma,$$

where, under model (2.2), k = p + 1 + q(q+1)/2. Therefore, without loss of generality, we can define directional derivatives for each $\boldsymbol{\theta}$ in terms of I_1, \ldots, I_k

and $d_1, ..., d_k$.

Definition A.1. Let $d_j \in \mathbb{R}^k$, j = 1, ..., k, be k linearly independent unit vectors and let I_j , be real (half closed or open) intervals containing 0.

The directional derivative of $\ell_n(\boldsymbol{\theta})$ with respect to $\theta_j \in I_j$ in the direction of \boldsymbol{d}_j is

$$D_{d_j}\ell_n(\boldsymbol{\theta}) = \lim_{t \to 0} \frac{\ell_n(\boldsymbol{\theta} + t\boldsymbol{d}_j) - \ell_n(\boldsymbol{\theta})}{t},$$

if $\tilde{\theta}_j$ is an interior of the real interval I_j , and

$$D_{\boldsymbol{d}_j}\ell_n(\boldsymbol{\theta}) = \lim_{h \to 0, h + \tilde{\theta}_j \in I_j} D_{\boldsymbol{d}_j}\ell_n(\boldsymbol{\theta} + h\boldsymbol{d}_j),$$

if $\tilde{\theta}_j$ is on the boundary of I_j .

The direction derivative of $D_{d_j}\ell_n(\boldsymbol{\theta})$ with respect to θ_l in the direction d_l is

$$D_{\boldsymbol{d}_l} D_{\boldsymbol{d}_j} \ell_n(\boldsymbol{\theta}) = \lim_{t \to 0} \frac{D_{\boldsymbol{d}_j} \ell_n(\boldsymbol{\theta} + t\boldsymbol{d}_l)) - D_{\boldsymbol{d}_j} \ell_n(\boldsymbol{\theta})}{t}$$

if $\tilde{\theta}_l$ is an interior of the real interval I_l , and

$$D_{\boldsymbol{d}_l} D_{\boldsymbol{d}_j} \ell_n(\boldsymbol{\theta}) = \lim_{h \to 0, h + \tilde{\theta}_l \in I_l} D_{\boldsymbol{d}_l} D_{\boldsymbol{d}_j} \ell_n(\boldsymbol{\theta} + h \boldsymbol{d}_l),$$

if $\tilde{\theta}_l$ is on the boundary of I_l .

If $\boldsymbol{\theta}$ is an interior point of Θ , then the direction derivatives coincide with the usual directional derivatives. The vector of directional derivatives is

$$\mathbf{S}_n(\boldsymbol{\theta}) = \boldsymbol{T}^{\top} [D_{\mathbf{d}_1} l_n(\boldsymbol{\theta}), \dots, D_{\mathbf{d}_k} l_n(\boldsymbol{\theta})]^{\top}$$

and

$$\mathbf{F}_{n}(\boldsymbol{\theta}) = -\boldsymbol{T}^{\top} \begin{pmatrix} D_{\mathbf{d}_{1}} D_{\mathbf{d}_{1}} l_{n}(\boldsymbol{\theta}) & \dots & D_{\mathbf{d}_{1}} D_{\mathbf{d}_{k}} l_{n}(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ D_{\mathbf{d}_{k}} D_{\mathbf{d}_{1}} l_{n}(\boldsymbol{\theta}) & \dots & D_{\mathbf{d}_{k}} D_{\mathbf{d}_{k}} l_{n}(\boldsymbol{\theta}) \end{pmatrix} \boldsymbol{T},$$

where $\boldsymbol{T} = (\boldsymbol{d}_1 \ \boldsymbol{d}_2 \ \dots \ \boldsymbol{d}_k).$

Let $D_{n,\theta} = E_{\theta}[S_n(\theta)S_n^{\top}(\theta)]$ and $G_{n,\theta} = E_{\theta}[F_n(\theta)]$ be the expectations of $S_n(\theta)S_n^{\top}(\theta)$ and $F_n(\theta)$ for each $\theta \in \Theta$. The matrices $D_{n,\theta}$ and $F_{n,\theta}$ depend on θ but the authors wrote as if there was a 'true value' θ_0 and, for this reason, they omitted the index. However, as we do not know the value of θ_0 , here we simply state the conditions for each $\theta_0 \in \Theta$, which is more precise and somewhat equivalent.

Assume, for each $\boldsymbol{\theta}_0 \in \Theta_0$, that

- (A1) There exists a neighborhood \mathcal{N} of $\boldsymbol{\theta}_0$, where ℓ_n is continuous on $\mathcal{N} \cap \Theta$ and the first two directional derivatives exist, are finite and continuous on $\mathcal{N} \cap \Theta$.
- (A2) A subset $\Omega \subset \Theta$ satisfies (A2) if there exists a closed cone \mathcal{C}_{Ω} with vertex at $\boldsymbol{\theta}_0$ such that $\forall \boldsymbol{v} \in \Omega$,

$$\inf_{\boldsymbol{w}\in\mathcal{C}_{\Omega}}\left|\boldsymbol{G}_{n,\boldsymbol{\theta}_{0}}^{\top/2}(\boldsymbol{w}-\boldsymbol{v})\right|\leq u(\boldsymbol{v})\left|\boldsymbol{G}_{n,\boldsymbol{\theta}_{0}}^{\top/2}(\boldsymbol{v}-\boldsymbol{\theta}_{0})\right|$$

and $\forall \boldsymbol{w} \in \mathcal{C}_{\Omega}$,

$$\inf_{\boldsymbol{v}\in\Omega} \left| \boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{\top/2}(\boldsymbol{w}-\boldsymbol{v}) \right| \leq v(\boldsymbol{w}) \left| \boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{\top/2}(\boldsymbol{w}-\boldsymbol{\theta}_0) \right|$$

where $\mathbf{G}^{\top/2} = (\mathbf{G}^{1/2})^{\top}$ is such that $\mathbf{G}^{1/2}\mathbf{G}^{\top/2} = \mathbf{G}$, the functions u : $\Omega \to \mathbb{R}$ and $v : \mathcal{C}_{\Omega} \to \mathbb{R}$ are such that $u(\mathbf{v}) \to 0$ as $\mathbf{v} \to \boldsymbol{\theta}_0$ and $v(\mathbf{w}) \to 0$ as $\mathbf{w} \to \boldsymbol{\theta}_0$.

Let \mathbf{T}_n be an arbitrary nonstochastic orthogonal matrix and define

$$\tilde{\mathcal{C}}_{\Omega_n}^{\boldsymbol{\theta}_0} = \left\{ \tilde{\boldsymbol{\theta}} : \tilde{\boldsymbol{\theta}} = \boldsymbol{T}_n \boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{\top/2} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0 \right), \boldsymbol{\theta} \in \mathcal{C}_{\Omega} \right\}.$$
(A.10)

(A3) A subset $\Omega \subset \Theta$ satisfies (A3) if there exists a closed cone $\tilde{\mathcal{C}}_{\Omega}^{\theta_0}$ with

vertex at $\mathbf{0} \in \mathbb{R}^k$, not depending on n, such that the sets $\tilde{\mathcal{C}}_{\Omega_n}^{\boldsymbol{\theta}_0}$, $n \geq 1$, asymptotically coincide with $\tilde{\mathcal{C}}_{\Omega}^{\boldsymbol{\theta}_0}$ in the sense that, as $n \to \infty$,

$$\sup_{|\boldsymbol{\beta}|=1} \left| \inf_{\boldsymbol{\theta} \in \tilde{\mathcal{C}}_{\Omega_n}^{\boldsymbol{\theta}_0}} |\boldsymbol{\beta} - \boldsymbol{\theta}|^2 - \inf_{\boldsymbol{\theta} \in \tilde{\mathcal{C}}_{\Omega}^{\boldsymbol{\theta}_0}} |\boldsymbol{\beta} - \boldsymbol{\theta}|^2 \right| \to 0.$$
(A.11)

