

**Modelos lineares parciais aditivos com
erros simétricos condicionais autorregressivos
e penalização com *splines* cúbicos**

Rodrigo Alves de Oliveira

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

Programa: Doutorado em Estatística
Orientador: Prof. Dr. Gilberto Alvarenga Paula

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

São Paulo, maio de 2021

Modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos e penalização com *splines* cúbicos

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 20/05/2021. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Gilberto Alvarenga Paula (orientador) - IME-USP
- Prof^a. Dr^a. Airlane Pereira Alencar - IME-USP
- Prof^a. Dr^a. Fernanda De Bastiani - DE-UFPE
- Prof. Dr. Manuel Galea - PUC-Chile
- Prof^a. Dr^a. Tatiane Ferreira do Nascimento Melo da Silva - IME-UFG

Agradecimentos

Agradeço antes de tudo aos meus orientadores, **Tatiane Melo** (IME-UFG), **Helton Saulo** (IME-UFG), **Getúlio Amaral** (DE-UFPE) e **Gilberto Paula** (IME-USP), por esses 10 anos de muito aprendizado e pelas reflexões científicas. Ao orientador e Professor Dr. **Gilberto Alvarenga Paula** por toda sua dedicação, seus ensinamentos e orientações que foram fundamentais para a conclusão deste trabalho e na minha formação como Estatístico. Pelo seu cuidado profissional e por sua presteza em todas as revisões deste trabalho, que foram fundamentais para o resultado obtido. Além do meu orientador, gostaria de agradecer a comissão julgadora da tese, por seus comentários e sugestões de diferentes perspectivas.

Agradeço à minha mãe, **Fernanda Alves**, e ao meu pai **Jávier Oliveira** (*in memoriam*), por todo o apoio, carinho e companheirismo fornecido durante toda a minha vida. À minha namorada **Suzan Silva** por toda a compreensão e carinho nos momentos mais difíceis para a conclusão deste trabalho. A toda minha família e amigos que sempre me animaram nos momentos mais complicados.

Agradeço aos professores e funcionários do Departamento de Estatística do Instituto de Matemática e Estatística da Universidade de São Paulo, por seus exemplos profissionais, orientação e apoio durante meu doutorado. A todos os meus amigos de graduação, pós-graduação e as pessoas que encontrei durante minha estada em Recife e São Paulo, que me ajudaram academicamente e emocionalmente, pelo apoio, amizade, e por suas contribuições em diferentes discussões teóricas, simulações e aplicações. E, por fim, agradeço à CAPES pelo suporte financeiro.

Resumo

Oliveira, R. A. **Modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos e penalização com *splines* cúbicos**. 2021. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2021.

Neste trabalho propomos os modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos e penalização com *splines* cúbicos de regressão para modelar tendência e sazonalidade em séries temporais. As funções de verossimilhança penalizada, escore penalizada e matriz de informação de Fisher penalizada são obtidas, bem como um processo iterativo do tipo *backfitting* (Gauss-Seidel) é desenvolvido para obter as estimativas de máxima verossimilhança penalizada dos componentes paramétrico e não paramétrico, alternando com o procedimento usual do algoritmo de Quase-Newton, em particular o método BFGS, e a generalização para problemas de grande porte, o chamado método L-BFGS para obter as estimativas dos parâmetros de dispersão e de autocorrelação. Em seguida, apresentamos métodos de diagnóstico, tais como análise de resíduos, e influência local sob cinco esquemas de perturbação: ponderação de casos, parâmetro de dispersão, variável resposta, variável explicativa contínua e coeficiente de autocorrelação. Estudos de simulação são desenvolvidos supondo diferentes distribuições para os erros e sob modelos condicionais autorregressivos de ordem um e dois, diante de diversos cenários. Avaliamos as propriedades empíricas dos estimadores dos parâmetros de localização, dispersão e autocorrelação. A abordagem apresentada é discutida em quatro aplicações. O primeiro exemplo refere-se ao banco de dados de anomalia da temperatura média global anual na superfície terra-mar no período de 1880 a 2020. O segundo exemplo refere-se ao banco de dados sobre a temperatura média diária na cidade de São Francisco nos Estados Unidos no período de janeiro de 1995 a abril de 2020. O terceiro exemplo refere-se aos dados diários de poluição e meteorológicos na estação da Marginal Tietê na Ponte dos Remédios na cidade de São Paulo no período de janeiro de 2014 a dezembro de 2020. E, o quarto, e último, aborda a mortalidade cardiovascular média semanal no condado de Los Angeles no período de 1970 a 1979.

Palavras-chave: erros autorregressivos, erros simétricos, métodos de diagnóstico, modelos semi-paramétricos, regressão penalizada, *splines* cúbicos de regressão.

Abstract

Oliveira, R. A. **Additive partial linear models with autoregressive conditional symmetric errors and penalized cubic spline**. 2021. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2021.

In this thesis we propose additive partial linear models with autoregressive conditional symmetric errors based on penalized cubic regression splines for modeling trend and seasonality in time series. A penalized log-likelihood function as well as the penalized score and penalized Fisher information matrix are derived. A backfitting (Gauss-Seidel) iterative process is developed for estimating the additive components jointly with the usual procedure of the Quasi-Newton algorithm, in particular the BFGS method, and the generalization for large problems, the so-called L-BFGS method to obtain the dispersion and autocorrelation estimates. Diagnostic methods, such residual analysis and the local influence approach with five perturbation schemes, namely case-weight, dispersion parameter, response variable, continuous explanatory variable and autocorrelation coefficient, are developed. Simulation studies are performed to assess the large sample behavior of the maximum penalized likelihood estimators. Finally, the methodology is illustrated with four applications. The first example refers to the anomaly database of annual global average temperature on land-sea surface in the period from 1880 to 2020. The second application is on the modeling of the average daily temperature in the city of San Francisco in the United States, from January 1995 to April 2010. The third application discusses the modeling of daily pollution data available from Ponte dos Remédios station in São Paulo city from January 2014 to December 2020. And, the fourth, and last, approach the average weekly cardiovascular mortality in Los Angeles County in the period from 1970 to 1979.

Keywords: autoregressive errors, cubic regression splines, diagnostic procedures, penalized regression, semiparametric models, symmetric errors.

