**Condições de regularidade para o modelo
de regressão com parametrização geral**

Laís Helen Loose

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

Programa: Estatística

Orientador: Prof. Dr. Alexandre Galvão Patriota

Durante o desenvolvimento deste trabalho a autora recebeu auxílio financeiro da CAPES (2 anos)
e do CNPq (1 ano e 2 meses)

São Paulo - SP, Maio de 2019

Condições de regularidade para o modelo de regressão com parametrização geral

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 24/05/2019. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Alexandre Galvão Patriota - IME-USP
- Prof^a. Dr^a. Silvia Lopes Ferrari - IME-USP
- Prof. Dr. Mário de Castro Andrade Filho - ICMC-USP
- Prof. Dr. Francisco Cribari Neto - UFPE
- Prof^a. Dr^a. Tatiane Ferreira do Nascimento Melo da Silva - UFG

Agradecimentos

À Deus, na certeza de que “em todo tempo Deus é bom”.

Ao Programa de Pós Graduação em Estatística do Instituto de Matemática e Estatística da Universidade de São Paulo pela oportunidade concedida.

Ao meu orientador Alexandre Galvão Patriota, pela sábia e precisa orientação. Por todo conhecimento compartilhado, pela dedicação, disponibilidade, paciência e por ter sido um “parceiro” no desenvolvimento dessa tese. És um excelente professor, pesquisador e orientador, um exemplo a ser seguido. Sou eternamente grata pelo privilégio de ter sido sua orientanda.

Ao meu namorado Moizés, por ter encarado comigo esse desafio que foi o doutorado. Por ser meu fiel companheiro, nas horas de alegrias e tristezas. Companheiro de estudos durante as disciplinas e de discussões durante a realização dessa tese. Obrigada por tudo o que compartilhamos até aqui, pela paciência, amor e carinho. Essa conquista é nossa!

Aos meus pais, Sidio e Haidi, que talvez nunca imaginaram que teriam uma filha, lá do Rincão Fundo, interior de Panambi - RS, doutora pela Universidade de São Paulo. Mas sempre deram importância ao estudo, me incentivaram e não mediram esforços nessa trajetória. Obrigada por todo apoio, pelo exemplo de caráter, honestidade e perseverança.

Aos meus irmãos Luís e Léo, minha cunhada Marta e vó Eli. Obrigada pelo amor, apoio e incentivo em todos os momentos.

Ao professor Fábio Bayer, meu orientador na graduação da UFSM. Foi e ainda és inspiração para minha carreira acadêmica. Obrigada pelo incentivo e amizade.

À minha família e à família do Moizés. Pelo carinho, amor e incentivo.

Aos amigos e colegas do IME-USP, em especial Andrés, Cátia, Francimário, Jaime, Magno, Rodrigo, Sitso, Shu e Terezinha. Obrigada pela amizade e carinho de vocês.

Aos funcionários do IME-USP, em especial a Regiane, secretária do programa. Obrigada pela competência e eficiência em tudo que faz.

Aos amigos e demais pessoas não mencionadas que contribuíram de alguma forma para a realização dessa tese. Obrigada.

À Banca Examinadora pelas sugestões.

À CAPES e CNPq pelo apoio financeiro concedido.

Resumo

LOOSE, L. H. **Condições de regularidade para o modelo de regressão com parametrização geral**. 2019. 86 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2019.

Este trabalho objetiva apresentar um estudo detalhado e sistemático de algumas condições de regularidade para inferências baseadas em máxima verossimilhança no modelo de regressão elíptico multivariado com parametrização geral proposto em [Lemonte e Patriota \(2011\)](#). O modelo em estudo tem vários modelos importantes como casos particulares, entre eles temos os modelos lineares e não lineares homocedásticos e heterocedásticos, modelos mistos, modelos heterocedásticos com erros nas variáveis e na equação, modelos multiníveis, entre outros. As condições de regularidade estudadas estão associadas à identificabilidade do modelo, à existência, à unicidade, à consistência e à normalidade assintótica dos estimadores de máxima verossimilhança (EMV) e à distribuição assintótica das estatísticas de testes. Para isso, são enunciadas as condições suficientes e formalizados os teoremas que garantem a existência, unicidade, consistência e normalidade assintótica dos EMV e a distribuição assintótica das estatísticas de teste usuais. Além disso, os resultados de cada teorema são comentados e as demonstrações são apresentadas com detalhes. Inicialmente, considerou-se o modelo sob a suposição de normalidade dos erros, para, na sequência, ser possível generalizar os resultados para o caso elíptico. A fim de exemplificar os resultados obtidos, foram verificadas, analiticamente, a validade de algumas condições e os resultados de alguns teoremas em casos particulares do modelo geral. Ademais, foi desenvolvido um estudo de simulação em que uma das condições é violada adotando o modelo heterocedástico com erros nas variáveis e na equação. Por meio de simulações de Monte Carlo foram avaliados os impactos sobre a consistência e normalidade assintótica dos EMV.

Palavras-chave: modelos de regressão, distribuições elípticas, teoria assintótica, propriedades assintóticas dos estimadores, estimador de máxima verossimilhança.

Abstract

LOOSE, L. H. **Regularity conditions for the regression model with general parameterization**. 2019. 86 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2019.

This work aims to present a detailed and systematic study of some regularity conditions for inferences based on maximum likelihood in the multivariate elliptic regression model with general parameterization proposed in [Lemonte e Patriota \(2011\)](#). The model under study has several important models as particular cases, among them we have the linear and non-linear homocedastic and heterocedastic models, mixed models, heterocedastic models with errors in the variables and in the equation, multilevel models, among others. The regularity conditions studied are associated with the identifiability of the model, existence, uniqueness, consistency and asymptotic normality of the maximum likelihood estimators (MLE) and the asymptotic distribution of some test statistics. Sufficient conditions are stated to guarantee the existence, unicity, consistency and asymptotic normality of the MLE and the asymptotic distribution of the usual test statistics. In addition, the results of each theorem are commented and the proof are presented in detail. Initially, the model was considered under the assumption of normality of the errors, and then the results were generalized for the elliptical case. In order to exemplify the attained results, some particular cases of the general model are analyzed analytically, the validity of some conditions and the results of some theorems are verified. In addition, a simulation study is developed with one of the conditions violated under the heterocedastic model with errors in the variables and in the equation. By means of Monte Carlo simulations, the impacts of this violation on the consistency and the asymptotic normality of the MLE are evaluated.

Keywords: regression models, elliptical distribution, asymptotic theory, asymptotic properties of estimators, maximum likelihood estimator.

Sumário

Lista de Abreviaturas	ix
Lista de Símbolos	xi
Lista de Figuras	xiii
Lista de Tabelas	xv
1 Introdução	1
1.1 Considerações preliminares	3
1.2 Objetivos	4
1.3 Contribuições	4
1.4 Organização do trabalho	5
2 Condições de regularidade para o modelo de regressão normal multivariado com parametrização geral	7
2.1 O modelo de regressão normal com parametrização geral	7
2.2 Condições de regularidade	9
2.2.1 Existência do EMV	11
2.2.2 Unicidade do EMV	11
2.2.3 Consistência do EMV	13
2.2.4 Normalidade assintótica do EMV	13
2.2.5 Distribuição assintótica das estatísticas de teste	14
2.3 Exemplos	16
2.3.1 Modelo linear normal	16
2.3.2 Modelo não linear heterocedástico	17
2.3.3 Modelo com erros nas variáveis e na equação	19
2.4 Conclusões	26
3 Condições de regularidade para o modelo de regressão elíptico multivariado com parametrização geral	29
3.1 O modelo de regressão elíptico com parametrização geral	29
3.2 Condições de regularidade	32
3.2.1 Existência do EMV	33
3.2.2 Unicidade do EMV	33
3.2.3 Consistência do EMV	35

3.2.4	Normalidade assintótica do EMV	35
3.2.5	Distribuição assintótica das estatísticas de teste	36
3.3	Exemplo	36
3.3.1	Modelo não linear heterocedástico	36
3.4	Conclusões	40
4	Conclusões	43
4.1	Considerações finais	43
4.2	Sugestões para pesquisas futuras	43
A	Provas dos Teoremas do Capítulo 2	45
A.1	Prova do Teorema 2.2.1 - Existência assintótica do EMV	45
A.2	Prova do Teorema 2.2.2 - Unicidade do EMV caso geral	47
A.3	Prova do Teorema 2.2.3 - Unicidade do EMV caso em que a média e a matriz de covariâncias não compartilham parâmetros	47
A.4	Prova do Teorema 2.2.4 - Consistência do EMV	50
A.5	Prova do Teorema 2.2.5 - Normalidade assintótica do EMV	50
A.6	Prova do Teorema 2.2.6 - Distribuição assintótica das estatísticas de teste	54
B	Provas dos Teoremas do Capítulo 3	57
B.1	Prova do Teorema 3.2.2 - Unicidade do EMV caso geral	57
B.2	Prova do Teorema 3.2.3 - Unicidade do EMV no caso em que a média e a matriz de covariâncias não compartilham parâmetros	57
B.3	Prova do Teorema 3.2.5 - Normalidade assintótica do EMV	59
	Referências Bibliográficas	65

Lista de Abreviaturas

EMV	Estimador de máxima verossimilhança (<i>maximum likelihood estimator</i>)
ind	Variáveis independentes (<i>independent variables</i>)
MLG	Modelos lineares generalizados (<i>generalized linear models</i>)
EP	Exponencial potência (<i>power exponential</i>)
L-BFGS	Limite-memória Broyden-Fletcher-Goldfarb-Shanno (<i>Limited-memory Broyden-Fletcher-Goldfarb-Shanno</i>)
EQM	Erro quadrático médio (<i>mean square error</i>)
DP	Desvio padrão (<i>standard deviation</i>)
CA	Coefficiente de assimetria (<i>skewness coefficient</i>)
C	Coefficiente de curtose (<i>kurtosis coefficient</i>)
KS	Kolmogorov-Smirnov (<i>Kolmogorov-Smirnov</i>)
SW	Shapiro-Wilk (<i>Shapiro-Wilk</i>)
NPD	Não positiva definida (<i>not positive definite</i>)
Nc	Não convergência (<i>non-convergence</i>)

Lista de Símbolos

$\mathbf{A}, \mathbf{B}, \dots$	Matrizes ou vetores
$\mathbf{a}, \mathbf{b}, \dots$	Vetores
$\ \mathbf{A}\ $	Norma euclidiana da matriz $\mathbf{A}$
$\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \dots$	Vetores de parâmetros desconhecidos
$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \dots$	Estimadores dos vetores de parâmetros desconhecidos
$\boldsymbol{\Theta}$	Espaço paramétrico
$\text{vec}(\cdot)$	Operador que transforma uma matriz em vetor, empilhando as colunas da matriz uma embaixo da outra
$\text{vech}(\cdot)$	Operador que transforma a parte triangular inferior de uma matriz em vetor
$\otimes$	Produto de Kronecker
$\text{bdiag}\{\cdot\}$	Operador que transforma matrizes em uma matriz diagonal em blocos
$\det(\mathbf{A})$	Determinante da matriz $\mathbf{A}$
$\text{tr}\{\mathbf{A}\}$	Operador que soma os elementos da diagonal principal da matriz $\mathbf{A}$
$\xrightarrow{P_{\boldsymbol{\theta}}}$	Convergência em probabilidade sob a medida $P_{\boldsymbol{\theta}}$
$\xrightarrow{\mathcal{D}}$	Convergência em distribuição
$\ell(\cdot)$	Logaritmo da função de verossimilhança, também chamaremos de função de log-verossimilhança
$\mathcal{H}_0$	Hipótese nula
$\mathcal{H}_1$	Hipótese alternativa

Lista de Figuras

1.1	Diagrama de trabalhos que abordam o modelo de regressão elíptico com parametrização geral.	3
2.1	Medidas avaliadas na simulação <i>versus</i> tamanho amostral quando todas as condições são satisfeitas associadas às estimativas de $\hat{\sigma}^2$ apresentadas na Tabela 2.1.	23
2.2	Histogramas acompanhados da função densidade da distribuição normal padrão, considerando as réplicas de Z_n , associados às estimativas de β_1 , cenário em que todas as condições são satisfeitas.	24
2.3	Medidas avaliadas na simulação <i>versus</i> tamanho amostral quando a condição C4 é violada associadas às estimativas de $\hat{\sigma}_x^2$	26
2.4	Histogramas acompanhados da função densidade da distribuição normal padrão, considerando as réplicas de Z_n , associados às estimativas de β_1 , cenário que viola a condição C4.	27

Lista de Tabelas

1.1	Distribuições elípticas multivariadas ¹	2
2.1	Resultados da simulação de Monte Carlo considerando o modelo dado em (2.13), com $\beta_0 = 2$, $\beta_1 = 1$, $\mu_x = -2$, $\sigma_x^2 = 4$, $\sigma^2 = 10$, $\sqrt{\tau_{xi}} \sim U(0.5; 1.5)$ e $\sqrt{\tau_{yi}} \sim U(0.5; 4)$. As colunas apresentam a média, o viés e a raiz quadrada do erro quadrático médio dos estimadores ($\sqrt{EQM}$), a média ($Média_{Z_n}$), o desvio-padrão (DP_{Z_n}), o coeficiente de assimetria (CA_{Z_n}), o coeficiente de curtose (C_{Z_n}) e os p-valores associados ao teste de normalidade de Kolmogorov-Smirnov (KS_{Z_n}) e de Shapiro-Wilk (SW_{Z_n}) da variável padronizada Z_n . Espera-se que os valores das medidas associadas a Z_n se encontrem próximos aos da normal padrão.	22
2.2	Resultados da simulação de Monte Carlo considerando o modelo dado em (2.13), com $\beta_0 = 2$, $\beta_1 = 0$, $\mu_x = 0$, $\sigma_x^2 = 4$, $\sigma^2 = 10$, $\sqrt{\tau_{yi}} \sim U(0.5; 4)$ e $\tau_{xi} = 2(i^{1.1})$. As colunas apresentam a média, o viés e a raiz quadrada do erro quadrático médio dos estimadores ($\sqrt{EQM}$), a média ($Média_{Z_n}$), o desvio-padrão (DP_{Z_n}), o coeficiente de assimetria (CA_{Z_n}), o coeficiente de curtose (C_{Z_n}) e os p-valores associados ao teste de normalidade de Kolmogorov-Smirnov (KS_{Z_n}) e de Shapiro-Wilk (SW_{Z_n}) da variável padronizada Z_n . Espera-se que os valores das medidas associadas a Z_n se encontrem próximos aos da normal padrão, quando satisfeitas as condições de regularidade.	25
3.1	Funções $\psi_i(2, 1)$, $\psi_i(2, 2)$, $\psi_i(3, 2)$ e $\psi_i(3, 3)$ para as distribuições normal, Cauchy, t de Student (t) e exponencial potência (EP).	30

Capítulo 1

Introdução

A análise de regressão é uma das técnicas estatísticas mais utilizadas para verificar hipóteses científicas e fazer previsões em diversas áreas. Em geral, o objetivo é investigar a relação entre uma variável de interesse (variável resposta) e variáveis explicativas (covariáveis). Os modelos de regressão baseados na suposição de normalidade são os mais conhecidos e por muito tempo foram os mais utilizados na prática. No entanto, a suposição de normalidade nem sempre é satisfeita. Em tais situações, podem ser utilizadas distribuições alternativas, sendo uma escolha possível a família de distribuições elípticas, a qual inclui a distribuição normal. Esta classe de distribuições tem recebido atenção crescente na literatura estatística, em especial, pelo fato de que abrange todas as distribuições contínuas simétricas. Nesta classe estão incluídas distribuições importantes como, por exemplo, t de Student, exponencial potência, normal contaminada, logística I e logística II, entre outras, com caudas mais pesadas ou mais leves do que as caudas da distribuição normal.

A seguir, adotamos a mesma notação de [Lemonte e Patriota \(2011\)](#), a fim de introduzir a distribuição elíptica multivariada. Seja $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$ um vetor aleatório, o qual tem distribuição elíptica multivariada, definida em [Fang et al. \(2018\)](#), com parâmetro de localização $\boldsymbol{\mu}$ ($d \times 1$) e uma matriz escala $\boldsymbol{\Sigma}$ ($d \times d$) positiva definida. Se a função densidade existe, é dada por

$$f_{\mathbf{Y}}(\mathbf{y}) = \det(\boldsymbol{\Sigma})^{-1/2} g \left[(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right], \quad \mathbf{y} \in \mathbb{R}^d,$$

em que $g : \mathbb{R} \rightarrow [0, \infty)$ é tal que $\int_0^\infty u^{\frac{d}{2}-1} g(u) du < \infty$. A função $g(\cdot)$ é conhecida como função geradora de densidade. Escrevemos $\mathbf{Y} \sim \mathcal{E}l_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$. A função característica é dada por $\psi(\mathbf{t}) = E(\exp(i\mathbf{t}^\top \mathbf{Y})) = \exp(i\mathbf{t}^\top \boldsymbol{\mu}) \varphi(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t})$ em que $\mathbf{t} \in \mathbb{R}^d$ e $\varphi : [0, \infty) \rightarrow \mathbb{R}$. Se φ é duas vezes diferenciável em zero, temos que $E(\mathbf{Y}) = \boldsymbol{\mu}$ e $\text{Var}(\mathbf{Y}) = \xi \boldsymbol{\Sigma}$, em que $\xi = \varphi'(0)$. Quando $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \mathbf{I}_d$, em que $\mathbf{I}_d$ é a matriz identidade, temos a família esférica de densidades. Mais detalhes sobre a classe de distribuições elípticas podem ser obtidos em [Fang et al. \(2018\)](#).

A Tabela 1.1, obtida de [Galea et al. \(2000\)](#) e também apresentada em [Lemonte e Patriota \(2011\)](#), considera alguns exemplos de distribuições da família elíptica com a forma da função geradora em cada caso. Notemos que distribuições bem conhecidas na literatura estatística pertencem à família de distribuições elípticas.

Assumamos que $g(\cdot)$ é uma função contínua e duas vezes diferenciável. No contexto elíptico, é usual definir as seguintes quantidades:

$$W_g(u) = \frac{d \log g(u)}{du} = \frac{g'(u)}{g(u)} \quad \text{e} \quad W_g^2(u) = \frac{d^2 \log g(u)}{du^2} = \frac{g''(u)}{g(u)} - \left(\frac{g'(u)}{g(u)} \right)^2.$$

A seguir, são exemplificadas essas quantidades para as distribuições normal, t de Student, Cauchy, logística e exponencial potência.

Para a distribuição normal, denotada por $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, temos:

$$W_g(u) = -\frac{1}{2} \quad \text{e} \quad W_g^2(u) = 0.$$

Tabela 1.1: *Distribuições elípticas multivariadas*¹.

Distribuição	Notação	Função geradora
Normal	$\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$g(u) = c \exp(-u/2), \quad u \geq 0$
t de Student	$t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	$g(u) = c(1 + u/\nu)^{-(\nu+d)/2}, \quad u \geq 0$
Cauchy	$\mathcal{C}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$g(u) = c(1 + u)^{-(d+1)/2}, \quad u \geq 0$
Logística	$\mathcal{L}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$g(u) = ce^{-u}/(1 + e^{-u})^2, \quad u \geq 0$
Exponencial potência	$\mathcal{EP}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha)$	$g(u) = c \exp(-u^\alpha/2), \quad u \geq 0$
Normal contaminada	$\mathcal{NC}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \delta, \tau)$	$g(u) = c [(1 - \delta) \exp(-u/2) + \delta \tau^{-d/2} \exp(-u/(2\tau))],$ $u \geq 0$

¹ c denota a constante normalizadora.

Para a distribuição t de Student, denotada por $t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, com ν graus de liberdade, $\nu > 0$, temos (Lange *et al.*, 1989):

$$W_g(u) = -\frac{1}{2} \left(\frac{\nu + d}{\nu + u} \right) \text{ e } W_g^2(u) = \frac{1}{2} \left(\frac{\sqrt{\nu + d}}{\nu + u} \right)^2.$$

Para a distribuição Cauchy, denotada por $\mathcal{C}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, temos:

$$W_g(u) = -\frac{1}{2} \left(\frac{1 + d}{1 + u} \right) \text{ e } W_g^2(u) = \frac{1}{2} \left(\frac{\sqrt{1 + d}}{1 + u} \right)^2.$$

Para a distribuição logística, denotada por $\mathcal{L}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, temos:

$$W_g(u) = \frac{-1 + e^{-u}}{1 + e^{-u}} \text{ e } W_g^2(u) = \frac{-2e^{-u}}{(1 + e^{-u})^2}.$$

Para a distribuição exponencial potência, denotada por $\mathcal{EP}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha)$, temos (Gómez *et al.*, 1998):

$$W_g(u) = -\frac{\alpha u^{\alpha-1}}{2} \text{ e } W_g^2(u) = -\frac{\alpha(\alpha-1)u^{\alpha-2}}{2}.$$

Os modelos de regressão cuja distribuição do erro aleatório pertence à classe de distribuições elípticas multivariadas têm recebido grande destaque na comunidade acadêmica nos últimos anos. Uma proposta inovadora é apresentada em Lemonte e Patriota (2011), em que o modelo de regressão elíptico multivariado com parametrização geral é introduzido. Nessa abordagem, o vetor de localização e a matriz de escala compartilham parâmetros. Como consequência dessa abordagem, ocorre a unificação de diversos modelos já conhecidos na literatura, tais como: os modelos heteroscedásticos com erros nas variáveis (Akritas e Bershad, 1996; Kulathinal *et al.*, 2002; Patriota *et al.*, 2011), os modelos mistos com efeito fixo não linear (Patriota, 2011; Russo *et al.*, 2009), modelos mistos com uma covariável sujeita a erros de medida (Borssoi *et al.*, 2017), modelo de regressão multivariado com erros aleatórios equicorrelacionados (Ibacache-Pulgar *et al.*, 2014) e modelos multiníveis (Manghi *et al.*, 2016), entre outros. A Figura 1.1 apresenta um diagrama que relaciona os trabalhos existentes na literatura sobre o modelo de regressão elíptico com parametrização geral.

No trabalho de Lemonte e Patriota (2011), as expressões da função escore e das matrizes de informação observada e esperada foram obtidas por meio de operações matriciais simples e interpretáveis. Os autores apresentam o algoritmo do método escore de Fisher para obtenção dos EMV nesse modelo. Usualmente, as inferências são baseadas em aproximações assintóticas da distribuição dos estimadores de máxima verossimilhança (EMV), sendo necessárias algumas suposições

Figura 1.1: Diagrama de trabalhos que abordam o modelo de regressão elíptico com parametrização geral.

para que essas aproximações sejam válidas. [Melo et al. \(2018\)](#) obtêm o viés de segunda ordem do EMV e propõem estimadores com viés reduzido. Já [Melo et al. \(2017\)](#) apresentam as estatísticas da razão de verossimilhanças ajustada via Skovgaard ([Skovgaard, 2001](#)) e via Barndorff-Nielsen ([Barndorff-Nielsen, 1986](#)) para inferências no modelo de regressão elíptico multivariado com parametrização geral.

Esses trabalhos consideram algumas suposições acerca do comportamento da função de log-verossimilhança e suas derivadas. Ademais, [Melo et al. \(2017\)](#) comentam acerca da necessidade de condições de regularidade para a validade do Teorema de Lindeberg-Feller (ou o Teorema de Liapounov), ver [Sen e Singer \(1993\)](#), para que a função escore convirja para a distribuição normal, o que permitiria conhecer a distribuição assintótica da estatística de teste considerada no trabalho de [Melo et al. \(2017\)](#), mas essas condições não são detalhadas. Nesse sentido, até o momento, não há publicações a respeito de resultados sobre o estudo das condições de regularidade para a identificabilidade do modelo, existência, unicidade, consistência e normalidade assintótica dos EMV, e distribuição assintótica das estatísticas de teste, por exemplo. Portanto, percebemos a necessidade de um estudo sistemático e detalhado das condições de regularidade para o modelo de regressão elíptico multivariado com parametrização geral.

1.1 Considerações preliminares

O estudo detalhado e sistemático das condições de regularidade nos modelos estatísticos não é um assunto muito abordado na literatura estatística de modelos de regressão. Em geral, as condições de regularidade não são verificadas com rigor em grande parte dos trabalhos estatísticos. Dessa forma, considerada a generalidade do modelo, as condições de regularidade estudadas nesta tese são mais específicas do que as especificadas na literatura, por exemplo em [Cox e Hinkley \(1974\)](#), as quais consideram condições sobre a função de verossimilhança e não sobre o modelo específico. Portanto, este estudo poderá servir de aporte teórico para outros estudos a fim de possibilitar o desenvolvimento da teoria assintótica de forma detalhada em uma grande variedade de modelos, que são casos especiais do modelo elíptico com parametrização geral.

Dentre os trabalhos que abordam condições de regularidade, um trabalho importante é o de [Fahrmeir e Kaufmann \(1985\)](#), o qual trata das propriedades assintóticas, a saber, existência, consistência e normalidade assintótica do EMV em modelos lineares generalizados (MLG). Esse trabalho tem sido referência no estudo da teoria assintótica em modelos com estrutura semelhante à dos MLGs. Adicionalmente, o artigo de [Krätschmer \(2007\)](#) aborda a questão da unicidade com probabilidade um de estimativas extremas em distribuições dominadas pela medida de Lebesgue. Os resultados de ambos trabalhos são utilizados como suporte para obtenção das condições de regularidade nesta tese.

Inicialmente, no Capítulo 2, é considerado o modelo de regressão normal multivariado com parametrização geral apresentado em [Patriota e Lemonte \(2009\)](#), o qual é um caso particular do

modelo elíptico, assumindo a distribuição normal para os erros. Obtidos os resultados para o modelo normal, no Capítulo 3, os resultados são generalizados para o modelo que considera a classe de distribuições elípticas para os erros.

A fim de uma melhor leitura e entendimento desta tese, destacamos que são necessários alguns requisitos básicos sob os quais o leitor tenha familiaridade. São eles:

- Modelos de regressão de forma geral;
- Operações, derivadas e propriedades matriciais;
- Teoria assintótica frequentista, a saber: convergências em probabilidade, em distribuição, ordens de magnitude, Teorema Central do Limite de Liapounov, Lei Fraca dos Grandes Números e demais teoremas que embasam a teoria assintótica.

1.2 Objetivos

O objetivo geral desse trabalho é estudar as condições de regularidade relacionadas à teoria assintótica frequentista, a saber: distribuição assintótica dos EMV, propriedades assintóticas dos EMV e distribuição assintótica das estatísticas de teste no modelo de regressão elíptico com parametrização geral.

Objetivos específicos

O objetivo geral é dividido nos seguintes objetivos específicos:

- Revisar o estado da arte acerca do modelo de regressão elíptico com parametrização geral;
- Identificar algumas condições de regularidade e, a partir delas, propor teoremas. Tais condições estão relacionadas à identificabilidade do modelo, à existência, à unicidade, à consistência e normalidade assintótica dos EMV e a distribuição assintótica das estatísticas de teste;
- Realizar estudos de simulação em que são violadas algumas condições de regularidade em modelos particulares do modelo geral;
- Especificar os resultados obtidos para o modelo de regressão elíptico com parametrização geral em alguns modelos particulares a fim de motivar o uso do modelo e demonstrar sua generalidade.

1.3 Contribuições

Por fim, as principais contribuições deste trabalho são as seguintes:

- Apresentação das condições de regularidade e formalização dos teoremas que garantem a existência dos EMV no modelo com parametrização geral assumindo distribuição normal e elíptica para os erros;
- Identificação das condições de regularidade e desenvolvimento dos teoremas que garantem a unicidade dos EMV, abordando também a questão da identificabilidade do modelo assumindo distribuição normal e elíptica para os erros;
- Apresentação das condições de regularidade e formalização dos teoremas que garantem a consistência e normalidade assintótica dos EMV no modelo com parametrização geral assumindo distribuição normal e elíptica para os erros;
- Descrição das condições de regularidade para a distribuição assintótica das estatísticas de teste, adicionada a formulação e demonstração de um teorema no caso de erros normais e elípticos;

- Exemplificação da teoria estudada em casos particulares do modelo com parametrização geral, facilitando assim o entendimento em relação à generalidade e à complexidade do modelo em estudo.

1.4 Organização do trabalho

O presente trabalho está organizado da seguinte forma. No Capítulo 2, apresentamos o modelo de regressão normal multivariado com parametrização geral, as condições de regularidade e teoremas para existência, unicidade, consistência e normalidade assintótica dos EMV e também para a distribuição assintótica das estatísticas de teste. Ainda, são apresentados exemplos em que os teoremas são aplicados em modelos específicos a fim de ilustrar analiticamente o uso dos teoremas. Também discutimos um exemplo em que uma das condições é violada com o propósito de ilustrar computacionalmente como essa condição afeta os resultados. Ao finalizar o Capítulo 2 ressaltamos os principais pontos estudados no seu desenvolvimento. No Capítulo 3, apresentamos o modelo de regressão elíptico multivariado com parametrização geral, generalizamos alguns resultados do Capítulo 2, entre eles, a existência, unicidade, consistência e normalidade assintótica do EMV, também generalizamos o resultado da distribuição assintótica das estatísticas de teste. Também apresentamos exemplos para ilustrar o uso dos teoremas e demonstrar a generalidade do modelo. Ao fim do Capítulo 3 destacamos as principais contribuições obtidas em seu desenvolvimento. Por fim, no Capítulo 4, discutimos algumas conclusões obtidas neste trabalho. Além disso, apresentamos ideias para a continuidade deste trabalho e para realização de trabalhos futuros.

As provas dos teoremas apresentados no Capítulo 2 são desenvolvidas no Apêndice A. As provas dos teoremas do Capítulo 3 são oferecidas no Apêndice B, sendo algumas delas análogas às do Capítulo 2 apresentadas no Apêndice A.

Capítulo 2

Condições de regularidade para o modelo de regressão normal multivariado com parametrização geral

O modelo de regressão normal com parametrização geral apresentado em [Patriota e Lemonte \(2009\)](#) é um caso particular do modelo de regressão elíptico que vem sendo discutido nesta tese. Nesse sentido, iniciamos o estudo das condições de regularidade para o modelo normal e, posteriormente, estendemos esses resultados para a classe de distribuições elípticas.

2.1 O modelo de regressão normal com parametrização geral

O modelo de regressão normal multivariado com parametrização geral tem a mesma estrutura do modelo elíptico. Seguiremos a mesma notação apresentada em artigos anteriores ([Lemonte e Patriota, 2011](#); [Patriota e Lemonte, 2009](#)). Consideremos n vetores aleatórios independentes $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$, n denota o número de unidades amostrais. O modelo normal multivariado com parametrização geral pode ser representado pela forma simplificada abaixo

$$\mathbf{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \mathbf{x}_i) + \mathbf{e}_i, \quad i = 1, \dots, n, \quad (2.1)$$

com $\mathbf{e}_i \stackrel{\text{ind}}{\sim} \mathcal{N}_{d_i}(\mathbf{0}, \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \mathbf{w}_i))$, em que “ $\stackrel{\text{ind}}{\sim}$ ” denota “independentemente distribuído”, $\mathbf{x}_i$ e $\mathbf{w}_i$ são vetores $m_i \times 1$ e $k_i \times 1$ de variáveis auxiliares conhecidas, respectivamente, associadas à i -ésima resposta observada $\mathbf{Y}_i$. Então,

$$\mathbf{Y}_i \stackrel{\text{ind}}{\sim} \mathcal{N}_{d_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad i = 1, \dots, n,$$

em que $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \mathbf{x}_i)$ é o vetor de médias e $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \mathbf{w}_i)$ é uma matriz de covariâncias $d_i \times d_i$ definida positiva. $\boldsymbol{\theta}$ é o vetor $p \times 1$ de parâmetros de interesse desconhecidos, com $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$. $\boldsymbol{\mu}_i$ e $\boldsymbol{\Sigma}_i$ têm forma conhecida e assumimos que são duas vezes diferenciáveis com respeito a cada elemento de $\boldsymbol{\theta}$.

A fim de simplificar a notação, sejam $\mathbf{z}_i = \mathbf{Y}_i - \boldsymbol{\mu}_i$ e $\mathbf{u}_i = \mathbf{z}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{z}_i$. A função de log-verossimilhança associada ao modelo, a menos de uma constante, é dada por

$$\ell(\mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}),$$

em que $\ell_i(\boldsymbol{\theta}) = -\frac{1}{2} \log \det(\boldsymbol{\Sigma}_i) - \frac{1}{2} u_i$. Definamos

$$\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\theta}^\top}, \quad \mathbf{V}_i = \frac{\partial \text{vec}(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{\theta}^\top} \quad \text{e} \quad \tilde{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i,$$

em que o operador $\text{vec}(\cdot)$ transforma uma matriz em um vetor, empilhando as colunas da matriz uma embaixo da outra e $\otimes$ denota o produto de Kronecker. Os cálculos para obtenção da função escore e da matriz informação foram realizados por meio de métodos de derivação matricial apresentados em Magnus e Neudecker (2019). Mais detalhes podem ser obtidos em Patriota e Lemonte (2009). O vetor escore e a matriz informação de Fisher após alguma álgebra matricial podem ser escritos, respectivamente, como (Patriota e Lemonte, 2009)

$$\mathbf{U}_\theta = \sum_{i=1}^n \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i \quad \text{e} \quad \mathbf{K}_\theta = \sum_{i=1}^n \mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i, \quad (2.2)$$

em que

$$\mathbf{F}_i = \begin{pmatrix} \mathbf{D}_i \\ \mathbf{V}_i \end{pmatrix}, \quad \mathbf{H}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i & \mathbf{0} \\ \mathbf{0} & 2\tilde{\boldsymbol{\Sigma}}_i \end{pmatrix}^{-1} \quad \text{e} \quad \mathbf{s}_i = \begin{pmatrix} \mathbf{z}_i \\ \text{vec}(\mathbf{z}_i \mathbf{z}_i^\top - \boldsymbol{\Sigma}_i) \end{pmatrix}.$$

A matriz de informação observada de $\boldsymbol{\theta}$ é dada por $\mathbf{J}_\theta = -\mathbf{L}_\theta$, obtida de (Patriota e Lemonte, 2009)

$$\mathbf{L}_\theta = \frac{\partial^2 \ell(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \sum_{i=1}^n \left(\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i \right), \quad (2.3)$$

em que $\mathbf{B}_i$ é uma matriz $p \times p$ com o elemento (s, r) , $s = 1, \dots, p$ e $r = 1, \dots, p$, dado por $\mathbf{r}_i^\top \mathbf{H}_i \mathbf{s}_i$, com o vetor $\mathbf{r}_i = \begin{pmatrix} \mathbf{a}_{i(sr)} \\ \text{vec}(\mathbf{C}_{i(sr)}) \end{pmatrix}$, $\mathbf{a}_{i(sr)} = \frac{\partial^2 \boldsymbol{\mu}_i}{\partial \theta_s \partial \theta_r}$, $\mathbf{C}_{i(sr)} = \frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial \theta_s \partial \theta_r}$ e

$$\ddot{\mathbf{M}}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i & 2\boldsymbol{\Sigma}_i \otimes \mathbf{z}_i^\top \\ 2\boldsymbol{\Sigma}_i \otimes \mathbf{z}_i & \mathbf{M}_{i;22} \end{pmatrix},$$

em que $\mathbf{M}_{i;22} = 2[(\mathbf{z}_i \mathbf{z}_i^\top) \otimes \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i \otimes (\mathbf{z}_i \mathbf{z}_i^\top) - \tilde{\boldsymbol{\Sigma}}_i]$. Adicionalmente, temos que $E(\ddot{\mathbf{M}}_i) = \mathbf{H}_i^{-1}$ e $E(\mathbf{B}_i) = \mathbf{0}$. Vale ressaltar que a obtenção do vetor escore, da matriz de informação de Fisher e da matriz de informação observada se dá mediante a condição de que $\boldsymbol{\mu}_i$ e $\boldsymbol{\Sigma}_i$ têm forma conhecida e são duas vezes diferenciáveis com respeito a cada elemento de $\boldsymbol{\theta}$. Na Seção 2.2, esta é uma das condições formalizada.