- (B1) $E_{\boldsymbol{\theta}_0}[\boldsymbol{S}_n(\boldsymbol{\theta}_0)] = \boldsymbol{0}$ and the matrices $\boldsymbol{D}_{n,\boldsymbol{\theta}_0}$ and $\boldsymbol{G}_{n,\boldsymbol{\theta}_0}$ are finite.
- (B2) The minimum eigenvalue of G_{n,θ_0} increases to infinity as $n \to \infty$.
- (B3) For any fixed A > 0, $\sup_{\boldsymbol{\theta} \in N_n^{\boldsymbol{\theta}_0}(A)} \left| \left| \mathbf{G}_{n,\boldsymbol{\theta}_0}^{-1/2} \mathbf{F}_n(\boldsymbol{\theta}) \mathbf{G}_{n,\boldsymbol{\theta}_0}^{-\top/2} \mathbf{I}_k \right| \right| \xrightarrow{P_{\boldsymbol{\theta}_0}} 0$ where $||\cdot||$ is the sum of the absolute values of the elements of a matrix, " $\xrightarrow{P_{\boldsymbol{\theta}_0}}$ " stands for "convergence in probability" according to the measure $P_{\boldsymbol{\theta}_0}$ and

$$N_n^{\boldsymbol{\theta}_0}(A) = \left\{ \boldsymbol{\theta} \in \Theta : (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{G}_{n,\boldsymbol{\theta}_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \le A^2 \right\}.$$

- (B4) There exists a matrix \boldsymbol{V} positive definite such that $\left| \left| \boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{-1/2} \boldsymbol{D}_{n,\boldsymbol{\theta}_0} \boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{-\top/2} \boldsymbol{V} \right| \right| \rightarrow 0.$
- (B5) $\boldsymbol{D}_{n,\boldsymbol{\theta}_0}^{-1/2} \boldsymbol{S}_{n,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0) \xrightarrow{d} N_k(\boldsymbol{0},\boldsymbol{I}_k)$, where $N_k(\boldsymbol{0},\boldsymbol{I}_k)$ is the standard k-variate normal distribution.

B Special results under a Linear Mixed Model

Consider model (2.2) and the log-likelihood function ℓ_n in Equation (2.3). As in Giampaoli and Singer (2009), we also write $\Sigma_i = \sum_{g=1}^{r+1} \tau_g H_{ig}$, where

$$\boldsymbol{H}_{ig} = \begin{cases} \boldsymbol{I}_{n_i} & \text{if } g = 1, \\ \boldsymbol{Z}_i \boldsymbol{A}_g \boldsymbol{Z}_i^\top & \text{if } g = 2, \dots, r+1, \end{cases}$$
$A_g, g = 2, \ldots, r + 1$, are known matrices, r = q(q+1)/2 and $\tau_1 = \sigma^2$. In the following, $\boldsymbol{u}_i = \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{e}_i \sim N_{n_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$. Giampaoli and Singer (2009) showed that condition (A1) is valid for all $\boldsymbol{\theta}$ on the boundary of Θ , therefore, it is also valid for all interior points of Θ by using the same arguments. They also computed the vector of directional derivatives which is given by

$$oldsymbol{S}_n(oldsymbol{ heta}) = \sum_{i=1}^n egin{pmatrix} oldsymbol{X}_i^ op oldsymbol{\Sigma}_i^{-1} oldsymbol{u}_i \ -rac{1}{2} ext{tr} ig\{ oldsymbol{\Sigma}_i^{-1} oldsymbol{H}_{i1} ig\} + rac{1}{2} oldsymbol{u}_i^ op oldsymbol{\Sigma}_i^{-1} oldsymbol{H}_{i1} oldsymbol{\Sigma}_i^{-1} oldsymbol{u}_i \ dots \ -rac{1}{2} ext{tr} ig\{ oldsymbol{\Sigma}_i^{-1} oldsymbol{H}_{i(r+1)} ig\} + rac{1}{2} oldsymbol{u}_i^ op oldsymbol{\Sigma}_i^{-1} oldsymbol{H}_{i(r+1)} oldsymbol{\Sigma}_i^{-1} oldsymbol{H}_{i(r+1)} oldsymbol{\Sigma}_i^{-1} oldsymbol{u}_i ig) \,.$$

Following the same matrix operations as in Patriota (2011), we can write this score vector as

$$oldsymbol{S}_n(oldsymbol{ heta}) = \sum_{i=1}^n egin{pmatrix} oldsymbol{X}_i^ op oldsymbol{\Sigma}_i^{-1}oldsymbol{u}_i \ rac{1}{2} ilde{oldsymbol{H}}_i^ op (oldsymbol{\Sigma}_i^{-1}\otimesoldsymbol{\Sigma}_i^{-1}) ext{vec}(oldsymbol{u}_ioldsymbol{u}_i^ op - oldsymbol{\Sigma}_i) \end{pmatrix}.$$

where $\tilde{H}_i = (\text{vec}(H_{i1}) \text{ vec}(H_{i2}) \dots \text{ vec}(H_{i(r+1)}))$ and \otimes is the Kronecker product and the vec operator stacks the columns of the matrix on the top of one another. From this matrix expression it is clear that $E_{\theta}(S_n(\theta)) = 0$ for each $\theta \in \Theta$.

Notice that, by independence and normality of u_i (its odd moments are zero), we have that

$$oldsymbol{D}_{n,oldsymbol{ heta}} = \sum_{i=1}^n egin{pmatrix} oldsymbol{X}_i^ op \mathbf{\Sigma}_i^{-1} oldsymbol{X}_i & oldsymbol{0} \ oldsymbol{0} & rac{1}{2} ilde{oldsymbol{H}}_i^ op (\mathbf{\Sigma}_i^{-1} \otimes \mathbf{\Sigma}_i^{-1}) ilde{oldsymbol{H}}_i \end{pmatrix}$$

for each $\boldsymbol{\theta} \in \Theta$. This is attained from the fact that $E_{\boldsymbol{\theta}} \left(\operatorname{vec}(\boldsymbol{u}_i \boldsymbol{u}_i^\top - \boldsymbol{\Sigma}_i) \operatorname{vec}(\boldsymbol{u}_i \boldsymbol{u}_i^\top - \boldsymbol{\Sigma}_i) \right)$ $\boldsymbol{\Sigma}_i)^\top = \boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i + \boldsymbol{J}(\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i)$, where \boldsymbol{J} is a permutation matrix such that $\operatorname{vec}(\boldsymbol{A}) = \boldsymbol{J}\operatorname{vec}(\boldsymbol{A}^\top)$ for a squared matrix \boldsymbol{A} . It should be noticed that, by properties of the permutation matrix, $D_{n,\theta}$ does not depend on J.

Also, by applying the same matrix operations in expressions (28)-(31) of Giampaoli and Singer (2009), we conclude that $G_{n,\theta} = D_{n,\theta}$ for each $\theta \in \Theta$. We assume that

(C1) the matrix X_i does not depend on n and is full column rank, for $i = 1, \ldots, n$ and there are at least np matrices of \tilde{H}_i , $i = 1, \ldots, n$, that are full column rank, where $p \in (0, 1]$.