Sumário

Lista de Abreviaturas	xi
Lista de Símbolos	xiii
Lista de Figuras	xv
Lista de Tabelas	xix
1 Introdução	1
1.1 Motivação	3
1.2 Organização do trabalho	3
2 Conceitos	5
2.1 Regressão paramétrica e não paramétrica	8
2.2 Erros simétricos	15
2.2.1 Modelo de regressão paramétrico linear com erros simétricos	17
2.2.2 Modelo parcialmente linear com erros simétricos	20
2.3 Modelo com erros autorregressivos de primeira ordem	23
3 Modelos aditivos com erros simétricos condicionais AR(1)	27
3.1 O modelo	28
3.2 Função de verossimilhança penalizada	31
3.3 Estimação dos parâmetros	31
3.3.1 Função escore penalizada	32
3.3.2 Matriz Hessiana penalizada	32
3.3.3 Matriz de informação de Fisher penalizada	33
3.3.4 Processo iterativo	34
3.4 Procedimentos inferenciais	35
3.5 Graus de liberdade efetivos	36
3.6 Comentários	37
4 Modelos lineares parciais aditivos com erros simétricos condicionais AR(p)	39
4.1 O modelo	40
4.2 Função de verossimilhança penalizada	43
4.3 Estimação dos parâmetros	43
4.3.1 Função escore penalizada	43

4.3.2	Matriz Hessiana penalizada	44
4.3.3	Matriz de informação de Fisher penalizada	45
4.3.4	Processo iterativo	53
4.4	Procedimentos inferenciais	55
4.4.1	Erro padrão aproximado	55
4.4.2	Intervalos de confiança simultâneos	55
4.4.3	Testes de hipóteses	56
4.5	Graus de liberdade efetivos	56
4.6	Comentários	57
5	Análise de Diagnóstico	59
5.1	Análise de resíduos	60
5.2	Influência local	61
5.2.1	Perturbação ponderação de casos	62
5.2.2	Perturbação no parâmetro de dispersão	63
5.2.3	Perturbação na variável resposta	63
5.2.4	Perturbação na variável explicativa	64
5.2.5	Perturbação no coeficiente de autocorrelação	64
5.3	Seleção de modelos	65
5.4	Estimação do parâmetro de suavização	65
6	O Pacote - APLMS	67
7	Simulações	75
7.1	Simulação AR(1)	76
7.2	Simulação AR(2)	78
8	Aplicações	81
8.1	Anomalia da temperatura média global na superfície terra-mar	81
8.1.1	Modelagem	82
8.2	Temperatura em São Francisco	86
8.2.1	Análise descritiva	87
8.2.2	Modelagem	90
8.3	Partículas de poluição no município de São Paulo	96
8.3.1	Análise descritiva	97
8.3.2	Modelagem	100
8.4	Mortalidade cardiovascular	108
8.4.1	Análise descritiva	108
8.4.2	Modelagem	110
9	Conclusões	117
9.1	Considerações finais	117
9.2	Sugestões para pesquisas futuras	117

A	Função escore, matriz Hessiana e informação de Fisher: caso AR(1)	119
A.1	Função escore penalizada	119
A.2	Matriz Hessiana penalizada	120
A.3	Matriz informação de Fisher penalizada	122
B	Função escore, Matriz Hessiana e informação de Fisher: caso AR(p)	125
B.1	Função escore penalizada	125
B.2	Matriz Hessiana penalizada	126
B.3	Matriz informação de Fisher penalizada	129
C	Análise de sensibilidade	133
C.1	Ponderação de casos	133
C.2	Perturbação ponderada no parâmetro de dispersão	135
C.3	Perturbação na variável resposta	136
C.4	Perturbação na variável explicativa	139
C.5	Perturbação no coeficiente de autocorrelação	142
D	Análises Simulações	145
D.1	Tabelas e gráficos da simulação AR(1)	145
D.2	Tabelas e gráficos da simulação AR(2)	154
	Referências Bibliográficas	165

Lista de Abreviaturas

VC	Validação cruzada
VCG	Validação cruzada generalizada
AIC	Critério de Informação de Akaike - <i>Akaike information criterion</i>
BIC	Critério de Informação Bayesiano - <i>Bayesian information criterion</i>
EMVP	Estimador de máxima verossimilhança penalizado
MLPAS	Modelos lineares parciais aditivos com erros simétricos
APLMS	<i>Additive partial linear models with symmetric errors</i>
CETESB	Companhia Ambiental do Estado de São Paulo
df_s	Graus de liberdade efetivos dos suavizadores
df	Graus de liberdade efetivos

Lista de Símbolos

$g(\cdot)$	Função geradora de densidade da distribuição simétrica
ϕ	Parâmetro de dispersão
ξ	Constante positiva que depende da distribuição simétrica
ϵ	Erro autocorrelacionado
θ	Parâmetros do modelo
$L(\theta)$	Logaritmo da função de verossimilhança regular associada a θ
$L_p(\theta, \lambda)$	Logaritmo da função de verossimilhança penalizada
U_θ	Função escore
$\ddot{L}_{\theta\theta}$	Matriz Hessiana
$\mathcal{I}_{\theta\theta}$	Matriz de informação de Fisher
$\Gamma(\cdot)$	Função gama
$\text{Beta}(\cdot, \cdot)$	Função beta
$\psi'(\cdot)$	Função trigama
Υ	matriz de variância-covariância de ϵ

Lista de Figuras

2.1	Ilustração de dados com padrão horizontal.	6
2.2	Ilustração de diferentes tipos de tendência em modelos de séries temporais.	7
2.3	Ilustração de série temporal com padrão sazonal.	7
2.4	Ilustração de séries temporais para diferentes valores do termo de autocorrelação ρ	24
4.1	Autocorrelção teórica ϱ_s para o processo AR(2) com $\epsilon_i = \rho_1\epsilon_{i-1} + \rho_2\epsilon_{i-2} + e_i$	50
4.2	Autocorrelação teórica ϱ_s para o processo AR(2) com $\epsilon_i = \rho_1\epsilon_{i-1} + \rho_2\epsilon_{i-2} + e_i$	51
7.1	Ilustração da distribuição dos erros e_i	76
7.2	Disposição das funções não paramétricas teóricas para a tendência (cima) e sazonalidade (baixo), para os tamanhos amostrais iguais a 100, 300 e 500.	77
7.3	Disposição das funções não paramétricas teóricas para os componentes f_1 (cima) e f_2 (baixo), para os tamanhos amostrais iguais a 100, 300 e 500.	78
8.1	Ilustração da série temporal da anomalia da temperatura média global ($^{\circ}\text{C}$) anual na superfície terra-mar.	82
8.2	Histograma (primeiro-esquerdo), gráfico de índices dos resíduos (primeiro-direito), função de autocorrelação (segundo-esquerdo), função de autocorrelação parcial (segundo-direito), probabilidade normal (terceiro-esquerdo), e peso \hat{v}_i contra os resíduos (terceiro-direito) do modelo ajustado (8.1) para a anomalia da temperatura média global na superfície terrestre-oceânica sob erros com distribuição exponencial potência com $\nu = -0,33$	84
8.3	Gráficos de índices C_i nos esquemas de perturbação, ponderação de casos (primeiro-esquerdo), parâmetro de dispersão (primeiro-direito), variável resposta (segundo-esquerdo) e coeficiente de autocorrelação (ρ_1) (segundo-direito) do modelo (8.1) para a anomalia da temperatura média global na superfície terra-mar sob erros com distribuição dos erros exponencial potência com $\nu = -0,33$	85
8.4	Bandas de confiança pontuais para a tendência do modelo (8.1) sob erros com distribuição exponencial potência com $\nu = -0,33$ ajustado à anomalia da temperatura média global anual na superfície terra-mar ($^{\circ}\text{C}$).	86
8.5	Série temporal e distribuição empírica da temperatura média ($^{\circ}\text{C}$) diária em São Francisco de Janeiro de 1995 a Abril de 2020.	87
8.6	<i>Boxplot</i> e gráfico de densidades por mês da série temporal da temperatura média ($^{\circ}\text{C}$) diária em São Francisco de Janeiro de 1995 a Abril de 2020.	88