O método escore de Fisher pode ser utilizado para estimar $\boldsymbol{\theta}$ por meio da solução iterativa da equação

$$\mathbf{F}_\theta^{(m)\top} \mathbf{H}^{(m)} \mathbf{F}_\theta^{(m)} \boldsymbol{\theta}^{(m+1)} = \mathbf{F}_\theta^{(m)\top} \mathbf{H}^{(m)} \mathbf{s}^{*(m)}, \quad m = 0, 1, \dots, \quad (2.4)$$

em que

$$\begin{aligned} \mathbf{F}_\theta^{(m)\top} &= (\mathbf{F}_1^{(m)\top}, \dots, \mathbf{F}_n^{(m)\top}), \quad \mathbf{H}^{(m)} = \text{bdiag}\{\mathbf{H}_1^{(m)}, \mathbf{H}_2^{(m)}, \dots, \mathbf{H}_n^{(m)}\}, \\ \mathbf{s}^{*(m)} &= \mathbf{F}_\theta^{(m)} \boldsymbol{\theta}^{(m)} + \mathbf{s}^{(m)}, \quad \mathbf{s}^{(m)} = (\mathbf{s}_1^{(m)\top}, \dots, \mathbf{s}_n^{(m)\top})^\top, \end{aligned}$$

e $\mathbf{H} = \text{bdiag}\{\mathbf{H}_1, \dots, \mathbf{H}_n\}$ denota uma matriz diagonal em blocos com as matrizes $\mathbf{H}_1, \dots, \mathbf{H}_n$ na diagonal e m indica o número da iteração. O esquema apresentado em (2.4) representa os passos iterativos de mínimos quadrados reponderados, sendo as iterações realizadas até que a convergência seja alcançada. Notemos pela equação (2.4) que a estimação de $\boldsymbol{\theta}$ via máxima verossimilhança pode ser realizada utilizando qualquer *software* com uma biblioteca de álgebra matricial. Vale destacar que

devido à complexidade das funções não lineares a convergência é sensível ao ponto inicial utilizado no procedimento de otimização. Neste trabalho, a estimação será realizada utilizando o *software* R (R Core Team, 2018), com auxílio da função `nlminb`, por meio de um método quasi-Newton, o algoritmo utilizado é o L-BFGS-B, o qual é uma extensão do algoritmo L-BFGS (Press *et al.*, 1992), para mais detalhes ver Zhu *et al.* (1997).

Qualquer sequência de estimadores $\{\hat{\boldsymbol{\theta}}_n\}_{n \geq 1}$ tal que

$$\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_n(\mathbf{y}) = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell(\mathbf{y}, \boldsymbol{\theta}),$$

é chamado EMV. Sob certas condições $\{\hat{\boldsymbol{\theta}}_n\}_{n \geq 1}$ é único e satisfaz a equação $\mathbf{U}_{\hat{\boldsymbol{\theta}}_n} = \mathbf{0}$.

No modelo (2.1), não é evidente que esse estimador é único. A seguir, apresentamos as condições de regularidade para existência, unicidade com probabilidade um, consistência e normalidade assintótica do EMV e para a distribuição assintótica das estatísticas de teste, além de alguns exemplos associados às condições estudadas.

2.2 Condições de regularidade

Definamos algumas quantidades necessárias a seguir. Para a matriz $\mathbf{A} = (a_{ij})$ a norma de Frobenius ou norma euclidiana, denotada por $\|\mathbf{A}\|$, é dada por $\|\mathbf{A}\| := \left(\sum_j \sum_i a_{ij}^2 \right)^{1/2}$.

Definamos

$$B_{\boldsymbol{\theta}}(\delta) := \left\{ \boldsymbol{\theta}' \in \boldsymbol{\Theta} : \|\mathbf{K}_{\boldsymbol{\theta}}^{1/2}(\boldsymbol{\theta}' - \boldsymbol{\theta})\| \leq \delta \right\},$$

uma sequência que depende de n através de $\mathbf{K}_{\boldsymbol{\theta}}$, da vizinhança de $\boldsymbol{\theta}$, com $\delta > 0$. As expressões da função de log-verossimilhança ($\ell(\mathbf{y}, \boldsymbol{\theta})$), do vetor escore ($\mathbf{U}_{\boldsymbol{\theta}}$), da matriz de informação observada ($\mathbf{J}_{\boldsymbol{\theta}}$) também são expressões que dependem de n . No entanto, a fim de simplificar a notação omitimos n .

Seja $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top, \dots, \mathbf{Y}_n^\top)$ um vetor, com n sendo o número de unidades amostrais, $\mathbf{Y}_i$ é um vetor $d_i \times 1$ e $N = d_1 + \dots + d_n$. Definamos a função $\mathbf{G} : \boldsymbol{\Theta} \rightarrow \mathbb{R}^{N + \frac{N(N-1)}{2}}$, tal que

$$\mathbf{G}(\boldsymbol{\theta}) = (\boldsymbol{\mu}(\boldsymbol{\theta})^\top, \text{vech}(\boldsymbol{\Sigma}(\boldsymbol{\theta}))^\top)^\top, \quad (2.5)$$

com $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\boldsymbol{\mu}_1(\boldsymbol{\theta})^\top, \dots, \boldsymbol{\mu}_n(\boldsymbol{\theta})^\top)^\top$ um vetor $N \times 1$, $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \text{bdiag}\{\boldsymbol{\Sigma}_1(\boldsymbol{\theta}), \dots, \boldsymbol{\Sigma}_n(\boldsymbol{\theta})\}$ uma matriz $N \times N$, em que a operação $\text{vech}(\boldsymbol{\Sigma})$ transforma a parte triangular inferior da matriz $\boldsymbol{\Sigma}$ em um vetor.

Considere as condições enunciadas a seguir

C1. O espaço paramétrico $\boldsymbol{\Theta} \subseteq \mathbb{R}^p$ é aberto e convexo.

Esta condição se faz necessária no desenvolvimento teórico que segue. Ela nos permite, por exemplo, aplicar o teorema do valor médio multivariado, o qual será utilizado na demonstração da normalidade assintótica dos EMV.

C2. As duas primeiras derivadas de $\boldsymbol{\mu}_i$ e $\boldsymbol{\Sigma}_i$ com respeito a cada elemento de $\boldsymbol{\theta}$ existem, são contínuas e limitadas, para todo $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

A condição C2 é uma condição necessária para que possamos estabelecer o vetor escore e a matriz informação de Fisher, por exemplo. Essa condição foi utilizada para obtenção de (2.2) e (2.3) na Seção 2.1.

C3. $\mathbf{F}_{\boldsymbol{\theta}} = (\mathbf{F}_1^\top, \dots, \mathbf{F}_n^\top)^\top$ tem posto coluna completo, ou seja, o número de colunas linearmente independentes (LI) é p , assim $\text{posto}(\mathbf{F}_{\boldsymbol{\theta}}) = p$, para todo $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Notemos que se a condição C3 for satisfeita, $\mathbf{K}_{\boldsymbol{\theta}}$ é definida positiva.

C4. $\lim_{\substack{n \rightarrow \infty \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \frac{1}{n} \mathbf{K}_{\boldsymbol{\theta}} = \boldsymbol{\Lambda}_{\boldsymbol{\theta}}$ existe, é uma matriz definida positiva e seus elementos são finitos para todo

Se a condição C4 for satisfeita temos $\text{ch}_p(\mathbf{K}_{\boldsymbol{\theta}}) \rightarrow \infty$, quando $n \rightarrow \infty$, em que $\text{ch}_p(\mathbf{K}_{\boldsymbol{\theta}})$ é o menor autovalor da matriz $\mathbf{K}_{\boldsymbol{\theta}}$. Esta condição é apresentada em [Fahrmeir e Kaufmann \(1985\)](#) e conhecida como condição de divergência.

C5. Para cada $\delta > 0$,

$$\lim_{n \rightarrow \infty} P_{\boldsymbol{\theta}} \left(\mathbf{J}_{\boldsymbol{\theta}'} - c \mathbf{K}_{\boldsymbol{\theta}} \text{ ser semidefinida positiva para todo } \boldsymbol{\theta}' \in B_{\boldsymbol{\theta}}(\delta) \right) = 1,$$

em que c é uma constante, tal que $c > 0$ é independente de δ .

A condição C5 estabelece que a matriz $\mathbf{J}_{\boldsymbol{\theta}'} - c \mathbf{K}_{\boldsymbol{\theta}}$ será semidefinida positiva com $\boldsymbol{\theta}' \in \partial B_{\boldsymbol{\theta}}(\delta)$, em que $\partial B_{\boldsymbol{\theta}}(\delta)$ denota a borda de $B_{\boldsymbol{\theta}}(\delta)$, o que é abordado em detalhes no Apêndice A, quando as equações abaixo são satisfeitas,

$$\begin{aligned} & \boldsymbol{\lambda}^\top (\mathbf{J}_{\boldsymbol{\theta}'} - c \mathbf{K}_{\boldsymbol{\theta}}) \boldsymbol{\lambda} \geq 0, \quad \boldsymbol{\lambda} \in \mathbb{R}^p, \quad \boldsymbol{\lambda} \neq \mathbf{0}, \\ \Leftrightarrow & \boldsymbol{\lambda}^\top \mathbf{J}_{\boldsymbol{\theta}'} \boldsymbol{\lambda} \geq c \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}} \boldsymbol{\lambda}, \quad \boldsymbol{\lambda} \in \mathbb{R}^p, \quad \boldsymbol{\lambda} \neq \mathbf{0}, \\ \Leftrightarrow & \text{ch}_1(\mathbf{J}_{\boldsymbol{\theta}'}) \geq \text{ch}_p(\mathbf{J}_{\boldsymbol{\theta}'}) \geq c \text{ch}_1(\mathbf{K}_{\boldsymbol{\theta}}) \geq c \text{ch}_p(\mathbf{K}_{\boldsymbol{\theta}}), \end{aligned}$$

em que $\text{ch}_1(\mathbf{A})$ denota o maior autovalor da matriz $\mathbf{A}$. Esse resultado é consequência da Proposição A.1.1 enunciada no Apêndice A, tal resultado é apresentado em [Horn e Johnson \(2013\)](#).

Notemos que se a condição C4 for satisfeita, existe c tal que a condição C5 é válida desde que $\text{ch}_p(\mathbf{J}_{\boldsymbol{\theta}'}) \rightarrow \infty$ quando $n \rightarrow \infty$.

C6. Consideremos a função $\mathbf{G} : \boldsymbol{\Theta} \rightarrow \mathbb{R}^{N + \frac{N(N-1)}{2}}$ definida em (2.5). Assumamos que existem subconjuntos $\boldsymbol{\Theta}^1, \boldsymbol{\Theta}^2, \dots$ abertos e convexos e $\mathbf{G}^1, \mathbf{G}^2, \dots$ tais que

$$\mathbf{G}(\boldsymbol{\Theta}) = \bigcup_{j=1}^{\infty} \mathbf{G}^j(\boldsymbol{\Theta}^j),$$

em que $\mathbf{G}^j(\boldsymbol{\Theta}^j)$ é uma função duas vezes diferenciável e o posto do jacobiano de $\mathbf{G}^j$ em relação a $\boldsymbol{\theta}$ é p_j (em que $\boldsymbol{\Theta} \subseteq \mathbb{R}^p$ é aberto e convexo), para cada $\boldsymbol{\theta} \in \boldsymbol{\Theta}^j$, $\boldsymbol{\Theta}^j$ um subconjunto de $\mathbb{R}^{p_j}$ e $j \geq 1$.

Em [Krätschmer \(2007\)](#), a função $\mathbf{G}^j(\boldsymbol{\Theta}^j)$ com as condições aqui descritas é dita “ C^2 -immersion”. Notemos que se as condições C1, C2 e C3 forem satisfeitas, então a condição C6 é satisfeita.

C7. A função $\mathbf{G}$ é injetiva.

Se $\mathbf{G}$ é injetiva, ou seja, $\mathbf{G}(\boldsymbol{\theta}_1) = \mathbf{G}(\boldsymbol{\theta}_2) \implies \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ para todo $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, então temos que $\boldsymbol{\theta}$ é identificável no modelo (2.1).

C8. Existe uma sequência $M_n = O_p(n)$ tal que

$$\|\mathbf{L}_{\boldsymbol{\theta}_n^*} - \mathbf{L}_{\boldsymbol{\theta}}\| \leq M_n \|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\|,$$

em que $\boldsymbol{\theta}_n^* = h \widehat{\boldsymbol{\theta}}_n + (1 - h) \boldsymbol{\theta}$, $h \in (0, 1)$, para todo $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

A condição C8 nos garante que a segunda derivada da função de log-verossimilhança é limitada em um determinado sentido com probabilidade um. Se a terceira derivada existir, for contínua e limitada, então temos que C8 é satisfeita.

As condições aqui enunciadas serão utilizadas nos teoremas de existência, unicidade, consistência e normalidade assintótica do EMV e no teorema que aborda a distribuição assintótica das estatísticas de teste.

2.2.1 Existência do EMV

Nesta subseção, discutimos a existência do EMV. O estudo das condições necessárias para a existência está baseado no artigo de [Fahrmeir e Kaufmann \(1985\)](#), sendo aqui realizado de forma semelhante, considerando as particularidades do modelo em estudo. A seguir, enunciamos o teorema e comentamos as condições estabelecidas.

Teorema 2.2.1. *Seja o modelo definido em (2.1). Sob as condições C1, C2, C3 e C5, existe uma sequência $\{\hat{\theta}_n\}_{n \geq 1}$ de vetores aleatórios tais que*

$$\lim_{n \rightarrow \infty} P_{\theta}(U_{\hat{\theta}_n} = 0) = 1, \text{ para cada } \theta \in \Theta(\text{existência assintótica}).$$

O resultado contido nesse teorema aponta para a existência assintótica do EMV. Ou seja, para cada $\theta \in \Theta$ e n suficientemente grande, existirá uma sequência $\{\hat{\theta}_n\}_{n \geq 1}$ que nos fornece o zero da função escore ($U_{\hat{\theta}_n}$) com probabilidade um. As condições C2 e C3 são necessárias para que possamos estabelecer o vetor escore, a matriz informação de Fisher e informação observada, o que foi feito na Seção 2.1. Já a condição C5 é uma condição suficiente para provarmos a existência assintótica do EMV no modelo normal com parametrização geral. Uma afirmação equivalente à condição C5 é que para todo $\lambda \in \mathbb{R}^p$ e qualquer $\delta > 0$,

$$\lim_{n \rightarrow \infty} P_{\theta}(\lambda^{\top} K_{\theta}^{-1/2} J_{\theta'} K_{\theta}^{-1/2} \lambda \geq c \lambda^{\top} \lambda, \quad \theta' \in B_{\theta}(\delta)) = 1, \quad (2.6)$$

Na prova do teorema, trabalhamos com a borda de $B_{\theta}(\delta)$ a fim de garantir que os resultados valem para qualquer $\theta' \in B_{\theta}(\delta)$. A condição C1 nos garante que, para qualquer combinação linear de θ' e θ , definida como θ^* , com $\theta' \in B_{\theta}(\delta)$, temos que $\theta^* \in B_{\theta}(\delta)$. O objetivo consiste em mostrar que existe θ tal que a função de log-verossimilhança é maior em θ do que em qualquer $\theta' \in B_{\theta}(\delta)$. Entre os passos da prova, expandimos a função de log-verossimilhança em série de Taylor, o que pode ser feito devido à condição C1. Em seguida, utilizamos a relação obtida pela expansão que considera o vetor escore e a matriz informação observada. Pela condição C5, mediante a relação (2.6) e a desigualdade de Markov, chegamos ao resultado. Cabe destacar que a prova se dá de maneira semelhante à apresentada em [Fahrmeir e Kaufmann \(1985\)](#), consideradas as particularidades do modelo aqui em estudo. A prova com detalhes é apresentada no Apêndice A, Seção A.1.

2.2.2 Unicidade do EMV

Nesta subseção, apresentamos condições suficientes para a unicidade do EMV de θ com probabilidade um. Aplicamos os resultados apresentados em [Krätschmer \(2007\)](#) em que são descritas condições suficientes que garantem a unicidade de estimativas de máxima verossimilhança em distribuições dominadas pela medida de Lebesgue. Aqui, mostramos que, no modelo (2.1), as condições suficientes são simplificadas.

Recordemos que a medida de Lebesgue em $\mathbb{R}$ é a distância entre dois pontos. No $\mathbb{R}^2$ é a área de um retângulo, no $\mathbb{R}^3$ a área de um cubo, e assim por diante. Ademais, se temos uma restrição em $\mathbb{R}^2$ tal que a dimensão é reduzida a $\mathbb{R}$ ($\mathbb{R}^2 \rightarrow \mathbb{R}$) a medida associada ao $\mathbb{R}^2$ neste caso é zero, ou seja, tem medida de Lebesgue nula.

Definamos o conjunto

$$R := \left\{ \mathbf{y} \in \mathbb{R}^N : U_{\hat{\theta}_n(\mathbf{y})} = 0 \right\} \cap \left\{ \mathbf{y} \in \mathbb{R}^N : \det(L_{\hat{\theta}_n(\mathbf{y})}) = 0 \right\}, \quad (2.7)$$

o qual corresponde ao conjunto dos vetores $\mathbf{y}$ tais que $\hat{\theta}_n(\mathbf{y})$ é ponto crítico degenerado da função de log-verossimilhança, ou seja, o determinante da matriz hessiana aplicada no ponto crítico é nulo. O teorema a seguir apresenta condições sob as quais a unicidade do EMV é garantida com probabilidade um.

Teorema 2.2.2. *Seja o modelo definido em (2.1). Sob as condições C7 e C6, se R tiver medida de Lebesgue nula, então o conjunto*

$$Q := \left\{ \mathbf{y} \in \mathbb{R}^N : \ell(\mathbf{y}, \boldsymbol{\theta}_1) = \ell(\mathbf{y}, \boldsymbol{\theta}_2) = \max_{\boldsymbol{\theta} \in \Theta} \ell(\mathbf{y}, \boldsymbol{\theta}), \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta \right\}$$

é mensurável e tem medida de Lebesgue nula.

Este teorema trata da unicidade do EMV com probabilidade um. A abordagem sobre a unicidade com probabilidade um é realizada através da medida nula do conjunto Q . Isso decorre do fato que o conjunto Q contém todo $\mathbf{y}$ tal que dois pontos distintos do espaço paramétrico ($\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$) maximizam a função de log-verossimilhança. Se o estimador for único, então o conjunto Q deverá ter medida nula. A condição de que o conjunto R tem medida nula está relacionada à matriz hessiana ser positiva ou negativa definida, se R tiver medida nula teríamos um ponto de máximo ou mínimo.

Para a prova desse teorema adaptamos as condições suficientes enunciadas em Krätschmer (2007) ao modelo em estudo. A condição C7, da injetividade da função $\mathbf{G}$, é utilizada inicialmente no Lema A.2.1, apresentado no Apêndice A, o qual aborda a derivada da função de log-verossimilhança em relação ao vetor $\mathbf{y}$. Pela injetividade, também temos a identificabilidade do modelo. Ainda, se as condições C7 e C6 forem satisfeitas, a função de log-verossimilhança será contínua e duas vezes diferenciável, esta é uma das condições enunciadas em Krätschmer (2007). A condição de que R tem medida nula foi estudada com mais detalhes no desenvolvimento dessa tese a fim de ser simplificada no modelo (2.1). No entanto, devido à generalidade do modelo, tal condição deve ser verificada em cada caso específico do modelo, conforme exemplo na Seção 2.3. Notemos ainda que as condições C1, C2 e C3 não são exigidas nesse teorema, porém, se forem satisfeitas temos que a condição C6 será satisfeita. A prova com detalhes é apresentada no Apêndice A, Seção A.2.

A seguir, apresentamos um corolário que nos garante a unicidade do EMV.

Corolário 2.2.1. *Se a matriz $\mathbf{J}_\theta$ for definida positiva quase certamente para todo $\theta \in \Theta$ e a condição C1 for satisfeita, então o EMV é único com probabilidade um.*

Percebamos que se a matriz $\mathbf{J}_\theta$ for definida positiva então ela admite inversa, logo $\det(\mathbf{L}_{\hat{\theta}_n(\mathbf{y})}) \neq 0$ e assim R tem medida de Lebesgue nula. A prova é trivial. Notemos ainda que se a matriz $\mathbf{J}_\theta$ for positiva definida a função de log-verossimilhança será concava, logo o EMV é único.

Vale destacar que Fahrmeir e Kaufmann (1985) discutem brevemente a unicidade do EMV nos MLG com função de ligação canônica. Nessa situação, temos que as informações observada e esperada coincidem. Fahrmeir e Kaufmann (1985) argumentam que, se a matriz de informação de Fisher for definida positiva e o espaço paramétrico convexo, o EMV é único. Isso corrobora com o Corolário 2.2.1.

Alternativamente, Krätschmer (2007) apresenta condições suficientes que substituem a condição de que o conjunto R tem medida de Lebesgue nula. Com o intuito de utilizarmos a abordagem adotada em Krätschmer (2007) consideramos o caso em que $\boldsymbol{\mu}(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ não compartilham parâmetros. Assim, escrevemos $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. Quando $\boldsymbol{\mu}(\boldsymbol{\theta})$ não depende de $\boldsymbol{\gamma}$ escrevemos $\boldsymbol{\mu}(\boldsymbol{\beta})$. De maneira análoga, quando $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ não depende de $\boldsymbol{\beta}$ escrevemos $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$. Adicionalmente, considere o vetor $\mathbf{x} = \mathbf{y} - \boldsymbol{\nu}$, em que $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N)^\top$ corresponde ao vetor de médias da variável aleatória $\mathbf{Y}$ sem estrutura de regressão na média. Definamos ainda $\hat{\boldsymbol{\gamma}}_{\mathbf{y}-\boldsymbol{\nu}} = \hat{\boldsymbol{\gamma}}_{\mathbf{x}}$ o EMV de $\boldsymbol{\gamma}$ para cada $\boldsymbol{\nu}$ fixado.

Teorema 2.2.3. *Seja o modelo definido em (2.1), com $\boldsymbol{\mu}(\boldsymbol{\beta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$. Sob as condições C1, C2, C3 e C7, se a função $f(\mathbf{x}) = \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}})\mathbf{x}$ for um difeomorfismo, então o conjunto*

$$Q := \left\{ \mathbf{y} \in \mathbb{R}^N : \ell(\mathbf{y}, \boldsymbol{\theta}_1) = \ell(\mathbf{y}, \boldsymbol{\theta}_2) = \max_{\boldsymbol{\theta} \in \Theta} \ell(\mathbf{y}, \boldsymbol{\theta}), \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta \right\}$$

é mensurável e tem medida de Lebesgue nula.

Esse teorema é uma alternativa ao Teorema 2.2.2, sendo aplicável no caso específico em que $\boldsymbol{\mu}(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ não compartilham parâmetros. Novamente, trabalhamos com o conjunto Q o qual

tem a mesma interpretação dada anteriormente, no entanto temos uma alteração nas condições. A principal diferença está associada à função $f(\mathbf{x})$, a qual deve ser um difeomorfismo, ou seja, $f(\mathbf{x})$ e $f^{-1}(\mathbf{x})$ devem ser diferenciáveis. Na elaboração desse teorema, adaptamos um teorema apresentado em Krätschmer (2007). Em Krätschmer (2007), são apresentadas algumas condições identificadas como 3.1 e 3.2, entre outras, aqui as escrevemos como C3.1 e C3.2, por exemplo. As condições C4.2 de Krätschmer (2007) e a condição de que $f(\mathbf{x})$ é um difeomorfismo substituem a condição de que R tem medida nula do Teorema 2.2.2. Mediante as condições C2 e C3, a condição C4.2 de Krätschmer (2007) no modelo (2.1) é satisfeita, tal fato provamos com detalhes no Apêndice A, Lema A.3.1. Ainda, sob as condições C1, C2 e C3, a condição C6 é satisfeita. Destacamos que a função $f(\mathbf{x}) = \Sigma^{-1}(\hat{\gamma}_{\mathbf{x}})\mathbf{x}$ foi definida de acordo com as condições apresentadas em Krätschmer (2007). Buscamos formas de verificar que $f(\mathbf{x})$ é um difeomorfismo, no entanto devido à generalidade do modelo, essa verificação deve ser feita em cada modelo específico. Uma possibilidade é apresentada abaixo como corolário, sendo que as particularidades de cada modelo específico são consideradas nas expressões utilizadas. Na Seção 2.3, apresentamos a utilização desse corolário em um modelo específico.

Corolário 2.2.2. *Seja $\mathbf{V}'_{\mathbf{x}} = \left(\frac{\partial \Sigma(\hat{\gamma}_{\mathbf{x}})}{\partial x_1}, \dots, \frac{\partial \Sigma(\hat{\gamma}_{\mathbf{x}})}{\partial x_N} \right)$ e $\mathbf{S}_{\mathbf{x}} = \text{bdiag} \{ \Sigma^{-1}(\hat{\gamma}_{\mathbf{x}})\mathbf{x}, \dots, \Sigma^{-1}(\hat{\gamma}_{\mathbf{x}})\mathbf{x} \}$. Se $\det(\mathbf{I}_N + \mathbf{V}'_{\mathbf{x}}\mathbf{S}_{\mathbf{x}}) \neq 0$, então $f(\mathbf{x})$ é um difeomorfismo.*

O Corolário 2.2.2 está relacionado a um dos resultados de análise matemática, o qual relaciona a inversa da derivada da função e o conceito de difeomorfismo.

As provas do Teorema 2.2.3 e do Corolário 2.2.2 são apresentadas com detalhes no Apêndice A, Seção A.3. Estas provas são desenvolvidas a fim de que o leitor consiga compreender a sequência e a necessidade de cada elemento que compõe a demonstração.

Vale ressaltar que as técnicas aqui estudadas associadas ao Teorema 2.2.3 não se aplicam ao modelo geral, apenas ao caso em que $\mu(\theta)$ e $\Sigma(\theta)$ não compartilham parâmetros. Em trabalhos futuros, buscaremos condições alternativas a fim de simplificar as condições do Teorema 2.2.2, em especial a condição de que o conjunto R tem medida nula, também para o modelo geral.

2.2.3 Consistência do EMV

A seguir, apresentamos o teorema sobre a consistência do EMV $\hat{\theta}_n$. Esse teorema está relacionado ao teorema da existência, sendo necessária apenas uma condição adicional.

Teorema 2.2.4. *Seja o modelo definido em (2.1). Sob as condições C1-C5, então existe uma sequência $\{\hat{\theta}_n\}_{n \geq 1}$ de EMV tal que*

$$\hat{\theta}_n \xrightarrow{P_{\theta}} \theta, \text{ para todo } \theta \in \Theta.$$

Lembremos que a definição de consistência fraca de um estimador está associada à convergência em probabilidade. Assim, sob as condições apresentadas existe uma sequência de estimadores que é consistente para estimar θ . Em comparação ao Teorema 2.2.1 adicionamos a condição C4. Novamente as condições C2 e C3 são necessárias para obtenção do vetor escore e matriz informação de Fisher, garantindo que essa é positiva definida. As condições C1 e C5 também foram utilizadas na prova da existência assintótica do EMV. Para a demonstração desse teorema, utilizamos um resultado obtido ao longo da prova do Teorema 2.2.1. Se satisfeita a condição C4, temos a consistência do EMV. As condições C4 e C5 são condições suficientes para provar a consistência do EMV.

O resultado da consistência do EMV é utilizado na prova do Teorema 2.2.5 que aborda a normalidade assintótica, o qual será apresentado a seguir. A prova do Teorema 2.2.4 é apresentada com detalhes no Apêndice A, Seção A.4.

2.2.4 Normalidade assintótica do EMV

A distribuição dos estimadores é uma característica importante para o desenvolvimento inferencial na modelagem estatística. Ela é fundamental para dimensionar o erro das estimativas, construir

intervalos de confiança e testar hipóteses de interesse. Nesse sentido, a normalidade assintótica é uma das propriedades dos EMV mais utilizadas. Em geral os trabalhos destacam que sob condições de regularidade o resultado da normalidade assintótica é válido, no entanto não explicitam quais são essas condições (Bayer e Cribari-Neto, 2013; Cribari-Neto e Queiroz, 2014; Melo *et al.*, 2018; Russo *et al.*, 2009). A seguir, apresentamos o teorema que garante a normalidade assintótica dos EMV de θ no modelo de regressão normal com parametrização geral.

Teorema 2.2.5. *Seja o modelo definido em (2.1). Sob as condições C1-C5 e C8,*

$$K_{\theta}^{1/2} \left(\hat{\theta}_n - \theta \right) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, I_p),$$

em que $\hat{\theta}_n$ é o EMV de θ .

Destacamos, a seguir, alguns passos que foram seguidos na demonstração do teorema.

1. O Teorema do Valor Médio é usado para reescrever o vetor escore, considerando a abordagem apresentada em Patriota (2019) e a condição C1.
2. Por meio do Teorema Central do Limite de Liapounov e das condições C3 e C4, mostra-se a convergência em distribuição $\frac{U_{\theta}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \Lambda_{\theta})$, para cada $\theta \in \Theta$.
3. Mediante aplicação da Lei Fraca dos Grandes Números de Markov e das condições C3 e C4, mostra-se a convergência em probabilidade $\frac{L_{\theta}}{n} + \frac{K_{\theta}}{n} \xrightarrow{P_{\theta}} \mathbf{0}$, para cada $\theta \in \Theta$.
4. A condição C8 e o Teorema 2.2.4 são utilizados para garantir que $(\hat{\theta}_n - \theta) = O_p(n^{-1/2})$, para cada $\theta \in \Theta$.
5. Pelas condições C3 e C4 e pelo Teorema de Cramér-Wold chegamos ao resultado.

Detalhes, condições e provas dos teoremas auxiliares utilizados para demonstrar o teorema principal podem ser verificados em Sen e Singer (1993) e Van der Vaart (2000). A prova completa do Teorema 2.2.5 pode ser obtida no Apêndice A, Seção A.5.

2.2.5 Distribuição assintótica das estatísticas de teste

Após a estimação dos parâmetros do modelo, em geral, são realizados testes a fim de verificar se hipóteses feitas sobre os parâmetros são suportadas por evidências obtidas a partir de dados amostrais.

Seja o vetor de parâmetros $\theta = (\psi^{\top}, \tau^{\top})^{\top}$, em que $\psi = (\psi_1, \dots, \psi_q)^{\top}$ representa o vetor de parâmetros de interesse e $\tau = (\tau_1, \dots, \tau_{p-q})^{\top}$ o vetor de parâmetros de perturbação. Suponha que o interesse esteja em testar

$$\mathcal{H}_0 : \psi = \psi_0 \quad \text{versus} \quad \mathcal{H}_1 : \psi \neq \psi_0,$$

em que ψ_0 é um vetor de constantes especificado de dimensão q .

As estatísticas de teste usuais para realização de testes de hipóteses são: (i) estatística da razão de verossimilhanças (RV_n)(Neyman e Pearson, 1928), (ii) estatística Wald (W_n)(Wald, 1943), (iii) estatística escore de Rao (S_n)(Rao, 1948) e (iv) estatística gradiente (G_n)(Terrell, 2002).

A fim de definir as estatísticas de teste, consideremos o estimador de máxima verossimilhança sob $\mathcal{H}_0$ (restrito) é denotado por $\tilde{\theta}_n = (\tilde{\psi}_n^{\top}, \tilde{\tau}_n^{\top})^{\top}$ e o estimador de máxima verossimilhança irrestrito $\hat{\theta}_n = (\hat{\psi}_n^{\top}, \hat{\tau}_n^{\top})^{\top}$. Conforme Cordeiro (1992), podemos particionar o vetor escore U_{θ} , a matriz informação de Fisher K_{θ} e sua inversa K_{θ}^{-1} da mesma maneira que particionamos θ , sendo $U_{\theta}^{\top} = (U_{\psi}(\theta)^{\top}, U_{\tau}(\theta)^{\top})$,

$$K_{\theta} = \begin{pmatrix} K_{\psi\psi}(\theta) & K_{\psi\tau}(\theta) \\ K_{\tau\psi}(\theta) & K_{\tau\tau}(\theta) \end{pmatrix} \quad \text{e} \quad K_{\theta}^{-1} = \begin{pmatrix} K^{\psi\psi}(\theta) & K^{\psi\tau}(\theta) \\ K^{\tau\psi}(\theta) & K^{\tau\tau}(\theta) \end{pmatrix}.$$

Considerando essas partições, as estatísticas da razão de verossimilhanças, Wald, escore e gradiente são escritas conforme segue:

$$RV_n = 2[\ell(\mathbf{y}, \hat{\boldsymbol{\theta}}_n) - \ell(\mathbf{y}, \tilde{\boldsymbol{\theta}}_n)], \quad (2.8)$$

$$W_n = (\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0)^\top [\mathbf{K}^{\boldsymbol{\psi}\boldsymbol{\psi}}(\hat{\boldsymbol{\theta}}_n)]^{-1} (\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0), \quad (2.9)$$

$$S_n = [\mathbf{U}_{\boldsymbol{\psi}}(\tilde{\boldsymbol{\theta}}_n)]^\top \mathbf{K}^{\boldsymbol{\psi}\boldsymbol{\psi}}(\tilde{\boldsymbol{\theta}}_n) \mathbf{U}_{\boldsymbol{\psi}}(\tilde{\boldsymbol{\theta}}_n) \quad (2.10)$$

$$\text{e } G_n = [\mathbf{U}_{\boldsymbol{\psi}}(\tilde{\boldsymbol{\theta}}_n)]^\top (\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0), \quad (2.11)$$

sendo $\mathbf{K}^{\boldsymbol{\psi}\boldsymbol{\psi}} = \left(\mathbf{K}_{\boldsymbol{\psi}\boldsymbol{\psi}} - \mathbf{K}_{\boldsymbol{\psi}\boldsymbol{\tau}} \mathbf{K}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{-1} \mathbf{K}_{\boldsymbol{\tau}\boldsymbol{\psi}}^\top \right)^{-1}$. Note que quando $\boldsymbol{\psi}$ e $\boldsymbol{\tau}$ são ortogonais, $\mathbf{K}^{\boldsymbol{\psi}\boldsymbol{\psi}} = \mathbf{K}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}$.