Therefore, (B1) and (B4) are valid for every $\boldsymbol{\theta} \in \Theta$. Furthermore, let $\lambda_{\min}(A)$ denote the minimum eigenvalue of A and #(A) denote the number of elements of A. Then, by (C1),

$$#\left(\left\{i:\lambda_{\min}\left(\tilde{\boldsymbol{H}}_{i}^{\top}(\boldsymbol{\Sigma}_{i}^{-1}\otimes\boldsymbol{\Sigma}_{i}^{-1})\tilde{\boldsymbol{H}}_{i}\right)>0,\ i=1,\ldots,n\right\}\right)\geq np,$$

for any $\boldsymbol{\theta} \in \Theta$. This shows that the minimum eigenvalue of $\boldsymbol{G}_{n,\boldsymbol{\theta}}$ is at least of order O(np) for any $\boldsymbol{\theta} \in \Theta$. Therefore, under (C1), condition (B2) is valid for any $\boldsymbol{\theta} \in \Theta$.

The matrix $F_n(\theta)$ derived in Giampaoli and Singer (2009) is given by

$$oldsymbol{F}_n(oldsymbol{ heta}) = \sum_{i=1}^n egin{pmatrix} oldsymbol{X}_i^ op oldsymbol{\Sigma}_i^{-1} oldsymbol{X}_i & oldsymbol{F}_{n,oldsymbol{eta},oldsymbol{\gamma}} \ oldsymbol{F}_{n,oldsymbol{eta},oldsymbol{\gamma}}(oldsymbol{ heta})^ op oldsymbol{F}_{n,oldsymbol{eta},oldsymbol{\gamma}}(oldsymbol{ heta}) \end{pmatrix},$$

where the element (j,g) of $F_{n,\beta,\gamma}(\boldsymbol{\theta})$ is $\sum_{i=1}^{n} \operatorname{tr} \{ \boldsymbol{x}_{ij}^{\top} \boldsymbol{\Sigma}_{i}^{\top} \boldsymbol{H}_{ig} \boldsymbol{\Sigma}_{i}^{-1} \boldsymbol{u}_{i} \}, \boldsymbol{x}_{ij}$ is the *j*th column of \boldsymbol{X}_{i} , and the element (j,g) of $F_{n,\gamma,\gamma}(\boldsymbol{\theta})$ is $\frac{1}{2} \sum_{i=1}^{n} \operatorname{tr} \{ \boldsymbol{\Sigma}_{i}^{-1} \boldsymbol{H}_{ij} \boldsymbol{\Sigma}_{i}^{-1} [2\boldsymbol{u}_{i} \boldsymbol{u}_{i}^{\top} - \boldsymbol{\Sigma}_{i}] \boldsymbol{\Sigma}_{i}^{-1} \boldsymbol{H}_{ig} \}$. The matrix version by using our notation is

$$m{F}_n(m{ heta}) = \sum_{i=1}^n egin{pmatrix} m{X}_i^ op m{\Sigma}_i^{-1} m{X}_i & m{X}_i^ op ig((m{u}_i^ op m{\Sigma}_i^{-1}) \otimes m{\Sigma}_i^{-1}ig) m{ ilde H}_i \ m{ ilde H}_i^ op ig((m{\Sigma}_i^{-1}m{u}_i) \otimes m{\Sigma}_i^{-1}ig) m{X}_i & m{ ilde H}_i^ op ig(m{\Sigma}_i^{-1} \otimes (2m{\Sigma}_i^{-1}m{u}_im{u}_i^ op m{\Sigma}_i^{-1} - m{\Sigma}_i^{-1})m{ ilde H}_i ig).$$

In order to prove that (B3) is valid, notice that for any fixed A > 0,

 $N_n^{\boldsymbol{\theta}}(A) \to \{\boldsymbol{\theta}\} \text{ as } n \to \infty \text{ for each } \boldsymbol{\theta} \in \Theta, \text{ because, as } n \to \infty, \lambda_{\min}(\boldsymbol{G}_{n,\boldsymbol{\theta}}) \to \infty \text{ for each } \boldsymbol{\theta} \in \Theta.$

Assume that

(C2)
$$||\frac{1}{n}\boldsymbol{G}_{n,\boldsymbol{\theta}}|| = O(1)$$
 for each $\boldsymbol{\theta} \in \Theta$

(C3) $||\frac{1}{n}\boldsymbol{F}_{n}(\boldsymbol{\theta}) - \frac{1}{n}\boldsymbol{G}_{n,\boldsymbol{\theta}}|| \xrightarrow{P_{\boldsymbol{\theta}}} 0$ for each $\boldsymbol{\theta} \in \Theta$.

As (C2) and (C3) are sufficient to attain the main results, we do not impose any further restriction on X_i and Z_i . However, it is possible to show that conditions (C2) and (C3) are implied by the following more basic conditionss: there exist two positive constants $M_X < \infty$ and $M_Z < \infty$ such that $\max_{1 \leq i \leq n} ||X_i|| \leq M_X$ and $\max_{1 \leq i \leq n} ||Z_i|| \leq M_Z$.

Notice that

$$\begin{aligned} \left| \left| \mathbf{G}_{n,\boldsymbol{\theta}}^{-1/2} \mathbf{F}_{n}(\boldsymbol{\theta}') \mathbf{G}_{n,\boldsymbol{\theta}}^{-1/2} - \mathbf{I}_{k} \right| \right| &= \left| \left| \mathbf{G}_{n,\boldsymbol{\theta}}^{-1/2} (\mathbf{F}_{n}(\boldsymbol{\theta}') - \mathbf{G}_{n,\boldsymbol{\theta}}) \mathbf{G}_{n,\boldsymbol{\theta}}^{-1/2} \right| \right| \\ &= \left| \left| \left(\frac{1}{n} \mathbf{G}_{n,\boldsymbol{\theta}} \right)^{-1/2} \left(\frac{1}{n} \mathbf{F}_{n}(\boldsymbol{\theta}') - \frac{1}{n} \mathbf{G}_{n,\boldsymbol{\theta}} \right) \left(\frac{1}{n} \mathbf{G}_{n,\boldsymbol{\theta}} \right)^{-1/2} \right| \right| \\ &\leq \left| \left| \left(\frac{1}{n} \mathbf{G}_{n,\boldsymbol{\theta}} \right) \right| \right|^{-1} \left| \left| \frac{1}{n} \mathbf{F}_{n}(\boldsymbol{\theta}') - \frac{1}{n} \mathbf{G}_{n,\boldsymbol{\theta}} \right| \right| \\ &\leq \left| \left| \left(\frac{1}{n} \mathbf{G}_{n,\boldsymbol{\theta}} \right) \right| \right|^{-1} \left(\left| \left| \frac{1}{n} \mathbf{F}_{n}(\boldsymbol{\theta}') - \frac{1}{n} \mathbf{G}_{n,\boldsymbol{\theta}'} \right| \right| + \left| \left| \frac{1}{n} \mathbf{G}_{n,\boldsymbol{\theta}'} - \frac{1}{n} \mathbf{G}_{n,\boldsymbol{\theta}} \right| \right| \end{aligned}$$