8.7	Média mensal da série de temperatura média diária de São Francisco, ao longo dos anos (cima), gráfico dos meses em relação aos anos individuais (meio) e dados de cada estação coletados juntos em mini gráficos de tempos separados (baixo).	89
8.8	Gráficos de dispersão defasados para a média mensal da temperatura média ($^{\circ}\text{C}$) diária em São Francisco de Janeiro de 1995 a Abril de 2020.	90
8.9	Gráficos de índices dos resíduos (primeiro), gráficos de probabilidade normal (segundo), função de autocorrelação - FACs (terceiro), função de autocorrelação parcial - FACP (quarto) e peso das observações contra os resíduos quantílicos condicionais do modelo (8.2) sob distribuições de erro normal (esquerda), t -Student (centro) e exponencial potência (direita) ajustadas à temperatura média diária de São Francisco.	94
8.10	Gráficos de índices de C_i no esquema de perturbação ponderação de casos para o modelo (8.2) e também pela partição particular de $\hat{\mathbf{f}}$, $\hat{\phi}$ e $\hat{\rho}$ sob erros com distribuições normal (esquerda), t -Student (centro) e exponencial potência (direita) ajustadas à temperatura média diária de São Francisco.	95
8.11	Bandas de confiança pontuais para os componentes de sazonalidade (esquerda) e de tendência (direita) do modelo (8.2) sob erros com distribuição t -Student com $\nu = 5$ graus de liberdade ajustado à temperatura média diária de São Francisco.	96
8.12	Histograma das variáveis $\text{MP}_{2,5}$ (cima) e MP_{10} (baixo) e respectiva transformação \log (direita).	98
8.13	Fator médio de impacto das variáveis $\text{MP}_{2,5}$ (cima) e MP_{10} (baixo) por ano(esquerda), mês(meio) e dia da semana(direita).	99
8.14	Matriz de correlação das variáveis de poluição e meteorológicas na estação da Marginal Tietê na Ponte dos Remédios.	100
8.15	Impacto médio do ano, meses do ano e dias da semana em relação aos materiais particulados $\text{MP}_{2,5}$ (esquerda) e MP_{10} (direita) sob erros com distribuição t -Student com $\nu = 5$ graus de liberdade.	104
8.16	Gráficos de índices dos resíduos (primeiro), probabilidade normal (segundo), função de autocorrelação (terceiro), função de autocorrelação parcial (quarto) e peso das observações contra os resíduos quantílicos condicionais dos modelos ajustados (8.3) e (8.4) para os materiais particulados $\log(\text{MP}_{2,5})$ (esquerda) e $\log(\text{MP}_{10})$ (direita) sob erros com distribuição t -Student com $\nu = 5$.	106
8.17	Gráficos de índices C_i nos esquemas de perturbação, ponderação de casos (primeiro), parâmetro de dispersão (segundo), variável resposta (terceiro), variável explicativa (CO) (quarto) e coeficiente de autocorrelação (ρ_1) (quinto) dos modelos (8.3) e (8.4) com distribuição dos erros t -Student com $\nu = 5$ graus de liberdade para $\log(\text{MP}_{2,5})$ (esquerda) e $\log(\text{MP}_{10})$ (direita).	107
8.18	Mortalidade cardiovascular média semanal (parte superior), temperatura (parte central) e poluição por partículas de poluição (parte inferior) no Condado de Los Angeles. Existem 508 médias suavizadas de seis dias obtidas pela filtragem de valores diários ao longo do período de 10 anos 1970-1979.	109
8.19	Diagrama de dispersão mostrando a relação entre mortalidade cardiovascular média semanal e as variáveis temperatura (primeiro) e partículas de poluição (segundo) no Condado de Los Angeles.	110

8.20 Histograma (primeiro-esquerdo), gráficos de índices dos resíduos (primeiro-direito), função de autocorrelação (segundo-esquerdo), função de autocorrelação parcial (segundo-direito), probabilidade normal (terceiro-esquerdo), e peso das observações contra os resíduos quantílicos condicionais do modelo ajustado (8.6) para a mortalidade cardiovascular média semanal sob erros com distribuição t -Student com $\nu = 16$ graus de liberdade. 113

8.21 Gráficos de índices C_i nos esquemas de perturbação, ponderação de casos (primeiro-esquerdo), parâmetro de dispersão (segundo-esquerdo), variável resposta (terceiro-esquerdo), variável explicativa tempo (quarto-esquerdo), variável explicativa temperatura (primeiro-direito), variável explicativa partículas (segundo-direito), coeficiente de autocorrelação (ρ_1) (terceiro-direito) e (ρ_2) (quarto-direito) do modelo (8.6) para a mortalidade cardiovascular sob erros com distribuição t -Student com $\nu = 16$ graus de liberdade. 115

8.22 Bandas de confiança pontuais para a sazonalidade do modelo (8.6) sob erros com distribuição t -Student com $\nu = 16$ graus de liberdade ajustado à mortalidade cardiovascular média semanal. 116

D.1 Gráficos da média das estimativas para a função $f_1(t)$ sob $\phi = 1$ com coeficientes de autocorrelação $\rho = -0,75, -0,25, 0,25$ e $0,75$ a partir do estudo de simulação a qual os dados foram gerados do modelo (7.1) sob os erros com distribuição normal (primeiro), t_3 (segundo), PE_1 (terceiro), PE_2 (quarto), NG (quinto) e L_2 (sexto) e ajustado sob o mesmo erro do modelo. Tamanho amostrais de $n = 100, 300$ e 500 ,na esquerda, meio e direita, respectivamente. 152

D.2 Gráficos da média das estimativas para a função $f_2(s)$ sob $\phi = 1$ com coeficientes de autocorrelação $\rho = -0,75, -0,25, 0,25$ e $0,75$ a partir do estudo de simulação a qual os dados foram gerados do modelo (7.1) sob os erros com distribuição normal (primeiro), t_3 (segundo), PE_1 (terceiro), PE_2 (quarto), NG (quinto) e L_2 (sexto) e ajustado sob o mesmo erro do modelo. Tamanho amostrais de $n = 100, 300$ e 500 ,na esquerda, meio e direita, respectivamente. 153

D.3 Gráficos da média das estimativas para a função $f_1(t)$ sob $\phi = 1,3$ com coeficientes de autocorrelação $\rho_1 = 0,656$ e $\rho_2 = -0,309$ a partir do estudo de simulação a qual os dados foram gerados do modelo (8.5) sob os erros com distribuição normal (primeiro), t_3 (segundo), PE_1 (terceiro), PE_2 (quarto), NG (quinto) e L_2 (sexto) e ajustado sob o mesmo erro do modelo. Tamanho amostrais de $n = 100, 300$ e 500 ,na esquerda, meio e direita, respectivamente. 162

D.4 Gráficos da média das estimativas para a função $f_2(s)$ sob $\phi = 1,3$ com coeficientes de autocorrelação $\rho_1 = 0,656$ e $\rho_2 = -0,309$ a partir do estudo de simulação a qual os dados foram gerados do modelo (8.5) sob os erros com distribuição normal (primeiro), t_3 (segundo), PE_1 (terceiro), PE_2 (quarto), NG (quinto) e L_2 (sexto) e ajustado sob o mesmo erro do modelo. Tamanho amostrais de $n = 100, 300$ e 500 ,na esquerda, meio e direita, respectivamente. 163