As estatísticas acima representam as técnicas mais importantes para avaliação e teste de modelos estatísticos. Um importante trabalho no contexto de testes de hipóteses é [Buse \(1982\)](#), o qual apresenta interpretações geométricas dos testes baseados nas estatísticas RV_n , W_n e S_n . No mesmo sentido, [Muggeo e Lovison \(2014\)](#) apresenta a interpretação geométrica para o teste baseado na estatística G_n . A estatística gradiente pode ser vista como uma derivação das estatísticas escore e Wald. A estatística escore mede o comprimento do quadrado do vetor escore sob $\mathcal{H}_0$ usando a inversa da matriz de informação de Fisher, enquanto a estatística Wald considera o quadrado da distância entre o estimador de máxima verossimilhança irrestrito e o restrito de $\boldsymbol{\theta}$, por meio da matriz informação de Fisher ([Vargas et al., 2013](#)). A estatística gradiente, apesar de recente, se comparada às demais estatísticas, tem recebido destaque na literatura estatística nos últimos anos, entre eles [Lemonte \(2013\)](#); [Lemonte e Ferrari \(2012\)](#); [Vargas et al. \(2013\)](#).

A seguir, enunciamos um teorema importante utilizado na realização de testes de hipóteses. O teorema traz as condições sob as quais conhecemos a distribuição assintótica das estatísticas de teste apresentadas.

Teorema 2.2.6. *Seja o modelo definido em (2.1). Sob as condições C1-C5, C8 e sob a hipótese nula*

$$RV_n, W_n, S_n, G_n \xrightarrow{\mathcal{D}} \chi_q^2$$

quando $n \rightarrow \infty$, em que χ_q^2 é uma variável aleatória qui-quadrado com q graus de liberdade.

Este teorema nos traz um resultado geralmente utilizado na estatística clássica para a realização de testes de hipóteses. Sob as condições de regularidade C1-C5, C8 e sob a hipótese nula, as quatro estatísticas apresentadas são assintoticamente equivalentes, possuindo aproximadamente distribuição χ_q^2 em grandes amostras, em que q é o número de parâmetros testados na hipótese nula. Assim, a hipótese $\mathcal{H}_0$ será rejeitada para valores grandes de RV_n , W_n , S_n e G_n comparados com o valor crítico $\chi_q^2(\alpha)$ obtido do quantil $1 - \alpha$ da distribuição χ_q^2 para um nível de significância α fixado.

O desenvolvimento da prova, a partir do resultado da normalidade assintótica do EMV (Teorema 2.2.5) e os resultados já explicitados nesse trabalho, é semelhante às provas apresentadas na literatura. Em [Sen e Singer \(1993\)](#) é provada a distribuição assintótica das estatísticas RV_n , W_n e S_n e para a estatística gradiente G_n a prova é feita em [Terrell \(2002\)](#) no caso em que o vetor $\boldsymbol{\theta}$ é testado conjuntamente. Já [Fahrmeir \(1988\)](#) considera testar uma partição do vetor $\boldsymbol{\theta}$ em modelos com observações não homogêneas. Os resultados existentes na literatura são utilizados como suporte no desenvolvimento da prova apresentada no Apêndice A, Seção A.6.

A seguir apresentamos alguns exemplos analíticos em que detalhamos a verificação de algumas condições de regularidade e um estudo de simulação em que violamos uma das condições enunciadas.

2.3 Exemplos

Nesta seção, apresentamos três exemplos a fim de ilustrar as condições de regularidade apresentadas e os resultados dos teoremas obtidos. Os dois primeiros exemplos tratam da questão da unicidade do EMV. Neles exemplificamos o uso dos dois teoremas apresentados. Ambos abordam casos particulares do modelo de regressão normal com parametrização geral em que $\boldsymbol{\mu}(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ não compartilham parâmetros. Nestes exemplos também especificamos algumas quantidades e condições de regularidade do modelo geral em cada caso específico. O terceiro exemplo envolve um modelo em que $\boldsymbol{\mu}(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ compartilham parâmetros. A fim de ilustrar o que ocorre quando uma das suposições é violada, optamos por realizar um estudo de simulação. Nosso objetivo é perturbar a condição C4. A violação dessa condição pode comprometer a consistência e normalidade assintótica do EMV.

2.3.1 Modelo linear normal

A fim de ilustrar o uso dos Teoremas 2.2.2 e 2.2.3, consideramos o modelo linear normal definido em forma matricial por

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + e_i, \quad i = 1, \dots, n,$$

em que Y_i é a variável resposta correspondente à i -ésima observação, $\mathbf{X}_i$ é um vetor $1 \times p$ das variáveis explicativas ou covariáveis associado à i -ésima observação, $\boldsymbol{\beta}$ é um vetor $p \times 1$ de parâmetros desconhecidos. Assumimos que e_i e e_j são independentes, para $i \neq j$, com $e_i \sim \mathcal{N}(0, \sigma^2)$. Então,

$$Y_i \stackrel{\text{ind}}{\sim} \mathcal{N}(\mathbf{X}_i \boldsymbol{\beta}, \sigma^2)$$

é um caso especial do modelo (2.1), com $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$, $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \sigma^2$, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)$ e $z_i = y_i - \boldsymbol{\mu}_i$. Seja $\mathbf{X}$ a matriz $n \times p$ das variáveis explicativas. A seguir exemplificamos algumas quantidades para o modelo linear normal:

$$\mathbf{G}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \text{vech}(\sigma^2 \mathbf{I}_n) \end{pmatrix}, \quad \mathbf{F}_i = \begin{pmatrix} \mathbf{X}_i & 0 \\ \mathbf{0} & 1 \end{pmatrix}, \quad \mathbf{H}_i = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}^{-1} \quad \text{e} \quad \mathbf{s}_i = \begin{pmatrix} z_i \\ z_i^2 - \sigma^2 \end{pmatrix}.$$

Dado que $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \mathbf{X}_i \boldsymbol{\beta}$ e $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \sigma^2$ são funções contínuas em $\boldsymbol{\theta}$, duas vezes diferenciáveis, então C2 é válida. Se $\mathbf{X}$ tiver posto completo, então C3 também é válida. Consequentemente, a condição C6 é satisfeita. A função $\mathbf{G}(\boldsymbol{\theta})$ é injetiva, logo C7 é satisfeita. Ainda, como $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ não compartilham parâmetros podemos utilizar os dois teoremas apresentados que abordam a unicidade do EMV, o Teorema 2.2.2 e o Teorema 2.2.3.

Exemplo da aplicação do Teorema 2.2.2

Notemos que $\mathbf{B}_i$ é uma matriz $p \times p$ de zeros. Então,

$$\mathbf{L}_\theta = - \sum_{i=1}^n \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i,$$

em que

$$\ddot{\mathbf{M}}_i = \begin{pmatrix} \sigma^2 & 2\sigma^2 z_i \\ 2\sigma^2 z_i & 4\sigma^2 z_i^2 + 2\sigma^4 \end{pmatrix}.$$

O EMV do vetor $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)^\top$ se $\mathbf{X}$ tem posto completo é dado por $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ e $\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n} = \frac{\mathbf{Y}^\top \mathbf{A} \mathbf{Y}}{n}$, em que $\mathbf{A} = \mathbf{I}_n - \mathbf{H}$, com $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ é a matriz de projeção, idempotente, simétrica e com posto completo, ou seja, $\text{posto}(\mathbf{H}) = p$. Assim, $\mathbf{A}$ é uma

matriz simétrica idempotente com $\text{posto}(\mathbf{A}) = n - p$. Após alguma álgebra, temos

$$\mathbf{L}_{\hat{\boldsymbol{\theta}}_n} = \begin{pmatrix} -\frac{\mathbf{X}^\top \mathbf{X}}{\hat{\sigma}^2} & -\frac{\mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{\hat{\sigma}^2} \\ -\frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \mathbf{X}}{\hat{\sigma}^2} & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix}.$$

Notemos que $\frac{\mathbf{Y}^\top \mathbf{A} \mathbf{Y}}{\hat{\sigma}^2} \sim \chi_q^2$, em que χ_q^2 denota a distribuição qui-quadrado com q graus de liberdade, $q = \text{posto}(\mathbf{A})$. Assim, $P\left(\frac{\mathbf{Y}^\top \mathbf{A} \mathbf{Y}}{\hat{\sigma}^2} > 0\right) = 1$, o que implica que o conjunto R dado em (2.7) tem medida nula. Portanto, se a matriz $\mathbf{X}$ no modelo linear normal tiver posto completo, temos que o EMV de $\boldsymbol{\theta}$ é único com probabilidade um.

Exemplo da aplicação do Teorema 2.2.3

Basta mostrar que $f(\mathbf{x})$ é um difeomorfismo. Neste modelo, $f_j(\mathbf{x})$ é escrita na forma

$$f_j(\mathbf{x}) = \frac{n}{\sum_{s=1}^n x_s^2} x_j.$$

Temos $f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_n(\mathbf{x}))$. Podemos mostrar que $f(\mathbf{x}) = f^{-1}(\mathbf{x})$:

$$f_j(f_j(\mathbf{x})) = \frac{n}{\sum_{s=1}^n f_j(\mathbf{x})_s^2} f_j(\mathbf{x}) = \frac{n^2}{\sum_{s=1}^n \left(\frac{n x_s}{\sum_{k=1}^n x_k^2} \right)^2} \frac{x_j}{\sum_{s=1}^n x_s^2} = x_j.$$

Então, temos que $f(\mathbf{x})$ é uma função bijetora infinitamente diferenciável com $f(\mathbf{x}) = f^{-1}(\mathbf{x})$. Logo, $f(\mathbf{x})$ é um difeomorfismo. Sendo assim, se a matriz $\mathbf{X}$ tiver posto completo, então o EMV de $\boldsymbol{\theta}$ é único com probabilidade um.

Vale salientar que nesse modelo a existência e a unicidade do EMV quando $\mathbf{X}$ tem posto completo são óbvias. No entanto, buscamos exemplificar os teoremas em um modelo em que o resultado já é conhecido, a fim de motivar o entendimento dos teoremas.

2.3.2 Modelo não linear heterocedástico

Consideremos agora o modelo não linear heterocedástico definido por

$$Y_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i) + e_i,$$

em que Y_i é a i -ésima resposta observada, $\mathbf{x}_i$ um vetor $1 \times p_1$ das variáveis explicativas, $\boldsymbol{\beta}$ é um vetor $p_1 \times 1$ de parâmetros desconhecidos e $\boldsymbol{\mu}_i$ é uma função não linear em $\boldsymbol{\beta}$. Assumimos que e_i e e_j , para $i \neq j$ são independentes, com $e_i \sim \mathcal{N}(0, \sigma_i^2)$. Ainda, assumimos que $\sigma_i^2 = h_i(\boldsymbol{\gamma}, \mathbf{w}_i)$, em que $\boldsymbol{\gamma}$ é um vetor $p_2 \times 1$ de parâmetros desconhecidos e $\mathbf{w}_i$ é um vetor $1 \times p_2$ das variáveis explicativas. Então,

$$Y_i \stackrel{\text{ind}}{\sim} \mathcal{N}(\boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i), \sigma_i^2),$$

é um caso especial do modelo (2.1), com $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i)$, $\boldsymbol{\Sigma}_i = \sigma_i^2$, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. Quando $\boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$ temos o modelo linear heteroscedástico. Sejam $z_i = y_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i)$, $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i)}{\partial \boldsymbol{\beta}^\top}$, $\mathbf{V}_i = \frac{\partial h_i(\boldsymbol{\gamma}, \mathbf{w}_i)}{\partial \boldsymbol{\gamma}^\top}$, $\mathbf{X}$ a matriz de variáveis explicativas associadas à função $\boldsymbol{\mu}(\boldsymbol{\beta})$, $\mathbf{W}$ a matriz de variáveis explicativas associadas à função $h(\boldsymbol{\gamma}, \mathbf{w})$ e $\boldsymbol{\Sigma}(\boldsymbol{\gamma}) = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$. Temos

$$\mathbf{G}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\beta}) \\ \text{vech}(\boldsymbol{\Sigma}(\boldsymbol{\gamma})) \end{pmatrix}, \quad \mathbf{F}_i = \begin{pmatrix} \mathbf{D}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_i \end{pmatrix}, \quad \mathbf{H}_i = \begin{pmatrix} \sigma_i^2 & 0 \\ 0 & 2\sigma_i^4 \end{pmatrix}^{-1} \quad \text{e} \quad \mathbf{s}_i = \begin{pmatrix} z_i \\ z_i^2 - \sigma_i^2 \end{pmatrix}.$$

Então,

$$\mathbf{L}_\theta = \sum_{i=1}^n (\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i),$$

em que

$$\ddot{\mathbf{M}}_i = \begin{pmatrix} \sigma_i^2 & 2\sigma_i^2 z_i \\ 2\sigma_i^2 z_i & 4\sigma_i^2 z_i^2 + 2\sigma_i^4 \end{pmatrix}.$$

Notemos que $\mathbf{B}_i$ depende da estrutura das funções $\mu_i(\boldsymbol{\beta}, \mathbf{x}_i)$ e $h_i(\boldsymbol{\gamma}, \mathbf{w}_i)$, pois envolve a segunda derivada dessas funções. A outra expressão de $\mathbf{L}_\theta$ é dada por

$$\mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i = \begin{pmatrix} \frac{\mathbf{D}_i \mathbf{D}_i^\top}{\sigma_i^2} & \frac{\mathbf{D}_i^\top \mathbf{V}_i z_i}{\sigma_i^4} \\ \frac{\mathbf{D}_i \mathbf{V}_i^\top z_i}{\sigma_i^4} & \mathbf{V}_i \mathbf{V}_i^\top \left(\frac{z_i^2}{\sigma_i^6} + \frac{1}{2\sigma_i^4} \right) \end{pmatrix}.$$

A fim de melhor exemplificar o uso do Teorema 2.2.3 e do Corolário 2.2.2 tomemos

$$\mu_i(\boldsymbol{\beta}, \mathbf{x}_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i})}{1 + \exp(\beta_0 + \beta_1 x_{1i})} \quad \text{e} \quad h_i(\boldsymbol{\gamma}, \mathbf{w}_i) = \exp(\gamma_0 + \gamma_1 w_{2i}). \quad (2.12)$$

O vetor de parâmetros é $\boldsymbol{\theta} = (\beta_0, \beta_1, \gamma_0, \gamma_1)^\top$. Neste caso específico temos

$$\mathbf{D}_i = \begin{pmatrix} \frac{\exp(\beta_0 + \beta_1 x_{1i})}{[1 + \exp(\beta_0 + \beta_1 x_{1i})]^2}, & \frac{\exp(\beta_0 + \beta_1 x_{1i}) x_{1i}}{[1 + \exp(\beta_0 + \beta_1 x_{1i})]^2} \end{pmatrix} \quad \text{e} \\ \mathbf{V}_i = \begin{pmatrix} \exp(\gamma_0 + \gamma_1 w_{2i}), & \exp(\gamma_0 + \gamma_1 w_{2i}) w_{2i} \end{pmatrix}.$$

As funções $\mu_i(\boldsymbol{\beta}, \mathbf{x}_i)$ e $h_i(\boldsymbol{\gamma}, \mathbf{w}_i)$ especificadas em (2.12) são funções contínuas em $\boldsymbol{\theta}$, duas vezes diferenciáveis e limitadas, então a condição C2 é satisfeita. Quando $\mathbf{X}$ e $\mathbf{W}$ tem posto completo, a condição C3 também é satisfeita. Ainda, temos que $\mathbf{G}$ é injetiva, portanto C7 é satisfeita. Para o resultado do Teorema 2.2.3 ser válido, ainda temos que mostrar que a função $f(\mathbf{x}) = \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}})\mathbf{x}$ é um difeomorfismo. Nesse caso, temos

$$f(\mathbf{x}) = \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}})\mathbf{x} = \left(\frac{x_1}{\exp(\hat{\gamma}_0 + \hat{\gamma}_1 w_{21})}, \quad \frac{x_2}{\exp(\hat{\gamma}_0 + \hat{\gamma}_1 w_{22})}, \quad \dots, \quad \frac{x_n}{\exp(\hat{\gamma}_0 + \hat{\gamma}_1 w_{2n})} \right)^\top.$$

Para mostrar que $f(\mathbf{x})$ é um difeomorfismo, analisamos a condição apresentada no Corolário 2.2.2. A expressão do i -ésimo vetor coluna da matriz $\mathbf{V}'_{\mathbf{x}} \mathbf{S}_{\mathbf{x}}$ é dado por

$$\left(x_1 \left[\frac{d\hat{\gamma}_0}{dx_i} + \frac{d\hat{\gamma}_1 w_{21}}{dx_i} \right], \quad x_2 \left[\frac{d\hat{\gamma}_0}{dx_i} + \frac{d\hat{\gamma}_1 w_{22}}{dx_i} \right], \quad \dots, \quad x_N \left[\frac{d\hat{\gamma}_0}{dx_i} + \frac{d\hat{\gamma}_1 w_{2n}}{dx_i} \right] \right)^\top.$$

Se $\mathbf{X}$ e $\mathbf{W}$ tem posto completo, a matriz $\mathbf{V}'_{\mathbf{x}} \mathbf{S}_{\mathbf{x}}$ também terá. Logo

$$\det(\mathbf{I}_n + \mathbf{V}'_{\mathbf{x}} \mathbf{S}_{\mathbf{x}}) \neq 0.$$

Então, $f(\mathbf{x})$ é um difeomorfismo. Considerando as funções (2.12) e, se $\mathbf{Z}$ tem posto completo, todas as condições do Teorema 2.2.3 são satisfeitas e o EMV de $\boldsymbol{\theta}$ é único com probabilidade um. As funções explicitadas em (2.12) são utilizadas a fim de exemplificação, sendo que qualquer outra função de interesse poderia ser utilizada.

Vale destacar que um dos exemplos apresentados em Krättschmer (2007) no qual é estudada a unicidade do EMV de modelos de regressão não lineares, pode ser visto como um caso particular do modelo não linear heterocedástico. Fazendo um paralelo, em Krättschmer (2007) temos a função da média dada por $F(\boldsymbol{\alpha})$ e da variância $\tilde{F}(\boldsymbol{\beta})$, assim teríamos $\beta_1 = \exp(\gamma_0)$, $\beta_2 = \exp(\gamma_0 + \gamma_1)$, ..., $\beta_q = \exp(\gamma_0 + \gamma_q)$ e $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 w_{1i} + \dots + \gamma_q w_{qi})$ em que w_{ji} é binária, com $i = 1, \dots, n$ e

$j = 1, \dots, q$. Sendo garantida a unicidade do EMV nesse caso com probabilidade um.

2.3.3 Modelo com erros nas variáveis e na equação

A fim de exemplificar a violação de uma das condições, consideramos o modelo discutido em [Patriota *et al.* \(2009\)](#), o qual é um caso particular de (2.1) e será utilizado no estudo de simulação. Admitamos o modelo com erros nas variáveis, com duas variáveis observadas, sendo elas Y_i e X_i relacionadas por meio das equações

$$Y_i = y_i + \eta_{yi} \quad \text{e} \quad X_i = x_i + \eta_{xi}, \quad (2.13)$$

em que $y_i|x_i \stackrel{\text{iid}}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = 1, \dots, n$, β_0 e β_1 são parâmetros desconhecidos e σ^2 é a variância do erro da equação. Os erros η_{yi} e η_{xi} são independentes de x_i e y_i e são distribuídos tais que

$$\begin{pmatrix} \eta_{yi} \\ \eta_{xi} \end{pmatrix} \stackrel{\text{iid}}{\sim} \mathcal{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{yi} & 0 \\ 0 & \tau_{xi} \end{pmatrix} \right), \quad i = 1, \dots, n,$$

em que as variâncias τ_{yi} e τ_{xi} são conhecidas para $i = 1, \dots, n$. Supondo que x_i e e_i são independentes e tais que $x_i \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu_x, \sigma_x^2)$ e $e_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, em que “ $\stackrel{\text{iid}}{\sim}$ ” denota independentes e identicamente distribuídas. Denotemos $\mathbf{Y}_i = (Y_i, X_i)^\top$ e $\boldsymbol{\theta} = (\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma^2)^\top$. Então, temos que

$$\mathbf{Y}_i \stackrel{\text{iid}}{\sim} \mathcal{N}_2(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}_i(\boldsymbol{\theta}))$$

é um caso particular do modelo (2.1) com

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \begin{pmatrix} \beta_1^2 \sigma_x^2 + \tau_{yi} + \sigma^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \tau_{xi} \end{pmatrix}. \quad (2.14)$$

Explicitamos ainda algumas quantidades do modelo geral para este caso particular:

$$\mathbf{G}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{1}_n \otimes \boldsymbol{\mu}(\boldsymbol{\theta}) \\ \text{vech}(\boldsymbol{\Sigma}(\boldsymbol{\theta})) \end{pmatrix}, \quad \mathbf{F}_i = \begin{pmatrix} \mathbf{D}_i \\ \mathbf{V}_i \end{pmatrix}, \quad \mathbf{H}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & 2\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) \end{pmatrix}^{-1} \quad \text{e}$$

$$\mathbf{s}_i = \begin{pmatrix} \mathbf{z}_i \\ \text{vec}(\mathbf{z}_i \mathbf{z}_i^\top - \boldsymbol{\Sigma}_i(\boldsymbol{\theta})) \end{pmatrix}, \quad \text{em que } \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \text{bdiag}\{\boldsymbol{\Sigma}_1(\boldsymbol{\theta}), \dots, \boldsymbol{\Sigma}_n(\boldsymbol{\theta})\},$$

$$\mathbf{D}_i = \begin{pmatrix} 1 & \mu_x & \beta_1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{V}_i = \begin{pmatrix} 0 & 2\beta_1 \sigma_x^2 & 0 & \beta_1^2 & 1 \\ 0 & \sigma_x^2 & 0 & \beta_1 & 0 \\ 0 & \sigma_x^2 & 0 & \beta_1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{e} \quad \mathbf{z}_i = \begin{pmatrix} y_i - \beta_0 - \beta_1 \mu_x \\ x_i - \mu_x \end{pmatrix}.$$

As funções dadas em (2.14) são duas vezes diferenciáveis, contínuas em $\boldsymbol{\theta}$ e limitadas. Assim, a condição C2 é satisfeita. Em relação à condição C3, as cinco colunas de $\mathbf{F}_i$ são linearmente independentes, como $p = 5$ temos que a condição C3 também é satisfeita.

A contribuição do i -ésimo indivíduo para a infominação de Fisher é dada por

$$\mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i = \begin{pmatrix} \mathbf{K}_{\beta_0 \beta_0 i} & \mathbf{K}_{\beta_0 \beta_1 i} & \mathbf{K}_{\beta_0 \mu_x i} & \mathbf{K}_{\beta_0 \sigma_x^2 i} & \mathbf{K}_{\beta_0 \sigma^2 i} \\ \mathbf{K}_{\beta_1 \beta_0 i} & \mathbf{K}_{\beta_1 \beta_1 i} & \mathbf{K}_{\beta_1 \mu_x i} & \mathbf{K}_{\beta_1 \sigma_x^2 i} & \mathbf{K}_{\beta_1 \sigma^2 i} \\ \mathbf{K}_{\mu_x \beta_0 i} & \mathbf{K}_{\mu_x \beta_1 i} & \mathbf{K}_{\mu_x \mu_x i} & \mathbf{K}_{\mu_x \sigma_x^2 i} & \mathbf{K}_{\mu_x \sigma^2 i} \\ \mathbf{K}_{\sigma_x^2 \beta_0 i} & \mathbf{K}_{\sigma_x^2 \beta_1 i} & \mathbf{K}_{\sigma_x^2 \mu_x i} & \mathbf{K}_{\sigma_x^2 \sigma_x^2 i} & \mathbf{K}_{\sigma_x^2 \sigma^2 i} \\ \mathbf{K}_{\sigma^2 \beta_0 i} & \mathbf{K}_{\sigma^2 \beta_1 i} & \mathbf{K}_{\sigma^2 \mu_x i} & \mathbf{K}_{\sigma^2 \sigma_x^2 i} & \mathbf{K}_{\sigma^2 \sigma^2 i} \end{pmatrix}.$$

Definamos $a_i = \beta_1^2 \sigma_x^2 + \tau_{yi} + \sigma^2$, $b = \beta_1 \sigma_x^2$ e $c_i = \sigma_x^2 + \tau_{xi}$. Os componentes da matriz $\mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i$ são dados por

$$\begin{aligned} K_{\beta_0 \beta_0 i} &= \frac{-c_i}{b^2 - a_i c_i}, & K_{\beta_1 \beta_1 i} &= \frac{a_i c_i^2 \mu_x^2 - b^2 c_i \mu_x^2 + b^2 \sigma_x^4 + a_i c_i \sigma_x^4 + 2c_i^2 \beta_1^2 \sigma_x^4 - 4bc_i \beta_1 \sigma_x^4}{b^4 + a_i^2 c_i^2 - 2a_i b^2 c_i}, \\ K_{\mu_x \mu_x i} &= \frac{-a_i - c_i \beta_1^2 + 2b \beta_1}{b^2 - a_i c_i}, & K_{\sigma_x^2 \sigma_x^2 i} &= \frac{a_i^2 + c_i^2 \beta_1^4 - 4bc_i \beta_1^3 + 4b^2 \beta_1^2 + 2a_i c_i \beta_1^2 - 4a_i b \beta_1}{2b^4 + 2a_i^2 c_i^2 - 4a_i b^2 c_i}, \\ K_{\sigma^2 \sigma^2 i} &= \frac{c_i^2}{2b^4 + 2a_i^2 c_i^2 - 4a_i b^2 c_i}, & K_{\beta_0 \beta_1 i} &= K_{\beta_1 \beta_0 i} = \frac{-c_i \mu_x}{b^2 - a_i c_i}, & K_{\beta_0 \mu_x i} &= K_{\mu_x \beta_0 i} = \frac{b - c_i \beta_1}{b^2 - a_i c_i}, \\ K_{\beta_1 \mu_x i} &= K_{\mu_x \beta_1 i} = \frac{b \mu_x - c_i \beta_1 \mu_x}{b^2 - a_i c_i}, & K_{\sigma_x^2 \beta_1 i} &= K_{\beta_1 \sigma_x^2 i} = \frac{-a_i b \sigma_x^2 + c_i^2 \beta_1^3 \sigma_x^2 - 3bc_i \beta_1^2 \sigma_x^2 + a_i c_i \beta_1 \sigma_x^2}{b^4 + a_i^2 c_i^2 - 2a_i b^2 c_i}, \\ K_{\sigma^2 \beta_1 i} &= K_{\beta_1 \sigma^2 i} = \frac{-bc_i \sigma_x^2 + c_i^2 \beta_1 \sigma_x^2}{b^4 + a_i^2 c_i^2 - 2a_i b^2 c_i}, & K_{\sigma^2 \sigma_x^2 i} &= K_{\sigma_x^2 \sigma^2 i} = \frac{b^2 c_i^2 \beta_1^2 - 2bc_i \beta_1}{2b^4 + 2a_i^2 c_i^2 - 4a_i b^2 c_i}, \\ \text{e } K_{\beta_0 \sigma_x^2 i} &= K_{\beta_0 \sigma^2 i} = K_{\sigma_x^2 \beta_0 i} = K_{\sigma^2 \beta_0 i} = K_{\sigma_x^2 \mu_x i} = K_{\mu_x \sigma_x^2 i} = K_{\sigma^2 \mu_x i} = K_{\mu_x \sigma^2 i} = 0. \end{aligned}$$

A condição C4 é satisfeita desde que os termos da matriz $\mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i$ sejam limitados. Se τ_{xi} e τ_{yi} são limitados temos que C4 é satisfeita. Na literatura é usual, em simulações, utilizar τ_{xi} e τ_{yi} provenientes da distribuição uniforme, assim C4 é satisfeita.

No estudo de simulação, consideramos inicialmente um cenário em que a condição C4 é satisfeita, esse mesmo cenário foi utilizado em [Patriota et al. \(2009\)](#). Para a avaliação numérica, foram consideradas 5000 réplicas de Monte Carlo, fixamos $\beta_0 = 2$, $\beta_1 = 1$, $\mu_x = -2$, $\sigma_x^2 = 4$ e $\sigma^2 = 10$, τ_{xi} e τ_{yi} foram gerados a partir de uma distribuição uniforme e são tais que $\sqrt{\tau_{xi}} \sim U(0.5; 1.5)$ e $\sqrt{\tau_{yi}} \sim U(0.5; 4)$. Os tamanhos amostrais utilizados foram $n = 50, 100, 500, 1000, 5000$. Para cada réplica gera-se uma amostra $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ da distribuição normal bivariada. Para a avaliação numérica, foram utilizadas algumas medidas baseadas nas 5000 réplicas das estimativas de máxima verossimilhança dos parâmetros do modelo. Foram calculadas a média, o viés e a raiz quadrada do erro quadrático médio (EQM). Consideramos o vetor padronizado $Z_n = \mathbf{K}_{\hat{\theta}}^{1/2} (\hat{\theta}_n - \theta)$, o qual, sob as condições de regularidade discutidas, converge em distribuição para a normal com vetor de médias zero e matriz de variâncias e covariâncias identidade. A partir das 5000 réplicas de Z_n , calculamos a média (Média_{Z_n}), o desvio-padrão (DP_{Z_n}), o coeficiente de assimetria (CA_{Z_n}), o coeficiente de curtose (C_{Z_n}) e os valores-p associados ao teste de normalidade de Kolmogorov-Smirnov (KS_{Z_n}) e de Shapiro-Wilk (SW_{Z_n}). Os resultados das medidas calculadas são comparados com a distribuição normal padrão e os valores-p dos testes de normalidade são comparados com o nível de significância adotado. Ainda, nas tabelas, incluímos a informação em relação ao número de não convergências no procedimento de otimização (Nc) e número de vezes em que a matriz de informação de Fisher não é definida positiva (NPD), para cada tamanho amostral.

Na Tabela 2.1, verifica-se que à medida que o tamanho amostral aumenta, os valores das medidas calculadas para Z_n tornam-se mais próximos aos valores observados na distribuição normal padrão. Esse resultado é esperado, dado que as condições estabelecidas para a normalidade assintótica são satisfeitas. Destacaremos alguns resultados a seguir:

- Verificamos que os estimadores são assintoticamente não viesados e consistentes, já que o viés diminui à medida que n aumenta, e a medida $\sqrt{\text{EQM}}$ está mais próxima de zero em amostras maiores.

- Consideramos as medidas $Média_{Z_n}$ e DP_{Z_n} , por exemplo, os resultados tendem a zero e um, respectivamente. Tal resultado é esperado quando as condições enunciadas são satisfeitas, pois temos a convergência em distribuição para a normal padrão (média zero e desvio-padrão um).
- Em relação às medidas de assimetria e curtose, para o tamanho amostral $n = 50$, associadas a $\hat{\sigma}_x^2$, temos $CA_{Z_n} = -0.607$ e $C_{Z_n} = 4.096$, quando se espera que esses valores sejam próximos a zero e três, respectivamente, na distribuição normal padrão. No entanto, para tamanhos amostrais maiores esses resultados são próximos aos esperados.
- Os valores de KS_{Z_n} e SW_{Z_n} associados ao teste de normalidade, quando superiores a 0.05 indicam que a hipótese de normalidade não é rejeitada. Notemos que quando $n = 5000$, com base no teste de Shapiro-Wilk ($SW_{Z_n} > 0.05$), a distribuição dos estimadores dos cinco parâmetros do modelo é normal.
- Temos o indicativo de que utilizar a aproximação à distribuição normal não é adequada para a realização de inferências associadas aos parâmetros σ_x^2 e σ^2 , quando $n < 500$. Para β_0 mesmo no menor tamanho amostral ($n = 50$) já temos indicativos da normalidade assintótica.
- Notamos ainda que a padronização associada aos estimadores $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\mu}_x$ converge para a distribuição normal padrão mais rapidamente do que a associada aos estimadores das variâncias ($\hat{\sigma}_x^2$ e $\hat{\sigma}_2$).
- Em relação às não convergências e número de vezes em que a matriz informação de Fisher não foi positiva definida, nesse cenário todos os tamanhos amostrais apresentaram $Nc=0$ e $NPD=0$, ou seja, não tivemos problemas de não convergência e de matriz não positiva definida.
- Os resultados indicam que a variável Z_n converge para a distribuição normal padrão, quando as condições estudadas são satisfeitas.

Apresentamos ainda algumas figuras a fim de ilustrar os resultados da Tabela 2.1. A Figura 2.1 apresenta as medidas que foram avaliadas nas simulações *versus* os tamanhos amostrais considerados, associadas às estimativas de $\hat{\sigma}^2$. Vale destacar que os piores resultados são os associados às figuras apresentadas. A análise gráfica corrobora com os resultados discutidos acerca da Tabela 2.1, confirmando que os EMV são consistentes e convergem para a distribuição normal à medida que o tamanho amostral aumenta.

Buscamos ainda ilustrar a distribuição assintótica dos EMV através da Figura 2.2, em que apresentamos o histograma das estimativas de $\hat{\beta}_1$ considerando as réplicas do estudo de simulação e a função densidade da distribuição normal padrão. Com isso, o que percebemos é que há indicativos da convergência da distribuição do EMV de β_1 para a distribuição normal. Os mesmos gráficos foram obtidos para os demais parâmetros, sendo observado comportamento semelhante nesse cenário.

A fim de avaliar o que ocorre quando uma das condições é violada, vamos considerar outro cenário. Analisaremos a violação da condição C4 enunciada na Seção 2.2. A fim de simplificar, analisaremos o caso em que $\mu_x = 0$ e $\beta_1 = 0$, a contribuição do i -ésimo indivíduo na informação de Fisher neste caso é dada por

$$\mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i = \begin{pmatrix} \frac{1}{\tau_{yi} + \sigma^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{\sigma_x^4}{(\tau_{yi} + \sigma^2)(\sigma_x^2 + \tau_{xi})} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_x^2 + \tau_{xi}} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2(\sigma_x^2 + \tau_{xi})^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2(\tau_{yi} + \sigma^2)^2} \end{pmatrix}.$$

Assumimos que σ_x^2 , σ^2 e τ_{yi} são finitos. Notemos que se τ_{xi} crescer na ordem de $i^{1+\delta}$ com

Tabela 2.1: Resultados da simulação de Monte Carlo considerando o modelo dado em (2.13), com $\beta_0 = 2$, $\beta_1 = 1$, $\mu_x = -2$, $\sigma_x^2 = 4$, $\sigma^2 = 10$, $\sqrt{\tau_{xi}} \sim U(0.5; 1.5)$ e $\sqrt{\tau_{yi}} \sim U(0.5; 4)$. As colunas apresentam a média, o viés e a raiz quadrada do erro quadrático médio dos estimadores ($\sqrt{EQM}$), a média ($Média_{Z_n}$), o desvio-padrão (DP_{Z_n}), o coeficiente de assimetria (CA_{Z_n}), o coeficiente de curtose (C_{Z_n}) e os p-valores associados ao teste de normalidade de Kolmogorov-Smirnov (KS_{Z_n}) e de Shapiro-Wilk (SW_{Z_n}) da variável padronizada Z_n . Espera-se que os valores das medidas associadas a Z_n se encontrem próximos aos da normal padrão.