Therefore, as $G_{n,\theta} = D_{n,\theta}$ is continuous, we have that, under (C2) and (C3), we only need to show that $\sup_{\theta' \in N_n^{\theta}(A)} || \frac{1}{n} G_{n,\theta'} - \frac{1}{n} G_{n,\theta} || \longrightarrow 0$ for any fixed A > 0and for each $\theta \in \Theta$. By continuity of $G_{n,\theta}$ and condition (C2), there must be a constant C such that

$$\left\| \frac{1}{n} \boldsymbol{G}_{n,\boldsymbol{\theta}'} - \frac{1}{n} \boldsymbol{G}_{n,\boldsymbol{\theta}} \right\| \leq C \left\| \boldsymbol{\theta}' - \boldsymbol{\theta} \right\|.$$

Recall that for each fixed A > 0, $\lim_{n \to \infty} N_n^{\theta}(A) = \{\theta\}$. Therefore, for each fixed A > 0 and $\epsilon > 0$ there exists $n_{A,\epsilon} \in \mathbb{N}$ such that for each $\theta' \in N_n^{\theta}(A)$, we have

that $||\boldsymbol{\theta}' - \boldsymbol{\theta}|| < \epsilon, \forall n \ge n_{A,\epsilon}$. Therefore, for each fixed A > 0 and $\epsilon > 0$,

$$\sup_{\boldsymbol{\theta}' \in N_n^{\boldsymbol{\theta}}(A)} \left| \left| \frac{1}{n} \boldsymbol{G}_{n,\boldsymbol{\theta}'} - \frac{1}{n} \boldsymbol{G}_{n,\boldsymbol{\theta}} \right| \right| \le C\epsilon, \quad \forall \ n \ge n_{A,\epsilon}$$

and, we conclude that $\sup_{\boldsymbol{\theta}' \in N_n^{\boldsymbol{\theta}}(A)} || \frac{1}{n} \boldsymbol{G}_{n,\boldsymbol{\theta}'} - \frac{1}{n} \boldsymbol{G}_{n,\boldsymbol{\theta}} || \longrightarrow 0 \text{ for any fixed } A > 0$ and $\boldsymbol{\theta} \in \Theta$. Hence, condition (B3) holds.

Notice that, under normality, we have that

$$E_{\theta} (|\boldsymbol{X}_{i} \boldsymbol{\Sigma}_{i}^{-1} \boldsymbol{u}_{i}|^{4}) < \infty \quad \text{and} \quad E_{\theta} (|\tilde{\boldsymbol{H}}_{i} (\boldsymbol{\Sigma}_{i}^{-1} \otimes \boldsymbol{\Sigma}_{i}^{-1}) \operatorname{vec}(\boldsymbol{u}_{i} \boldsymbol{u}_{i}^{\top} - \boldsymbol{\Sigma}_{i})|^{4}) < \infty$$

for each $\boldsymbol{\theta} \in \Theta$. Then, as $\lambda_{\min}(\boldsymbol{D}_{n,\boldsymbol{\theta}}) \geq O(np)$ and by (C2) $||\frac{1}{n}\boldsymbol{D}_{n,\boldsymbol{\theta}}|| = O(1)$, the Liapounov's condition for convergence in distribution is valid and, therefore, (B5) holds.

Conditions (A2) and (A3) depend on the specification of the null parameter space Θ_0 . After enunciating the main theorem, Vu and Zhou (1997) discussed that condition (A2) is equivalent to Chernoff regularity (Chernoff, 1954) if

$$\lim_{n \to \infty} \frac{\lambda_{\min}(\boldsymbol{G}_{n,\boldsymbol{\theta}})}{\lambda_{\max}(\boldsymbol{G}_{n,\boldsymbol{\theta}})} > 0$$
(B.12)

for each $\boldsymbol{\theta} \in \Theta$, see Geyer (1994) for more details. Drton (2009) showed that any set or union of sets defined by polynomial equalities or inequalities on $\boldsymbol{\theta}$ (algebraic or semialgebraic set, respectively) is Chernoff regular. By condition (C2), we have that (B.12) is satisfied. Therefore, any set formed by polynomial equalities or inequalities on $\boldsymbol{\theta}$ must satisfy (A2). For instance, consider $p, q \geq 4$, then the sets $\Theta_{01} = \{\boldsymbol{\theta} \in \Theta : \beta_1^3 = \beta_4^2, \gamma_2 \geq \gamma_3, \gamma_4 = 0\},$ $\Theta_{02} = \{\boldsymbol{\theta} \in \Theta : 2\beta_3 = \beta_2, \gamma_2 = \gamma_3, \gamma_4 = 0\}$ and $\Theta_{03} = \{\boldsymbol{\theta} \in \Theta : C\boldsymbol{\theta} = d\},$ where *C* is a known matrix and *d* a known vector, satisfy (A2).

The cones that satisfy (A3) will define the asymptotic distribution and they must be proved by the analyst for each specific Θ_0 and Θ .

B.1 Proof of Theorem 3.2

Proof. From conditions (C1)–(C3), conditions (A1) and (B1)–(B5) are valid for the linear mixed model (2.2) as we show in Appendix B. Condition (A2) is valid because Θ'_0 and Θ are Chernoff regular. Let $\mathbb{B} = \{ \boldsymbol{\beta} \in \mathbb{R}^p : \beta_j = 0, \forall j \in J \}$ and notice that $\Theta'_0 = \mathbb{B} \times \mathbb{R}_+ \times \{0\}$ and $\Theta = \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_0$.

In the sequel, we follow same steps as in Giampaoli and Singer (2009) to show that the cones $C_{\Theta'_0}^{\theta_0} = \mathbb{B} \times \mathbb{R} \times \{0\}$ and $C_{\Theta}^{\theta_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}_0$ with vertex at $\theta_0 \in \Theta'_0$ satisfy (A2). Note that $\Theta'_0 \subset C_{\Theta'_0}^{\theta_0}$ and $\Theta \subset C_{\Theta}^{\theta_0}$, then the first part of condition (A2) holds by taking $u(\boldsymbol{v}) = |\boldsymbol{v} - \boldsymbol{\theta}_0|$, since

$$\inf_{\boldsymbol{w}\in\mathcal{C}_{\Omega}^{\boldsymbol{\theta}_{0}}}|\boldsymbol{G}_{n,\boldsymbol{\theta}_{0}}^{1/2}(\boldsymbol{w}-\boldsymbol{v})|=0,\quad\forall\;\boldsymbol{v}\in\Omega,$$

where $\Omega = \Theta'_0$ or $\Omega = \Theta$. For the second part of condition (A2), we need to analyze the (p+1)th element of \boldsymbol{w} for each $\boldsymbol{w} \in \mathcal{C}^{\boldsymbol{\theta}_0}_{\Omega}$, where $\Omega = \Theta'_0$ or $\Omega = \Theta$. If this (p+1)th element is nonnegative, then $\inf_{\boldsymbol{v}\in\Theta'_0} |\boldsymbol{G}^{1/2}_{n,\boldsymbol{\theta}_0}(\boldsymbol{w}-\boldsymbol{v})| = 0$. Therefore, the only challenge occurs when the p+1 element of \boldsymbol{w} is negative. In this case,

$$\inf_{\boldsymbol{v}\in\Theta_0'}|\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{v})|=|\boldsymbol{g}_{p+1}||w_{p+1}|,$$

where \boldsymbol{g}_{p+1} is the (p+1)th column of $\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}$. As this infimum is not zero, we have to analyze it further.