Lista de Tabelas

2.1	Definição das funções base e matrizes usadas para definir o <i>spline</i> cúbico de regressão. $h_j = x_{j+1} - x_j$.	13
2.2	Função geradora de densidade de probabilidade e ξ_g para algumas distribuições simétricas	17
2.3	Expressões para $W_g(\delta)$ e $W'_g(\delta)$ para algumas distribuições simétricas	18
2.4	Valores de d_g e f_g para algumas distribuições simétricas.	20
6.1	Resumo das funções no pacote APLMS	67
8.1	Qualidade de ajuste para o modelo (8.1) sob erros com distribuição normal, normal generalizada e exponencial potência ajustados aos dados de anomalia da temperatura média global anual da superfície terra-mar.	82
8.2	Qualidade de ajuste para o modelo (8.2) sob erros com distribuição normal, t -Student e Exponencial Potência ajustado à temperatura média diária de São Francisco.	91
8.3	Estimativas dos parâmetros (erros padrão aproximados) do modelo (8.2) sob erros com distribuições normais, t -Student e exponencial potência com estruturas AR(1), AR(2) e AR(3) ajustado à temperatura média diária de São Francisco.	92
8.4	Estatísticas resumidas dos poluentes e meteorológicos do ar na estação Marginal Tietê na Ponte dos Remédios, São Paulo, de janeiro de 2014 a dezembro de 2020.	97
8.5	Qualidade de ajuste para os modelos (8.3) e (8.4) sob erros com distribuição normal, t -Student e Exponencial Potência ajustados aos dados de materiais particulados na estação marginal Tietê, Ponte dos Remédios.	101
8.6	Estatísticas resumidas da mortalidade cardiovascular média semanal, temperatura e partículas de poluição no condado de Los Angeles no período de 1970 a 1979.	108
8.7	Qualidade de ajuste para o modelo (8.6) sob erros com distribuição normal, t -Student, exponencial potência e Logística-II ajustados à mortalidade cardiovascular média semanal no condado de Los Angeles, entre 1970 e 1979.	111
8.8	Estimativas dos parâmetros (erros padrão aproximados) do modelo (8.6) sob erros com distribuições normal, t -Student, exponencial potência e Logística-II com estrutura AR(2) ajustado à mortalidade cardiovascular média semanal no condado de Los Angeles, entre 1970 e 1979.	111

D.1 Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro normal e ajustados sobre a distribuição dos erros normal, t_3 , PE₁ e PE₂, NG e L₂. 146

D.2 Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro t -Student ($\nu = 3$) e ajustados sobre a distribuição dos erros normal, t_3 , PE₁ e PE₂, NG e L₂. . . 147

D.3 Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro exponencial potência ($\nu = -0,3$) e ajustados sobre a distribuição dos erros normal, t_3 , PE₁ e PE₂, NG e L₂. 148

D.4 Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro exponencial potência ($\nu = 0,5$) e ajustados sobre a distribuição dos erros normal, t_3 , PE₁ e PE₂, NG e L₂. 149

D.5 Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro normal generalizada e ajustados sobre a distribuição dos erros normal, t_3 , PE₁ e PE₂, NG e L₂. 150

D.6 Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro L₂ e ajustados sobre a distribuição dos erros normal, t_3 , PE₁ e PE₂, NG e L₂. 151

D.7 Percentual de acerto pelo método AIC no estudo de simulação no qual os dados são gerados sob a distribuição dos erros, normal, t_3 , PE₁ e PE₂, NG e L₂. 155

D.8 Média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro normal e ajustados sobre a distribuição dos erros normal, t_3 , PE₁, PE₂, NG e L₂. 156

D.9 Média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro t -Student ($\nu = 3$) e ajustados sobre a distribuição dos erros normal, t_3 , PE₁, PE₂, NG e L₂. 157

D.10 Média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro exponencial potência ($\nu = -0,3$) - PE₁ e ajustados sobre a distribuição dos erros normal, t_3 , PE₁, PE₂, NG e L₂. 158

- D.11 Média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro exponencial potência ($\nu = 0,5$) - PE₂ e ajustados sobre a distribuição dos erros normal, t_3 , PE₁, PE₂, NG e L₂. 159
- D.12 Média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro normal generalizada e ajustados sobre a distribuição dos erros normal, t_3 , PE₁, PE₂, NG e L₂. 160
- D.13 Média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro L₂ e ajustados sobre a distribuição dos erros normal, t_3 , PE₁, PE₂, NG e L₂. 161

Capítulo 1

Introdução

Os avanços na tecnologia revolucionaram a maneira como processamos as informações e elaboramos as pesquisas ambientais, socioeconômicas e de negócios. Esses avanços na teoria e na prática são uma resposta à crescente complexidade e competitividade nos diversos campos de estudo. A complexidade aumenta o risco associado às decisões, tornando importante ter uma boa base de informações.

O objetivo das pesquisas científicas é fornecer aos estudiosos, formadores de políticas e empresários, de maneira semelhante, informações que facilitem a tomada de decisão. Praticamente toda organização, pública ou privada, opera em um ambiente incerto e dinâmico, com conhecimento imperfeito do futuro. A pesquisa científica é parte integrante do sistema de planejamento e controle, e as organizações precisam de um procedimento de estudo que lhes permitam entender e prever o futuro de maneira eficaz e oportuna.

Modelos de regressão são técnicas estatísticas aplicadas com frequência que servem de base para o estudo e caracterização de um sistema de interesse. Utiliza-se análise de regressão para formular um modelo matemático razoável da relação entre as variáveis preditoras e a variável resposta do sistema. Existem diversos tipos de modelos, e a escolha de um pode ser baseada no conhecimento prévio do sistema ou em considerações como suavidade e continuidade da resposta e das variáveis preditoras.

No contexto em que as observações são obtidas ou medidas ao longo do tempo, a ordem dos dados é fundamental. Neste tipo de dados existe de certa forma uma dependência com os acontecimentos anteriores ou posteriores. Logo, é de interesse modelar essa dependência. Não obstante, tem-se a modelagem e previsão de séries temporais as quais são um problema matemático com aplicações em diversos setores a fim de realizar um melhor planejamento e auxiliar na tomada de decisões de diretores empresariais e gestores públicos. O principal objetivo da modelagem de séries temporais é coletar e estudar as observações passadas de uma série temporal para desenvolver um modelo apropriado que descreva suas características. Esse modelo é usado para prever valores futuros ou entender a relação com outras séries.

Uma série temporal é um conjunto sequencial de pontos de dados, medido tipicamente em tempos sucessivos. Esta é dita ser contínua quando as observações são feitas continuamente no tempo, define-se o conjunto $T = \{t : t_1 < t < t_2\}$ a série temporal será denotada por $\{x(t) : t \in T\}$. Uma série temporal é dita ser discreta quando as observações são feitas em tempos específicos, geralmente equiespaçados. Matematicamente define-se o conjunto $T = \{t_1, \dots, t_n\}$ a série temporal será denotada por $\{x(t) : t \in T\}$. Por simplicidade considera-se $T = \{0, 1, 2, \dots\}$, ou seja, $\{x(t), t = 0, 1, 2, \dots\}$, em que t representa o tempo decorrido. Nesta tese, consideramos o tempo discretizado em que as observações são feitas em tempos específicos. As medidas em uma série temporal são organizadas em uma ordem cronológica adequada. Os sistemas do mundo real exibem comportamentos não lineares e não estacionários. Atualmente, existem muitos objetos descritos por séries temporais com tendências de comportamento desconhecidas, componentes sazonais e estocásticos, o que complica significativamente a aquisição de um modelo preditivo eficaz.