Estimador	Média	Viés	$\sqrt{EQM}$	$Média_{Z_n}$	DP_{Z_n}	CA_{Z_n}	C_{Z_n}	KS_{Z_n}	SW_{Z_n}
$n = 50, Nc=0, NPD=0$									
$\hat{\beta}_0$	2.041	0.041	0.897	-0.032	1.037	-0.034	3.043	0.081	0.565
$\hat{\beta}_1$	1.017	0.017	0.342	-0.100	1.025	-0.090	3.278	0.000	0.000
$\hat{\mu}_x$	-2.004	-0.004	0.316	-0.011	1.034	-0.057	3.164	0.448	0.079
$\hat{\sigma}_x^2$	3.895	-0.105	0.980	-0.309	1.128	-1.170	7.490	0.000	0.000
$\hat{\sigma}^2$	9.281	-0.719	3.134	-0.497	1.249	-1.377	8.338	0.000	0.000
$n = 100, Nc=0, NPD=0$									
$\hat{\beta}_0$	2.022	0.022	0.618	-0.016	1.015	0.003	3.125	0.329	0.216
$\hat{\beta}_1$	1.010	0.010	0.237	-0.046	1.019	-0.058	3.085	0.033	0.106
$\hat{\mu}_x$	-2.000	0.000	0.224	0.001	1.019	-0.040	3.019	0.532	0.487
$\hat{\sigma}_x^2$	3.946	-0.054	0.696	-0.213	1.056	-0.670	4.096	0.000	0.000
$\hat{\sigma}^2$	9.688	-0.312	2.176	-0.300	1.075	-0.619	3.800	0.000	0.000
$n = 500, Nc=0, NPD=0$									
$\hat{\beta}_0$	2.002	0.002	0.270	-0.018	1.006	0.013	2.888	0.216	0.103
$\hat{\beta}_1$	1.002	0.002	0.100	-0.019	0.999	-0.046	2.967	0.616	0.266
$\hat{\mu}_x$	-1.999	0.001	0.100	0.009	1.006	-0.045	3.021	0.214	0.256
$\hat{\sigma}_x^2$	3.990	-0.010	0.315	-0.093	1.012	-0.290	3.228	0.001	0.000
$\hat{\sigma}^2$	9.912	-0.088	0.991	-0.159	1.032	-0.227	3.051	0.000	0.000
$n = 1000, Nc=0, NPD=0$									
$\hat{\beta}_0$	1.999	-0.001	0.190	-0.017	1.007	-0.032	2.933	0.200	0.652
$\hat{\beta}_1$	1.000	0.000	0.071	-0.019	1.008	-0.052	2.956	0.256	0.165
$\hat{\mu}_x$	-1.999	0.001	0.071	0.012	0.990	0.000	2.908	0.236	0.617
$\hat{\sigma}_x^2$	3.994	-0.006	0.224	-0.072	1.016	-0.200	3.099	0.012	0.000
$\hat{\sigma}^2$	9.977	-0.023	0.691	-0.081	1.010	-0.188	3.018	0.008	0.000
$n = 5000, Nc=0, NPD=0$									
$\hat{\beta}_0$	2.000	0.000	0.084	-0.006	1.006	-0.010	2.922	0.802	0.837
$\hat{\beta}_1$	1.000	0.000	0.032	-0.010	1.005	-0.003	3.086	0.828	0.394
$\hat{\mu}_x$	-1.999	0.001	0.032	0.016	0.998	0.013	3.051	0.492	0.829
$\hat{\sigma}_x^2$	3.997	-0.003	0.100	-0.051	1.001	-0.065	2.934	0.001	0.294
$\hat{\sigma}^2$	9.994	-0.006	0.310	-0.040	1.007	-0.058	2.959	0.090	0.154

$i = 1, \dots, n$ e $\delta > 0$ temos que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{\sigma_x^4}{(\tau_{yi} + \sigma^2)(\sigma_x^2 + \tau_{xi})} = 0, \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma_x^2 + \tau_{xi}} = 0$$

$$\text{e} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{1}{2(\sigma_x^2 + \tau_{xi})^2} = 0. \quad (2.15)$$

Dessa forma, a matriz $\mathbf{\Lambda}_{\theta}$ não será definida positiva, logo, a condição C4 será violada. Notemos que a condição C4 está associada à variância dos estimadores, a violação dessa condição refletirá em especial nas estimativas da variância, o que verificaremos por meio da medida DP_{Z_n} . Espera-se

Figura 2.1: Medidas avaliadas na simulação versus tamanho amostral quando todas as condições são satisfeitas associadas às estimativas de $\hat{\sigma}^2$ apresentadas na Tabela 2.1.

que os resultados mais críticos em amostras finitas estejam relacionados às estimativas da variância de $\hat{\sigma}_x^2$, pois a última sequência de (2.15) converge mais rapidamente.

Na Tabela 2.2, apresentamos os resultados do estudo de simulação considerando a configuração descrita. Fixamos $\beta_0 = 2$, $\beta_1 = 0$, $\mu_x = 0$, $\sigma_x^2 = 4$ e $\sigma^2 = 10$, τ_{yi} foi gerado a partir de uma distribuição uniforme tal que $\sqrt{\tau_{yi}} \sim U(0.5; 4)$ e τ_{xi} da relação $\tau_{xi} = 2(i^{1.1})$. As mesmas medidas da Tabela 2.1 foram calculadas. A seguir, destacamos alguns resultados:

- Em relação ao viés dos estimadores, notamos que o mesmo diminui conforme o tamanho amostral aumenta, sendo mais próximo a zero quando $n = 5000$.
- A raiz quadrada do EQM dos estimadores de β_0 , β_1 , μ_x e σ^2 diminui à medida que n aumenta. Já para $\hat{\sigma}_x^2$, a medida $\sqrt{\text{EQM}}$ permanece em torno de 3.5 para todos os tamanhos amostrais, indicando que quando a condição C4 é violada esse estimador não é consistente.
- Em relação às medidas calculadas para a variável Z_n é possível observar que o desvio-padrão aumenta nos estimadores de β_1 e μ_x . Tal fato já era esperado, devido à perturbação na condição C4 e os resultados em (2.15). Os valores mais discrepantes são para $\hat{\sigma}_x^2$, em que os valores de DP_{Z_n} são superiores a 15 em todos os tamanhos amostrais.
- Os valores de assimetria e curtose observados também diferem dos esperados na distribuição normal padrão. As maiores distorções são verificadas para o estimador de σ_x^2 , para o tamanho amostral $n = 5000$, por exemplo, temos $\text{CA}_{Z_n} = -15.583$ e $\text{C}_{Z_n} = 316.351$.
- Os testes de normalidade da variável Z_n , indicam a rejeição da hipótese de normalidade. Sendo assim, não há indícios para concluir que Z_n tem distribuição normal à medida que o tamanho amostral aumenta.

Figura 2.2: Histogramas acompanhados da função densidade da distribuição normal padrão, considerando as réplicas de Z_n , associados às estimativas de β_1 , cenário em que todas as condições são satisfeitas.

- Nesse cenário em que perturbamos a condição da matriz positiva definida, no tamanho amostral $n = 5000$, por exemplo, observamos que em 527 réplicas, das 5000 calculadas, a matriz não foi definida positiva. Adicionalmente, o valor de NPD aumenta à medida que o tamanho amostral aumenta.
- Além disso, verificamos algumas falhas na convergência nesse cenário, as quais aumentam conforme aumentamos o tamanho da amostra.

Na Figura 2.3, ilustramos os resultados da Tabela 2.2. Apresentamos novamente as medidas associadas às estimativas com o pior desempenho. Na situação em que violamos a condição C4, os piores resultados estão associados às estimativas de $\hat{\sigma}_x^2$, como já havíamos comentado. O que podemos destacar é que a medida da raiz do EQM não decresce para zero, mesmo no maior tamanho amostral considerado. As outras medidas que merecem destaque são a assimetria e curtose, as quais

Tabela 2.2: Resultados da simulação de Monte Carlo considerando o modelo dado em (2.13), com $\beta_0 = 2$, $\beta_1 = 0$, $\mu_x = 0$, $\sigma_x^2 = 4$, $\sigma^2 = 10$, $\sqrt{\tau_{yi}} \sim U(0.5; 4)$ e $\tau_{xi} = 2(i^{1.1})$. As colunas apresentam a média, o viés e a raiz quadrada do erro quadrático médio dos estimadores ($\sqrt{EQM}$), a média ($Média_{Z_n}$), o desvio-padrão (DP_{Z_n}), o coeficiente de assimetria (CA_{Z_n}), o coeficiente de curtose (C_{Z_n}) e os p-valores associados ao teste de normalidade de Kolmogorov-Smirnov (KS_{Z_n}) e de Shapiro-Wilk (SW_{Z_n}) da variável padronizada Z_n . Espera-se que os valores das medidas associadas a Z_n se encontrem próximos aos da normal padrão, quando satisfeitas as condições de regularidade.

Estimador	Média	Viés	$\sqrt{EQM}$	$Média_{Z_n}$	DP_{Z_n}	CA_{Z_n}	C_{Z_n}	KS_{Z_n}	SW_{Z_n}
$n = 50, Nc=2, NPD=172$									
$\hat{\beta}_0$	1.972	-0.028	1.691	0.004	1.032	0.064	3.216	0.820	0.026
$\hat{\beta}_1$	0.012	0.012	2.457	-0.005	1.726	-0.005	4.058	0.000	0.000
$\hat{\mu}_x$	0.001	0.001	0.706	0.006	2.968	0.451	32.848	0.000	0.000
$\hat{\sigma}_x^2$	3.914	-0.086	3.499	-6.872	41.488	-16.776	424.978	0.000	0.000
$\hat{\sigma}^2$	6.804	-3.196	5.390	-0.744	1.089	-1.103	10.087	0.000	0.000
$n = 100, Nc=3, NPD=287$									
$\hat{\beta}_0$	1.968	-0.032	1.420	0.019	1.099	0.071	3.597	0.015	0.000
$\hat{\beta}_1$	-0.002	-0.002	2.043	-0.016	1.950	-0.087	5.160	0.000	0.000
$\hat{\mu}_x$	-0.007	-0.007	0.641	-0.038	3.634	-2.800	137.064	0.000	0.000
$\hat{\sigma}_x^2$	3.953	-0.047	3.375	-6.525	44.999	-19.807	579.772	0.000	0.000
$\hat{\sigma}^2$	7.582	-2.418	4.468	-0.618	1.027	-0.306	3.517	0.000	0.000
$n = 500, Nc=15, NPD= 389$									
$\hat{\beta}_0$	1.985	-0.015	0.910	0.005	1.522	0.079	6.539	0.000	0.000
$\hat{\beta}_1$	-0.028	-0.028	1.593	0.038	3.569	0.007	7.863	0.000	0.000
$\hat{\mu}_x$	0.012	0.012	0.554	-0.011	5.195	-1.710	91.361	0.000	0.000
$\hat{\sigma}_x^2$	4.070	0.070	3.395	-8.075	50.671	-14.746	309.734	0.000	0.000
$\hat{\sigma}^2$	8.211	-1.789	3.486	-0.598	1.163	-0.951	4.495	0.000	0.000
$n = 1000, Nc=42, NPD= 417$									
$\hat{\beta}_0$	1.990	-0.010	0.685	0.052	1.904	-0.086	8.528	0.000	0.000
$\hat{\beta}_1$	0.016	0.016	1.030	-0.064	4.672	-0.093	9.787	0.000	0.000
$\hat{\mu}_x$	0.003	0.003	0.524	0.022	5.381	0.338	72.959	0.000	0.000
$\hat{\sigma}_x^2$	4.215	0.215	3.536	-6.096	33.332	-11.812	201.317	0.000	0.000
$\hat{\sigma}^2$	8.466	-1.534	3.122	-0.600	1.358	-1.581	6.561	0.000	0.000
$n = 5000, Nc=153, NPD= 527$									
$\hat{\beta}_0$	1.993	-0.007	0.316	0.047	2.640	0.156	12.645	0.000	0.000
$\hat{\beta}_1$	-0.004	-0.004	0.675	-0.235	6.177	-0.305	19.384	0.000	0.000
$\hat{\mu}_x$	0.005	0.005	0.477	-0.033	5.071	-1.745	67.708	0.000	0.000
$\hat{\sigma}_x^2$	4.264	0.264	3.508	-2.386	18.612	-15.583	316.351	0.000	0.000
$\hat{\sigma}^2$	9.184	-0.816	1.940	-0.361	1.739	-4.133	26.103	0.000	0.000

apresentam valores discrepantes e não há indicativo de convergência para o valor esperado na distribuição normal padrão.

A Figura 2.4 apresenta alguns histogramas com a função densidade da distribuição normal padrão associados ao cenário em que a condição C4 é violada. Podemos notar que há uma discrepância entre o histograma e a curva da função densidade normal padrão, há indícios de que a variância aumenta em tamanhos amostrais maiores. A análise gráfica confirma o que verificamos por meio da análise da Tabela 2.2.

Consideramos ainda simulações em que τ_{xi} foi gerado do valor absoluto da distribuição Cauchy padrão. Os resultados são semelhantes aos da Tabela 2.1, ou seja, a distribuição dos estimadores também convergiu para a distribuição normal. Por questão de brevidade, omitimos esses resultados.

Figura 2.3: Medidas avaliadas na simulação versus tamanho amostral quando a condição $C4$ é violada associadas às estimativas de $\hat{\sigma}_x^2$.

2.4 Conclusões

Neste capítulo estudamos as condições de regularidade para inferências via máxima verossimilhança no modelo de regressão normal multivariado com parametrização geral. Na Seção 2.1, apresentamos o modelo de regressão normal multivariado discutido inicialmente em [Patriota e Lemonte \(2009\)](#). Apresentamos as quantidades do vetor escore, matriz de informação de Fisher e matriz de informação observada em forma matricial. Na Seção 2.2, enunciamos as condições de regularidade e formalizamos os teoremas que garantem a existência assintótica, unicidade com probabilidade um, consistência e normalidade assintótica. Além de um teorema que aborda a distribuição assintótica das estatísticas de teste usuais. Na Seção 2.3, apresentamos três exemplos a fim de ilustrar a verificação das condições de regularidade e o resultado dos teoremas em alguns modelos específicos. No exemplo que aborda o modelo linear normal buscamos exemplificar o uso dos Teoremas 2.2.2 e 2.2.3. Os resultados obtidos estão de acordo com resultados já descritos na literatura. No exemplo considerando o modelo não linear heterocedástico exemplificamos o uso do Teorema 2.2.3 com auxílio do Corolário 2.2.2. No último exemplo abordamos o modelo com erros nas variáveis e na equação, verificamos as condições de regularidade dos Teoremas 2.2.4 e 2.2.5. Além disso, realizamos um estudo de simulação em que buscamos violar uma das condições de regularidade, a fim de verificar qual o impacto sobre a consistência e normalidade assintótica. O estudo de simulação indicou que quando violamos a condição $C4$, no sentido de que $\mathbf{\Lambda}_\theta$ existe, mas não é uma matriz definida positiva, a consistência e normalidade assintótica são comprometidas no modelo com erros nas variáveis e na equação. Vale salientar que os exemplos apresentados tem por objetivo possibilitar ao leitor um melhor entendimento das condições de regularidade apresentadas, sendo possível um estudo mais detalhado. Destacamos ainda que há uma grande variedade de exemplos de aplicações das condições de regularidade e resultados obtidos, devido à generalidade do modelo. Nesse sentido, em trabalhos

(a) $n = 50$ (b) $n = 100$ (c) $n = 500$ (d) $n = 1000$

Figura 2.4: Histogramas acompanhados da função densidade da distribuição normal padrão, considerando as réplicas de Z_n , associados às estimativas de β_1 , cenário que viola a condição $C4$.

futuros buscaremos explicitar as condições de regularidade estudadas em outros casos particulares do modelo.

Capítulo 3

Condições de regularidade para o modelo de regressão elíptico multivariado com parametrização geral

Neste capítulo trabalhamos com o modelo elíptico com parametrização geral a fim de enunciar condições de regularidade que garantam as propriedades já citadas dos EMV e a distribuição assintótica das estatísticas de teste no modelo geral. Definimos, na Seção 3.1, o modelo elíptico com parametrização geral e quantidades importantes como a função de log-verossimilhança e a matriz de informação de Fisher. Na Seção 3.2, enunciamos e discutimos as condições de regularidade, também apresentamos os teoremas que abordam as propriedades dos EMV e a distribuição assintótica das estatísticas de teste. Vale destacar que as condições apresentadas estão relacionadas com às apresentadas no Capítulo 2, no sentido de condições mais gerais, que incluem a classe de distribuições elípticas. Por fim, na Seção 3.3, apresentamos um exemplo em que explicitamos as condições de regularidade em um modelo específico.

3.1 O modelo de regressão elíptico com parametrização geral

Seguiremos a mesma notação apresentada em artigos anteriores como em [Lemonte e Patriota \(2011\)](#) e [Melo et al. \(2018\)](#). Consideremos n vetores aleatórios independentes $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$. O modelo elíptico multivariado com parametrização geral pode ser representado como

$$\mathbf{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \mathbf{x}_i) + \mathbf{e}_i, \quad i = 1, \dots, n, \quad (3.1)$$

em que $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ é um vetor de parâmetros ($p \times 1$) e $\mathbf{e}_1, \dots, \mathbf{e}_n$ são vetores aleatórios independentes tais que $\mathbf{e}_i$ tem distribuição elíptica d_i -variada com vetor de locação de zeros e uma matriz escala $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \mathbf{w}_i)$ ($d_i \times d_i$), ou seja, $\mathbf{e}_i \stackrel{\text{ind}}{\sim} \mathcal{E}l_{d_i}(\mathbf{0}, \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \mathbf{w}_i), g)$, em que $\mathbf{x}_i$ e $\mathbf{w}_i$ são vetores $m_i \times 1$ e $k_i \times 1$ de covariáveis conhecidas associadas à i -ésima resposta observada $\mathbf{Y}_i$. Então,

$$\mathbf{Y}_i \stackrel{\text{ind}}{\sim} \mathcal{E}l_{d_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, g), \quad i = 1, \dots, n,$$

em que $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \mathbf{x}_i)$ e $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}, \mathbf{x}_i)$. Assumimos que $\boldsymbol{\mu}_i$ e $\boldsymbol{\Sigma}_i$ têm forma conhecida e são duas vezes diferenciáveis com respeito a cada elemento de $\boldsymbol{\theta}$.

Seja $\mathbf{z}_i = \mathbf{Y}_i - \boldsymbol{\mu}_i$ e $\mathbf{u}_i = \mathbf{z}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{z}_i$. A função de log-verossimilhança associada ao modelo (3.1), a menos de uma constante, é dada por

$$\ell(\mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}),$$

em que $\ell_i(\boldsymbol{\theta}) = -\frac{1}{2} \log \det(\boldsymbol{\Sigma}_i) + \log g(u_i)$. Em [Melo *et al.* \(2017\)](#), assume-se que $\ell(\mathbf{y}, \boldsymbol{\theta})$ é uma função de log-verossimilhança regular com respeito a $\boldsymbol{\theta}$. No entanto, nenhuma condição de regularidade é especificada. Definamos adicionalmente

$$v_i = -2W_g(u_i).$$

O vetor escore e a matriz informação de Fisher após alguma álgebra matricial podem ser escritos, respectivamente, como ([Lemonte e Patriota, 2011](#))

$$\mathbf{U}_{\boldsymbol{\theta}} = \sum_{i=1}^n \mathbf{F}_i^{\top} \mathbf{H}_i \mathbf{s}_i \quad \text{e} \quad \mathbf{K}_{\boldsymbol{\theta}} = \sum_{i=1}^n \mathbf{F}_i^{\top} \mathbf{H}_i \mathbf{M}_i \mathbf{H}_i \mathbf{F}_i, \quad (3.2)$$

em que

$$\mathbf{F}_i = \begin{pmatrix} \mathbf{D}_i \\ \mathbf{V}_i \end{pmatrix}, \quad \mathbf{H}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i & \mathbf{0} \\ \mathbf{0} & 2\tilde{\boldsymbol{\Sigma}}_i \end{pmatrix}^{-1}, \quad \mathbf{s}_i = \begin{pmatrix} v_i \mathbf{z}_i \\ \text{vec}(v_i \mathbf{z}_i \mathbf{z}_i^{\top} - \boldsymbol{\Sigma}_i) \end{pmatrix} \quad \text{e}$$

$$\mathbf{M}_i = \begin{pmatrix} \frac{4\psi_i(2,1)}{d_i} \boldsymbol{\Sigma}_i & \mathbf{0} \\ \mathbf{0} & 2c_i \tilde{\boldsymbol{\Sigma}}_i \end{pmatrix} + (c_i - 1) \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \text{vec}(\boldsymbol{\Sigma}_i) \text{vec}(\boldsymbol{\Sigma}_i)^{\top} \end{pmatrix},$$

em que $c_i = 4\psi_i(2,2)/\{d_i(d_i+2)\}$, $\psi_i(2,1) = E(W_g^2(U_i)U_i)$ e $\psi_i(2,2) = E(W_g^2(U_i)U_i^2)$, com $U_i = \|\mathbf{E}_i\|^2$, $\mathbf{E}_i \sim \mathcal{E}l_{d_i}(\mathbf{0}, \mathbf{I}_{d_i}, g)$. Os valores de $\psi_i(l, k)$ podem ser obtidos a partir da resolução da integral ([Lange *et al.*, 1989](#); [Melo *et al.*, 2017](#))

$$\psi_i(l, k) = \int_0^{\infty} W_g(s^2)^l g(s^2) r^{d_i+2k-1} c_{d_i} ds, \quad (3.3)$$

em que $c_{d_i} = 2\pi^{\frac{d_i}{2}}/\Gamma\left(\frac{d_i}{2}\right)$ é a área da superfície da esfera unitária em $\mathbb{R}^{d_i}$ e $\Gamma(\cdot)$ é a função gama. As quantidades para as diferentes distribuições podem ser obtidas pela solução algébrica ou numérica da integral (3.3). A Tabela 3.1 foi obtida de [Melo *et al.* \(2017\)](#) e apresenta algumas quantidades para as distribuições normal, Cauchy, t de Student e exponencial potência (EP).

Tabela 3.1: Funções $\psi_i(2, 1)$, $\psi_i(2, 2)$, $\psi_i(3, 2)$ e $\psi_i(3, 3)$ para as distribuições normal, Cauchy, t de Student (t) e exponencial potência (EP).

	$\psi_i(2, 1)$	$\psi_i(2, 2)$	$\psi_i(3, 2)$	$\psi_i(3, 3)$	Restrição
Normal	$\frac{d_i}{4}$	$\frac{d_i(d_i+2)}{4}$	$-\frac{d_i(d_i+2)}{8}$	$-\frac{d_i(d_i+2)(d_i+4)}{8}$	$d_i \geq 1$
Cauchy	$\frac{d_i(d_i+1)}{4(d_i+3)}$	$\frac{d_i(d_i+2)(d_i+1)}{4(d_i+3)}$	$-\frac{d_i(d_i+2)(d_i+1)^2}{8(d_i+3)(d_i+5)}$	$-\frac{d_i(d_i+2)(d_i+4)(d_i+1)^2}{8(d_i+3)(d_i+5)}$	$d_i \geq 1$
t	$\frac{d_i(d_i+\nu)}{4(d_i+\nu+2)}$	$\frac{d_i(d_i+2)(d_i+\nu)}{4(d_i+\nu+2)}$	$-\frac{d_i(d_i+2)(d_i+\nu)^2}{8(d_i+\nu+2)(d_i+\nu+4)}$	$-\frac{d_i(d_i+2)(d_i+4)(d_i+\nu)^2}{8(d_i+\nu+2)(d_i+\nu+4)}$	$d_i \geq 1$
EP	$\frac{\lambda^2 \Gamma\left(\frac{4\lambda-1}{2\lambda}\right)}{2^{1/\lambda} \Gamma\left(\frac{1}{2\lambda}\right)}$	$\frac{2\lambda+1}{4}$	$-\frac{\lambda^3 \Gamma\left(\frac{6\lambda-1}{2\lambda}\right)}{2^{1/\lambda} \Gamma\left(\frac{1}{2\lambda}\right)}$	$-\frac{2(\lambda+1)(4\lambda+1)}{8}$	$d_i = 1, \lambda > \frac{1}{4}$
EP	$\frac{\lambda^2 \Gamma\left(\frac{d_i-2}{2\lambda}+2\right)}{2^{1/\lambda} \Gamma\left(\frac{d_i}{2\lambda}\right)}$	$\frac{d_i(2\lambda+d_i)}{4}$	$-\frac{\lambda^3 \Gamma\left(\frac{d_i-2}{2\lambda}+3\right)}{2^{1/\lambda} \Gamma\left(\frac{d_i}{2\lambda}\right)}$	$-\frac{d_i(2\lambda+d_i)(4\lambda+d_i)}{8}$	$d_i \geq 2, \lambda > 0$

Notemos que no caso da distribuição normal, considerando as quantidades da Tabela 3.1, chegamos a $\mathbf{M}_i = \mathbf{H}_i^{-1}$. Adicionalmente temos $v_i = 1$ e assim as expressões de $\mathbf{U}_{\boldsymbol{\theta}}$ e $\mathbf{K}_{\boldsymbol{\theta}}$ são simplificadas e coincidem com as apresentadas no Capítulo 2 e em [Patriota e Lemonte \(2009\)](#).

Cabe destacar que as expressões de $\psi_i(l, k)$ podem envolver parâmetros adicionais das distribuições. Na Tabela 3.1, por exemplo, identificamos os graus de liberdade ν no caso da distribuição t de Student e o parâmetro de forma λ na distribuição exponencial potência. Na distribuição t de Student, por exemplo, conforme [Lucas \(1997\)](#), ao estimarmos os graus de liberdade ν via verossimilhança perdemos a propriedade de robustez para observações extremas da distribuição t de Student. Em geral, os trabalhos assumem que essas quantidades são fixas na maximização da verossimilhança e determinam o valor por meio da minimização de alguma medida de informação, como o critério de Akaike (AIC) ou o critério de informação bayesiano (BIC), por exemplo. Dentre os trabalhos que utilizam essa forma para determinar os valores dessas quantidades temos [Russo et al. \(2009\)](#), [Ibacache-Pulgar e Paula \(2011\)](#), [Matos et al. \(2013\)](#) e [Massuia et al. \(2015\)](#), entre outros. No entanto, vale salientar que esse procedimento ainda utiliza a verossimilhança, em vista que esses critérios utilizam a função de log-verossimilhança. Uma alternativa sugerida em [Vanegas e Paula \(2015\)](#) e [Vanegas e Paula \(2016\)](#) é escolher para essa quantidade assumida fixa, a qual definimos de forma geral por ξ , o valor que minimiza uma medida de qualidade de ajuste dada por

$$\Upsilon_\xi = n^{-1} \sum_{i=1}^n |\Phi^{-1}[F_T(\hat{t}^{(k)})] - \nu^{(k)}|,$$

em que $F_T(\cdot)$ é a função de distribuição acumulada de T , $\hat{t}^{(k)}$ é a k -ésima estatística de ordem de $\hat{t}$, $\nu^{(k)}$ é a esperança da k -ésima estatística de ordem da amostra aleatória de tamanho n da distribuição normal padrão e $\Phi^{-1}(\cdot)$ é a inversa da função de distribuição acumulada da distribuição normal padrão. Para mais detalhes ver [Vanegas e Paula \(2015\)](#) e [Vanegas e Paula \(2016\)](#).

A matriz informação observada de θ é dada por $\mathbf{J}_\theta = -\mathbf{L}_\theta$, obtida de [Lemonte e Patriota \(2011\)](#),

$$\mathbf{L}_\theta = \frac{\partial^2 \ell(\mathbf{y}, \theta)}{\partial \theta \partial \theta^\top} = \sum_{i=1}^n \left(\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i \right),$$

em que $\mathbf{B}_i$ é uma matriz $p \times p$ com o elemento (s, r) , $s = 1, \dots, p$ e $r = 1, \dots, p$, dado por $\mathbf{r}_i^\top \mathbf{H}_i \mathbf{s}_i$, com o vetor $\mathbf{r}_i = \begin{pmatrix} \mathbf{a}_{i(sr)} \\ \text{vec}(\mathbf{C}_{i(sr)}) \end{pmatrix}$, $\mathbf{a}_{i(sr)} = \frac{\partial^2 \boldsymbol{\mu}_i}{\partial \theta_s \partial \theta_r}$, $\mathbf{C}_{i(sr)} = \frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial \theta_s \partial \theta_r}$ e

$$\begin{aligned} \ddot{\mathbf{M}}_i = & 2W_g(u_i) \begin{pmatrix} \boldsymbol{\Sigma}_i & 2\mathbf{z}_i^\top \otimes \boldsymbol{\Sigma}_i \\ 2\boldsymbol{\Sigma}_i \otimes \mathbf{z}_i & 2(\boldsymbol{\Sigma}_i \otimes (\mathbf{z}_i \mathbf{z}_i^\top) + (\mathbf{z}_i \mathbf{z}_i^\top) \otimes \boldsymbol{\Sigma}_i) \end{pmatrix} \\ & + 4W_g^2(u_i) \begin{pmatrix} \mathbf{z}_i \mathbf{z}_i^\top & \mathbf{z}_i^\top \otimes (\mathbf{z}_i \mathbf{z}_i^\top) \\ (\mathbf{z}_i \mathbf{z}_i^\top) \otimes \mathbf{z}_i & \text{vec}(\mathbf{z}_i \mathbf{z}_i^\top) \text{vec}(\mathbf{z}_i \mathbf{z}_i^\top)^\top \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 2\tilde{\boldsymbol{\Sigma}}_i \end{pmatrix}. \end{aligned}$$

O método escore de Fisher pode ser utilizado para estimar θ por meio da solução iterativa da equação

$$\mathbf{F}_\theta^{(m)\top} \mathbf{W}^{(m)} \mathbf{F}_\theta^{(m)} \theta^{(m+1)} = \mathbf{F}_\theta^{(m)\top} \mathbf{W}^{(m)} \mathbf{s}^{*(m)}, \quad m = 0, 1, \dots,$$

em que

$$\begin{aligned} \mathbf{F}_\theta^{(m)\top} &= (\mathbf{F}_1^{(m)\top}, \mathbf{F}_2^{(m)\top}, \dots, \mathbf{F}_n^{(m)\top}), \quad \mathbf{W}^{(m)} = \mathbf{H}^{(m)} \mathbf{M}^{(m)} \mathbf{H}^{(m)} \\ \mathbf{H}^{(m)} &= \text{bdiag}\{\mathbf{H}_1^{(m)}, \mathbf{H}_2^{(m)}, \dots, \mathbf{H}_n^{(m)}\}, \quad \mathbf{M}^{(m)} = \text{bdiag}\{\mathbf{M}_1^{(m)\top}, \mathbf{M}_2^{(m)\top}, \dots, \mathbf{M}_n^{(m)\top}\}, \\ \mathbf{s}^{*(m)} &= \mathbf{F}_\theta^{(m)} \theta^{(m)} + \mathbf{H}^{-1(m)} \mathbf{M}^{-1(m)} \mathbf{s}^{(m)}, \quad \mathbf{s}^{(m)} = (\mathbf{s}_1^{(m)\top}, \dots, \mathbf{s}_n^{(m)\top})^\top, \end{aligned}$$

e m indica o número da iteração. As iterações são realizadas até que a convergência seja alcançada.

Assim como já definido no Capítulo 2, qualquer sequência de estimadores $\{\hat{\theta}_n\}_{n \geq 1}$ tal que

$$\hat{\theta}_n = \hat{\theta}_n(\mathbf{y}) = \arg \max_{\theta \in \Theta} \ell(\mathbf{y}, \theta),$$

é chamado EMV. Sob certas condições $\{\hat{\theta}_n\}_{n \geq 1}$ satisfaz a equação $\mathbf{U}_{\hat{\theta}_n} = \mathbf{0}$.

No modelo (3.1), não é evidente que esse estimador é único, o que será estudado na Seção 3.2.

Mais detalhes sobre o modelo elíptico com parametrização geral e as expressões apresentadas podem ser obtidos em Lemonte e Patriota (2011). Ainda Lemonte e Patriota (2011) abordam a alavancagem generalizada e apresentam uma aplicação.

3.2 Condições de regularidade

Nesta seção, buscamos generalizar os resultados obtidos na Seção 2.2. As condições que serão aqui apresentadas estão diretamente ligadas às apresentadas no Capítulo 2, a generalização dar-se-á em termos de condições adicionais sobre a função geradora de densidades g . No caso da distribuição normal, estudada no Capítulo 2, as condições sobre a função g que serão aqui adicionadas são satisfeitas devido às características da distribuição normal. Destacamos ainda que os parâmetros adicionais contidos nas expressões de $\psi_i(l, k)$ serão considerados fixos.

Considere as condições enunciadas a seguir

- C1. O espaço paramétrico $\Theta \subseteq \mathbb{R}^p$ é aberto e convexo.
- C2. As funções μ_i e Σ_i são contínuas e as duas primeiras derivadas com respeito a cada elemento de θ existem, são contínuas e limitadas, para todo $\theta \in \Theta$.
- C3. $\mathbf{F}_\theta = (\mathbf{F}_1^\top, \dots, \mathbf{F}_n^\top)^\top$ tem posto coluna completo, assim $\text{posto}(\mathbf{F}_\theta) = p$, para todo $\theta \in \Theta$.
- C4. $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{K}_\theta = \mathbf{A}_\theta$ existe, é uma matriz positiva definida e seus elementos são finitos para todo $\theta \in \Theta$.
- C5. Para cada $\delta > 0$,

$$\lim_{n \rightarrow \infty} P_\theta \left(\mathbf{J}_{\theta'} - c \mathbf{K}_\theta \text{ ser positiva semidefinida para todo } \theta' \in B_\theta(\delta) \right) = 1,$$

com alguma constante c , tal que $c > 0$ e independente de δ .

- C6. Consideremos a função $\mathbf{G} : \Theta \rightarrow \mathbb{R}^{N + \frac{N(N-1)}{2}}$, definida em (2.5), em que $N = d_1 + \dots + d_n$. Assumamos que existem subconjuntos $\Theta^1, \Theta^2, \dots$ abertos e convexos e $\mathbf{G}^1, \mathbf{G}^2, \dots$ tais que

$$\mathbf{G}(\Theta) = \bigcup_{j=1}^{\infty} \mathbf{G}^j(\Theta^j),$$

em que $\mathbf{G}^j(\Theta^j)$ é uma função duas vezes diferenciável e o posto do jacobiano de $\mathbf{G}^j$ em relação a θ é p_j (em que $\Theta \subseteq \mathbb{R}^p$ é aberto e convexo), para cada $\theta \in \Theta^j$, Θ^j um subconjunto de $\mathbb{R}^{p_j}$ e $j \geq 1$.