Notice that $|\boldsymbol{G}_{n,\boldsymbol{\theta}_{0}}^{1/2}(\boldsymbol{w}-\boldsymbol{\theta}_{0})| = c_{1} + |\boldsymbol{g}_{p+1}||w_{p+1} - \sigma_{0}^{2}|$, where $c_{1} \geq 0$. By assumption, $w_{p+1} < 0$ and $\sigma_{0}^{2} > 0$, this implies that $|w_{p+1} - \sigma_{0}^{2}| > |w_{p+1}|$ and, therefore,

$$\inf_{\boldsymbol{v}\in\Theta_0'} |\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{v})| = |\boldsymbol{g}_{p+1}||w_{p+1}| < c_1 + |\boldsymbol{g}_{p+1}||w_{p+1} - \sigma_0^2| = |\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{\theta}_0)|.$$

Let

$$w(\boldsymbol{w}) = \begin{cases} |\boldsymbol{G}_{n,\boldsymbol{ heta_0}}^{1/2}(\boldsymbol{w} - \boldsymbol{ heta_0})|, & ext{if } w_{p+1} > 0, \\ 1, & ext{otherwise.} \end{cases}$$

As $\boldsymbol{w} \to \boldsymbol{\theta}_0$, $v(\boldsymbol{w}) \to \boldsymbol{0}$, since when \boldsymbol{w} is sufficiently close to $\boldsymbol{\theta}_0$, w_{p+1} is positive. Hence, $\mathcal{C}_{\Theta'_0}^{\boldsymbol{\theta}_0}$ satisfies condition (A2). By employing the same argument we can show that $\mathcal{C}_{\Theta}^{\boldsymbol{\theta}_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}_0$ also satisfies (A2).

Following almost the same steps as in Giampaoli and Singer (2009), we know that, for a large n, G_{n,θ_0} is positive definite for each $\theta_0 \in \Theta_0$. Therefore, for any $(p+2) \times (p+2)$ orthogonal nonstochastic matrix T_n , we have that $\tilde{C}_{\Omega_n}^{\theta_0} = T_n G_{n,\theta_0}^{1/2} (C_{\Omega}^{\theta_0} - \theta_0) = C_{\Omega}^{\theta_0}$ for any fixed $\theta_0 \in \Theta_0$ and $\Omega = \Theta'_0$ or $\Omega = \Theta$. Therefore, (A3) holds. We conclude that the cones $\tilde{C}_{\Theta'_0}^{\theta_0} = C_{\Theta'_0}^{\theta_0}$ and $\tilde{C}_{\Theta}^{\theta_0} = C_{\Theta}^{\theta_0}$ are the ones to be employed in Theorem 3.1 and, hence, we have that

$$\inf_{\boldsymbol{\theta}\in\tilde{\mathcal{C}}_{\Theta_{0}}^{\boldsymbol{\theta}_{0}}}|\boldsymbol{N}-\boldsymbol{\theta}|^{2}-\inf_{\boldsymbol{\theta}\in\tilde{\mathcal{C}}_{\Theta}^{\boldsymbol{\theta}_{0}}}|\boldsymbol{N}-\boldsymbol{\theta}|^{2}=\sum_{j\in J}N_{j}^{2}+N_{p+2}^{2}I(N_{p+2}\geq 0),$$

where $\mathbf{N} = (N_1, \dots, N_p, N_{p+1}, N_{p+2})$ is a random vector following a standard multivariate normal distribution. That is, it is a 50-50% mixture of chisquared distributions with g and g + 1 degrees-of-freedom.

B.2 Proof of Theorem 5.1

Proof. From conditions (C1)–(C3), conditions (A1) and (B1)–(B5) are valid for the linear mixed model (2.2) as we show in Appendix B. For each $\theta_0 \in \Theta$ the condition (A2) is valid because $\{\theta_0\}$ and Θ are Chernoff regular. The cone $C_{\{\theta_0\}}^{\theta_0} = \{\theta_0\}$ satisfies (A2), because

$$\inf_{\boldsymbol{w}\in\mathcal{C}_{\Omega}^{\boldsymbol{\theta}_{0}}}|\boldsymbol{G}_{n,\boldsymbol{\theta}_{0}}^{1/2}(\boldsymbol{w}-\boldsymbol{v})|=\inf_{\boldsymbol{v}\in\Omega}|\boldsymbol{G}_{n,\boldsymbol{\theta}_{0}}^{1/2}(\boldsymbol{w}-\boldsymbol{v})|=0\quad\forall\;\boldsymbol{v}\in\Omega,\boldsymbol{u}\in\mathcal{C}_{\Omega}^{\boldsymbol{\theta}_{0}},$$

where $\Omega = \{\boldsymbol{\theta}_0\}$. The first and second part of condition (A2) hold by taking $u(\boldsymbol{v}) = |\boldsymbol{v} - \boldsymbol{\theta}_0|$ and $v(\boldsymbol{w}) = |\boldsymbol{w} - \boldsymbol{\theta}_0|$, then $\mathcal{C}^{\boldsymbol{\theta}_0}_{\{\boldsymbol{\theta}_0\}}$ satisfies (A2).

For each fixed $\boldsymbol{\theta}_0 \in \Theta = \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_0$, we have two possible cases, namely, (a) $\boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$ or (b) $\boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \{0\}$. For each case, we follow the same ideas employed in the proof of Theorem 3.2. Note that the case (b) was proved in Appendix B.1, then the cone $\mathcal{C}_{\Theta}^{\boldsymbol{\theta}_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}_0$ with vertex in $\boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \{0\}$ satisfies (A2).

In case (a), where the parameter vector is an interior point of Θ , consider the cone $\mathcal{C}_{\Theta}^{\boldsymbol{\theta}_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}$ with vertex at $\boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$. The first part of condition (A2) holds by taking $u(\boldsymbol{v}) = |\boldsymbol{v} - \boldsymbol{\theta}_0|$, since $\Theta \subseteq \mathcal{C}_{\Theta}^{\boldsymbol{\theta}_0}$ which implies

$$\inf_{\boldsymbol{w}\in\mathcal{C}_{\Theta}^{\boldsymbol{\theta}_{0}}} |\boldsymbol{G}_{n,\boldsymbol{\theta}_{0}}^{1/2}(\boldsymbol{w}-\boldsymbol{v})| = 0, \quad \boldsymbol{\theta}_{0}\in\mathbb{R}^{p}\times\mathbb{R}_{+}\times\mathbb{R}_{+}, \forall \ \boldsymbol{v}\in\Theta.$$

For the second part of condition (A2), we need to analyze the (p + 1)th and (p + 2)th elements of $\boldsymbol{w} \in \mathcal{C}_{\Theta}^{\boldsymbol{\theta}_0}$:

- If the (p + 1)th and (p + 2)th elements of \boldsymbol{w} are nonnegatives, then $\inf_{\boldsymbol{v}\in\Theta} |\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w} - \boldsymbol{v})| = 0.$
- From Appendix B.1, for both cases, namely, "if (p + 1)th element of \boldsymbol{w} is negative and (p + 2)th element of \boldsymbol{w} is nonnegative" or "if (p + 1)th element of \boldsymbol{w} is nonnegative and (p + 2)th element of \boldsymbol{w} is negative", we conclude that $\inf_{\boldsymbol{v}\in\Theta} |\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w} \boldsymbol{v})| = 0.$
- If the (p+1)th and (p+2)th elements of \boldsymbol{w} are negatives, then $\inf_{\boldsymbol{v}\in\Theta} |\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{v})| \leq |\boldsymbol{g}_{p+1}||w_{p+1}| + |\boldsymbol{g}_{p+2}||w_{p+2}|$, where \boldsymbol{g}_{p+1} and \boldsymbol{g}_{p+2} are the (p+1)th and (p+2)th columns of $\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}$. Moreover, notice that $|\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{\theta}_0)| = c_2 + |\boldsymbol{g}_{p+1}||w_{p+1} \sigma_0^2| + |\boldsymbol{g}_{p+2}||w_{p+2} \gamma_0|$, where $c_2 \geq 0$. As $w_{p+1} < 0$ and $w_{p+2} < 0$, we have that $|w_{p+1}| < |w_{p+1} \sigma_0^2|$ and $|w_{p+2}| < |w_{p+2} \gamma_0|$.