Em diversas situações os dados apresentam atipicidade que, de certa forma, são inconsistentes com as demais observações. Deste modo, os modelos tradicionais que lidam com as suposições de normalidade dos erros não são satisfeitos. Isso pode comprometer as inferências usuais, como, por exemplo, estimativas viesadas dos parâmetros, teste de hipóteses errôneos, ou identificação incorreta do modelo. Nesse momento, cabe ao pesquisador adotar ou abordar análises mais robustas nos estudos em desenvolvimento. No contexto da análise de regressão, existem os modelos com erros simétricos (Cysneiros e Paula, 2005).

A família simétrica de distribuições permite uma extensão da distribuição normal para modelagem estatística de dados reais envolvendo distribuições com caudas mais pesadas e mais leves do que as da distribuição normal. As distribuições que possuem caudas pesadas apresentam valores distantes das medidas de localização (medidas de tendência central, ou medidas de posição como média, mediana e moda) com probabilidade maior do que a distribuição normal. Caudas pesadas ocorrem, por exemplo, quando a variância (medidas de dispersão indicam a variabilidade dos dados como variância, desvio padrão) é alta, até mesmo infinita como é o caso da distribuição Cauchy. Também estão associadas a ocorrência de valores discrepantes ou extremos (*outliers*), entendidos como valores muito distantes daqueles das demais observações. Embora um valor discrepante possa resultar de um erro de medida, também pode ser um resultado genuíno, indicando um comportamento extremo da variável, que merece ser estudado, e não removido.

Dentre os principais trabalhos realizados sobre modelos simétricos temos Galea *et al.* (2003, 2005), Vanegas e Cysneiros (2010) que estudaram diagnósticos estatísticos em modelos lineares e não lineares com erros simétricos, respectivamente, enquanto que Cysneiros *et al.* (2007, 2010) desenvolveram métodos de diagnóstico para modelos lineares e não lineares heterocedásticos simétricos. Considerando erros simétricos autoregressivos, por exemplo, Paula *et al.* (2009) discutiram estimação e diagnóstico em modelos lineares com erros elípticos AR(1), Liu (2004) desenvolveu métodos de diagnóstico em modelos de séries temporais heterocedásticos condicionais sob erros elípticos e Cao *et al.* (2010) apresentaram procedimentos para avaliar a heterocedasticidade e autocorrelação, bem como métodos de diagnóstico em modelos não lineares com erros simétricos AR(1).

No entanto, pode-se melhorar a qualidade do ajuste dos modelos lineares com erros AR(1), levando em consideração a relação entre tempo e resposta no componente sistemático, e como esse tipo de forma geralmente não é linear, componentes não paramétricos parecem ser mais flexíveis que os paramétricos, sendo assim, indicados. Tais modelos, denominados modelos parcialmente lineares, foram investigados sob erros normais (ver, por exemplo, Greene (2012)) e, mais recentemente, sob erros simétricos independentes (ver, por exemplo, Ibacache-Pulgar e Paula (2011); Ibacache-Pulgar *et al.* (2013)), com ênfase especial nos procedimentos de estimação e diagnóstico.

Os resultados obtidos em Ibacache-Pulgar e Paula (2011) e Ibacache-Pulgar *et al.* (2013) foram estendidos por Relvas e Paula (2016) para o caso em que os erros são representados pela forma AR(1). Embora a definição do modelo esteja correta, em nosso trabalho damos ênfase na decomposição da série temporal, no primeiro momento, em dois componentes, tendência e sazonalidade. No que se refere à parte não paramétrica $f(\cdot)$ do modelo, ao invés de utilizar *spline* cúbico natural, ao qual devemos estimar r parâmetros referentes aos valores distintos da variável explicativa, representamos $f(\cdot)$ em uma forma aditiva. Desse modo, estamos ajustando um modelo paramétrico em que $f(\cdot)$ assume uma forma flexível paramétrica. Quando temos muitas variáveis explicativas contínuas é praticamente impossível ajustar uma função $f(\cdot)$ para cada covariável. Assim, a principal aplicabilidade de usar $f(\cdot)$ é controlar efeitos importantes tais como tempo, temperatura, umidade, e outros em modelos com componentes paramétricos. Uma vantagem com essa abordagem é diminuir o número de parâmetros a serem estimados pela função $f(\cdot)$. No segundo momento, apresentamos um procedimento de regressão paramétrico e não paramétrico combinado que acomoda facilmente transformações lineares dos dados e, portanto, fornece uma estrutura conveniente para a análise desse problema.

Os modelos lineares parciais aditivos são uma extensão natural do modelo não paramétrico com erros simétricos, além disso, generalizar a ordem 1 para ordem p dos erros autorregressivos permite uma maior flexibilização da modelagem estatística. São adotadas funções não paramétricas

para estudar a relação entre a variável resposta e as variáveis explicativas permitindo, assim, uma maior flexibilidade e não impondo uma forma rígida de dependência na modelagem das variáveis em questão.

1.1 Motivação

A influência das atividades humanas no clima da Terra é bastante discutida e pode trazer mudanças neste século. Na verdade, de acordo com dados da indústria de resseguros, o número de desastres relacionados ao clima aumentou significativamente desde 1970 (Cheema *et al.*, 2011). Mudanças climáticas têm sido para muitos países uma das maiores ameaças ambientais à produção de alimentos, disponibilidade de água, biodiversidade florestal e aos meios de subsistências (Chung *et al.*, 2011).

Por um lado, entender a natureza e a escala de possíveis mudanças climáticas nas cidades pode ser de interesse dos formuladores de políticas e das pessoas que nelas trabalham. Isso pode dar a chance de estarem preparados para melhores medidas de mitigação e adaptação. Por outro lado, a análise e previsão de séries temporais tornou-se uma ferramenta ampla em várias aplicações hidrometeorológicas para estudar tendências e mudanças de chuvas, umidade, temperatura, vazão e muitas outras variáveis ambientais (Machiwal e Jha (2009); Soltani *et al.* (2007)). Trabalhos recentes na determinação de probabilidades de mudança de temperatura global incluem, por exemplo, Cahill *et al.* (2015), Mikkonen *et al.* (2015) e Rahmstorf *et al.* (2017).

Considerando o trabalho realizado por Relvas e Paula (2016) em que os autores modelam a relação da série dependente de temperatura média da cidade de Ubatuba e a série de Cananéia, cidades do Estado de São Paulo, e uma função do tempo para controlar possíveis efeitos do tempo, pode-se observar que não há o cuidado em decompor a série de acordo com uma possível tendência e sazonalidade. Além disso, a função é definida pelo *spline* cúbico natural, em grandes amostras, o número de nós que representa r valores distintos da variável explicativa cresce indefinidamente. Sendo assim, o custo computacional aumenta e adotar um *spline* que tenha o número de nós fixos é mais coerente.