- C7. A função $\mathbf{G}$ é injetiva.

Notemos que se $\mathbf{G}$ é injetiva, então temos que θ é identificável no modelo (3.1).

- C8. Existe uma sequência $M_n = O_p(n)$ tal que

$$\|\mathbf{L}_{\theta_n^*} - \mathbf{L}_\theta\| \leq M_n \|\theta_n^* - \theta\|,$$

em que $\theta_n^* = h \hat{\theta}_n + (1 - h) \theta$, $h \in (0, 1)$, para todo $\theta \in \Theta$.

As condições C1-C8 são análogas às apresentadas no Capítulo 2 considerando as quantidades definidas associadas ao modelo (3.1). As condições apresentadas a seguir são condições adicionais que são necessárias no caso em que assumimos a distribuição elíptica para os erros, as quais são satisfeitas no caso da normalidade dos erros.

- C9. A função g é contínua, com as duas primeiras derivadas contínuas e limitadas e o conjunto $\{u \in \mathbb{R} : g(u) > 0\}$ não depende de θ .
- C10. $E([W_g(U_i)]^l U_i^s)$ e $E([W_g^2(U_i)]^l U_i^s)$, para $l, s = 1, 2$ e $i = 1, \dots, n$, são finitas, em que $U_i = \|\mathbf{E}_i\|^2$, com $\mathbf{E}_i \sim \mathcal{E}_{d_i}(\mathbf{0}, \mathbf{I}_{d_i}, g)$.
- C11. $W_g(u) \neq 0$ com probabilidade um, com $u = [\mathbf{Y} - \boldsymbol{\mu}(\theta)]^\top \boldsymbol{\Sigma}^{-1}(\theta) [\mathbf{Y} - \boldsymbol{\mu}(\theta)]$.

As condições C2, C9 e C10 são condições naturais para obtenção do vetor escore e da matriz de informação de Fisher. Em relação às condições C1-C8, os mesmos comentários do Capítulo 2 acerca dessas condições são aplicados aqui.

As condições aqui enunciadas serão utilizadas nos teoremas de existência, unicidade, consistência e normalidade assintótica do EMV e distribuição assintótica das estatísticas de teste.

3.2.1 Existência do EMV

Nesta subseção, discutimos a existência do EMV no modelo elíptico com parametrização geral. Nosso estudo está relacionado com o resultado obtido no Capítulo 2, o qual foi desenvolvido a partir do trabalho de Fahrmeir e Kaufmann (1985). A ideia consiste em adicionar condições suficientes que são atendidas no caso da distribuição normal, mas que podem não ser satisfeitas dependendo da função geradora g no caso das distribuições elípticas. A seguir, enunciamos o teorema e comentamos as condições estabelecidas.

Teorema 3.2.1. *Seja o modelo definido em (3.1). Sob as condições C1-C3, C5, C9 e C10, existe uma sequência $\{\hat{\theta}_n\}_{n \geq 1}$ de vetores aleatórios tais que*

$$\lim_{n \rightarrow \infty} P_{\theta}(\mathbf{U}_{\hat{\theta}_n} = \mathbf{0}) = 1, \text{ para cada } \theta \in \Theta \text{ (existência assintótica)}.$$

As condições C2, C3, C9 e C10 são necessárias para que possamos estabelecer o vetor escore, a matriz de informação de Fisher e informação observada, o que foi feito na Seção 3.1, além de garantir que $\mathbf{K}_{\theta}$ é uma matriz definida positiva. Relacionando com o Capítulo 2, estabelecidas as quantidades da Seção 3.1, as condições C1 e C5 são suficientes para provarmos a existência assintótica do EMV no modelo elíptico com parametrização geral, da mesma forma que no modelo normal.

Assim, estabelecidas as quantidades da Seção 3.1, a prova do teorema é dada de maneira análoga à do Teorema 2.2.1, a qual é apresentada no Apêndice A, Seção A.1.

3.2.2 Unicidade do EMV

Nesta subseção, discutimos as condições suficientes para garantir a unicidade do EMV de θ com probabilidade um no modelo (3.1). Novamente aplicamos os resultados apresentados em Krätschmer (2007).

Consideremos o conjunto R definido no Capítulo 2 por

$$R := \left\{ \mathbf{y} \in \mathbb{R}^N : \mathbf{U}_{\hat{\theta}_n(\mathbf{y})} = \mathbf{0} \right\} \cap \left\{ \mathbf{y} \in \mathbb{R}^N : \det(\mathbf{L}_{\hat{\theta}_n(\mathbf{y})}) = 0 \right\}. \quad (3.4)$$

Este conjunto contém vetores $\mathbf{y}$ tais que $\hat{\theta}_n(\mathbf{y})$ é ponto crítico degenerado da função de log-verossimilhança, ou seja, o determinante da matriz hessiana aplicada no ponto crítico é nulo. A seguir, apresentamos as condições sob as quais a unicidade do EMV é garantida com probabilidade um no modelo elíptico multivariado com parametrização geral.

Teorema 3.2.2. *Seja o modelo dado em (3.1). Sob as condições C6, C7, C9 e C11, se R tiver medida de Lebesgue nula, então o conjunto*

$$Q := \left\{ \mathbf{y} \in \mathbb{R}^N : \ell(\mathbf{y}, \theta_1) = \ell(\mathbf{y}, \theta_2) = \max_{\theta \in \Theta} \ell(\mathbf{y}, \theta), \theta_1 \neq \theta_2, \theta_1, \theta_2 \in \Theta \right\}$$

é mensurável e tem medida de Lebesgue nula.

A abordagem aqui também é feita em termos da unicidade do EMV com probabilidade um. O mesmo conjunto Q introduzido e interpretado no Capítulo 2 é aqui definido, quando este apresentar medida nula temos a unicidade do EMV com probabilidade um. Assim como já mencionado no Capítulo 2, se o conjunto R tem medida nula a matriz $\mathbf{L}_\theta$ é positiva definida ou negativa definida e assim teríamos, um ponto de mínimo ou de máximo, respectivamente.

O procedimento para a prova desse teorema está embasado na sequência adotada para o caso em que supomos que os erros seguem distribuição normal, sendo necessária apenas a inclusão das condições C9 e C11. Notemos que a condição C11 está associada à quantidade $W_g(u)$, a qual é definida no Capítulo 1 e é uma quantidade importante quando trabalhamos com a classe de distribuições elípticas. Utilizamos como base as condições suficientes enunciadas em Krätschmer (2007), adaptamos essas condições ao modelo (3.1). A condição C7 aborda a injetividade da função $\mathbf{G}$ e é utilizada no Lema B.1.1, apresentado no Apêndice B, o qual está associado à derivada da função de log-verossimilhança em relação ao vetor $\mathbf{y}$. As condições C7 e C11 nos garantem o resultado do Lema B.1.1 para toda a classe de distribuições elípticas. A condição C11 é satisfeita no caso da distribuição normal, garantindo o resultado do Lema A.2.1, apresentado no Apêndice A, para o modelo estudado no Capítulo 2. Satisfeitas as condições C6, C7 e C9 a função de log-verossimilhança é duas vezes diferenciável. Pela condição C7, também temos a identificabilidade do modelo. A condição de que o conjunto R , definido em (3.4), tem medida nula é uma condição suficiente para completar a prova. Esta condição deve ser verificada em cada caso específico do modelo, conforme exemplo na Seção 3.3. Notemos ainda que as condições C1-C3 não são exigidas nesse teorema, no entanto, se forem satisfeitas temos que a condição C6 é satisfeita. A prova com detalhes é apresentada no Apêndice B, Seção B.1.

Uma forma de mostrar que o conjunto R tem medida de Lebesgue nula é utilizar o Corolário 2.2.1 apresentado no Capítulo 2. Ou ainda, a seguir apresentamos um teorema com condições alternativas a de que o conjunto R tem medida nula. Novamente restringiremos o nosso estudo ao caso em que $\boldsymbol{\mu}(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ não compartilham parâmetros. Assim escrevemos $\boldsymbol{\mu}(\boldsymbol{\beta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$.

Teorema 3.2.3. *Seja o modelo dado em (3.1), com $\boldsymbol{\mu}(\boldsymbol{\beta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$. Sob as condições C1-C3, C7 e C9-C11, se a função $f(\mathbf{x}) = -2W_g(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}})\mathbf{x})\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}})\mathbf{x}$ for um difeomorfismo, então o conjunto*

$$Q := \left\{ \mathbf{y} \in \mathbb{R}^N : \ell(\mathbf{y}, \boldsymbol{\theta}_1) = \ell(\mathbf{y}, \boldsymbol{\theta}_2) = \max_{\boldsymbol{\theta} \in \Theta} \ell(\mathbf{y}, \boldsymbol{\theta}), \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta \right\}$$

é mensurável e tem medida de Lebesgue nula.

Este teorema é uma alternativa ao Teorema 3.2.2, aplicável no caso específico em que $\boldsymbol{\mu}(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ não compartilham parâmetros, diretamente relacionado ao Teorema 2.2.3. A alteração em relação ao Teorema 2.2.3 se dá em termos do acréscimo de algumas condições sobre a função geradora de densidade g e suas derivadas. A função $f(\mathbf{x})$ tem uma pequena alteração que se dá por meio de W_g , mantidas algumas quantidades relacionadas a $f(\mathbf{x})$ estudada no Capítulo 2. O teorema supõe que $f(\mathbf{x})$ é um difeomorfismo, tal condição deve ser verificada em cada modelo e distribuição específica.

Na elaboração desse teorema objetivamos estender o resultado do Teorema 2.2.3, apresentado no Capítulo 2, o qual está fundamentado no trabalho de Krätschmer (2007). As condições C2, C3 e de que $f(\mathbf{x})$ é um difeomorfismo substituem a condição de que R tem medida nula do Teorema 3.2.2. Mediante as condições C2 e C3 temos que a condição C4.2 de Krätschmer (2007) no modelo (3.1) é satisfeita, tal fato provamos com detalhes no Apêndice B, Lema B.2.1. Ainda, sob as condições C1-C3, temos que a condição C6 é satisfeita. Adicionalmente obtivemos a função $f(\mathbf{x}) = -2W_g(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}})\mathbf{x})\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}})\mathbf{x}$, a qual foi definida de acordo com as condições apresentadas em Krätschmer (2007). A prova do Teorema 3.2.3 é apresentada com detalhes no Apêndice B, Seção B.2.

Para a verificação de que $f(\mathbf{x})$ é um difeomorfismo, uma possibilidade é verificar se $f(\mathbf{x})$ e $f^{-1}(\mathbf{x})$ são diferenciáveis. Outra possibilidade é verificar se a derivada de $f(\mathbf{x})$ é inversível. Ambos resultados implicam que $f(\mathbf{x})$ é um difeomorfismo. Essa verificação deve ser feita em cada modelo específico.

Vale ressaltar que o Teorema 3.2.3 não se aplica ao modelo geral, apenas ao caso em que $\boldsymbol{\mu}(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ não compartilham parâmetros. O modelo geral será objeto de estudo em trabalhos futuros.

3.2.3 Consistência do EMV

A seguir, apresentamos o teorema sobre a consistência do EMV $\hat{\boldsymbol{\theta}}_n$. Em relação ao Teorema 2.2.4, que trata da consistência para o modelo normal com parametrização geral, temos a inclusão das condições C9 e C10 descritas neste capítulo. Tais condições no modelo normal com parametrização geral são facilmente verificadas.

Teorema 3.2.4. *Seja o modelo definido em (3.1). Sob as condições C1-C5, C9 e C10, existe uma sequência $\{\hat{\boldsymbol{\theta}}_n\}_{n \geq 1}$ de EMV tais que*

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{P_{\boldsymbol{\theta}}} \boldsymbol{\theta}, \text{ para todo } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Novamente, as condições C2, C3, C9 e C10 são necessárias para obtenção do vetor escore, da matriz de informação observada e matriz de informação de Fisher, garantindo que essa é positiva definida. Para a prova desse teorema utilizamos uma expressão obtida mediante as condições C1 e C5 no decorrer da prova do Teorema 2.2.1, apresentada no Apêndice A, Seção A.1. Quando a condição C4 for satisfeita, temos a consistência do EMV.

Após definidas as quantidades da Seção 3.1, a prova do Teorema 3.2.4 é igual à do Teorema 2.2.4, apresentada com detalhes no Apêndice A, Seção A.4.

3.2.4 Normalidade assintótica do EMV

A seguir, enunciamos e discutimos as condições do teorema que garantem a normalidade assintótica dos EMV de $\boldsymbol{\theta}$ do modelo elíptico com parametrização geral.

Teorema 3.2.5. *Seja o modelo definido em (3.1). Sob as condições C1-C5 e C8-C10,*

$$\mathbf{K}_{\boldsymbol{\theta}}^{1/2} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p).$$

As condições aqui consideradas estão relacionadas com as utilizadas no Teorema 2.2.5. Os passos que foram seguidos na realização da prova do teorema são os mesmos enunciados no Capítulo 2, sendo consideradas duas condições adicionais, as condições C9 e C10. A condição C9 está associada à função geradora de densidades g , no caso da distribuição normal essa função é a exponencial. A condição C10 nos garante que alguns valores esperados são finitos, o que no caso do modelo normal com parametrização geral ocorre como consequência da distribuição dos erros. Novamente o teorema do valor médio multivariado e a condição C1 são usadas para reescrever o vetor escore, sendo utilizada a abordagem apresentada em Patriota (2019). Utilizamos o Teorema Central do Limite de Liapounov e as condições C3, C4 e C10, para mostrar a convergência em distribuição $\frac{\mathbf{U}_{\boldsymbol{\theta}}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \mathbf{\Lambda}_{\boldsymbol{\theta}})$, para cada $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Aplicando a Lei Fraca dos Grandes Números de Markov e as condições C3, C4 e C10, mostramos a convergência em probabilidade $\frac{\mathbf{L}_{\boldsymbol{\theta}}}{n} + \frac{\mathbf{K}_{\boldsymbol{\theta}}}{n} \xrightarrow{P_{\boldsymbol{\theta}}} \mathbf{0}$, para cada $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Vale destacar a inclusão da condição C10, a qual garante que alguns valores esperados são finitos e assim podemos aplicar a Lei Fraca dos Grandes Números de Markov. A condição C8 juntamente com o resultado do Teorema 3.2.4 garantem que $(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = O_p(n^{-1/2})$, para cada $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Por fim, por meio das condições C3 e C4, e pelo Teorema de Cramér-Wold chegamos ao resultado.

Notemos a semelhança na sequência da prova da normalidade assintótica dos EMV para o modelo em que os erros têm distribuição elíptica e o modelo em que os erros têm distribuição

normal. Para o caso elíptico comparado ao normal, apenas se faz necessária a inclusão de algumas condições sobre a função geradora de densidades g e W_g , então o resultado segue.

3.2.5 Distribuição assintótica das estatísticas de teste

Da mesma forma que no Capítulo 2, consideremos o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\tau}^\top)^\top$, em que $\boldsymbol{\psi} = (\psi_1, \dots, \psi_q)^\top$ representa o vetor de parâmetros de interesse e $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{p-q})^\top$ o vetor de parâmetros de perturbação. Suponha que o interesse esteja em testar

$$\mathcal{H}_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0 \quad \text{versus} \quad \mathcal{H}_1 : \boldsymbol{\psi} \neq \boldsymbol{\psi}_0,$$

em que $\boldsymbol{\psi}_0$ é um vetor de constantes especificado de dimensão q .

Consideremos as estatísticas já definidas no Capítulo 2, nas equações (2.8) - (2.11). O teorema a seguir trata das condições de regularidade sob as quais conhecemos a distribuição assintótica das estatísticas de teste apresentadas quando consideramos o modelo (3.1).

Teorema 3.2.6. *Seja o modelo definido em (2.1). Sob as condições C1-C5, C8 e sob a hipótese nula*

$$RV_n, W_n, S_n, G_n \xrightarrow{\mathcal{D}} \chi_q^2$$

quando $n \rightarrow \infty$, em que χ_q^2 representa a distribuição de uma variável aleatória qui-quadrado com q graus de liberdade.

Sob as condições de regularidade enunciadas no teorema e sob a hipótese nula, as quatro estatísticas apresentadas são assintoticamente equivalentes, possuindo aproximadamente distribuição χ_q^2 em grandes amostras, em que q é o número de parâmetros testados na hipótese nula. As condições C1-C5 e C8-C10 nos garantem a normalidade assintótica do EMV. Com o resultado do Teorema 3.2.5, a prova do Teorema 3.2.6 é análoga à do Teorema 2.2.6 apresentada no Apêndice A, Seção A.6.

Na Seção 3.3, buscamos exemplificar a verificação das condições de regularidade apresentadas e alguns teoremas.

3.3 Exemplo

A seguir, apresentamos um exemplo em que consideramos o modelo não linear heterocedástico. Associado a esse modelo e assumindo uma distribuição específica da classe das elípticas, a distribuição t de Student, verificamos a validade das condições enunciadas e os resultados dos teoremas nesse modelo específico.

3.3.1 Modelo não linear heterocedástico

Consideremos agora o modelo não linear heterocedástico definido por

$$Y_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i) + e_i,$$

em que Y_i é a i -ésima observação da variável resposta, $\mathbf{x}_i$ um vetor $1 \times p_1$ de variáveis explicativas, $\boldsymbol{\beta}$ é um vetor $p_1 \times 1$ de parâmetros desconhecidos e $\boldsymbol{\mu}_i$ é uma função não linear em $\boldsymbol{\beta}$. Assumimos que e_i e e_j , para $i \neq j$ são independentes, com $e_i \sim \mathcal{E}l(0, \sigma_i^2, g)$. Ainda, assumimos que $\sigma_i^2 = h_i(\boldsymbol{\gamma}, \mathbf{w}_i)$, em que $\boldsymbol{\gamma}$ é um vetor $p_2 \times 1$ de parâmetros desconhecidos e $\mathbf{w}_i$ um vetor $1 \times p_2$ das variáveis explicativas. Então

$$Y_i \sim \mathcal{E}l(\boldsymbol{\mu}_i(\boldsymbol{\theta}), \sigma_i^2, g),$$

é um caso especial do modelo (3.1), com $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i)$, $\boldsymbol{\Sigma}_i = \sigma_i^2$, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. Quando $\boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$ temos o modelo elíptico linear heteroscedástico. Sejam $z_i = y_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i)$, $\mathbf{D}_i =$

$\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i)}{\partial \boldsymbol{\beta}^\top}$, $\mathbf{V}_i = \frac{\partial h_i(\boldsymbol{\gamma}, \mathbf{w}_i)}{\partial \boldsymbol{\gamma}^\top}$, $\mathbf{X}$ a matriz que contém as variáveis explicativas associadas à função $\boldsymbol{\mu}(\boldsymbol{\beta}, \mathbf{x})$, $\mathbf{W}$ a matriz que contém as variáveis explicativas associadas à função $h(\boldsymbol{\gamma}, \mathbf{w})$, $\boldsymbol{\Sigma}(\boldsymbol{\gamma}) = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$. Temos

$$\mathbf{G}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\beta}, \mathbf{X}) \\ \text{vech}(\boldsymbol{\Sigma}(\boldsymbol{\gamma})) \end{pmatrix}, \quad \mathbf{F}_i = \begin{pmatrix} \mathbf{D}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_i \end{pmatrix}, \quad \mathbf{H}_i = \begin{pmatrix} \sigma_i^2 & 0 \\ 0 & 2\sigma_i^4 \end{pmatrix}^{-1}, \quad \mathbf{s}_i = \begin{pmatrix} v_i z_i \\ v_i z_i^2 - \sigma_i^2 \end{pmatrix}$$

$$\text{e } \mathbf{M}_i = \begin{pmatrix} 4\psi_i(2, 1)\sigma_i^2 & 0 \\ 0 & 2c_i\sigma_i^4 \end{pmatrix} + (1 - c_i) \begin{pmatrix} 0 & 0 \\ 0 & \sigma_i^4 \end{pmatrix}.$$

A matriz de segundas derivadas é dada por

$$\mathbf{L}_{\boldsymbol{\theta}} = \sum_{i=1}^n (\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i),$$

em que

$$\ddot{\mathbf{M}}_i = 2W_g(u_i) \begin{pmatrix} \sigma_i^2 & 2\sigma_i^2 z_i \\ 2\sigma_i^2 z_i & 4\sigma_i^2 z_i^2 + 2\sigma_i^4 \end{pmatrix} + 4W_g^2(u_i) \begin{pmatrix} z_i^2 & z_i^3 \\ z_i^3 & z_i^4 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 2\sigma_i^4 \end{pmatrix}$$

e $\mathbf{B}_i$ depende da estrutura das funções $\boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i)$ e $h(\boldsymbol{\gamma}, \mathbf{w}_i)$, pois envolve a segunda derivada dessas funções. Notemos ainda que as matrizes dependem de quantidades específicas da distribuição associada aos erros.

Modelo não linear homocedástico

A fim de melhor exemplificar o uso dos teoremas e das condições de regularidade, especifiquemos uma distribuição para os erros e as funções $\boldsymbol{\mu}_i$ e $\boldsymbol{\Sigma}_i$. Consideremos um exemplo, motivado por uma aplicação, apresentada em [Tiede e Pagano \(1979\)](#), a qual é retomada em [Lemonte e Patriota \(2011\)](#), sob o enfoque de modelos com parametrização geral. A distribuição considerada é a t de Student univariada ($d_i = 1$), com

$$\boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i) = \beta_1 + \frac{\beta_2}{1 + \beta_3 x_i^{\beta_4}} \quad \text{e} \quad \boldsymbol{\Sigma}_i = h_i(\boldsymbol{\gamma}, \mathbf{w}_i) = \sigma^2. \quad (3.5)$$

O vetor de parâmetros é $\boldsymbol{\theta} = (\beta_1, \beta_2, \beta_3, \beta_4, \sigma^2)^\top$, em que $\beta_1, \beta_2, \beta_3, \beta_4 \in \mathbb{R}$ e $\sigma^2 \in \mathbb{R}_+$. Notemos que trata-se de um modelo não linear homocedástico. As funções apresentadas e a distribuição t de Student foram também consideradas no estudo de simulação apresentado em [Melo *et al.* \(2018\)](#). A seguir, vamos explicitar algumas quantidades para esse modelo e discutir as condições de regularidade. Em relação ao espaço paramétrico, considerando o suporte dos parâmetros descrito, temos $\boldsymbol{\theta} \in \mathbb{R}^4 \times \mathbb{R}_+$, notemos que é aberto e convexo, logo a condição C1 é satisfeita.

Neste caso específico temos

$$\mathbf{D}_i = \begin{pmatrix} 1, & \frac{1}{1 + \beta_3 x_i^{\beta_4}}, & -\frac{\beta_2 x_i^{\beta_4}}{(1 + \beta_3 x_i^{\beta_4})^2}, & -\frac{\beta_2 \beta_3 x_i^{\beta_4} \log(x_i)}{(1 + \beta_3 x_i^{\beta_4})^2} \end{pmatrix} \quad \text{e} \quad \mathbf{V}_i = 1.$$

A função $g(u)$ no caso da distribuição t de Student univariada é dada por $g(u) = c(1 + u/\nu)^{-(\nu+1)/2}$. Notemos que a condição C9 é satisfeita. Conforme já exemplificamos no Capítulo 1, $W_g(u)$ e $W_g^2(u)$ são dados por

$$W_g(u) = -\frac{1}{2} \left(\frac{\nu+1}{\nu+u} \right) \quad \text{e} \quad W_g^2(u) = \frac{1}{2} \left(\frac{\sqrt{\nu+1}}{\nu+u} \right)^2.$$

Assim, a condição C11 é satisfeita. Ainda, da Tabela 3.1 temos que a condição C10 é satisfeita.

As funções $\boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{x}_i)$ e $h_i(\boldsymbol{\gamma}, \mathbf{w}_i)$ especificadas em (3.5) são funções contínuas em $\boldsymbol{\theta}$, duas vezes di-

ferenciáveis e limitadas, então a condição C2 é satisfeita. Para analisar a condição C3, consideremos a matriz $\mathbf{F}_\theta$ dada por

$$\mathbf{F}_\theta = \begin{pmatrix} 1 & \frac{1}{1+\beta_3 x_1^{\beta_4}} & -\frac{\beta_2 x_1^{\beta_4}}{(1+\beta_3 x_1^{\beta_4})^2} & -\frac{\beta_2 \beta_3 x_1^{\beta_4} \log(x_1)}{(1+\beta_3 x_1^{\beta_4})^2} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & \frac{1}{1+\beta_3 x_2^{\beta_4}} & -\frac{\beta_2 x_2^{\beta_4}}{(1+\beta_3 x_2^{\beta_4})^2} & -\frac{\beta_2 \beta_3 x_2^{\beta_4} \log(x_2)}{(1+\beta_3 x_2^{\beta_4})^2} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{1}{1+\beta_3 x_n^{\beta_4}} & -\frac{\beta_2 x_n^{\beta_4}}{(1+\beta_3 x_n^{\beta_4})^2} & -\frac{\beta_2 \beta_3 x_n^{\beta_4} \log(x_n)}{(1+\beta_3 x_n^{\beta_4})^2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Notemos que duas condições necessárias sobre o vetor das covariáveis são tais que $x_i > 0, \forall i$ e $\exists x_i : x_i \neq \bar{x}$ para que a matriz $\mathbf{F}_\theta$ não tenha colunas linearmente dependentes. Vamos supor que o posto coluna da matriz $\mathbf{F}_\theta$ é $p = 5$. Então a condição C3 é satisfeita. Em situações práticas, dadas as covariáveis, é possível verificar em algum *software* como o R, por exemplo, a validade dessa condição. Satisfeitas as condições C1, C2 e C3, então C6 também é satisfeita. Notemos ainda que a função $\mathbf{G}$ é injetiva, então C7 é satisfeita.

Para analisarmos a condição C4, faz-se necessária a análise da matriz de informação de Fisher, a qual é dada por

$$\mathbf{K}_\theta = \begin{pmatrix} \mathbf{K}_{\beta_1 \beta_1} & \mathbf{K}_{\beta_1 \beta_2} & \mathbf{K}_{\beta_1 \beta_3} & \mathbf{K}_{\beta_1 \beta_4} & 0 \\ \mathbf{K}_{\beta_2 \beta_1} & \mathbf{K}_{\beta_2 \beta_2} & \mathbf{K}_{\beta_2 \beta_3} & \mathbf{K}_{\beta_2 \beta_4} & 0 \\ \mathbf{K}_{\beta_3 \beta_1} & \mathbf{K}_{\beta_3 \beta_2} & \mathbf{K}_{\beta_3 \beta_3} & \mathbf{K}_{\beta_3 \beta_4} & 0 \\ \mathbf{K}_{\beta_4 \beta_1} & \mathbf{K}_{\beta_4 \beta_2} & \mathbf{K}_{\beta_4 \beta_3} & \mathbf{K}_{\beta_4 \beta_4} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{K}_{\sigma^2 \sigma^2} \end{pmatrix},$$

em que

$$\begin{aligned} \mathbf{K}_{\beta_1 \beta_1} &= \frac{(1+\nu)n}{(3+\nu)\sigma^2}, \quad \mathbf{K}_{\beta_1 \beta_2} = \mathbf{K}_{\beta_2 \beta_1} = \frac{(1+\nu)}{(3+\nu)\sigma^2} \sum_{i=1}^n \frac{1}{1+\beta_3 x_i^{\beta_4}}, \\ \mathbf{K}_{\beta_1 \beta_3} &= \mathbf{K}_{\beta_3 \beta_1} = -\frac{(1+\nu)\beta_2}{(3+\nu)\sigma^2} \sum_{i=1}^n \frac{x_i^{\beta_4}}{(1+\beta_3 x_i^{\beta_4})^2}, \quad \mathbf{K}_{\beta_2 \beta_2} = \frac{(1+\nu)}{(3+\nu)\sigma^2} \sum_{i=1}^n \frac{1}{(1+\beta_3 x_i^{\beta_4})^2}, \\ \mathbf{K}_{\beta_2 \beta_3} &= \mathbf{K}_{\beta_3 \beta_2} = -\frac{(1+\nu)\beta_2}{(3+\nu)\sigma^2} \sum_{i=1}^n \frac{x_i^{\beta_4}}{(1+\beta_3 x_i^{\beta_4})^3}, \quad \mathbf{K}_{\beta_3 \beta_3} = \frac{(1+\nu)\beta_2^2}{(3+\nu)\sigma^2} \sum_{i=1}^n \frac{x_i^{2\beta_4}}{(1+\beta_3 x_i^{\beta_4})^4}, \\ \mathbf{K}_{\beta_1 \beta_4} &= \mathbf{K}_{\beta_4 \beta_1} = -\frac{(1+\nu)\beta_2 \beta_3}{(3+\nu)\sigma^2} \sum_{i=1}^n \frac{x_i^{\beta_4} \log(x_i)}{(1+\beta_3 x_i^{\beta_4})^2}, \quad \mathbf{K}_{\beta_4 \beta_4} = \frac{(1+\nu)\beta_2^2 \beta_3^2}{(3+\nu)\sigma^2} \sum_{i=1}^n \frac{x_i^{2\beta_4} \log(x_i)^2}{(1+\beta_3 x_i^{\beta_4})^4}, \\ \mathbf{K}_{\beta_2 \beta_4} &= \mathbf{K}_{\beta_4 \beta_2} = -\frac{(1+\nu)\beta_2 \beta_3}{(3+\nu)\sigma^2} \sum_{i=1}^n \frac{x_i^{\beta_4} \log(x_i)}{(1+\beta_3 x_i^{\beta_4})^3}, \\ \mathbf{K}_{\beta_3 \beta_4} &= \mathbf{K}_{\beta_4 \beta_3} = \frac{(1+\nu)\beta_2^2 \beta_3}{(3+\nu)\sigma^2} \sum_{i=1}^n \frac{x_i^{2\beta_4} \log(x_i)}{(1+\beta_3 x_i^{\beta_4})^4} \quad \text{e} \quad \mathbf{K}_{\sigma^2 \sigma^2} = \frac{(2+\nu)n}{(3+\nu)2\sigma^4}. \end{aligned}$$

Notemos que a condição C4 é satisfeita, desde que a sequência de covariáveis $\{x_i\}_{i=1}^\infty$ seja limitada. Em situações em que as covariáveis crescem ou decrescem em i , por exemplo, eventualmente a condição C4 pode não ser satisfeita. Em simulações, em geral, é utilizada a distribuição uniforme, a qual é limitada, satisfazendo a condição C4.