Therefore

$$\inf_{\boldsymbol{v}\in\Theta}|\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{v})|<|\boldsymbol{G}_{n,\boldsymbol{\theta}_0}^{1/2}(\boldsymbol{w}-\boldsymbol{\theta}_0)|.$$

Let

$$v(\boldsymbol{w}) = \begin{cases} |\boldsymbol{G}_{n,\boldsymbol{\theta_0}}^{1/2}(\boldsymbol{w} - \boldsymbol{\theta}_0)|, & \text{if } w_{p+1} > 0 \text{ or } w_{p+2} > 0, \\ 1, & \text{otherwise.} \end{cases}$$

As $\boldsymbol{w} \to \boldsymbol{\theta}_0$, $v(\boldsymbol{w}) \to \boldsymbol{0}$, since when \boldsymbol{w} is sufficiently close to $\boldsymbol{\theta}_0$, w_{p+1} and w_{p+2} are positive. Hence, $\mathcal{C}_{\Theta}^{\boldsymbol{\theta}_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}$ with vertex at $\boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$ satisfies condition (A2).

By the same arguments used in Appendix B.1, for any $(p+2) \times (p+2)$ orthogonal nonstochastic matrix \mathbf{T}_n , we have that $\tilde{\mathcal{C}}_{\{\theta_0\}_n}^{\theta_0} = \mathbf{T}_n \mathbf{G}_{n,\theta_0}^{1/2} (\mathcal{C}_{\{\theta_0\}}^{\theta_0} - \mathbf{\theta}_0) = \{\mathbf{0}\}$ and $\tilde{\mathcal{C}}_{\Theta_n}^{\theta_0} = \mathbf{T}_n \mathbf{G}_{n,\theta_0}^{1/2} (\mathcal{C}_{\Theta}^{\theta_0} - \mathbf{\theta}_0) = \mathcal{C}_{\Theta}^{\theta_0}$ for each $\theta_0 \in \Theta$. Therefore, (A3) holds. We conclude that the cones $\tilde{\mathcal{C}}_{\{\theta_0\}}^{\theta_0} = \{\mathbf{0}\}$ and $\tilde{\mathcal{C}}_{\Theta}^{\theta_0} = \mathcal{C}_{\Theta}^{\theta_0}$ are the ones to be employed in Theorem 3.1.

Let $\mathbf{N} = (N_1, \dots, N_p, N_{p+1}, N_{p+2})$ be a random vector following a standard multivariate normal distribution. Consider the two cases:

Case (a) If $\boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$, we have that $\tilde{\mathcal{C}}_{\{\boldsymbol{\theta}_0\}}^{\boldsymbol{\theta}_0} = \{\mathbf{0}\}$ and $\tilde{\mathcal{C}}_{\Theta}^{\boldsymbol{\theta}_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}_+$ then

$$\inf_{\boldsymbol{\theta}\in\tilde{\mathcal{C}}_{\{\boldsymbol{\theta}_{0}\}}}|\boldsymbol{N}-\boldsymbol{\theta}|^{2}-\inf_{\boldsymbol{\theta}\in\tilde{\mathcal{C}}_{\Theta}^{\boldsymbol{\theta}_{0}}}|\boldsymbol{N}-\boldsymbol{\theta}|^{2}=\sum_{j=1}^{p+2}N_{j}^{2}.$$

That is, it is a chisquared distribution with p + 2 degrees-of-freedom.

Case (b) If $\boldsymbol{\theta}_0 \in \mathbb{R}^p \times \mathbb{R}_+ \times \{0\}$, we have that $\tilde{\mathcal{C}}_{\{\boldsymbol{\theta}_0\}}^{\boldsymbol{\theta}_0} = \{\mathbf{0}\}$ and $\tilde{\mathcal{C}}_{\Theta}^{\boldsymbol{\theta}_0} = \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}_0$, then

$$\inf_{\boldsymbol{\theta}\in\tilde{\mathcal{C}}_{\{\boldsymbol{\theta}_0\}}^{\boldsymbol{\theta}_0}} |\boldsymbol{N}-\boldsymbol{\theta}|^2 - \inf_{\boldsymbol{\theta}\in\tilde{\mathcal{C}}_{\Theta}^{\boldsymbol{\theta}_0}} |\boldsymbol{N}-\boldsymbol{\theta}|^2 = \sum_{j=1}^{p+1} N_j^2 + N_{p+2}^2 I(N_{p+2} \ge 0).$$

That is, it is a 50-50% mixture of chisquared distributions with p+1 and p+2 degrees-of-freedom.

The proof is completed.

C Proofs of Theorems in Section 5

C.1 Proof of Theorem 5.2

Proof. For any $\boldsymbol{\theta} \in \Theta_0 \subseteq \Theta^{(j)}$, the asymptotic distribution of $\lambda_n(\{\boldsymbol{\theta}\})$ is continuous and its cumulative distribution F_j is a strictly increasing function. As $\Theta_0 \neq \emptyset$, by (5.9), we have that

$$\Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_{0} = \left\{ \boldsymbol{\theta} \in \Theta_{0} : F_{j}(\lambda_{n}^{obs}(\{\boldsymbol{\theta}\})) \leq 1 - \alpha \right\}$$
$$= \left\{ \boldsymbol{\theta} \in \Theta_{0} : 1 - F_{j}(\lambda_{n}^{obs}(\{\boldsymbol{\theta}\})) \geq \alpha \right\}$$

and, if there exists $\boldsymbol{\theta}_0 \in \Theta_0$ such that $\inf_{\boldsymbol{\theta} \in \Theta_0} F_j(\lambda_n^{obs}(\{\boldsymbol{\theta}\})) = F_j(\lambda_n^{obs}(\{\boldsymbol{\theta}_0\}))$, then

$$\begin{split} \left\{ \alpha \in (0,1) : \left\{ \boldsymbol{\theta} \in \Theta_0 : 1 - F_j(\lambda_n^{obs}(\{\boldsymbol{\theta}\})) \ge \alpha \right\} \neq \varnothing \right\} &= \left(0, \sup_{\boldsymbol{\theta} \in \Theta_0} \left(1 - F_j(\lambda_n^{obs}(\{\boldsymbol{\theta}\})) \right) \right] \\ &= \left(0, \left(1 - \inf_{\boldsymbol{\theta} \in \Theta_0} F_j(\lambda_n^{obs}(\{\boldsymbol{\theta}\})) \right) \right] \\ &= \left(0, \left(1 - F_j(\inf_{\boldsymbol{\theta} \in \Theta_0} \lambda_n^{obs}(\{\boldsymbol{\theta}\})) \right) \right] \\ &= \left(0, \left(1 - F_j(\lambda_n^{obs}(\Theta_0)) \right) \right], \end{split}$$

otherwise, the above intervals should be opened. In any case,

$$\sup\left\{\alpha \in (0,1): \left\{\boldsymbol{\theta} \in \Theta_0: 1 - F_j(\lambda_n^{obs}(\{\boldsymbol{\theta}\})) \ge \alpha\right\} = 1 - F_j(\lambda_n^{obs}(\Theta_0))$$

and the result follows.