O objetivo desta tese é criar e discutir estratégias robustas para analisar uma série temporal e entender os fenômenos que a acometem. No primeiro momento, estendemos os modelos lineares parciais com erros simétricos condicionais autorregressivos de primeira ordem por considerar o preditor linear com dois componentes não paramétricos, para modelar a tendência e a sazonalidade como sugerido por Wood (2017, na Seção 7.7.2). Ao contrário do trabalho proposto por Relvas e Paula (2016), os componentes não paramétricos são formados por *splines* cúbicos e cíclicos com nós fixos ao invés de *splines* cúbicos naturais. Logo, em seguida, realizamos uma generalização do modelo, ou seja, desenvolvemos os modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p . Apresentamos as análises dos resíduos e, por último, uma análise de diagnóstico, com uma nova concepção dos coeficientes de autocorrelação.

Os resultados deste trabalho visam auxiliar pesquisadores de qualquer área, mesmo aqueles com pouca experiência em modelagem estatística, na análise de dados com dependência temporal. O desenvolvimento adequado desse tipo de estudo é essencial para criação de políticas públicas e de uma gestão empresarial que possam promover melhorias na tomada de decisão de governos e empresários.

1.2 Organização do trabalho

O trabalho foi desenvolvido na seguinte sequência. No capítulo 2, apresentamos os conceitos de *spline* cúbico penalizado de regressão e sua versão cíclica, regressão com erros simétricos e com erros simétricos autorregressivos de primeira ordem. No capítulo 3, propomos os modelos lineares aditivos com erros simétricos condicionais autorregressivos de primeira ordem decompondo a série em dois componentes: tendência e sazonalidade com o uso de *splines* com nós fixos. No capítulo 4, estendemos para os modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de

ordem p . No capítulo 5, discutimos métodos de seleção de modelos e métodos de diagnóstico, análise de resíduos e análise de sensibilidade. No capítulo 6, descrevemos a biblioteca `aplms` desenvolvida em R, a qual será disponibilizada no *Comprehensive R Archive Network* (CRAN). No capítulo 7, realizamos simulações de dois modelos AR(1) e AR(2) para diferentes distribuições dos erros, e discutimos o problema de má especificação na escolha do modelo. No capítulo 8, realizamos quatro aplicações para ilustrar o uso dos modelos em estudo. No capítulo 9, apresentamos as conclusões do trabalho e nos apêndices resumimos algumas metodologias utilizadas no texto.

Capítulo 2

Conceitos

Modelos estatísticos são representações simplificadas de dados, geralmente construídos a partir de alguns relacionamentos matemáticos ou numericamente definidos. A análise de dados fornece uma escolha extremamente valiosa de técnicas de modelagem. Todas essas técnicas se beneficiam de algumas ideias gerais sobre dados e modelos, assim, podemos expressar quais dados devem ser usados nos modelos e quais relacionamentos o modelo postula sobre os dados.

Modelos estatísticos reproduzem as propriedades de outros objetos “reais”, mas de uma forma mais simples ou conveniente. Fazemos inferências a partir dos modelos e as aplicamos aos verdadeiros objetos, para os quais as mesmas inferências seriam impossíveis ou inconvenientes. As diferenças entre os modelos e a realidade são os resíduos, geralmente, são a chave para alcançar maior entendimento e um melhor modelo. Um bom modelo reproduz com maior precisão possível as propriedades relevantes do verdadeiro objeto, além de ser conveniente de usar e apresentar interpretações do fenômeno em estudo. Qualquer bom modelo deve facilitar inferências precisas e convenientes. O custo de criar ou usar um modelo também nos limita em alguns casos. Os modelos estatísticos permitem inferências sobre um objeto, atividade ou processo, modelando alguns dados observáveis associados.

A conveniência dos modelos estatísticos depende, é claro, da aplicação e dos tipos de inferência que os pesquisadores necessitam. Os critérios geralmente aplicados incluem simplicidade; por exemplo, um modelo é mais simples se requer menos parâmetros ou variáveis explicativas. Um modelo que usa muitas variáveis, além de peso e deslocamento, teria que nos recompensar com previsões substancialmente mais precisas, especialmente se as variáveis adicionais fossem mais difíceis de medir. Menos quantificável, mas extremamente importante é que o modelo corresponda, tanto quanto possível, a conceitos ou teorias que o pesquisador possui sobre o verdadeiro objeto.

A modelagem estatística é um processo de vários estágios que envolve, frequentemente, o uso repetido dos seguintes passos: 1- obtenção de dados adequados para representar o processo a ser modelado; 2- escolha de um modelo o qual será utilizado para descrever alguma relação nos dados; 3- ajuste do modelo, geralmente, estimando os parâmetros; 4- avaliação do modelo para ver o que diz sobre os dados; 5- uso de diagnósticos para ver de que maneira relevante o modelo não se encaixa tão bem quanto deveria. Os resumos e diagnósticos podem envolver tabelas, descrições verbais e exibições gráficas. Isso pode sugerir que o modelo falha em prever ou explicar todas as propriedades relevantes dos dados, ou podemos considerar um modelo mais simples que pode ser quase tão bom quanto um ajuste mais completo. Em qualquer um dos casos, o modelo será modificado, e o ajuste e a análise serão executados novamente.

Ao analisar observações obtidas sequencialmente ao longo do tempo, deve-se atentar aos modelos e ajustes de séries temporais, uma vez que a ordem dos dados é essencial. Além disso, as observações vizinhas são dependentes, e o interesse é analisar e modelar essa dependência. De um modo geral, os padrões dos dados servem como um guia com duas finalidades diferentes. Primeiro, uma simples observação dos dados mostrará ao pesquisador como eles se comportam. A natureza desse comportamento serve como um guia para especular sobre o padrão dos dados no futuro. Segundo, o padrão dos dados pode sugerir a existência de um relacionamento entre duas ou mais

variáveis. Nesse caso, os dados históricos de uma única variável não fornecem informações sobre o padrão subjacente. Um exemplo desse padrão de dados é a relação entre o consumo de bens em uma economia e o nível de renda da população. À medida que o nível de renda aumenta, espera-se que o consumo de bens e serviços também aumente.

Ao coletar dados para fazer uma previsão, o pesquisador geralmente enfrenta um dos seguintes padrões nos dados: horizontal, tendência, sazonal e cíclico. Quando não há tendência nos dados, o pesquisador está lidando com um padrão de dados horizontal. Isso implica que as observações não tendem a aumentar ou diminuir de forma sistemática. Em termos estatísticos, nos referimos a isso como um padrão estacionário. Em tal situação, o pesquisador tem a mesma probabilidade de observar o próximo valor da série acima ou abaixo do valor estacionário. A duração do tempo é o fator mais crítico ao decidir se um padrão horizontal é observado nos dados. Quanto menor o horizonte de tempo, maior a probabilidade de observar um padrão horizontal nos dados. A Figura 2.1 mostra um padrão horizontal típico para uma variável Y_t ao longo do tempo t .