Para verificarmos se a condição C5 é satisfeita é necessário analisar a matriz de informação observada, a qual é dada por $\mathbf{J}_\theta = -\mathbf{L}_\theta$, em que

$$\mathbf{L}_\theta = \begin{pmatrix} \mathbf{L}_{\beta_1\beta_1} & \mathbf{L}_{\beta_1\beta_2} & \mathbf{L}_{\beta_1\beta_3} & \mathbf{L}_{\beta_1\beta_4} & \mathbf{L}_{\beta_1\sigma^2} \\ \mathbf{L}_{\beta_2\beta_1} & \mathbf{L}_{\beta_2\beta_2} & \mathbf{L}_{\beta_2\beta_3} & \mathbf{L}_{\beta_2\beta_4} & \mathbf{L}_{\beta_2\sigma^2} \\ \mathbf{L}_{\beta_3\beta_1} & \mathbf{L}_{\beta_3\beta_2} & \mathbf{L}_{\beta_3\beta_3} & \mathbf{L}_{\beta_3\beta_4} & \mathbf{L}_{\beta_3\sigma^2} \\ \mathbf{L}_{\beta_4\beta_1} & \mathbf{L}_{\beta_4\beta_2} & \mathbf{L}_{\beta_4\beta_3} & \mathbf{L}_{\beta_4\beta_4} & \mathbf{L}_{\beta_4\sigma^2} \\ \mathbf{L}_{\sigma^2\beta_1} & \mathbf{L}_{\sigma^2\beta_2} & \mathbf{L}_{\sigma^2\beta_3} & \mathbf{L}_{\sigma^2\beta_4} & \mathbf{L}_{\sigma^2\sigma^2} \end{pmatrix},$$

em que

$$\mathbf{L}_{\beta_1\beta_1} = \frac{(1+\nu)}{\sigma^4} \sum_{i=1}^n \left[\frac{2z_i^2}{(\nu+u_i)^2} - \frac{\sigma^2}{\nu+u_i} \right],$$

$$\mathbf{L}_{\beta_2\beta_1} = \mathbf{L}_{\beta_1\beta_2} = \frac{(1+\nu)}{\sigma^4} \sum_{i=1}^n \frac{1}{1+\beta_3x_i^{\beta_4}} \left[\frac{2z_i^2}{(\nu+u_i)^2} - \frac{\sigma^2}{\nu+u_i} \right],$$

$$\mathbf{L}_{\beta_2\beta_2} = \frac{(1+\nu)}{\sigma^4} \sum_{i=1}^n \frac{1}{(1+\beta_3x_i^{\beta_4})^2} \left[\frac{2z_i^2}{(\nu+u_i)^2} - \frac{\sigma^2}{\nu+u_i} \right],$$

$$\mathbf{L}_{\beta_1\beta_3} = \mathbf{L}_{\beta_3\beta_1} = \frac{\beta_2(1+\nu)}{\sigma^4} \sum_{i=1}^n \frac{x_i^{\beta_4}}{(1+\beta_3x_i^{\beta_4})^2} \left[-\frac{2z_i^2}{(\nu+u_i)^2} + \frac{\sigma^2}{\nu+u_i} \right],$$

$$\mathbf{L}_{\beta_1\beta_4} = \mathbf{L}_{\beta_4\beta_1} = \frac{\beta_2\beta_3(1+\nu)}{\sigma^4} \sum_{i=1}^n \frac{x_i^{\beta_4} \log(x_i)}{(1+\beta_3x_i^{\beta_4})^2} \left[-\frac{2z_i^2}{(\nu+u_i)^2} + \frac{\sigma^2}{\nu+u_i} \right],$$

$$\mathbf{L}_{\beta_2\beta_3} = \mathbf{L}_{\beta_3\beta_2} = \frac{\beta_2(1+\nu)}{\sigma^4} \sum_{i=1}^n \frac{x_i^{\beta_4}}{(1+\beta_3x_i^{\beta_4})^3} \left[-\frac{2z_i^2}{(\nu+u_i)^2} + \frac{\sigma^2}{\nu+u_i} - \frac{z_i\sigma^2(1+\beta_3x_i^{\beta_4})}{\beta_2(\nu+u_i)} \right],$$

$$\mathbf{L}_{\beta_3\beta_3} = \frac{\beta_2^2(1+\nu)}{\sigma^4} \sum_{i=1}^n \frac{x_i^{2\beta_4}}{(1+\beta_3x_i^{\beta_4})^4} \left[\frac{2z_i^2}{(\nu+u_i)^2} - \frac{\sigma^2}{\nu+u_i} + \frac{2z_i\sigma^2(1+\beta_3x_i^{\beta_4})}{\beta_2(\nu+u_i)} \right],$$

$$\mathbf{L}_{\beta_2\beta_4} = \mathbf{L}_{\beta_4\beta_2} = \frac{\beta_2\beta_3(1+\nu)}{\sigma^4} \sum_{i=1}^n \frac{x_i^{\beta_4} \log(x_i)}{(1+\beta_3x_i^{\beta_4})^3} \left[-\frac{2z_i^2}{(\nu+u_i)^2} + \frac{\sigma^2}{\nu+u_i} - \frac{z_i\sigma^2(1+\beta_3x_i^{\beta_4})}{\beta_2(\nu+u_i)} \right],$$

$$\mathbf{L}_{\beta_3\beta_4} = \frac{\beta_2^2\beta_3(1+\nu)}{\sigma^4} \sum_{i=1}^n \frac{x_i^{2\beta_4} \log(x_i)}{(1+\beta_3x_i^{\beta_4})^4} \left[\frac{2z_i^2}{(\nu+u_i)^2} - \frac{\sigma^2}{\nu+u_i} + \frac{z_i\sigma^2(1+\beta_3x_i^{\beta_4})(\beta_3x_i^{\beta_4}-1)}{\beta_2\beta_3x_i^{\beta_4}(\nu+u_i)} \right],$$

$$\begin{aligned} \mathbf{L}_{\beta_4\beta_4} &= \frac{\beta_2^2\beta_3^2(1+\nu)}{\sigma^4} \sum_{i=1}^n \frac{x_i^{2\beta_4} \log(x_i)^2}{(1+\beta_3x_i^{\beta_4})^4} \left[\frac{2z_i^2}{(\nu+u_i)^2} - \frac{\sigma^2}{\nu+u_i} + \frac{z_i\sigma^2(1+\beta_3x_i^{\beta_4})(\beta_3x_i^{\beta_4}-1)}{\beta_2\beta_3x_i^{\beta_4}(\nu+u_i)} \right], \\ \mathbf{L}_{\beta_4\beta_3} &= \mathbf{L}_{\beta_3\beta_4}, \quad \mathbf{L}_{\beta_1\sigma^2} = \mathbf{L}_{\sigma^2\beta_1} = \frac{(1+\nu)}{\sigma^6} \sum_{i=1}^n \left[\frac{z_i^3}{(\nu+u_i)^2} - \frac{z_i\sigma^2}{\nu+u_i} \right], \\ \mathbf{L}_{\beta_2\sigma^2} &= \mathbf{L}_{\sigma^2\beta_2} = \frac{(1+\nu)}{\sigma^6} \sum_{i=1}^n \frac{1}{1+\beta_3x_i^{\beta_4}} \left[\frac{z_i^3}{(\nu+u_i)^2} - \frac{z_i\sigma^2}{\nu+u_i} \right], \\ \mathbf{L}_{\beta_3\sigma^2} &= \mathbf{L}_{\sigma^2\beta_3} = \frac{(1+\nu)\beta_2}{\sigma^6} \sum_{i=1}^n \frac{x_i^{\beta_4}}{(1+\beta_3x_i^{\beta_4})^2} \left[-\frac{z_i^3}{(\nu+u_i)^2} + \frac{z_i\sigma^2}{\nu+u_i} \right], \\ \mathbf{L}_{\beta_4\sigma^2} &= \mathbf{L}_{\sigma^2\beta_4} = \frac{(1+\nu)\beta_2\beta_3}{\sigma^6} \sum_{i=1}^n \frac{x_i^{\beta_4} \log(x_i)}{(1+\beta_3x_i^{\beta_4})^2} \left[-\frac{z_i^3}{(\nu+u_i)^2} + \frac{z_i\sigma^2}{\nu+u_i} \right] \\ \text{e } \mathbf{L}_{\sigma^2\sigma^2} &= \frac{(1+\nu)}{2\sigma^8} \sum_{i=1}^n \left[\frac{z_i^4}{(\nu+u_i)^2} - \frac{2z_i\sigma^2}{\nu+u_i} - \frac{\sigma^4}{\nu+u_i} + \frac{\sigma^2(\nu+u_i)}{1+\nu} \right]. \end{aligned}$$

Ainda, obtivemos os valores da matriz hessiana por meio da derivada numérica da função de log-verossimilhança. Utilizamos o *software* R por meio da função `optim`, utilizando o algoritmo BFGS (Press *et al.*, 1992). Os resultados numéricos e os resultados analíticos apresentaram valores iguais ou muito próximos. O fato de nem todos os elementos da matriz fornecerem resultados iguais se dá em razão de que os resultados baseados na derivada numérica são obtidos por meio de aproximações.

Considerando a matriz de informação observada, analisando cada termo, percebemos que assim como para a condição C4 ser satisfeita, a sequência de covariáveis deve ser limitada. Então existirá uma constante c tal que C5 é satisfeita.

A condição C8 é verificada mediante a existência da terceira derivada, sendo contínua e limitada, o que ocorre quando consideramos a função de log-verossimilhança da distribuição t-Student e as funções definidas em 3.5. Logo, a condição C8 é satisfeita.

Assim, temos que as condições dos Teoremas 3.2.1, 3.2.4, 3.2.5 e 3.2.6 são satisfeitas. Então, assintoticamente temos a existência, consistência e normalidade do EMV. Ainda, as estatísticas de teste apresentam distribuição assintótica qui-quadrado com $p = 5$ graus de liberdade.

Para mostrar que o resultado do Teorema 3.2.2, o qual aborda a unicidade do EMV com probabilidade um, é válido, temos que mostrar que o conjunto R tem medida nula. Consideremos a matriz $\mathbf{L}_\theta$, as covariáveis são tais que $x_i > 0, \forall i$ e $\exists x_i : x_i \neq \bar{x}$, assumamos que a matriz $\mathbf{L}_\theta$ tem posto $p = 5$, então é positiva definida e assim o conjunto R tem medida nula. Logo o EMV de θ no modelo com 3.5 é único com probabilidade um.

3.4 Conclusões

Neste capítulo estudamos as condições de regularidade para inferências via máxima verossimilhança no modelo de regressão elíptico com parametrização geral. Na Seção 3.1, apresentamos o modelo e quantidades importantes para a realização de inferências. Destacamos que essas quantidades foram introduzidas em Lemonte e Patriota (2011), sendo algumas quantidades adicionais apresentadas em Melo *et al.* (2017). Na Seção 3.2, enunciamos as condições de regularidade e os teoremas para a existência, unicidade, consistência e normalidade assintótica do EMV. Além de um teorema que garante a distribuição assintótica das estatísticas de teste da razão de verossimilhanças, Wald, escore e gradiente. Em relação às condições apresentadas no Capítulo 2 se fez necessária

apenas a inclusão de algumas condições sobre a função geradora de densidades g . Na Seção 3.3, apresentamos um exemplo no qual estudamos as condições de regularidade em um modelo específico. Consideramos o modelo não linear homocedástico discutido em Tiede e Pagano (1979), estudado novamente em Lemonte e Patriota (2011). A função geradora de densidade foi a da distribuição t de Student. Especificamos a matriz de informação de Fisher e a matriz de informação observada para então analisar as condições de regularidade. Após definirmos a função geradora de densidades e as funções $\boldsymbol{\mu}(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ obtemos algumas quantidades importantes. Essas quantidades estão associadas às condições de regularidade, nesse sentido podemos obter condições mais específicas do que as enunciadas, considerando a matriz de regressores. Destacamos ainda que outra função geradora de densidades poderia ter sido escolhida. Por meio da Tabela 3.1 a condição C10 pode ser analisada nos modelos normal, Cauchy, t de Student e exponencial potência. Notemos que nesses modelos as condições sobre a função geradora de densidades g são satisfeitas.

Capítulo 4

Conclusões

4.1 Considerações finais

Nesta tese, estudamos condições de regularidade para o modelo de regressão com parametrização geral. No Capítulo 2, estudamos as condições de regularidade para o modelo de regressão normal com parametrização geral sob as quais temos a existência, unicidade, consistência e normalidade assintótica dos EMV, além das condições sob as quais conhecemos a distribuição assintótica das estatísticas de teste. Enunciamos as condições de regularidade e formalizamos os teoremas que garantem as propriedades citadas. Ainda, apresentamos exemplos que ilustram a verificação das condições e os resultados dos teoremas em alguns casos particulares do modelo. No Capítulo 3, generalizamos os resultados obtidos para o modelo com parametrização geral no caso dos erros normais para o modelo que considera a classe de distribuição elíptica para os erros. Ademais, formalizamos os teoremas que garantem as propriedades assintóticas dos EMV e a distribuição assintótica das estatísticas de teste. Exemplificamos, ainda, a verificação das condições de regularidade em alguns modelos específicos.

Nesse sentido, vale destacar que o estudo sobre condições de regularidade realizado nesta tese servirá de suporte para estudos futuros em casos particulares do modelo de regressão elíptico com parametrização geral. Esperamos que este trabalho forneça ao leitor um suporte amplo e detalhado de condições de regularidade. Por fim, esperamos alavancar o desenvolvimento da teoria assintótica de forma detalhada em diferentes modelos estatísticos.

4.2 Sugestões para pesquisas futuras

- Especificar as condições de regularidade em modelos específicos existentes na literatura que não desenvolveram com detalhes a questão da teoria assintótica frequentista.
- Refinar as condições de regularidade apresentadas, no sentido de obtenção de condições mais específicas, sobre a matriz de regressores, por exemplo.
- Obter condições alternativas para verificação da unicidade do EMV também quando $\boldsymbol{\mu}(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ compartilham parâmetros.
- Avaliar, por meio de simulações, a violação de algumas condições de regularidade especificadas e sua influência sobre as propriedades assintóticas dos EMV e o tamanho dos testes, considerando as estatísticas apresentadas.

Apêndice A

Provas dos Teoremas do Capítulo 2

A.1 Prova do Teorema 2.2.1 - Existência assintótica do EMV

Apresentamos a seguir um lema e uma proposição que serão utilizados na prova do Teorema 2.2.1.

Lema A.1.1. *Sejam $\mathbf{x}$ e $\mathbf{y}$ dois vetores $p \times 1$ tais que $\mathbf{y}^\top \mathbf{y} = 1$ e a função $f(\mathbf{y}) = \mathbf{y}^\top \mathbf{x}$. Então, $\max_{\mathbf{y}^\top \mathbf{y}=1} f(\mathbf{y}) = \|\mathbf{x}\|$.*

Demonstração. Notemos que maximizar $f(\mathbf{y}) = \mathbf{y}^\top \mathbf{x}$ restrito a condição $\mathbf{y}^\top \mathbf{y} = 1$ é equivalente a maximizar $f(\mathbf{y}) = \mathbf{y}^\top \mathbf{x} + c(1 - \mathbf{y}^\top \mathbf{y})$, em que c é uma constante. O ponto crítico é dado em $\mathbf{y} = \mathbf{x}/2c$ em que $c = \|\mathbf{x}\|/2$. Assim, $\mathbf{y} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$. Então, $\max_{\mathbf{y}^\top \mathbf{y}=1} f(\mathbf{y}) = f\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) = \|\mathbf{x}\|$. $\square$

Proposição A.1.1. *(Rayleigh-Ritz de Horn e Johnson (2013), pag.176) Seja $\boldsymbol{\lambda} \neq \mathbf{0}$ um vetor, $\mathbf{A}$ uma matriz hermitiana $p \times p$ e os autovalores de $\mathbf{A}$ ordenados $ch_1(\mathbf{A}) \geq \dots \geq ch_p(\mathbf{A})$. Então*

$$ch_p(\mathbf{A}) = \min_{\boldsymbol{\lambda} \neq \mathbf{0}} \frac{\boldsymbol{\lambda}^\top \mathbf{A} \boldsymbol{\lambda}}{\boldsymbol{\lambda}^\top \boldsymbol{\lambda}} = \min_{\boldsymbol{\lambda} \boldsymbol{\lambda}^\top = 1} \boldsymbol{\lambda}^\top \mathbf{A} \boldsymbol{\lambda} \quad e \quad ch_1(\mathbf{A}) = \max_{\boldsymbol{\lambda} \neq \mathbf{0}} \frac{\boldsymbol{\lambda}^\top \mathbf{A} \boldsymbol{\lambda}}{\boldsymbol{\lambda}^\top \boldsymbol{\lambda}} = \max_{\boldsymbol{\lambda} \boldsymbol{\lambda}^\top = 1} \boldsymbol{\lambda}^\top \mathbf{A} \boldsymbol{\lambda}.$$

A prova pode ser encontrada em Horn e Johnson (2013), pag. 176. Adicionalmente, se $\mathbf{A}$ for uma matriz simétrica, a Proposição A.1.1 também é aplicável, pois uma matriz simétrica é uma matriz hermitiana.

Prova do Teorema 2.2.1

Demonstração. Para algum n e $\delta > 0$, o evento

$$\ell(\mathbf{y}, \boldsymbol{\theta}') - \ell(\mathbf{y}, \boldsymbol{\theta}) < 0, \quad \forall \quad \boldsymbol{\theta}' \in \partial B_{\boldsymbol{\theta}}(\delta),$$

em que $\partial B_{\boldsymbol{\theta}}(\delta)$ denota a borda de $B_{\boldsymbol{\theta}}(\delta)$, implica que existe um máximo local no interior de $B_{\boldsymbol{\theta}}(\delta)$. Portanto, para mostrar que existe um máximo, basta mostrar que para algum $\eta > 0$ fixado, existem $\delta > 0$ e n_1 tais que

$$P_{\boldsymbol{\theta}}\left(\ell(\mathbf{y}, \boldsymbol{\theta}') - \ell(\mathbf{y}, \boldsymbol{\theta}) < 0, \quad \forall \quad \boldsymbol{\theta}' \in \partial B_{\boldsymbol{\theta}}(\delta)\right) \geq 1 - \eta, \quad \text{para todo } n \geq n_0, \quad (\text{A.1})$$

para cada $\boldsymbol{\theta} \in \Theta$.

Na análise abaixo, nos restringimos ao caso em que $\boldsymbol{\theta}' \in \partial B_{\boldsymbol{\theta}}(\delta)$. Nesse contexto, expandindo em série de Taylor a função de log-verossimilhança $\ell(\mathbf{y}, \boldsymbol{\theta}')$ em torno de $\boldsymbol{\theta}$ e satisfeita a condição C1, temos

$$\ell(\mathbf{y}, \boldsymbol{\theta}') - \ell(\mathbf{y}, \boldsymbol{\theta}) = (\boldsymbol{\theta}' - \boldsymbol{\theta})^\top \mathbf{U}_{\boldsymbol{\theta}} - \frac{1}{2}(\boldsymbol{\theta}' - \boldsymbol{\theta})^\top \mathbf{J}_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}' - \boldsymbol{\theta}), \quad (\text{A.2})$$

com $\boldsymbol{\theta}^* = h\boldsymbol{\theta}' + (1-h)\boldsymbol{\theta}$ e $h \in (0, 1)$. Tomando $\boldsymbol{\lambda} = \mathbf{K}_{\boldsymbol{\theta}}^{1/2}(\boldsymbol{\theta}' - \boldsymbol{\theta})/\delta$, reescrevemos (A.2) e assim segue que

$$\ell(\mathbf{y}, \boldsymbol{\theta}') - \ell(\mathbf{y}, \boldsymbol{\theta}) = \delta \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}} - \frac{1}{2} \delta^2 \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}^*} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\lambda}. \quad (\text{A.3})$$

Notemos que de (A.3) obtemos algumas relações que serão úteis, explicitadas a seguir:

$$\begin{aligned} & \ell(\mathbf{y}, \boldsymbol{\theta}') - \ell(\mathbf{y}, \boldsymbol{\theta}) < 0, \\ \Leftrightarrow & \delta \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}} - \frac{1}{2} \delta^2 \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}^*} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\lambda} < 0, \\ \Leftrightarrow & \max_{\boldsymbol{\lambda} \neq 0} \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}} < \frac{1}{2} \delta \inf_{\boldsymbol{\lambda} \neq 0} \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}^*} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\lambda}. \end{aligned}$$

Notemos ainda que

$$P_{\boldsymbol{\theta}}\left(\ell(\mathbf{y}, \boldsymbol{\theta}') - \ell(\mathbf{y}, \boldsymbol{\theta}) < 0\right) \geq P_{\boldsymbol{\theta}}\left(\max_{\boldsymbol{\lambda} \neq 0} \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}} < \frac{1}{2} \delta \inf_{\boldsymbol{\lambda} \neq 0} \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}^*} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\lambda}\right).$$

Então, para mostrar (A.1), basta mostrar que

$$P_{\boldsymbol{\theta}}\left(\max_{\boldsymbol{\lambda} \neq 0} \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}} < \frac{1}{2} \delta \inf_{\boldsymbol{\lambda} \neq 0} \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}^*} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\lambda}\right) \geq 1 - \eta, \text{ para algum } n > n_1.$$

Como $\boldsymbol{\theta}' \in \partial B_{\boldsymbol{\theta}}(\delta)$, temos $\boldsymbol{\lambda}^\top \boldsymbol{\lambda} = 1$, pois $\boldsymbol{\lambda}^\top \boldsymbol{\lambda} = (\boldsymbol{\theta}' - \boldsymbol{\theta})^\top \mathbf{K}_{\boldsymbol{\theta}}^{1/2} \mathbf{K}_{\boldsymbol{\theta}}^{1/2} (\boldsymbol{\theta}' - \boldsymbol{\theta}) / \delta^2$ e $\mathbf{K}_{\boldsymbol{\theta}}^{1/2} (\boldsymbol{\theta}' - \boldsymbol{\theta}) = \delta$. Utilizando o resultado da Proposição A.1.1, segue que

$$\inf_{\boldsymbol{\lambda}^\top \boldsymbol{\lambda} = 1} \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}^*} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\lambda} = \text{ch}_p(\mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}^*} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2}).$$

Do Lema A.1.1 decorre que $\max_{\boldsymbol{\lambda}^\top \boldsymbol{\lambda} = 1} \boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}} = \|\mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}}\|$. Então, para mostrar (A.1) é suficiente mostrar que para algum $\eta > 0$ e $\delta > 0$,

$$P_{\boldsymbol{\theta}}\left(\|\mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}}\|^2 < \delta^2 \left(\text{ch}_p(\mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}^*} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2})\right)^2 / 4, \quad \forall \quad \boldsymbol{\theta}' \in \partial B_{\boldsymbol{\theta}}(\delta)\right) \geq 1 - \eta, \quad (\text{A.4})$$

para algum $n > n_1$ suficientemente grande.

Da condição C5 temos a afirmação equivalente (2.6). Da Proposição A.1.1 decorre que

$$\begin{aligned} & \lim_{n \rightarrow \infty} P_{\boldsymbol{\theta}}\left(\boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}'} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{\lambda} \geq c, \quad \boldsymbol{\theta}' \in \partial B_{\boldsymbol{\theta}}(\delta)\right) = 1 \\ \Leftrightarrow & \lim_{n \rightarrow \infty} P_{\boldsymbol{\theta}}\left(\text{ch}_p(\mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}'} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2}) \geq c, \quad \boldsymbol{\theta}' \in \partial B_{\boldsymbol{\theta}}(\delta)\right) = 1 \\ \Leftrightarrow & \lim_{n \rightarrow \infty} P_{\boldsymbol{\theta}}\left(\text{ch}_p(\mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{J}_{\boldsymbol{\theta}^*} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2}) \geq c, \quad \boldsymbol{\theta}^* \in B_{\boldsymbol{\theta}}(\delta)\right) = 1. \end{aligned}$$

Adicionalmente, notemos que

$$\mathbb{E}\left(\mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}} \mathbf{U}_{\boldsymbol{\theta}}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1/2}\right) = \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbb{E}\left(\mathbf{U}_{\boldsymbol{\theta}} \mathbf{U}_{\boldsymbol{\theta}}^\top\right) \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} = \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{K}_{\boldsymbol{\theta}} \mathbf{K}_{\boldsymbol{\theta}}^{-1/2} = \mathbf{I}_p.$$

Logo, resulta que $\mathbb{E}\left(\|\mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}}\|^2\right) = p$. Pela desigualdade de Markov, temos

$$P_{\boldsymbol{\theta}}\left(\|\mathbf{K}_{\boldsymbol{\theta}}^{-1/2} \mathbf{U}_{\boldsymbol{\theta}}\|^2 < (\delta c)^2 / 4\right) \geq 1 - \frac{4p}{(\delta c)^2} = 1 - \eta,$$

para $\delta^2 = 4p/c^2\eta$ e $n > n_0$ suficientemente grande, implica (A.4) e (A.1), como queríamos mostrar. $\square$

A.2 Prova do Teorema 2.2.2 - Unicidade do EMV caso geral

A seguir, enunciaremos e provamos um lema que demonstra que no modelo (2.1) uma das condições necessárias apresentadas em Krättschmer (2007) é satisfeita. Este é um resultado auxiliar e corresponde à condição C3.2 de Krättschmer (2007).

Lema A.2.1. *Sob o modelo (2.1), se a condição C7 for satisfeita, então o conjunto*

$$D_1 := \{\mathbf{y} \in \mathbb{R}^N : \mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}_1) = \mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}_2), \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta\}$$

tem medida de Lebesgue nula, em que $N = d_1 + \dots + d_n$ e

$$\mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}) = \frac{\partial \ell(\mathbf{y}, \boldsymbol{\theta})}{\partial \mathbf{y}} = -\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})).$$

Demonstração. Consideremos os três casos a seguir:

- (a) $\boldsymbol{\mu}(\boldsymbol{\theta}_1) \neq \boldsymbol{\mu}(\boldsymbol{\theta}_2)$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$,
- (b) $\boldsymbol{\mu}(\boldsymbol{\theta}_1) = \boldsymbol{\mu}(\boldsymbol{\theta}_2)$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) \neq \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$ e
- (c) $\boldsymbol{\mu}(\boldsymbol{\theta}_1) \neq \boldsymbol{\mu}(\boldsymbol{\theta}_2)$ e $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) \neq \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$.

Consideremos inicialmente (a). O conjunto $\{\mathbf{y} \in \mathbb{R}^N : \mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}_1) = \mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}_2), \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta\}$ tem medida de Lebesgue nula. Para os casos (b) e (c), notemos que

$$\begin{aligned} \{\mathbf{y} \in \mathbb{R}^N : \mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}_1) = \mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}_2), \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta\} = \\ \{\mathbf{y} \in \mathbb{R}^N : \mathbf{A}\mathbf{y} = \mathbf{d}, \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta\} = D_2, \end{aligned}$$

em que $\mathbf{A} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_1) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_2)$ e $\mathbf{d} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_1)\boldsymbol{\mu}(\boldsymbol{\theta}_1) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_2)\boldsymbol{\mu}(\boldsymbol{\theta}^*)$ com $\boldsymbol{\theta}^* = \boldsymbol{\theta}_1$, sob o caso (b), e $\boldsymbol{\theta}^* = \boldsymbol{\theta}_2$, sob o caso (c). Quando $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) \neq \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$, $\mathbf{A}$ tem pelo menos um autovalor diferente de zero e então o conjunto D_2 tem medida de Lebesgue nula e D_1 também. $\square$

Prova do Teorema 2.2.2

Demonstração. Para provar este teorema, temos que mostrar que as condições suficientes do Teorema 3.1 apresentado em Krättschmer (2007) são satisfeitas. Notemos que se $\mathbf{G}$ for injetiva, então $\boldsymbol{\theta}$ é identificável no modelo (2.1). Ainda pelo Lema A.2.1 decorre a condição C3.2 de Krättschmer (2007). Satisfeita a condição C6, $\mathbf{G}$ é uma função duas vezes diferenciável e assim $\ell(\cdot, \cdot)$ também é. Agora, notemos que se o conjunto $R = \{\mathbf{y} \in \mathbb{R}^N : \mathbf{U}_{\hat{\boldsymbol{\theta}}_n(\mathbf{y})} = \mathbf{0}\} \cap \{\mathbf{y} \in \mathbb{R}^N : \det(\mathbf{L}_{\hat{\boldsymbol{\theta}}_n(\mathbf{y})}) = 0\}$ tiver medida de Lebesgue nula, implica que $\ell(\cdot, \cdot)$ tem ponto crítico não degenerado com probabilidade um. Assim, todas as condições suficientes do Teorema 3.1 estabelecidas em Krättschmer (2007) são satisfeitas e concluímos que Q é mensurável e tem medida de Lebesgue nula. $\square$

A.3 Prova do Teorema 2.2.3 - Unicidade do EMV caso em que a média e a matriz de covariâncias não compartilham parâmetros

Para a prova desse teorema, definamos algumas quantidades que serão necessárias. O modelo que denominamos modelo estruturado (no sentido de uma estrutura de regressão na média e na matriz de covariâncias) é definido por

$$\mathbf{Y} \sim \mathcal{N}_N(\boldsymbol{\mu}(\boldsymbol{\beta}), \boldsymbol{\Sigma}(\boldsymbol{\gamma})), \quad (\text{A.5})$$

com $\boldsymbol{\mu}(\boldsymbol{\beta}) = (\boldsymbol{\mu}_1(\boldsymbol{\beta})^\top, \dots, \boldsymbol{\mu}_n(\boldsymbol{\beta})^\top)^\top$ e $\boldsymbol{\Sigma}(\boldsymbol{\gamma}) = \text{bdiag}\{\boldsymbol{\Sigma}_1(\boldsymbol{\gamma}), \dots, \boldsymbol{\Sigma}_n(\boldsymbol{\gamma})\}$.

O modelo não estruturado na média é definido por

$$\mathbf{W} \sim \mathcal{N}_N(\boldsymbol{\nu}, \boldsymbol{\Sigma}(\boldsymbol{\gamma})) \quad (\text{A.6})$$

com $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N)^\top$. A função de log-verossimilhança associada ao modelo não estruturado na média, a menos de uma constante, é dada por

$$\ell_0(\mathbf{y}; \boldsymbol{\nu}, \boldsymbol{\gamma}) := -\frac{1}{2} \log \det(\boldsymbol{\Sigma}(\boldsymbol{\gamma})) - \frac{1}{2}(\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\gamma})(\mathbf{y} - \boldsymbol{\nu}).$$

A condição que é utilizada como alternativa à condição de que o conjunto R tem medida nula é descrita em Krätschmer (2007) como condição C4.2, aqui apresentada a seguir em forma de lema.

Lema A.3.1. *Sob as condições C2 e C3, considere o subconjunto aberto H de $\mathbb{R}^N \times \mathbb{R}^N$, definido como*

$$H := \{(\mathbf{y}, \boldsymbol{\nu}) \in \mathbb{R}^N \times \mathbb{R}^N : |\mathbf{y} - \boldsymbol{\nu}| \neq 0\}.$$

Seja $\ell^ : H \rightarrow \mathbb{R}$ uma função tal que*

$$\ell^*(\mathbf{y}; \boldsymbol{\nu}) := -\frac{1}{2} \log \det(\boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathbf{y}-\boldsymbol{\nu}})) - \frac{1}{2}(\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{y}-\boldsymbol{\nu}})(\mathbf{y} - \boldsymbol{\nu}),$$

em que $\hat{\boldsymbol{\gamma}}_{\mathbf{y}-\boldsymbol{\nu}}$ é o EMV de $\boldsymbol{\gamma}$ para cada $\boldsymbol{\nu}$ fixado no modelo (A.6), ou seja, deve satisfazer a equação

$$\mathbf{V}_{\mathbf{x}}^\top \tilde{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}}) \text{vec}(\boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}})) = \mathbf{V}_{\mathbf{x}}^\top \tilde{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}}) \text{vec}(\mathbf{x}\mathbf{x}^\top), \quad (\text{A.7})$$

em que $\mathbf{x} = \mathbf{y} - \boldsymbol{\nu}$ e $\mathbf{V}_{\mathbf{x}} = \left(\text{vec} \left(\frac{\partial \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}})}{\partial x_1} \right), \dots, \text{vec} \left(\frac{\partial \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathbf{x}})}{\partial x_N} \right) \right)$. Adicionalmente, considere a função $\boldsymbol{\mu} : \mathbb{R}^k \rightarrow \mathbb{R}^N$. Então,

- (a) *$\boldsymbol{\mu}$ é uma função contínua duas vezes diferenciável em relação ao vetor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^\top$ com posto k .*
- (b) *Se para $\mathbf{y} \in \mathbb{R}^N$, a função $\ell_0(\mathbf{y}; \cdot, \boldsymbol{\gamma}) \circ \boldsymbol{\mu}$ tiver pontos críticos degenerados, então a função $\ell^*(\mathbf{y}; \cdot) \circ \boldsymbol{\mu}$ para todo $\boldsymbol{\beta} \in \boldsymbol{\mu}^{-1}(H_{\mathbf{y}})$ tem ponto crítico degenerado também, em que $H_{\mathbf{y}} := \{\boldsymbol{\nu} \in \mathbb{R}^N : (\mathbf{y}, \boldsymbol{\nu}) \in H\}$.*

Demonstração. Satisfeitas as condições C2 e C3, então temos que $\boldsymbol{\mu}$ é uma função contínua e (a) é satisfeita. Notemos que $\ell^*(\mathbf{y}; \boldsymbol{\nu})$ é a função de log-verossimilhança perfilada para $\boldsymbol{\nu}$ do modelo (A.6) a menos de uma constante. Considere agora a função $\boldsymbol{\mu}$ satisfazendo C2 temos que $\ell_0(\mathbf{y}; \cdot, \boldsymbol{\gamma}) \circ \boldsymbol{\mu}$ corresponde a função de log-verossimilhança e $\ell^*(\mathbf{y}; \cdot) \circ \boldsymbol{\mu}$ para todo $\boldsymbol{\beta} \in \boldsymbol{\mu}^{-1}(H_{\mathbf{y}})$ corresponde à função de log-verossimilhança perfilada, ambas associadas ao modelo (A.5). Note que uma das propriedades da função de log-verossimilhança é que os pontos de máximo da função de log-verossimilhança e da função de log-verossimilhança perfilada coincidem. Logo, temos que se $\ell_0(\mathbf{y}; \cdot, \boldsymbol{\gamma}) \circ \boldsymbol{\mu}$ tiver um ponto crítico degenerado, então $\ell^*(\mathbf{y}; \cdot) \circ \boldsymbol{\mu}$ para todo $\boldsymbol{\beta} \in \boldsymbol{\mu}^{-1}(H_{\mathbf{y}})$ também terá. $\square$

Se $\boldsymbol{\mu}$ não for contínua, não é possível obter o vetor escore e a matriz de informação de Fisher, por exemplo. No entanto, o resultado do Lema A.3.1 continua valendo desde que exista uma sequência $(k_j)_j$ de inteiros positivos $k_j \leq N$, uma sequência $(A^j)_j$ de subconjuntos abertos, em que $A^j \subset \mathbb{R}^{k_j}$ e uma sequência $\boldsymbol{\mu}^j(\boldsymbol{\Theta}_1)_j$ de funções $\boldsymbol{\mu}^j(\boldsymbol{\Theta}_1) : A^j \rightarrow \mathbb{R}^N$, em que $\boldsymbol{\Theta}_1$ é o espaço paramétrico correspondente à função $\boldsymbol{\mu}$, tais que cada $\boldsymbol{\mu}^j(\boldsymbol{\Theta}_1)$ é duas vezes diferenciável e o posto é igual à dimensão do conjunto A^j e $\boldsymbol{\mu}(\boldsymbol{\Theta}_1) = \bigcup_{j=1}^{\infty} \boldsymbol{\mu}^j(\boldsymbol{\Theta}_1)$.

Prova do Teorema 2.2.3

Demonstração. Para provar este teorema, basta mostrar que as condições do Teorema 4.3 apresentado em Krätschmer (2007) são satisfeitas. Se $\mathbf{G}$ for injetiva, então $\boldsymbol{\theta}$ é identificável no modelo (2.1).