C.2 Proof of Theorem 5.3

Proof. As $\Theta_0 \subseteq \Theta^{(j)}$, by Theorem 5.2, $s(\Theta_0, \boldsymbol{y}) = 1 - F_j(\lambda_n^{obs}(\Theta_0))$ and its random version is $s(\Theta_0, \boldsymbol{Y}) = 1 - F_j(\lambda_n(\Theta_0))$. Then, for each $\boldsymbol{\theta} \in \Theta_0$ and each $\alpha \in (0, 1)$,

$$P_{\theta}\left(s(\Theta_0, \mathbf{Y}) \le \alpha\right) = P_{\theta}\left(F_j(\lambda_n(\Theta_0)) \ge 1 - \alpha\right)$$
$$= P_{\theta}\left(\lambda_n(\Theta_0) \ge F_j^{-1}(1 - \alpha)\right).$$

As $\lambda_n(\Theta_0) = \inf_{\boldsymbol{\theta}\in\Theta_0} \lambda_n(\{\boldsymbol{\theta}\}) \leq \lambda_n(\{\boldsymbol{\theta}\})$ for each $\boldsymbol{\theta}\in\Theta_0$, we have that $\lambda_n(\Theta_0) \geq a \implies \lambda_n(\{\boldsymbol{\theta}\}) \geq a$ for each $\boldsymbol{\theta}\in\Theta_0$ and a > 0, then

$$\lim_{n \to \infty} P_{\boldsymbol{\theta}} \left(s(\Theta_0, \boldsymbol{Y}) \le \alpha \right) = \lim_{n \to \infty} P_{\boldsymbol{\theta}} \left(\lambda_n(\Theta_0) \ge F_j^{-1}(1 - \alpha) \right)$$
$$\le \lim_{n \to \infty} P_{\boldsymbol{\theta}} \left(\lambda_n(\{\boldsymbol{\theta}\}) \ge F_j^{-1}(1 - \alpha) \right)$$
$$= 1 - F_j \left(F_j^{-1}(1 - \alpha) \right)$$
$$= \alpha$$

for each $\boldsymbol{\theta} \in \Theta_0$. Therefore, the result follows.

C.3 Proof of Theorem 5.4

Proof. Let $\Theta_{0j} = \Theta_0 \cap \Theta^{(j)}$, for j = 1, 2. Then, by (5.9),

$$\begin{split} \Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_{0} &= (\Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_{01}) \cup (\Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_{02}) \\ &= \left\{ \boldsymbol{\theta} \in \Theta_{01} : F_{1}(\lambda_{n}^{obs}(\{\boldsymbol{\theta}\})) \leq 1 - \alpha \right\} \cup \left\{ \boldsymbol{\theta} \in \Theta_{02} : F_{2}(\lambda_{n}^{obs}(\{\boldsymbol{\theta}\})) \leq 1 - \alpha \right\} \\ &= \underbrace{\left\{ \boldsymbol{\theta} \in \Theta_{01} : 1 - F_{1}(\lambda_{n}^{obs}(\{\boldsymbol{\theta}\})) \geq \alpha \right\}}_{\Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_{01}} \cup \underbrace{\left\{ \boldsymbol{\theta} \in \Theta_{02} : 1 - F_{2}(\lambda_{n}^{obs}(\{\boldsymbol{\theta}\})) \geq \alpha \right\}}_{\Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_{02}}. \end{split}$$

Let
$$A_0 = \left\{ \alpha \in (0,1) : \Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_0 \neq \emptyset \right\}$$
 and $A_j = \left\{ \alpha \in (0,1) : \Lambda_{\alpha}(\boldsymbol{y}) \cap \Theta_{0j} \neq \emptyset \right\}$, $j = 1, 2$. Then,

$$A_0 \subseteq A_1 \cup A_2 \implies \sup(A_0) \le \sup(A_1 \cup A_2) \le \max\{\sup(A_1), \sup(A_2)\}$$

Notice also that, as $A_1 \subset A_0$ and $A_2 \subset A_0$, we have that

 $\sup(A_1) \le \sup(A_0)$ and $\sup(A_2) \le \sup(A_0) \implies \sup(A_0) \ge \max\{\sup(A_1), \sup(A_2)\}.$

Therefore, the equality holds and the result follows, since $\sup(A_0) = s(\Theta_0, \boldsymbol{y})$ and $\sup(A_j) = s(\Theta_0 \cap \Theta^{(j)}, \boldsymbol{y}), j = 1, 2.$

C.4 Proof of Theorem 5.5

Proof. By definition, $s(\emptyset, \boldsymbol{y}) = 0$. Let $\Theta^{(1)}$ and $\Theta^{(2)}$ be the sets defined in Equation (5.9). Therefore, by Theorem 5.4,

$$s(\Theta, \boldsymbol{y}) = \max\{s(\Theta^{(1)}, \boldsymbol{y}), s(\Theta^{(2)}, \boldsymbol{y})\}.$$

We have two cases for the maximum likelihood:

$$\sup_{\boldsymbol{\theta}\in\Theta}\ell_n(\boldsymbol{\theta})=\sup_{\boldsymbol{\theta}\in\Theta^{(1)}}\ell_n(\boldsymbol{\theta})\quad\text{or}\quad \sup_{\boldsymbol{\theta}\in\Theta}\ell_n(\boldsymbol{\theta})=\sup_{\boldsymbol{\theta}\in\Theta^{(2)}}\ell_n(\boldsymbol{\theta}).$$

Notice that, in any case, $\sup_{\theta \in \Theta} \ell_n(\boldsymbol{\theta}) = \sup_{\theta \in \Theta^{(j)}} \ell_n(\boldsymbol{\theta})$ implies $s(\Theta^{(j)}, \boldsymbol{y}) = 1$ and, therefore, $s(\Theta, \boldsymbol{y}) = 1$.

Let $\Theta_1 \subseteq \Theta_2 \subseteq \Theta$ be two sets and define $\Theta_{1j} = \Theta_1 \cap \Theta^{(j)}$ and $\Theta_{2j} = \Theta_2 \cap \Theta^{(j)}$, for j = 1, 2. By Theorem 5.4, we know that

$$s(\Theta_1, \boldsymbol{y}) = \max\{s(\Theta_{11}, \boldsymbol{y}), s(\Theta_{12}, \boldsymbol{y})\} \text{ and } s(\Theta_2, \boldsymbol{y}) = \max\{s(\Theta_{21}, \boldsymbol{y}), s(\Theta_{22}, \boldsymbol{y})\}.$$

Notice that, in order to show that $s(\Theta_1, \boldsymbol{y}) \leq s(\Theta_2, \boldsymbol{y})$, it is sufficient to show that $s(\Theta_{1j}, \boldsymbol{y}) \leq s(\Theta_{2j}, \boldsymbol{y})$ for each j = 1, 2. By Theorem 5.2, we have

$$s(\Theta_{1j}, \boldsymbol{y}) = 1 - F_j(\lambda_n^{obs}(\Theta_{1j}))$$
 and $s(\Theta_{2j}, \boldsymbol{y}) = 1 - F_j(\lambda_n^{obs}(\Theta_{2j}))$

As $\Theta_{1j} \subseteq \Theta_{2j}$ for each j = 1, 2, we have that $\lambda_n^{obs}(\Theta_{2j}) \leq \lambda_n^{obs}(\Theta_{1j})$. As F_j is a strictly continuous function, $F_j(\lambda_n^{obs}(\Theta_{2j})) \leq F_j(\lambda_n^{obs}(\Theta_{1j}))$ and $1 - F_j(\lambda_n^{obs}(\Theta_{2j})) \geq 1 - F_j(\lambda_n^{obs}(\Theta_{1j}))$ for each j = 1, 2. Therefore, $s(\Theta_{1j}, \boldsymbol{y}) \leq s(\Theta_{2j}, \boldsymbol{y})$ for each j = 1, 2. Hence, $s(\Theta_1, \boldsymbol{y}) \leq s(\Theta_2, \boldsymbol{y})$.