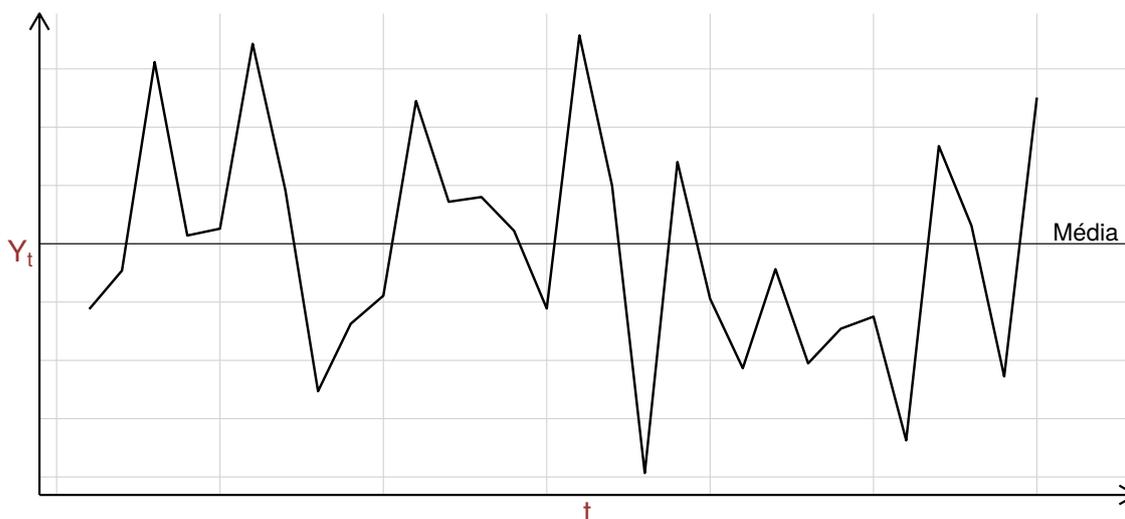


Figura 2.1: Ilustração de dados com padrão horizontal.

O padrão de tendência é um movimento de crescimento a longo prazo de uma série temporal. A tendência pode ser ascendente, descendente ou estável. Pode ser representada como um gráfico linear ou não linear suave. Nos negócios e na economia, associamos padrões de tendência à evolução das preferências do consumidor, às mudanças nos desenvolvimentos tecnológicos, ao produto nacional bruto, às mudanças na renda, aos preços das ações, à expansão industrial e às mudanças nas políticas fiscais do governo. Dependendo da natureza da variável, o padrão varia de linear, como mostrado na Figura 2.2(a), para outras formas não lineares, como mostrado nas Figuras 2.2(b)-2.2(f). Em um padrão de tendência linear, a variável de interesse cresce uma quantidade absoluta constante em cada período de tempo (t).

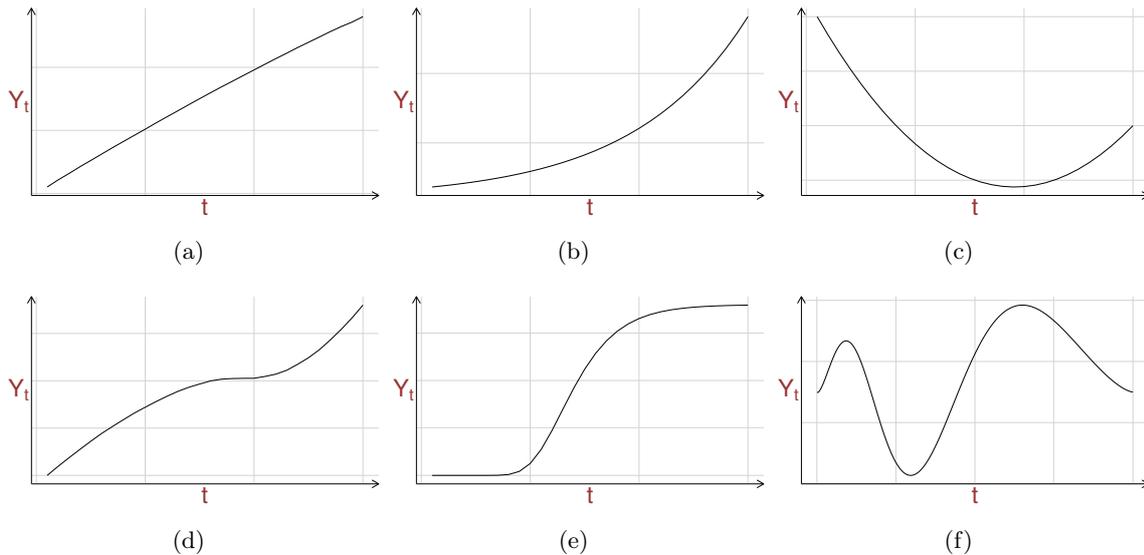


Figura 2.2: Ilustração de diferentes tipos de tendência em modelos de séries temporais.

Um padrão sazonal nos dados é observado quando um movimento previsível e repetitivo é notado em torno de uma linha de tendência dentro de um determinado período (anos, meses, horas). Isso significa que, para podermos analisar as variações sazonais, precisamos ter dados relatados semanalmente, mensalmente, trimestralmente, etc. Os padrões sazonais existem por vários motivos. Quando uma empresa decide reportar lucros e perdas trimestralmente, nos referimos a isso como um padrão sazonal induzido internamente, enquanto fatores como o clima provocam uma “estação” induzida externamente. Exemplos de padrões sazonais de dados estão associados a vendas de roupas, óleo para aquecimento e o número de carros novos vendidos durante um determinado período do ano, devido a mudanças no modelo. A Figura 2.3 ilustra um padrão no qual uma variação sazonal é observada.

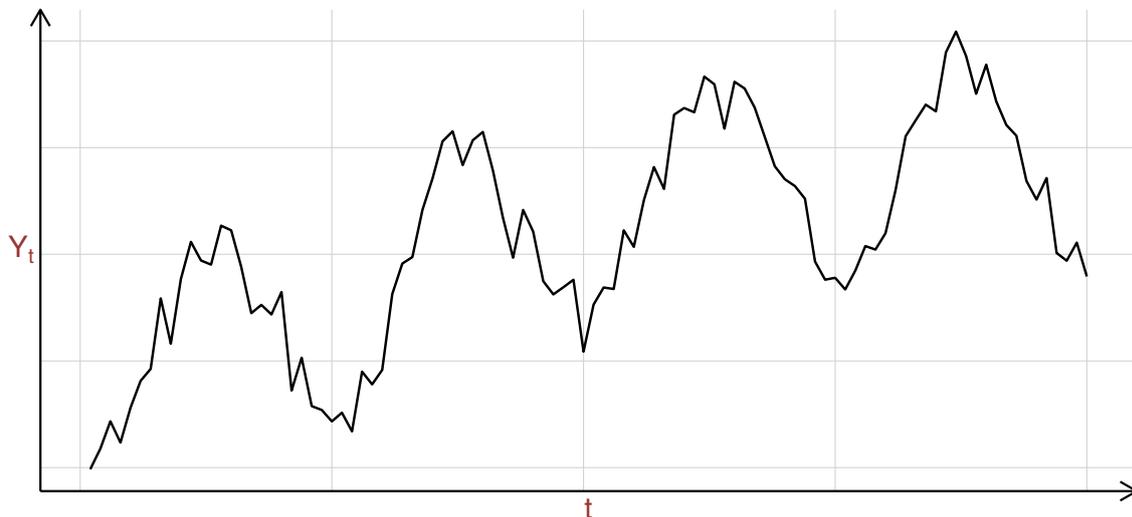


Figura 2.3: Ilustração de série temporal com padrão sazonal.