Como $\mathbf{G}$ pela condição C2 é uma função duas vezes diferenciável, então $\ell(\cdot, \cdot)$ também é. A condição C4.2 de Krätschmer (2007) é provada no Lema A.3.1 mediante as condições C2 e C3. A condição C3.2 é provada no Lema A.2.1. Adicionalmente temos que mostrar que existe um difeomorfismo $f : U \rightarrow \mathbb{R}^N$ em que $U \subset \mathbb{R}^N$ é um aberto, tal que

$$\mathbf{y} - \boldsymbol{\nu} \in U \quad \text{e} \quad \frac{\partial \ell^*(\mathbf{y}; \cdot)}{\partial \boldsymbol{\nu}} = f(\mathbf{y} - \boldsymbol{\nu}),$$

para $(\mathbf{y}, \boldsymbol{\nu}) \in H$ arbitrário. Notemos que

$$\begin{aligned} \frac{\partial \ell^*(\mathbf{y}; \cdot)}{\partial \nu_j} &= -\frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}}) \frac{\partial \boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}})}{\partial \nu_j} \right\} + (\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}}) \frac{d\boldsymbol{\nu}}{d\nu_j} \\ &\quad + \frac{1}{2} (\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}}) \frac{\partial \boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}})}{\partial \nu_j} \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}}) (\mathbf{y} - \boldsymbol{\nu}). \end{aligned}$$

Tomando $\mathbf{x} = \mathbf{y} - \boldsymbol{\nu}$, podemos escrever uma função de $\mathbf{x}$ dada por

$$f_j(\mathbf{x}) = -\frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \frac{\partial \boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{x}})}{\partial x_j} \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \left(\boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{x}}) - \mathbf{x} \mathbf{x}^\top \right) \right\} + \text{tr} \left\{ \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \frac{d\mathbf{x}}{dx_j} \mathbf{x}^\top \right\}.$$

Dessa forma, definamos $U := \{\mathbf{x} \in \mathbb{R}^N : |\mathbf{x}| \neq 0\}$ e a função $f : U \rightarrow \mathbb{R}^N$ baseada na derivada do vetor $\boldsymbol{\nu}$ pode ser escrita em termos de uma função de $\mathbf{x}$ em forma matricial, conforme segue,

$$f(\mathbf{x}) = \mathbf{D}_{\mathbf{x}} \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x} - \frac{1}{2} \mathbf{V}_{\mathbf{x}}^\top \tilde{\boldsymbol{\Sigma}}^{-1}(\hat{\gamma}_{\mathbf{x}}) \text{vec} \left(\boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{x}}) - \mathbf{x} \mathbf{x}^\top \right), \quad (\text{A.8})$$

em que $\mathbf{D}_{\mathbf{x}}$ é a matriz identidade. Pela condição (A.7), a equação (A.8) é simplificada e pode ser reescrita como

$$f(\mathbf{x}) = \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x}. \quad (\text{A.9})$$

Para $(\mathbf{y}, \boldsymbol{\nu}) \in H$ arbitrário,

$$\mathbf{y} - \boldsymbol{\nu} \in U, \quad \frac{\partial \ell^*(\mathbf{y}; \cdot)}{\partial \boldsymbol{\nu}} = f(\mathbf{y} - \boldsymbol{\nu}) \quad \text{e} \quad \frac{\partial \ell^*(\mathbf{y}; \cdot)}{\partial \mathbf{y}} = -f(\mathbf{y} - \boldsymbol{\nu}).$$

Como a função $f(\mathbf{x}) = \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x}$ é um difeomorfismo, por suposição temos que todas as condições necessárias do Teorema 4.3 estabelecidas em Krätschmer (2007) são satisfeitas e concluímos que Q é mensurável e tem medida de Lebesgue nula. $\square$

Prova do Corolário 2.2.2

Demonstração. Conforme Rudin (1976, Teorema 9.25, pag. 241), se a derivada de $f(\mathbf{x})$ for inversível para todo $\mathbf{x} \in U$, então $f(\mathbf{x})$ é um difeomorfismo. A derivada de $f(\mathbf{x})$ em relação a cada elemento j de $\mathbf{x}$ é dada por

$$\frac{\partial f(\mathbf{x})}{\partial x_j} = \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \left(\frac{\partial \mathbf{x}}{\partial x_j} - \frac{\partial \boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{x}})}{\partial x_j} \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x} \right).$$

A derivada de $f(\mathbf{x})$ em relação ao vetor $\mathbf{x}$ em forma matricial é dada por

$$\begin{aligned} \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^\top} &= \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_N} \right] = \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \left[\mathbf{D}_{\mathbf{x}} - \left(\frac{\partial \boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{x}})}{\partial x_1}, \dots, \frac{\partial \boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{x}})}{\partial x_N} \right) \mathbf{S}_{\mathbf{x}} \right] \\ &= \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) - \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{V}_{\mathbf{x}}' \mathbf{S}_{\mathbf{x}}. \end{aligned}$$

Utilizando o determinante, analisamos quando a derivada de $f(\mathbf{x})$ é inversível. Basta verificar

sob quais condições

$$\det \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^\top} \right) \neq 0.$$

Dessa forma,

$$\begin{aligned} \det \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^\top} \right) \neq 0 &\Leftrightarrow \det \left(\boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) - \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{V}_{\mathbf{x}}' \mathbf{S}_{\mathbf{x}} \right) \neq 0 \\ &\Leftrightarrow \det \left(\mathbf{I}_N + \mathbf{V}_{\mathbf{x}}'^{-1} \mathbf{S}_{\mathbf{x}} (\boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}))^{-1} \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \right) \det \left(\boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \right) \neq 0. \end{aligned}$$

Notemos que como $\boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}})$, por definição, admite inversa, então $(\boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}))^{-1} \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) = \mathbf{I}_N$ e $\det(\boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}})) \neq 0$. Assim,

$$\det \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^\top} \right) \neq 0 \Leftrightarrow \det \left(\mathbf{I}_N + \mathbf{V}_{\mathbf{x}}' \mathbf{S}_{\mathbf{x}} \right) \neq 0.$$

□

A.4 Prova do Teorema 2.2.4 - Consistência do EMV

Demonstração. Retomemos a prova do Teorema 2.2.1 do EMV, dada na Seção A.1. Notemos que (A.1) implica que para algum $\eta > 0$, existem $\delta > 0$ e n_1 tais que

$$P_{\boldsymbol{\theta}} \left(\left\| \mathbf{K}_{\boldsymbol{\theta}}^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \right\| \leq \delta \right) \geq 1 - \eta.$$

Por suposição, a condição C4 é satisfeita. Então, $\hat{\boldsymbol{\theta}}_n$ é um estimador consistente para $\boldsymbol{\theta}$. □

A.5 Prova do Teorema 2.2.5 - Normalidade assintótica do EMV

Para a prova desse teorema enunciaremos e provaremos três lemas que compõem a prova do teorema. Tomemos inicialmente $\boldsymbol{\lambda} \neq \mathbf{0}$. Adicionalmente, notemos que $E[(\mathbf{z}_i \mathbf{z}_i^\top \mathbf{z}_i \mathbf{z}_i^\top)] < \infty$, pois $\mathbf{z}_i$ pela formulação do modelo tem distribuição normal com média $\mathbf{0}$. Assim, os momentos de quarta ordem são finitos.

Lema A.5.1. *Se as condições C3 e C4 forem satisfeitas, então*

$$\frac{\mathbf{U}_{\boldsymbol{\theta}}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Lambda}_{\boldsymbol{\theta}}), \text{ para cada } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Demonstração. Utilizando o Teorema Central do Limite de Liapounov, temos

1. $E(\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}_i}) = \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i E(\mathbf{s}_i) = \mathbf{0}$.
2. $E[(\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}_i})^2] < \infty$. Pelas condições C3 e C4, temos

$$\begin{aligned} E[(\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}_i})^2] &= E(\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i \mathbf{s}_i^\top \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda}) \\ &= \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i E(\mathbf{s}_i \mathbf{s}_i^\top) \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \\ &= \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} < \infty. \end{aligned}$$

3. $E(\|\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}_i}\|^{2+\delta}) < \infty$. Pela condição C4 e pela desigualdade de Cauchy-Schwarz para matri-

zes (ver Magnus e Neudecker (2019), pag. 227-228), temos

$$\begin{aligned} E\left(\|\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}i}\|^{2+\delta}\right) &= E\left(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i\|^{2+\delta}\right) \\ &\leq E\left(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i\|^{2+\delta} \|\mathbf{s}_i\|^{2+\delta}\right) \\ &= \|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i\|^{2+\delta} E\left(\|\mathbf{s}_i\|^{2+\delta}\right) \\ &< \infty, \end{aligned}$$

em que $E\left(\|\mathbf{s}_i\|^{2+\delta}\right) < \infty$, já que os momentos de ordem quatro da variável aleatória $\mathbf{z}_i$ são finitos.

4. $\lim_{n \rightarrow \infty} \left(\sum_{i=1}^n \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \right)^{-\left(\frac{2+\delta}{2}\right)} \sum_{i=1}^n E\left(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i\|^{2+\delta}\right) = 0$, pois pela condição C4 e pelo provado no item 3, temos

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n E\left(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i\|^{2+\delta}\right)}{\left(\sum_{i=1}^n \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \right)^{\frac{2+\delta}{2}}} &= \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n E\left(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i\|^{2+\delta}\right)}{n^{\frac{2+\delta}{2}} \left(\boldsymbol{\lambda}^\top \frac{\sum_{i=1}^n \mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i}{n} \boldsymbol{\lambda} \right)^{\frac{2+\delta}{2}}} \\ &= \frac{1}{(\boldsymbol{\lambda}^\top \boldsymbol{\Lambda}_{\boldsymbol{\theta}} \boldsymbol{\lambda})^{\frac{2+\delta}{2}}} \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n E\left(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i\|^{2+\delta}\right)}{n^{\frac{2+\delta}{2}}} = 0. \end{aligned}$$

Então,

$$\frac{\boldsymbol{\lambda}^\top \sum_{i=1}^n \mathbf{U}_{\boldsymbol{\theta}i}}{\sqrt{\boldsymbol{\lambda}^\top \sum_{i=1}^n \mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

De forma alternativa, podemos escrever

$$\frac{\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \boldsymbol{\lambda}^\top \boldsymbol{\Lambda}_{\boldsymbol{\theta}} \boldsymbol{\lambda}\right),$$

e, pelo Teorema de Cramér-Wold,

$$\frac{\mathbf{U}_{\boldsymbol{\theta}}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}_p\left(\mathbf{0}, \boldsymbol{\Lambda}_{\boldsymbol{\theta}}\right), \text{ para cada } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

□

Lema A.5.2. *Se as condições C3 e C4 forem satisfeitas, então*

$$\frac{\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}}}{n} + \frac{\boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}}{n} \xrightarrow{P_{\boldsymbol{\theta}}} \mathbf{0}, \text{ para cada } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Demonstração. Aplicando a Lei Fraca dos Grandes Números de Markov, temos que

1. $E(\|\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i}\|) < \infty$, pois $E(\|\mathbf{s}_i\|) < \infty$ e, assim,

$$\begin{aligned} E(\|\mathbf{B}_i\|) &= E\left(\|\mathbf{r}_i^\top \otimes \mathbf{H}_i \mathbf{s}_i\|\right) \\ &\leq \|\mathbf{r}_i^\top \otimes \mathbf{H}_i\| E(\|\mathbf{s}_i\|) < \infty. \end{aligned}$$

Notando que o termo aleatório de maior ordem de $\ddot{\mathbf{M}}_i$ é $\mathbf{z}_i \mathbf{z}_i^\top$, temos que $E(\|\ddot{\mathbf{M}}_i\|) < \infty$. Adicionalmente, satisfeita a condição C4 temos que

$$\begin{aligned} E\left(\|\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i}\|\right) &= E\left(\|\boldsymbol{\lambda}^\top \mathbf{B}_i - \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i\|\right) \\ &\leq E\left(\|\boldsymbol{\lambda}^\top \mathbf{B}_i\|\right) + E\left(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i\|\right) \\ &\leq \|\boldsymbol{\lambda}^\top\| E(\|\mathbf{B}_i\|) + \|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i\| E(\|\ddot{\mathbf{M}}_i\|) \|\mathbf{H}_i \mathbf{F}_i\| < \infty. \end{aligned}$$

2. $E\left(\|\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i} - E(\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i})\|^{1+\delta}\right) < \infty$, para algum $\delta \in (0, 1]$. Temos

$$\begin{aligned} \varrho &= E\left(\|\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i} - E(\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i})\|^{1+\delta}\right) = E\left(\|\boldsymbol{\lambda}^\top (\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i + \mathbf{F}_i^\top \mathbf{H}_i \mathbf{F}_i)\|^{1+\delta}\right) \\ &= E\left(\|\boldsymbol{\lambda}^\top (\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i)\|^{1+\delta}\right), \end{aligned}$$

em que $\mathbf{Q}_i = \ddot{\mathbf{M}}_i - \mathbf{H}_i^{-1}$. Notemos que

$$\begin{aligned} \varrho &\leq E\left(\|\boldsymbol{\lambda}^\top (\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i)\|^2\right) \\ &= E\left(\text{tr}\left\{\left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i\right) \boldsymbol{\lambda} \boldsymbol{\lambda}^\top (\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i)\right\}\right) \\ &= E\left(\text{tr}\left\{\mathbf{B}_i^\top \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{B}_i\right\}\right) + E\left(\text{tr}\left\{\mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i\right\}\right) \\ &\quad - E\left(\text{tr}\left\{\mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{B}_i\right\}\right) - E\left(\text{tr}\left\{\mathbf{B}_i^\top \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i\right\}\right) \\ &< \infty. \end{aligned}$$

Analisando separadamente as esperanças da equação anterior, temos

$$E\left(\text{tr}\{\mathbf{B}_i^\top \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{B}_i\}\right) = E\left(\|\boldsymbol{\lambda}^\top \mathbf{B}_i\|^2\right) \leq \|\boldsymbol{\lambda}^\top\|^2 \|\mathbf{r}_i^\top \otimes \mathbf{H}_i\|^2 E(\|\mathbf{s}_i\|^2) < \infty,$$

pois em $E(\|\mathbf{s}_i\|^2)$ o maior termo aleatório é da ordem 4 e na distribuição normal esse momento existe e é finito. Logo, $E(\|\mathbf{s}_i\|^2) < \infty$. Em $E(\ddot{\mathbf{M}}_i \otimes \ddot{\mathbf{M}}_i^\top)$, $E(\ddot{\mathbf{M}}_i \otimes \mathbf{B}_i^\top)$ e $E(\ddot{\mathbf{M}}_i^\top \otimes \mathbf{B}_i^\top)$ também o maior termo aleatório é de ordem 4. Então, $E(\ddot{\mathbf{M}}_i \otimes \ddot{\mathbf{M}}_i^\top) < \infty$, $E(\ddot{\mathbf{M}}_i^\top \otimes \mathbf{B}_i^\top) < \infty$ e $E(\ddot{\mathbf{M}}_i \otimes \mathbf{B}_i^\top) < \infty$. Assim, temos os resultados

$$\begin{aligned} E\left(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda}\|^2\right) &= E\left(\text{tr}\{\mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i\}\right) \\ &= \text{vec}(\mathbf{H}_i \mathbf{F}_i)^\top E\left(\mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \otimes \mathbf{Q}_i^\top\right) \text{vec}(\mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i) \\ &= \text{vec}(\mathbf{H}_i \mathbf{F}_i)^\top \mathbf{F}_i^\top \mathbf{H}_i \left(E(\ddot{\mathbf{M}}_i \otimes \ddot{\mathbf{M}}_i^\top) - \mathbf{H}_i^{-1} \otimes \mathbf{H}_i^{-1}\right) \\ &\quad \times \text{vec}(\mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i) \\ &< \infty, \end{aligned}$$

$$\begin{aligned} E \left(\text{tr} \left\{ \mathbf{B}_i^\top \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right\} \right) &= \text{vec}(\mathbf{F}_i^\top \mathbf{H}_i)^\top E \left(\mathbf{Q}_i^\top \otimes \mathbf{B}_i^\top \right) \text{vec}(\boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i) \\ &= \text{vec}(\mathbf{F}_i^\top \mathbf{H}_i)^\top E \left(\ddot{\mathbf{M}}_i^\top \otimes \mathbf{B}_i^\top \right) \text{vec}(\boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i) < \infty, \end{aligned}$$

e

$$\begin{aligned} E \left(\text{tr} \left\{ \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{B}_i \right\} \right) &= E \left(\text{tr} \left\{ \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \mathbf{B}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \right\} \right) \\ &= \text{vec}(\boldsymbol{\lambda} \boldsymbol{\lambda}^\top)^\top \mathbf{F}_i^\top \mathbf{H}_i E \left(\mathbf{Q}_i \otimes \mathbf{B}_i^\top \right) \text{vec}(\mathbf{F}_i^\top \mathbf{H}_i) \\ &= \text{vec}(\boldsymbol{\lambda} \boldsymbol{\lambda}^\top)^\top \mathbf{F}_i^\top \mathbf{H}_i E \left(\ddot{\mathbf{M}}_i \otimes \mathbf{B}_i^\top \right) \text{vec}(\mathbf{F}_i^\top \mathbf{H}_i) < \infty. \end{aligned}$$

3. $\lim_{n \rightarrow \infty} \frac{1}{n^{1+\delta}} \sum_{i=1}^n E \left(\left\| \left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right) \boldsymbol{\lambda} \right\|^{1+\delta} \right) = 0$, tomando $\delta = 1$, temos que

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n E \left(\left\| \left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right) \boldsymbol{\lambda} \right\|^2 \right) \leq v,$$

em que $v = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{1 \leq i \leq n} E \left(\left\| \left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right) \boldsymbol{\lambda} \right\|^2 \right)$, como

$$E \left(\left\| \left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right) \boldsymbol{\lambda} \right\|^2 \right) < \infty \quad \forall i, \text{ então}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n E \left(\left\| \left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right) \boldsymbol{\lambda} \right\|^2 \right) = 0.$$

Logo,

$$\frac{\boldsymbol{\lambda}^\top \mathbf{L}_\theta}{n} + \frac{\boldsymbol{\lambda}^\top \mathbf{K}_\theta}{n} \xrightarrow{P_\theta} \mathbf{0}, \text{ para cada } \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

□

Lema A.5.3. *Se as condições C1-C5 e C8 forem satisfeitas, então*

$$\boldsymbol{\lambda}^\top \frac{1}{n} (\mathbf{L}_{\boldsymbol{\theta}_n^*} - \mathbf{L}_\theta) (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = o_p \left(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \right).$$

Demonstração. Consideremos a expressão $\left\| \frac{\mathbf{L}_{\boldsymbol{\theta}_n^*}}{n} - \frac{\mathbf{L}_\theta}{n} \right\|$. Pela condição C8, temos que

$$\begin{aligned} \left\| \frac{\mathbf{L}_{\boldsymbol{\theta}_n^*}}{n} - \frac{\mathbf{L}_\theta}{n} \right\| &= \frac{1}{n} \left\| \sum_{i=1}^n \mathbf{L}_{\boldsymbol{\theta}_n^* i} - \sum_{i=1}^n \mathbf{L}_{\boldsymbol{\theta} i} \right\| \\ &\leq \frac{1}{n} \sum_{i=1}^n \|\mathbf{U}_{(\boldsymbol{\theta}_n^* i)} - \mathbf{L}_{\boldsymbol{\theta} i}\| \\ &\leq \frac{1}{n} \sum_{i=1}^n M_i \|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\| = \frac{1}{n} \sum_{i=1}^n M_i h \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|, \end{aligned}$$

em que $M_n = h \sum_{i=1}^n M_i$ e M_i é tal que $M_n = O_p(n)$. Notemos ainda que $\boldsymbol{\theta}_n^* = h \hat{\boldsymbol{\theta}}_n + (1-h)\boldsymbol{\theta}$. Assim

$\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\| = \|h \hat{\boldsymbol{\theta}}_n + \boldsymbol{\theta} - h\boldsymbol{\theta} - \boldsymbol{\theta}\| = h \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|$. Pelas condições C1-C5, temos o resultado do Teorema 2.2.4 e assim $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| = o_p(1)$. Como $M_n = O_p(n)$, temos

$$\boldsymbol{\lambda}^\top \frac{1}{n} [\mathbf{L}_{\boldsymbol{\theta}_n^*} - \mathbf{L}_\theta] (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = o_p \left(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \right).$$

□

Prova do Teorema 2.2.5

Demonstração. Tomemos $\lambda \neq \mathbf{0}$. Aplicando o Teorema do Valor Médio para a função $\mathbf{U}_\theta$, a qual foi definida em (2.2), considerando a abordagem apresentada em Patriota (2019) e satisfeita a condição C1, temos

$$\lambda^\top \mathbf{U}_\theta = \lambda^\top \mathbf{U}_{\hat{\theta}_n} + \lambda^\top \mathbf{L}_{\theta_n^*} (\theta - \hat{\theta}_n), \quad (\text{A.10})$$

em que $\theta_n^* = \theta_{n,\lambda}^* = h\hat{\theta}_n + (1-h)\theta$, com $h = h_\lambda \in (0, 1)$. Como $\hat{\theta}_n$ é o estimador de máxima verossimilhança, então $\mathbf{U}_{\hat{\theta}_n} = \mathbf{0}$. Dividindo ambos os lados por n , reescrevemos (A.10) como

$$\lambda^\top \frac{\mathbf{U}_\theta}{n} = -\lambda^\top \frac{\mathbf{L}_\theta}{n} (\hat{\theta}_n - \theta) - \lambda^\top \frac{1}{n} (\mathbf{L}_{\theta_n^*} - \mathbf{L}_\theta) (\hat{\theta}_n - \theta). \quad (\text{A.11})$$

Para concluir a prova do teorema, observemos que $\frac{\lambda^\top \mathbf{U}_\theta}{n} = O_p(n^{-1/2})$. Dos resultados dos Lemas A.5.2 e A.5.3, a expressão (A.11) pode ser reescrita como

$$O_p(n^{-1/2}) = \lambda^\top \frac{\mathbf{K}_\theta}{n} (\hat{\theta}_n - \theta) - o_p(\|\hat{\theta}_n - \theta\|).$$

Logo,

$$\hat{\theta}_n - \theta = O_p(n^{-1/2}).$$

Multiplicando a expressão (A.11) por $\sqrt{n}$ temos

$$\lambda^\top \frac{\mathbf{U}_\theta}{\sqrt{n}} = \lambda^\top \frac{\mathbf{K}_\theta}{\sqrt{n}} (\hat{\theta}_n - \theta) + o_p(1). \quad (\text{A.12})$$

Do Lema A.5.1, temos que

$$\frac{\lambda^\top \mathbf{U}_\theta}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda^\top \Lambda_\theta \lambda).$$

Então,

$$\frac{\lambda^\top \mathbf{K}_\theta (\hat{\theta}_n - \theta)}{\sqrt{\lambda^\top \mathbf{K}_\theta \lambda}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

para cada $\theta \in \Theta$. Pelo Teorema de Cramér-Wold

$$\mathbf{K}_\theta^{1/2} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_p(0, \mathbf{I}_p).$$

□

A.6 Prova do Teorema 2.2.6 - Distribuição assintótica das estatísticas de teste

Primeiro enunciamos um teorema que será utilizados na prova do teorema que trata da distribuição assintótica das estatísticas de teste.

Teorema A.6.1. (*Teorema de Courant - Sen e Singer (1993), pg.28*) Sejam $\mathbf{A}$ e $\mathbf{B}$ duas matrizes

semidefinidas positivas com $\mathbf{B}$ não singular. Então,

$$ch_p(\mathbf{AB}^{-1}) = \inf_{\mathbf{x}} \frac{\mathbf{x}^\top \mathbf{Ax}}{\mathbf{x}^\top \mathbf{Bx}} \leq \max_{\mathbf{x}} \frac{\mathbf{x}^\top \mathbf{Ax}}{\mathbf{x}^\top \mathbf{Bx}} = ch_1(\mathbf{AB}^{-1}),$$

em que $ch_p(\mathbf{A})$ e $ch_1(\mathbf{A})$ são, respectivamente, o menor e o maior autovalor da matriz $\mathbf{A}$.

Prova do Teorema 2.2.6

Demonstração. A prova desse teorema será realizada para cada uma das estatísticas. Iniciamos com a estatística Wald (W_n). Satisfeitas as condições C1-C5 e C8 pelo Teorema 2.2.5, temos a normalidade assintótica do EMV,

$$\mathbf{K}_\theta^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p).$$

Seguindo os mesmos passos da prova da normalidade assintótica, sob as condições C1-C5 e C8 pelo Teorema 2.2.5 considerando uma partição de θ dada por $\psi = (\psi_1, \dots, \psi_q)^\top$ e a abordagem da Seção 2.2.5, temos que

$$[\mathbf{K}^{\psi\psi}(\theta)]^{-1/2}(\hat{\psi}_n - \psi) \xrightarrow{\mathcal{D}} \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q).$$

Então,

$$W_n^* = (\hat{\psi}_n - \psi_0)^\top [\mathbf{K}^{\psi\psi}(\tilde{\theta}_n)]^{-1}(\hat{\psi}_n - \psi_0) \rightarrow \chi_q^2.$$

Consideremos a quantidade

$$\frac{W_n}{W_n^*} = \frac{(\hat{\psi}_n - \psi_0)^\top [\mathbf{K}^{\psi\psi}(\hat{\theta}_n)]^{-1}(\hat{\psi}_n - \psi_0)}{(\hat{\psi}_n - \psi_0)^\top [\mathbf{K}^{\psi\psi}(\tilde{\theta}_n)]^{-1}(\hat{\psi}_n - \psi_0)},$$

usando o Teorema A.6.1 temos que

$$\text{ch}_q([\mathbf{K}^{\psi\psi}(\hat{\theta}_n)]^{-1} \mathbf{K}^{\psi\psi}(\tilde{\theta}_n)) \leq \frac{W_n}{W_n^*} \leq \text{ch}_1([\mathbf{K}^{\psi\psi}(\hat{\theta}_n)]^{-1} \mathbf{K}^{\psi\psi}(\tilde{\theta}_n)).$$

Notemos que como $\hat{\theta}_n \xrightarrow{P_\theta} \theta$, do Teorema 2.2.4 temos que $[\mathbf{K}_{\hat{\theta}_n}]^{-1} \mathbf{K}_\theta \xrightarrow{P_\theta} \mathbf{I}_p$. Consequentemente temos que

$$\text{ch}_q([\mathbf{K}^{\psi\psi}(\hat{\theta}_n)]^{-1} \mathbf{K}^{\psi\psi}(\tilde{\theta}_n)) \xrightarrow{P_\theta} 1 \quad \text{e} \quad \text{ch}_1([\mathbf{K}^{\psi\psi}(\hat{\theta}_n)]^{-1} \mathbf{K}^{\psi\psi}(\tilde{\theta}_n)) \xrightarrow{P_\theta} 1.$$

Segue que $\frac{W_n}{W_n^*} \xrightarrow{P_\theta} 1$. Reescrevendo $W_n = \frac{W_n W_n^*}{W_n^*}$, pelo Teorema de Slutsky concluímos que $W_n \xrightarrow{\mathcal{D}} \chi_q^2$.

Para provar a distribuição assintótica da estatística RV_n , inicialmente consideramos $\theta_0 = (\psi_0, \tau)^\top$ e a expansão

$$\ell(\mathbf{y}, \theta_0) = \ell(\mathbf{y}, \hat{\theta}_n) + (\theta_0 - \hat{\theta}_n)^\top \ell'(\mathbf{y}, \hat{\theta}_n) + \frac{1}{2}(\theta_0 - \hat{\theta}_n)^\top \ell''(\mathbf{y}, \hat{\theta}_n^*)(\theta_0 - \hat{\theta}_n), \quad (\text{A.13})$$

em que $\hat{\theta}_n^* = h\hat{\theta}_n + (1-h)\theta_0$ para algum $h \in (0, 1)$. Reescrevendo temos

$$\begin{aligned} \ell(\mathbf{y}, \hat{\theta}_n) - \ell(\mathbf{y}, \theta_0) &= -\frac{1}{2}(\hat{\theta}_n - \theta_0)^\top \ell''(\mathbf{y}, \hat{\theta}_n^*)(\hat{\theta}_n - \theta_0) \\ &= -\frac{n}{2}(\hat{\theta}_n - \theta_0)^\top \left[\frac{1}{n} \ell''(\mathbf{y}, \hat{\theta}_n^*) \right] (\hat{\theta}_n - \theta_0). \end{aligned}$$

Do Lema A.5.2 temos que

$$\frac{1}{n} \boldsymbol{\lambda}^\top \ell''(\mathbf{y}, \hat{\boldsymbol{\theta}}_n^*) + \frac{1}{n} \boldsymbol{\lambda}^\top \mathbf{K}_\theta \xrightarrow{P_\theta} \mathbf{0},$$

Assim,

$$RV_n^* = 2[\ell(\mathbf{y}, \hat{\boldsymbol{\theta}}_n) - \ell(\mathbf{y}, \boldsymbol{\theta}_0)] = (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \mathbf{K}_{\boldsymbol{\theta}_0} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1) = [\mathbf{U}_{\boldsymbol{\theta}_0}]^\top \mathbf{K}_{\boldsymbol{\theta}_0}^{-1} \mathbf{U}_{\boldsymbol{\theta}_0} + o_p(1).$$

Então, segue que $RV_n^* \xrightarrow{\mathcal{D}} \chi_p^2$. Notemos que a estatística da razão de verossimilhanças sob a hipótese nula pode ser decomposta como

$$\begin{aligned} RV_n &= 2[\ell(\mathbf{y}, \hat{\boldsymbol{\theta}}_n) - \ell(\mathbf{y}, \boldsymbol{\theta}_0)] - 2[\ell(\mathbf{y}, \tilde{\boldsymbol{\theta}}_n) - \ell(\mathbf{y}, \boldsymbol{\theta}_0)] \\ &= [\mathbf{U}_\theta]^\top \mathbf{K}_\theta^{-1} \mathbf{U}_\theta - [\mathbf{U}_\psi(\tilde{\boldsymbol{\theta}}_n)]^\top [\mathbf{K}_{\psi\psi}(\tilde{\boldsymbol{\theta}}_n)]^{-1} \mathbf{U}_\psi(\tilde{\boldsymbol{\theta}}_n) + o_p(1) \\ &= [\mathbf{U}_\theta]^\top \mathbf{K}_\theta^{-1/2} \mathbf{P}_\theta \mathbf{K}_\theta^{-1/2} \mathbf{U}_\theta + o_p(1), \end{aligned}$$

em que

$$\mathbf{P}_\theta = \mathbf{I} - \mathbf{K}_\theta^{-1/2} \begin{pmatrix} [\mathbf{K}_{\psi\psi}(\tilde{\boldsymbol{\theta}}_n)]^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{K}_\theta^{-1/2}.$$

Dos resultados obtidos para a normalidade assintótica chegamos a

$$\mathbf{P}_\theta = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}.$$

Notemos que RV_n é uma forma quadrática, o posto da matriz $\mathbf{P}$ é q , do resultado do Teorema 2.2.5 temos que $RV_n \xrightarrow{\mathcal{D}} \chi_q^2$.

Para a estatística S_n , lembrando o resultado do Lema A.5.1, considerando a partição de $\mathbf{U}_\theta$ temos que

$$\frac{\mathbf{U}_\theta}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Lambda}_\theta) \quad \text{e assim} \quad \frac{\mathbf{U}_{\boldsymbol{\theta}_0}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Lambda}_{\boldsymbol{\theta}_0}).$$

Então, como $\tilde{\boldsymbol{\theta}}_n$ é um estimador consistente para $\boldsymbol{\theta}_0$ segue que

$$[\mathbf{K}^{\psi\psi}(\tilde{\boldsymbol{\theta}}_n)]^{1/2} \mathbf{U}_\psi(\tilde{\boldsymbol{\theta}}_n) \xrightarrow{\mathcal{D}} \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q).$$

Logo

$$S_n = [\mathbf{U}_\psi(\tilde{\boldsymbol{\theta}}_n)]^\top [\mathbf{K}^{\psi\psi}(\tilde{\boldsymbol{\theta}}_n)]^{1/2} [\mathbf{K}^{\psi\psi}(\tilde{\boldsymbol{\theta}}_n)]^{1/2} \mathbf{U}_\psi(\tilde{\boldsymbol{\theta}}_n) \xrightarrow{\mathcal{D}} \chi_q^2.$$

Assim, temos que $S_n \xrightarrow{\mathcal{D}} \chi_q^2$.

Para a estatística gradiente G_n , retomando a expressão (A.12), podemos reescrever como

$$[\mathbf{K}^{\psi\psi}(\tilde{\boldsymbol{\theta}}_n)] \mathbf{U}_\psi(\tilde{\boldsymbol{\theta}}_n) = (\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) + o_p(1),$$

e ainda

$$[\mathbf{U}_\psi(\tilde{\boldsymbol{\theta}}_n)]^\top [\mathbf{K}^{\psi\psi}(\tilde{\boldsymbol{\theta}}_n)] \mathbf{U}_\psi(\tilde{\boldsymbol{\theta}}_n) = [\mathbf{U}_\psi(\tilde{\boldsymbol{\theta}}_n)]^\top (\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) + o_p(1).$$

Notemos que a expressão do lado esquerdo corresponde à estatística escore S_n , a qual já provamos a distribuição assintótica. Logo, segue que $G_n \xrightarrow{\mathcal{D}} \chi_q^2$. $\square$

Apêndice B

Provas dos Teoremas do Capítulo 3

B.1 Prova do Teorema 3.2.2 - Unicidade do EMV caso geral

A seguir enunciamos e provamos um lema que demonstra que sob certas condições no modelo (3.1) uma das condições necessárias apresentadas em Krätschmer (2007) é satisfeita. Este é um resultado auxiliar e corresponde à condição C3.2 de Krätschmer (2007).

Lema B.1.1. *Seja o modelo (3.1). Se as condições C7 e C11 são satisfeitas, então o conjunto*

$$D_1 := \{\mathbf{y} \in \mathbb{R}^N : \mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}_1) = \mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}_2), \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}\}$$

tem medida de Lebesgue nula, em que

$$\mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}) = \frac{\partial \ell(\mathbf{y}, \boldsymbol{\theta})}{\partial \mathbf{y}} = 2W_g(\mathbf{u})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})[\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})],$$

com $\mathbf{u} = [\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})]^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})[\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})]$

Demonstração. Da condição C7 e do Lema A.2.1, temos que $\mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}) = -\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})[\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})]$. Então, D_1 tem medida de Lebesgue nula. Notemos que

$$\mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}_1) = \mathbf{T}_{\mathbf{y}}(\boldsymbol{\theta}_2) \Leftrightarrow W_g(\mathbf{u}) = 0.$$

Assim, se $W_g(\mathbf{u}) = 0$, o conjunto D_1 tem medida de Lebesgue igual a um. Logo, sob a condição C11, temos que D_1 tem medida de Lebesgue nula. □

Prova do Teorema 3.2.2

Demonstração. Para provar este teorema, basta mostrar que as condições suficientes do Teorema 3.1 apresentado em Krätschmer (2007) são satisfeitas. Das condições C7 e C11, considerando o modelo (3.1), o resultado do Lema B.1.1 é válido. Então, a condição C3.2 de Krätschmer (2007) é satisfeita. Os demais passos da prova são iguais à prova apresentada no Apêndice A, Seção A.2. □

B.2 Prova do Teorema 3.2.3 - Unicidade do EMV no caso em que a média e a matriz de covariâncias não compartilham parâmetros

Consideremos novamente o modelo estruturado escrito como

$$\mathbf{Y} \sim \mathcal{E}l_N(\boldsymbol{\mu}(\boldsymbol{\beta}), \boldsymbol{\Sigma}(\boldsymbol{\gamma}), g), \tag{B.1}$$

com $\boldsymbol{\mu}(\boldsymbol{\beta}) = (\boldsymbol{\mu}_1(\boldsymbol{\beta})^\top, \dots, \boldsymbol{\mu}_n(\boldsymbol{\beta})^\top)^\top$ e $\boldsymbol{\Sigma}(\boldsymbol{\gamma}) = \text{bdiag}\{\boldsymbol{\Sigma}_1(\boldsymbol{\gamma}), \dots, \boldsymbol{\Sigma}_n(\boldsymbol{\gamma})\}$.