C.5 Proof of Theorem 5.6

Proof. For any nonempty set $\Theta_0 \subseteq \Theta$, by Theorem 5.4, $s(\Theta_0, \boldsymbol{y}) = \max\{s(\Theta_0 \cap \Theta^{(1)}, \boldsymbol{y}), s(\Theta_0 \cap \Theta^{(2)}, \boldsymbol{y})\}$, where $\Theta^{(1)}$ and $\Theta^{(2)}$ are defined in Equation (5.9). By Theorem 5.3, we have that

$$\lim_{n \to \infty} P_{\boldsymbol{\theta}} \left(s(\Theta_0 \cap \Theta^{(j)}, \boldsymbol{Y}) \le \alpha \right) \le \alpha,$$

for each $\boldsymbol{\theta} \in \Theta_0 \cap \Theta^{(j)}$, j = 1, 2 and $\alpha \in (0, 1)$.

Notice that $\max\{s(\Theta_0 \cap \Theta^{(1)}, \boldsymbol{Y}), s(\Theta_0 \cap \Theta^{(2)}, \boldsymbol{Y})\} \leq \alpha \implies s(\Theta_0 \cap \Theta^{(j)}, \boldsymbol{Y}) \leq \alpha \text{ for } j = 1, 2 \text{ and } \alpha \in (0, 1).$ Therefore,

$$\lim_{n \to \infty} P_{\boldsymbol{\theta}} \left(s(\Theta_0, \boldsymbol{Y}) \le \alpha \right) \le \lim_{n \to \infty} P_{\boldsymbol{\theta}} \left(s(\Theta_0 \cap \Theta^{(j)}, \boldsymbol{Y}) \le \alpha \right) \le \alpha$$

for each $\boldsymbol{\theta} \in \Theta_0$. Therefore, the result follows.

Referências Bibliográficas

- Bates et al.(2015) Douglas Bates, Martin Mächler, Ben Bolker e Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*, 67(1): 1–48. ISSN 1548-7660. doi: 10.18637/jss.v067.i01. URL: https://www.jstatsoft.org/v067/i01. (Citado na pg. 8)
- **Chernoff(1954)** Herman Chernoff. On the distribution of the likelihood ratio. *The Annals* of Mathematical Statistics, 25(3):573–578. (Citado nas pgs. 9 e 10)
- **Cox e Hinkley(1979)** David Roxbee Cox e David Victor Hinkley. *Theoretical statistics*. Chapman and Hall/CRC. (Citado na pg. 2)
- **Davis(2002)** Charles S Davis. Statistical methods for the analysis of repeated measurements. Springer Science & Business Media. (Citado na pg. 2)
- **De Andrade(1984)** Dalton Francisco De Andrade. Ml estimation and lr tests for the multivariate normal distribution with patterned mean and covariance matrix. ccmplete and incomplete data cases, 1984.
- **Demidenko(2013)** Eugene Demidenko. *Mixed models: theory and applications with R.* John Wiley & Sons. (Citado nas pgs. 2 e 8)
- Dempster et al.(1977) A. P. Dempster, N. M. Laird e D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38. ISSN 00359246. URL: http://www.jstor.org/stable/2984875. (Citado na pg. 7)
- Drton(2009) Mathias Drton. Likelihood ratio tests and singularities. The Annals of Statistics, 37(2):979-1012. doi: 10.1214/07-AOS571. (Citado na pg. 9)
- **Friedman e Halpern(2001)** Nir Friedman e Joseph Y Halpern. Plausibility measures and default reasoning. *Journal of the ACM*, 48(4):648–685. (Citado na pg. 21)
- **Giampaoli e Singer(2009)** Viviana Giampaoli e Julio M Singer. Likelihood ratio tests for variance components in linear mixed models. *Journal of Statistical Planning and Inference*, 139(4):1435–1448. (Citado nas pgs. i, iii, 2, 3, 10, 11, 13 e 25)
- Khuri et al.(2011) André I Khuri, Thomas Mathew e Bimal K Sinha. Statistical tests for mixed linear models, volume 906. John Wiley & Sons. (Citado na pg. 2)

- **Kowalski(1993)** Charles J Kowalski. Data analysis in craniofacial biology with special emphasis on longitudinal studies. *The Cleft palate-craniofacial journal*, 30(2):111–120. (Citado na pg. 1)
- Laird e Ware(1982) Nan M. Laird e James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974. ISSN 0006341X, 15410420. URL: http://www.jstor.org/stable/2529876. (Citado nas pgs. 5, 6 e 7)
- **Nobre** *et al.*(**2013**) Juvêncio S Nobre, Julio M Singer e Pranab K Sen. U-tests for variance components in linear mixed models. *Test*, 22(4):580–605. (Citado na pg. 2)
- **Patriota(2013)** Alexandre G Patriota. A classical measure of evidence for general null hypotheses. *Fuzzy Sets and Systems*, 233:74–88. (Citado nas pgs. i, iii, 3, 15, 16, 18 e 21)
- **Pinheiro e Bates(2006)** José Pinheiro e Douglas Bates. *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media. (Citado na pg. 1)
- Schervish(1996) Mark J Schervish. P values: what they are and what they are not. *The American Statistician*, 50(3):203–206. (Citado nas pgs. 3 e 15)
- **Self e Liang(1987)** Steven G Self e Kung-Yee Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610. (Citado nas pgs. 2, 10, 11 e 16)
- Sen e Singer(1994) Pranab K. Sen e Julio M. Singer. Large Sample Methods in Statistics: An Introduction with Applications. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis. ISBN 9780412042218. (Citado na pg. 2)
- **Shao e Tu(2012)** Jun Shao e Dongsheng Tu. *The jackknife and bootstrap*. Springer Science & Business Media. (Citado na pg. 2)
- **Silvapulle e Sen(2005)** Mervyn J Silvapulle e Pranab Kumar Sen. *Constrained statistical inference inequality, order, and shape restrictions*. Wiley series in probability and statistics. John Wiley & Sons, Hoboken, N.J. ISBN 0471208272. (Citado na pg. 2)
- Stram e Lee(1994) Daniel O. Stram e Jae Won Lee. Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50(4):1171–1177. ISSN 0006341X, 15410420. doi: 10.2307/2533455. URL: http://www.jstor.org/stable/2533455. (Citado nas pgs. i, iii, 2, 3, 10 e 11)
- Verbeke e Molenberghs(2000) Geert Verbeke e Geert Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer series in statistics. Springer Science & Business Media, New York, NY. ISBN 0-387-95027-3. (Citado na pg. 2)
- Vu e Zhou(1997) H.T.V. Vu e S. Zhou. Generalization of likelihood ratio tests under nonstandard conditions. *The Annals of Statistics*, 25(2):897–916. (Citado nas pgs. 2, 3, 9, 10, 11 e 25)
- **Wilks(1938)** Samuel S Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The annals of mathematical statistics*, 9(1):60–62. (Citado na pg. 9)