O modelo não estruturado na média é definido por

$$\mathbf{W} \sim \mathcal{E}l_N(\boldsymbol{\nu}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}), g) \quad (\text{B.2})$$

com $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N)^\top$. A função de log-verossimilhança associada ao modelo não estruturado na média é dada por

$$\ell_0(\mathbf{y}; \boldsymbol{\nu}, \boldsymbol{\gamma}) = -\frac{1}{2} \log \det(\boldsymbol{\Sigma}(\boldsymbol{\gamma})) + \log g\left((\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\gamma})(\mathbf{y} - \boldsymbol{\nu})\right).$$

Lema B.2.1. *Sob as condições C2, C3 e C9, considere o subconjunto aberto H de $\mathbb{R}^N \times \mathbb{R}^N$, definido como*

$$H := \{(\mathbf{y}, \boldsymbol{\nu}) \in \mathbb{R}^N \times \mathbb{R}^N : |\mathbf{y} - \boldsymbol{\nu}| \neq 0\}.$$

Seja $\ell^* : H \rightarrow \mathbb{R}$ uma função tal que

$$\ell^*(\mathbf{y}; \boldsymbol{\nu}) := -\frac{1}{2} \log \det(\boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_{\mathbf{y}-\boldsymbol{\nu}})) + \log g\left((\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\gamma}}_{\mathbf{y}-\boldsymbol{\nu}})(\mathbf{y} - \boldsymbol{\nu})\right),$$

em que $\hat{\boldsymbol{\gamma}}_{\mathbf{y}-\boldsymbol{\nu}}$ é o EMV de $\boldsymbol{\gamma}$ para cada $\boldsymbol{\nu}$ fixado no modelo (B.2), ou seja, deve satisfazer a equação

$$\mathbf{V}_x^\top \tilde{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\gamma}}_x) \text{vec}(\boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_x)) = -2W_g\left(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\gamma}}_x)\mathbf{x}\right) \mathbf{V}_x^\top \tilde{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\gamma}}_x) \text{vec}(\mathbf{x}\mathbf{x}^\top), \quad (\text{B.3})$$

em que $\mathbf{x} = \mathbf{y} - \boldsymbol{\nu}$ e $\mathbf{V}_x = \left(\text{vec}\left(\frac{\partial \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_x)}{\partial x_1}\right), \dots, \text{vec}\left(\frac{\partial \boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}}_x)}{\partial x_N}\right)\right)$. Adicionalmente, considere a função $\boldsymbol{\mu} : \mathbb{R}^k \rightarrow \mathbb{R}^N$. Então,

- (a) $\boldsymbol{\mu}$ é uma função contínua duas vezes diferenciável em relação ao vetor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^\top$ com posto k .
- (b) Se, para $\mathbf{y} \in \mathbb{R}^N$, a função $\ell_0(\mathbf{y}; \cdot, \boldsymbol{\gamma}) \circ \boldsymbol{\mu}$ tiver pontos críticos degenerados, então a função $\ell^*(\mathbf{y}; \cdot) \circ \boldsymbol{\mu}$ para todo $\boldsymbol{\beta} \in \boldsymbol{\mu}^{-1}(H_y)$ tem ponto crítico degenerado também, em que $H_y := \{\boldsymbol{\nu} \in \mathbb{R}^N : (\mathbf{y}, \boldsymbol{\nu}) \in H\}$.

Demonstração. $\boldsymbol{\mu}$ é uma função contínua tal que C2 e C3 são satisfeitas. Então temos que (a) é satisfeita. Note que $\ell^*(\mathbf{y}; \boldsymbol{\nu})$ é a função de log-verossimilhança perfilada para $\boldsymbol{\nu}$ do modelo (B.2) a menos de uma constante. Pelas condições C9 e C2, a função de log-verossimilhança é diferenciável. Considere agora a função $\boldsymbol{\mu}$ satisfazendo C2 temos que $\ell_0(\mathbf{y}; \cdot, \boldsymbol{\gamma}) \circ \boldsymbol{\mu}$ corresponde à função de log-verossimilhança e $\ell^*(\mathbf{y}; \cdot) \circ \boldsymbol{\mu}$ para todo $\boldsymbol{\beta} \in \boldsymbol{\mu}^{-1}(H_y)$ corresponde à função de log-verossimilhança perfilada, ambas associadas ao modelo (B.1). Note que uma das propriedades da função de log-verossimilhança é que os máximos das funções de log-verossimilhança e log-verossimilhança perfilada coincidem. Logo, temos que se $\ell_0(\mathbf{y}; \cdot, \boldsymbol{\gamma}) \circ \boldsymbol{\mu}$ tiver um ponto crítico degenerado, então $\ell^*(\mathbf{y}; \cdot) \circ \boldsymbol{\mu}$ para todo $\boldsymbol{\beta} \in \boldsymbol{\mu}^{-1}(H_y)$ também terá. $\square$

Prova do Teorema 3.2.3

Demonstração. Para provar este teorema, basta mostrar que as condições do Teorema 4.3 apresentado em Krätschmer (2007) são satisfeitas. Se a condição C7 é válida então $\boldsymbol{\theta}$ é identificável no modelo (3.1). Como $\mathbf{G}$, pela condição C2, é uma função duas vezes diferenciável, mediante a condição C9, a função de log-verossimilhança também é duas vezes diferenciável. A condição C4.2 de Krätschmer (2007) é provada no Lema B.2.1 mediante as condições C2, C3 e C9. A condição C3.2 é provada no Lema B.1.1. Adicionalmente, temos que mostrar que existe um difeomorfismo $f : U \rightarrow \mathbb{R}^N$ em que $U \subset \mathbb{R}^N$ é um aberto, tal que

$$\mathbf{y} - \boldsymbol{\nu} \in U \quad \text{e} \quad \frac{\partial \ell^*(\mathbf{y}; \cdot)}{\partial \boldsymbol{\nu}} = f(\mathbf{y} - \boldsymbol{\nu}),$$

para $(\mathbf{y}, \boldsymbol{\nu}) \in H$ arbitrário. Notemos que

$$\begin{aligned} \frac{\partial \ell^*(\mathbf{y}; \cdot)}{\partial \nu_j} &= -W_g \left((\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}})(\mathbf{y} - \boldsymbol{\nu}) \right) (\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}}) \frac{\partial \boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}})}{\partial \nu_j} \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}})(\mathbf{y} - \boldsymbol{\nu}) \\ &\quad - \frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}}) \frac{\partial \boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}})}{\partial \nu_j} \right\} - 2W_g \left((\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}})(\mathbf{y} - \boldsymbol{\nu}) \right) (\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{y}-\boldsymbol{\nu}}) \frac{d\boldsymbol{\nu}}{d\nu_j}. \end{aligned}$$

Tomando $\mathbf{x} = \mathbf{y} - \boldsymbol{\nu}$, podemos escrever uma função de $\mathbf{x}$ dada por

$$\begin{aligned} f_j(\mathbf{x}) &= -\frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \frac{\partial \boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{x}})}{\partial x_j} \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \left(\boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{x}}) + 2W_g \left(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x} \right) \mathbf{x} \mathbf{x}^\top \right) \right\} \\ &\quad - 2 \text{tr} \left\{ \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \frac{d\mathbf{x}}{dx_j} \mathbf{x}^\top W_g \left(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x} \right) \right\}. \end{aligned}$$

Dessa forma, definamos $U := \{\mathbf{x} \in \mathbb{R}^N : |\mathbf{x}| \neq 0\}$ e a função $f : U \rightarrow \mathbb{R}^N$ baseada na derivada do vetor $\boldsymbol{\nu}$ pode ser escrita em termos de uma função de $\mathbf{x}$ em forma matricial, conforme segue:

$$\begin{aligned} f(\mathbf{x}) &= -2W_g(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x}) \mathbf{D}_{\mathbf{x}} \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x} \\ &\quad - \frac{1}{2} \mathbf{V}_{\mathbf{x}}^\top \tilde{\boldsymbol{\Sigma}}^{-1}(\hat{\gamma}_{\mathbf{x}}) \left[\text{vec}(\boldsymbol{\Sigma}(\hat{\gamma}_{\mathbf{x}})) + 2W_g(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x}) \text{vec}(\mathbf{x} \mathbf{x}^\top) \right], \end{aligned} \quad (\text{B.4})$$

em que $\mathbf{D}_{\mathbf{x}}$ é a matriz identidade. Note que, pela condição (B.3), a equação (B.4) é simplificada e pode ser reescrita como

$$f(\mathbf{x}) = -2W_g(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x}) \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x}.$$

Para $(\mathbf{y}, \boldsymbol{\nu}) \in H$ arbitrário,

$$\mathbf{y} - \boldsymbol{\nu} \in U, \quad \frac{\partial \ell^*(\mathbf{y}; \cdot)}{\partial \boldsymbol{\nu}} = f(\mathbf{y} - \boldsymbol{\nu}) \quad \text{e} \quad \frac{\partial \ell^*(\mathbf{y}; \cdot)}{\partial \mathbf{y}} = -f(\mathbf{y} - \boldsymbol{\nu}).$$

Como a função $f(\mathbf{x}) = -2W_g(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x}) \boldsymbol{\Sigma}^{-1}(\hat{\gamma}_{\mathbf{x}}) \mathbf{x}$ é um difeomorfismo por suposição, temos que todas as condições suficientes do Teorema 4.3 estabelecidas em Krätschmer (2007) são satisfeitas e concluímos que Q é mensurável e tem medida de Lebesgue nula. $\square$

B.3 Prova do Teorema 3.2.5 - Normalidade assintótica do EMV

Para a prova desse teorema enunciaremos e provaremos três lemas que compõem a prova do teorema. Tomemos inicialmente $\boldsymbol{\lambda} \neq \mathbf{0}$.

Lema B.3.1. *Se as condições C3, C4 e C10 forem satisfeitas, então*

$$\frac{\mathbf{U}_{\boldsymbol{\theta}}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Lambda}_{\boldsymbol{\theta}}), \text{ para cada } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Demonstração. Utilizando o Teorema Central do Limite de Liapounov, temos

1. $E(\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}_i}) = \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i E(\mathbf{s}_i) = \mathbf{0}$.
2. $E[(\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}_i})^2] < \infty$. Pelas condições C3, C4 e C10 temos

$$\begin{aligned} E[(\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}_i})^2] &= E(\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i \mathbf{s}_i^\top \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda}) \\ &= \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i E(\mathbf{s}_i \mathbf{s}_i^\top) \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \\ &= \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{M}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} < \infty. \end{aligned}$$

3. $E(\|\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}i}\|^{2+\delta}) < \infty$. Pelas condições C4, C10 e pela desigualdade de Cauchy-Schwarz para matrizes (ver Magnus e Neudecker (2019), pg. 227-228) temos

$$\begin{aligned} E(\|\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}i}\|^{2+\delta}) &= E(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i\|^{2+\delta}) \\ &\leq E(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i\|^{2+\delta} \|\mathbf{s}_i\|^{2+\delta}) \\ &= \|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i\|^{2+\delta} E(\|\mathbf{s}_i\|^{2+\delta}) \\ &< \infty, \end{aligned}$$

em que $E(\|\mathbf{s}_i\|^{2+\delta}) < \infty$, pela condição C10.

4. $\lim_{n \rightarrow \infty} \left(\sum_{i=1}^n \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{M}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \right)^{-\left(\frac{2+\delta}{2}\right)} \sum_{i=1}^n E(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i\|^{2+\delta}) = 0$, pois pela condição C4 e pelo provado no item 3, temos

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n E(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i\|^{2+\delta})}{\left(\sum_{i=1}^n \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{M}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \right)^{\frac{2+\delta}{2}}} &= \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n E(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i\|^{2+\delta})}{n^{\frac{2+\delta}{2}} \left(\boldsymbol{\lambda}^\top \frac{\sum_{i=1}^n \mathbf{F}_i^\top \mathbf{H}_i \mathbf{M}_i \mathbf{H}_i \mathbf{F}_i}{n} \boldsymbol{\lambda} \right)^{\frac{2+\delta}{2}}} \\ &= \frac{1}{(\boldsymbol{\lambda}^\top \boldsymbol{\Lambda}_{\boldsymbol{\theta}} \boldsymbol{\lambda})^{\frac{2+\delta}{2}}} \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n E(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{s}_i\|^{2+\delta})}{n^{\frac{2+\delta}{2}}} = 0. \end{aligned}$$

Então,

$$\frac{\boldsymbol{\lambda}^\top \sum_{i=1}^n \mathbf{U}_{\boldsymbol{\theta}i}}{\sqrt{\boldsymbol{\lambda}^\top \sum_{i=1}^n \mathbf{F}_i^\top \mathbf{H}_i \mathbf{M}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

De forma alternativa, podemos escrever

$$\frac{\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{\lambda}^\top \boldsymbol{\Lambda}_{\boldsymbol{\theta}} \boldsymbol{\lambda})$$

e, pelo Teorema de Cramér-Wold,

$$\frac{\mathbf{U}_{\boldsymbol{\theta}}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Lambda}_{\boldsymbol{\theta}}), \text{ para cada } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

□

Lema B.3.2. *Se as condições C3, C4 e C10 forem satisfeitas, então*

$$\frac{\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}}}{n} + \frac{\boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}}}{n} \xrightarrow{P_{\boldsymbol{\theta}}} \mathbf{0}, \text{ para cada } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Demonstração. Aplicando a Lei Fraca dos Grandes Números de Markov, temos que

1. $E(\|\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i}\|) < \infty$, pois $E(\|\mathbf{s}_i\|) < \infty$ pela condição C10 e assim

$$\begin{aligned} E(\|\mathbf{B}_i\|) &= E(\|\mathbf{r}_i^\top \otimes \mathbf{H}_i \mathbf{s}_i\|) \\ &\leq \|\mathbf{r}_i^\top \otimes \mathbf{H}_i\| E(\|\mathbf{s}_i\|) \\ &< \infty. \end{aligned}$$

Notemos que $E(\|\ddot{\mathbf{M}}_i\|) < \infty$ pela condição 10. Adicionalmente,

$$\begin{aligned} E(\|\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i}\|) &= E(\|\boldsymbol{\lambda}^\top \mathbf{B}_i - \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i\|) \\ &\leq E(\|\boldsymbol{\lambda}^\top \mathbf{B}_i\|) + E(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i\|) \\ &\leq \|\boldsymbol{\lambda}^\top\| E(\|\mathbf{B}_i\|) + \|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i\| E(\|\ddot{\mathbf{M}}_i\|) \|\mathbf{H}_i \mathbf{F}_i\| \\ &< \infty. \end{aligned}$$

2. $E(\|\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i} - E(\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i})\|^{1+\delta}) < \infty$, para algum $\delta \in (0, 1]$. Temos

$$\begin{aligned} \varrho &= E(\|\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i} - E(\boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_i})\|^{1+\delta}) = E(\|\boldsymbol{\lambda}^\top (\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{M}}_i \mathbf{H}_i \mathbf{F}_i + \mathbf{F}_i^\top \mathbf{H}_i \mathbf{M}_i \mathbf{H}_i \mathbf{F}_i)\|^{1+\delta}) \\ &= E(\|\boldsymbol{\lambda}^\top (\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i)\|^{1+\delta}), \end{aligned}$$

em que $\mathbf{Q}_i = \ddot{\mathbf{M}}_i - \mathbf{M}_i$. Notemos que

$$\begin{aligned} \varrho &\leq E(\|\boldsymbol{\lambda}^\top (\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i)\|^2) \\ &= E(\text{tr}\{(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i) \boldsymbol{\lambda} \boldsymbol{\lambda}^\top (\mathbf{B}_i - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i)\}) \\ &= E(\text{tr}\{\mathbf{B}_i^\top \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{B}_i\}) + E(\text{tr}\{\mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i\}) \\ &\quad - E(\text{tr}\{\mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{B}_i\}) - E(\text{tr}\{\mathbf{B}_i^\top \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i\}) \\ &< \infty. \end{aligned}$$

Analisando separadamente as esperanças da equação anterior, temos

$$E(\text{tr}\{\mathbf{B}_i^\top \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{B}_i\}) = E(\|\boldsymbol{\lambda}^\top \mathbf{B}_i\|^2) \leq \|\boldsymbol{\lambda}^\top\|^2 \|\mathbf{r}_i^\top \otimes \mathbf{H}_i\|^2 E(\|\mathbf{s}_i\|^2) < \infty.$$

Da condição C10, temos que $E([W_g(\mathbf{z}_i)]^l (\mathbf{z}_i^\top \mathbf{z}_i)^s) < \infty$ e $E([W_g^2(\mathbf{z}_i)]^l (\mathbf{z}_i^\top \mathbf{z}_i)^s) < \infty$ para $l, s = 1, 2$. Então, $E(\ddot{\mathbf{M}}_i \otimes \ddot{\mathbf{M}}_i^\top) < \infty$, $E(\ddot{\mathbf{M}}_i^\top \otimes \mathbf{B}_i^\top) < \infty$ e $E(\ddot{\mathbf{M}}_i \otimes \mathbf{B}_i^\top) < \infty$. Assim temos os resultados a seguir:

$$\begin{aligned} E(\|\boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i\|^2) &= E(\text{tr}\{\mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i\}) \\ &= \text{vec}(\mathbf{H}_i \mathbf{F}_i)^\top E(\mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \otimes \mathbf{Q}_i^\top) \text{vec}(\mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i) \\ &= \text{vec}(\mathbf{H}_i \mathbf{F}_i)^\top \mathbf{F}_i^\top \mathbf{H}_i (E(\ddot{\mathbf{M}}_i \otimes \ddot{\mathbf{M}}_i^\top) - \mathbf{M}_i \otimes \mathbf{M}_i^\top) \\ &\quad \times \text{vec}(\mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i) \\ &< \infty, \end{aligned}$$

$$\begin{aligned}
E \left(\text{tr} \left\{ \mathbf{B}_i^\top \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right\} \right) &= \text{vec}(\mathbf{F}_i^\top \mathbf{H}_i)^\top E \left(\mathbf{Q}_i^\top \otimes \mathbf{B}_i^\top \right) \text{vec}(\boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i) \\
&= \text{vec}(\mathbf{F}_i^\top \mathbf{H}_i)^\top E \left(\ddot{\mathbf{M}}_i^\top \otimes \mathbf{B}_i^\top \right) \text{vec}(\boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{F}_i^\top \mathbf{H}_i) \\
&< \infty
\end{aligned}$$

e

$$\begin{aligned}
E \left(\text{tr} \left\{ \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \mathbf{B}_i \right\} \right) &= E \left(\text{tr} \left\{ \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \mathbf{B}_i \boldsymbol{\lambda} \boldsymbol{\lambda}^\top \right\} \right) \\
&= \text{vec}(\boldsymbol{\lambda} \boldsymbol{\lambda}^\top)^\top \mathbf{F}_i^\top \mathbf{H}_i E \left(\mathbf{Q}_i \otimes \mathbf{B}_i^\top \right) \text{vec}(\mathbf{F}_i^\top \mathbf{H}_i) \\
&= \text{vec}(\boldsymbol{\lambda} \boldsymbol{\lambda}^\top)^\top \mathbf{F}_i^\top \mathbf{H}_i E \left(\ddot{\mathbf{M}}_i \otimes \mathbf{B}_i^\top \right) \text{vec}(\mathbf{F}_i^\top \mathbf{H}_i) \\
&< \infty.
\end{aligned}$$

3. $\lim_{n \rightarrow \infty} \frac{1}{n^{1+\delta}} \sum_{i=1}^n E \left(\left\| \left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right) \boldsymbol{\lambda} \right\|^{1+\delta} \right) = 0$. Tomando $\delta = 1$, temos que

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n E \left(\left\| \left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right) \boldsymbol{\lambda} \right\|^2 \right) \leq v,$$

em que $v = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{1 \leq i \leq n} E \left(\left\| \left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right) \boldsymbol{\lambda} \right\|^2 \right)$. Como

$$E \left(\left\| \left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right) \boldsymbol{\lambda} \right\|^2 \right) < \infty \quad \forall i, \text{ então}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n E \left(\left\| \left(\mathbf{B}_i^\top - \mathbf{F}_i^\top \mathbf{H}_i \mathbf{Q}_i \mathbf{H}_i \mathbf{F}_i \right) \boldsymbol{\lambda} \right\|^2 \right) = 0.$$

Logo,

$$\frac{\boldsymbol{\lambda}^\top \mathbf{L}_\theta}{n} + \frac{\boldsymbol{\lambda}^\top \mathbf{K}_\theta}{n} \xrightarrow{P_\theta} \mathbf{0}, \text{ para cada } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

□

Lema B.3.3. *Se as condições C1-C5, C8 e C10 forem satisfeitas, então*

$$\boldsymbol{\lambda}^\top \frac{1}{n} [\mathbf{L}_{\boldsymbol{\theta}_n^*} - \mathbf{L}_\theta] (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = o_p \left(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \right).$$

Demonstração. Considere a expressão $\left\| \frac{\mathbf{L}_{\boldsymbol{\theta}_n^*}}{n} - \frac{\mathbf{L}_\theta}{n} \right\|$. Pela condição C8, temos que

$$\begin{aligned}
\left\| \frac{\mathbf{L}_{\boldsymbol{\theta}_n^*}}{n} - \frac{\mathbf{L}_\theta}{n} \right\| &= \frac{1}{n} \left\| \sum_{i=1}^n \mathbf{L}_{\boldsymbol{\theta}_n^* i} - \sum_{i=1}^n \mathbf{L}_{\boldsymbol{\theta} i} \right\| \\
&\leq \frac{1}{n} \sum_{i=1}^n \|\mathbf{U}_{\boldsymbol{\theta}_n^* i} - \mathbf{L}_{\boldsymbol{\theta} i}\| \\
&\leq \frac{1}{n} \sum_{i=1}^n M_i \|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\| \\
&= \frac{1}{n} \sum_{i=1}^n M_i h \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|,
\end{aligned}$$

em que $M_n = h \sum_{i=1}^n M_i$ e M_i é tal que $M_n = O_p(n)$. Notemos ainda que $\boldsymbol{\theta}_n^* = h \hat{\boldsymbol{\theta}}_n + (1-h)\boldsymbol{\theta}$. Assim,

$\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\| = \|h\widehat{\boldsymbol{\theta}}_n + \boldsymbol{\theta} - h\boldsymbol{\theta} - \boldsymbol{\theta}\| = h\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|$. Pelas condições C1-C5, C9 e C10, temos o resultado do Teorema 3.2.4 e assim $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| = o_p(1)$. E como $M_n = O_p(n)$, temos

$$\boldsymbol{\lambda}^\top \frac{1}{n} [\mathbf{L}_{\boldsymbol{\theta}_n^*} - \mathbf{L}_{\boldsymbol{\theta}}] (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = o_p(\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|).$$

□

Prova do Teorema 3.2.5

Demonstração. Tomemos $\boldsymbol{\lambda} \neq \mathbf{0}$. Aplicando o Teorema do Valor Médio para a função $\mathbf{U}_{\boldsymbol{\theta}}$, a qual foi definida em (3.2), considerando a abordagem apresentada em Patriota (2019), temos

$$\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}} = \boldsymbol{\lambda}^\top \mathbf{U}_{\widehat{\boldsymbol{\theta}}_n} + \boldsymbol{\lambda}^\top \mathbf{L}_{\boldsymbol{\theta}_n^*} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n), \quad (\text{B.5})$$

em que $\boldsymbol{\theta}_n^* = \boldsymbol{\theta}_{n,\boldsymbol{\lambda}}^* = h\widehat{\boldsymbol{\theta}}_n + (1-h)\boldsymbol{\theta}$, com $h = h_{\boldsymbol{\lambda}} \in (0, 1)$. Como $\widehat{\boldsymbol{\theta}}_n$ é o estimador de máxima verossimilhança, então $\mathbf{U}_{\widehat{\boldsymbol{\theta}}_n} = \mathbf{0}$. Dividindo ambos os lados por n , reescrevemos (B.5) como

$$\boldsymbol{\lambda}^\top \frac{\mathbf{U}_{\boldsymbol{\theta}}}{n} = -\boldsymbol{\lambda}^\top \frac{\mathbf{L}_{\boldsymbol{\theta}}}{n} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) - \boldsymbol{\lambda}^\top \frac{1}{n} [\mathbf{L}_{\boldsymbol{\theta}_n^*} - \mathbf{L}_{\boldsymbol{\theta}}] (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}). \quad (\text{B.6})$$

Para concluir a prova do Teorema, observemos que $\frac{\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}}}{n} = O_p(n^{-1/2})$. Dos resultados dos Lemas B.3.2 e B.3.3, a expressão (B.6) pode ser reescrita como

$$O_p(n^{-1/2}) = \boldsymbol{\lambda}^\top \frac{\mathbf{K}_{\boldsymbol{\theta}}}{n} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) - o_p(\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|).$$

Logo,

$$(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = O_p(n^{-1/2}).$$

Multiplicando a expressão (B.6) por $\sqrt{n}$, temos

$$\boldsymbol{\lambda}^\top \frac{\mathbf{U}_{\boldsymbol{\theta}}}{\sqrt{n}} = \boldsymbol{\lambda}^\top \frac{\mathbf{K}_{\boldsymbol{\theta}}}{\sqrt{n}} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) + o_p(1).$$

Do Lema B.3.1, temos que

$$\frac{\boldsymbol{\lambda}^\top \mathbf{U}_{\boldsymbol{\theta}}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{\lambda}^\top \boldsymbol{\Lambda}_{\boldsymbol{\theta}} \boldsymbol{\lambda}),$$

Então,

$$\frac{\boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})}{\sqrt{\boldsymbol{\lambda}^\top \mathbf{K}_{\boldsymbol{\theta}} \boldsymbol{\lambda}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

para cada $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Pelo Teorema de Cramér-Wold,

$$\mathbf{K}_{\boldsymbol{\theta}}^{1/2} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p).$$

□

Referências Bibliográficas

- Akritis e Bershadsky(1996)** Michael G. Akritis e Matthew A. Bershadsky. Linear regression for astronomical data with measurement errors and intrinsic scatter. *The Astrophysical Journal*, 470(20):706–714. Citado na pág. [2](#)
- Barndorff-Nielsen(1986)** Ole E. Barndorff-Nielsen. Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, 73(2):307–322. Citado na pág. [3](#)
- Bayer e Cribari-Neto(2013)** Fábio M. Bayer e Francisco Cribari-Neto. Bartlett corrections in beta regression models. *Journal of Statistical Planning and Inference*, 143(3):531–547. Citado na pág. [14](#)
- Borssoi et al.(2017)** Joelmir A. Borssoi, Gilberto A. Paula e Manuel Galea. Elliptical linear mixed models with a covariate subject to measurement error. *Statistical Papers*, páginas 1–39. Citado na pág. [2](#)
- Buse(1982)** A. Buse. The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *The American Statistician*, 36(3a):153–157. Citado na pág. [15](#)
- Cordeiro(1992)** G. Cordeiro. Introdução à teoria de verossimilhança. Em *X Simpósio Nacional de Probabilidade e Estatística*, Rio de Janeiro - Associação Brasileira de Estatística. Citado na pág. [14](#)
- Cox e Hinkley(1974)** David R. Cox e David V. Hinkley. *Theoretical Statistics*. Chapman and Hall/CRC, New York. Citado na pág. [3](#)
- Cribari-Neto e Queiroz(2014)** Francisco Cribari-Neto e Marcela P. F. Queiroz. On testing inference in beta regressions. *Journal of Statistical Computation and Simulation*, 84(1):186–203. Citado na pág. [14](#)
- Fahrmeir(1988)** L Fahrmeir. A note on asymptotic testing theory for nonhomogeneous observations. *Stochastic processes and their applications*, 28(2):267–273. Citado na pág. [15](#)
- Fahrmeir e Kaufmann(1985)** Ludwig Fahrmeir e Heinz Kaufmann. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, 13(1):342–368. Citado na pág. [3](#), [10](#), [11](#), [12](#), [33](#)
- Fang et al.(2018)** Kai-Tang Fang, Samuel Kotz e Kai W. Ng. *Symmetric Multivariate and Related Distributions*. Chapman and Hall/CRC, New York. ISBN 9781351085496. Citado na pág. [1](#)
- Galea et al.(2000)** Manuel Galea, Marco Riquelme e Gilberto A. Paula. Diagnostic methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistics*, 14(2):167–184. Citado na pág. [1](#)
- Gómez et al.(1998)** E. Gómez, Miguel A. Gomez-Viilegas e J. M. Marin. A multivariate generalization of the power exponential family of distributions. *Communications in Statistics-Theory and Methods*, 27(3):589–600. Citado na pág. [2](#)

- Horn e Johnson(2013)** Roger A. Horn e Charles R. Johnson. *Matrix analysis*. Cambridge University Press, New York. Citado na pág. [10](#), [45](#)
- Ibacache-Pulgar e Paula(2011)** Germán Ibacache-Pulgar e Gilberto A. Paula. Local influence for Student-t partially linear models. *Computational Statistics & Data Analysis*, 55(3):1462–1478. Citado na pág. [31](#)
- Ibacache-Pulgar et al.(2014)** Germán Ibacache-Pulgar, Gilberto A. Paula e Manuel Galea. On influence diagnostics in elliptical multivariate regression models with equicorrelated random errors. *Statistical Methodology*, 16:14–31. Citado na pág. [2](#)
- Krätschmer(2007)** Volker Krätschmer. The uniqueness of extremum estimation. *Statistics & Probability Letters*, 77(10):942–951. Citado na pág. [3](#), [10](#), [11](#), [12](#), [13](#), [18](#), [33](#), [34](#), [47](#), [48](#), [49](#), [57](#), [58](#), [59](#)
- Kulathinal et al.(2002)** S. B. Kulathinal, Kari Kuulasmaa e Dario Gasbarra. Estimation of an errors-in-variables regression model when the variances of the measurement errors vary between the observations. *Statistics in Medicine*, 21(8):1089–1101. Citado na pág. [2](#)
- Lange et al.(1989)** Kenneth L. Lange, Roderick J. A. Little e Jeremy M. G. Taylor. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408):881–896. Citado na pág. [2](#), [30](#)
- Lemonte(2013)** Artur J. Lemonte. On the gradient statistic under model misspecification. *Statistics & Probability Letters*, 83(1):390–398. Citado na pág. [15](#)
- Lemonte e Ferrari(2012)** Artur J. Lemonte e Silvia L. P. Ferrari. Local power and size properties of the LR, Wald, score and gradient tests in dispersion models. *Statistical Methodology*, 9(5):537–554. Citado na pág. [15](#)
- Lemonte e Patriota(2011)** Artur J. Lemonte e Alexandre G. Patriota. Multivariate elliptical models with general parameterization. *Statistical Methodology*, 8(4):389–400. Citado na pág. [iii](#), [v](#), [1](#), [2](#), [7](#), [29](#), [30](#), [31](#), [32](#), [37](#), [40](#), [41](#)
- Lucas(1997)** André Lucas. Robustness of the Student t based M-estimator. *Communications in Statistics-Theory and Methods*, 26(5):1165–1182. Citado na pág. [31](#)
- Magnus e Neudecker(2019)** Jan R. Magnus e Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 3^a edição. Citado na pág. [8](#), [51](#), [60](#)
- Manghi et al.(2016)** Roberto F. Manghi, Gilberto A. Paula e Francisco J. A. Cysneiros. On elliptical multilevel models. *Journal of Applied Statistics*, 43(12):2150–2171. Citado na pág. [2](#)
- Massuia et al.(2015)** Monique B. Massuia, Celso R. B. Cabral, Larissa A. Matos e Víctor H. Lachos. Influence diagnostics for Student-t censored linear regression models. *Statistics*, 49(5):1074–1094. Citado na pág. [31](#)
- Matos et al.(2013)** Larissa A. Matos, Marcos O. Prates, Ming-Hui Chen e Victor H. Lachos. Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica*, 23(3):1323–1345. Citado na pág. [31](#)
- Melo et al.(2017)** Tatiane F. N. Melo, Silvia L. P. Ferrari e Alexandre G. Patriota. Improved hypothesis testing in a general multivariate elliptical model. *Journal of Statistical Computation and Simulation*, 87(7):1416–1428. Citado na pág. [3](#), [30](#), [40](#)
- Melo et al.(2018)** Tatiane F. N. Melo, Silvia L. P. Ferrari e Alexandre G. Patriota. Improved estimation in a general multivariate elliptical model. *Brazilian Journal of Probability and Statistics*, 32(1):44–68. Citado na pág. [3](#), [14](#), [29](#), [37](#)

- Muggeo e Lovison(2014)** Vito MR Muggeo e Gianfranco Lovison. The “three plus one” likelihood-based test statistics: unified geometrical and graphical interpretations. *The American Statistician*, 68(4):302–306. Citado na pág. 15
- Neyman e Pearson(1928)** J. Neyman e E. S. Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, 20(1/2):175–240. Citado na pág. 14
- Patriota(2011)** Alexandre G. Patriota. A note on influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics & Data Analysis*, 55(1):218–225. Citado na pág. 2
- Patriota(2019)** Alexandre G. Patriota. On the mean value theorem for estimating functions. *The American Statistician*. doi: 10.1080/00031305.2018.1558110. Citado na pág. 14, 35, 54, 63
- Patriota e Lemonte(2009)** Alexandre G. Patriota e Artur J. Lemonte. Bias correction in a multivariate normal regression model with general parameterization. *Statistics & Probability Letters*, 79(15):1655–1662. Citado na pág. 3, 7, 8, 26, 30
- Patriota et al.(2009)** Alexandre G. Patriota, Heleno Bolfarine e Mário de Castro. A heteroscedastic structural errors-in-variables model with equation error. *Statistical Methodology*, 6(4):408–423. Citado na pág. 19, 20
- Patriota et al.(2011)** Alexandre G. Patriota, Artur J. Lemonte e Heleno Bolfarine. Improved maximum likelihood estimators in a heteroskedastic errors-in-variables model. *Statistical Papers*, 52(2):455–467. Citado na pág. 2
- Press et al.(1992)** W. Press, S. Teukolsky, W. Vetterling e B. Flannery. *Numerical recipes in C: The art of scientific computing*. Cambridge University Press, Cambridge, England. Citado na pág. 9, 40
- R Core Team(2018)** R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL <https://www.R-project.org/>. Citado na pág. 9
- Rao(1948)** Calyampudi R. Rao. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 44(1):50–57. Citado na pág. 14
- Rudin(1976, Teorema 9.25, pag. 241)** Walter Rudin. *Principles of mathematical analysis*. Number 4.2. New York: McGraw-Hill Book Company. Citado na pág. 49
- Russo et al.(2009)** Cibele M. Russo, Gilberto A. Paula e Reiko Aoki. Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics & Data Analysis*, 53(12):4143–4156. Citado na pág. 2, 14, 31
- Sen e Singer(1993)** Pranab K. Sen e Julio M. Singer. *Large Sample Methods in Statistics: An Introduction with Applications*. Chapman & Hall, New York. Citado na pág. 3, 14, 15, 54
- Skovgaard(2001)** Ib M. Skovgaard. Likelihood asymptotics. *Scandinavian Journal of Statistics*, 28:3–32. Citado na pág. 3
- Terrell(2002)** George R. Terrell. The gradient statistic. *Computing Science and Statistics*, 34: 206–215. Citado na pág. 14, 15
- Tiede e Pagano(1979)** James J. Tiede e Marcello Pagano. The application of robust calibration to radioimmunoassay. *Biometrics*, 35(3):567–574. Citado na pág. 37, 41
- Van der Vaart(2000)** Aad W. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge. Citado na pág. 14

- Vanegas e Paula(2015)** Luis H. Vanegas e Gilberto A. Paula. A semiparametric approach for joint modeling of median and skewness. *Test*, 24(1):110–135. Citado na pág. [31](#)
- Vanegas e Paula(2016)** Luis H. Vanegas e Gilberto A. Paula. Log-symmetric distributions: statistical properties and parameter estimation. *Brazilian Journal of Probability and Statistics*, 30(2):196–220. Citado na pág. [31](#)
- Vargas et al.(2013)** Tiago M. Vargas, Silvia L. P. Ferrari e Artur J. Lemonte. Gradient statistic: Higher-order asymptotics and Bartlett-type correction. *Electronic Journal of Statistics*, 7:43–61. Citado na pág. [15](#)
- Wald(1943)** Abraham Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3): 426–482. Citado na pág. [14](#)
- Zhu et al.(1997)** Ciyou Zhu, Richard H. Byrd, Peihuang Lu e Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560. Citado na pág. [9](#)