Um padrão cíclico ocorre com expansões e contrações comerciais e econômicas. Embora existam semelhanças com um padrão sazonal, os movimentos cíclicos variam em duração, geralmente durando mais de um ano. Os padrões cíclicos diferem em intensidade ou amplitude e estão correlacionados com ciclos de negócios em uma economia nacional. Na previsão, os padrões cíclicos são mais difíceis de prever, porque os ciclos não se repetem a intervalos constantes de tempo. Portanto, é impossível distinguir entre comportamento cíclico e tendência em suave evolução. Os analistas dedicaram quantidades substanciais de tempo e esforço para prever ciclos com mais precisão. No

entanto, esses esforços não foram tão frutíferos quanto se esperava. Pesquisas e desenvolvimentos metodológicos contínuos podem lançar mais luz sobre o assunto.

Além dos padrões acima, os dados de séries temporais também podem exibir o que é chamado de padrão autocorrelacionado. Esse padrão mostra simplesmente que os dados em um período de tempo estão relacionados ao seu valor em períodos anteriores. Quando confrontado com um padrão de dados autocorrelacionados, o analista sabe que existe uma correlação automática entre as observações em uma série. Isso significa que, se houver uma autocorrelação positiva alta, o valor em junho estará positivamente relacionado ao valor em maio, para dados observados mensalmente. Fatores como comportamento do consumidor, tendências e sazonalidade são importantes e devem ser considerados na observação de padrões de dados autocorrelacionados.

2.1 Regressão paramétrica e não paramétrica

A análise de regressão é um ramo da estatística que caracteriza os relacionamentos de forma paramétrica ou não paramétrica entre diferentes variáveis aleatórias de uma amostra ou população. A regressão paramétrica assume várias suposições a respeito dos dados brutos, como um relacionamento linear subjacente e uma distribuição específica para o erro, frequentemente, assume-se a distribuição normal ou gaussiana. A regressão não paramétrica reduz ou elimina estas suposições de linearidade dos dados e permite que conclusões mais gerais sejam feitas para uma variedade maior de dados que pode levar a um melhor ajuste.

Embora a regressão paramétrica seja amplamente usada nas pesquisas científicas e, também, seja o ramo da estatística mais frequentemente ensinado nas aulas introdutórias, nem sempre é o mais apropriado. Se os dados falharem em seguir a distribuição normal, as conclusões realizadas serão baseadas em suposições falsas e frequentemente imprecisas e enganosas. Se os dados, por exemplo, não forem suficientemente “normais”, existem algumas opções. Se a distribuição verdadeira, como, por exemplo, uma exponencial, for conhecida, existem testes e modelos paramétricos específicos para essa distribuição que podem ser aplicados.

Outra opção comum é fazer uma transformação dos dados e ver se esses são normais ou seguem outra distribuição conhecida. Caso os dados ainda não sejam normais ou facilmente identificáveis como uma distribuição conhecida específica, serão necessárias estatísticas não paramétricas para a análise. Como o nome indica, esse ramo da estatística usa menos suposições sobre os dados, como assumir apenas uma distribuição contínua.

No geral, as conclusões são mais rigorosas devido às suposições enfraquecidas sobre a distribuição dos dados. No entanto, como consequência dessas suposições, a distribuição e os modelos de regressão aplicados são geralmente mais complicados e difíceis de interpretar. Em outras palavras, para dados que estão próximos de uma distribuição conhecida, métodos não paramétricos adicionam complexidade desnecessária. Isso é indesejável, uma vez que o pesquisador almeja um modelo simples e capaz de interpretar termos e partes de seu modelo. Apesar dessa complexidade aumentada, os modelos não paramétricos geralmente são uma boa escolha devido ao aumento da flexibilidade e aplicabilidade aos dados.

Regressão linear é uma das técnicas estatísticas mais antigas e amplamente usadas. Considere um par de dados $(x_i, y_i), i = 1, \dots, n$, então tem-se o modelo da forma

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (2.1)$$

para os dados observados. A regressão linear é frequentemente aplicada aos dados sem um objetivo específico em mente. É útil, no entanto, identificar dois dos principais objetivos para os quais a regressão linear é vantajosa. A distinção entre estes não é de todo rígida e, regularmente, os dois se aplicam. O primeiro objetivo da regressão é fornecer um resumo ou redução dos dados observados, a fim de explorar e explicar a relação entre a variável independente x e a variável resposta y . É óbvio e natural, quando determinado gráfico de dados exibe uma tendência linear aproximada, desenhar uma linha reta para enfatizar essa tendência. A regressão linear automatiza esse procedimento e garante comparabilidade e consistência dos resultados, além de permitir uma interpretação dos

parâmetros, em particular da inclinação β_1 . Outro objetivo da regressão é usar o modelo (2.1) para predição: dado qualquer ponto x dentro do intervalo amostral, uma estimativa do valor esperado de uma nova observação y no ponto x é dada por $\hat{\beta}_0 + \hat{\beta}_1 x$, em que $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimativas de β_0 e β_1 . Embora a predição seja, sem dúvida, um aspecto importante da regressão, é provavelmente uma reflexão muito mais precisa da prática estatística considerar a regressão principalmente como um método baseado em modelos para resumo de dados.

A base de regressão fundamentada em modelos a distingue um pouco das ideias mais exclusivamente baseadas em dados subjacentes a técnicas mais puramente exploratórias, como as apresentadas por Tukey (1977). A adoção de uma abordagem pautada em um modelo tem vantagens e desvantagens. Uma vantagem é que a metodologia pode ser estendida a uma ampla variedade de estruturas de dados alternativos, por exemplo, ao longo das linhas de modelos lineares generalizados, como discutido por McCullagh e Nelder (1989).

Existem muitos conjuntos de dados em que é claramente inapropriado ajustar um modelo de linha reta da forma (2.1), sendo assim, um modelo mais apropriado é dado por

$$y_i = f(x_i) + \epsilon_i \quad (2.2)$$

em que f é alguma curva qualquer. Em geral, um modelo de regressão paramétrico assume que a forma de f é conhecida, exceto para muitos parâmetros desconhecidos. A forma específica de f pode vir de teorias científicas e/ou aproximações sob algumas suposições simplificadas. As suposições podem ser restritivas e as aproximações podem ser grosseiras para algumas aplicações. Um modelo inadequado pode levar a um viés sistemático e a conclusões enganosas. Na prática, deve-se sempre verificar a forma assumida para a função f .

Muitas vezes é difícil, se não impossível, obter uma forma funcional específica para f . Um modelo de regressão não paramétrico, em geral, não assume uma forma predeterminada. Em vez disso, faz-se suposições sobre propriedades qualitativas de f . Por exemplo, alguém pode estar disposto a assumir que f é “suave”, o que não se reduz a uma forma específica com número finito de parâmetros. Desse modo, geralmente leva a algumas coleções dimensionais infinitas de funções. A ideia básica da regressão não paramétrica é deixar os dados falarem por si mesmo. Isso permite que os dados possam se encaixar melhor em alguma função sem impor nenhuma forma específica a f . Consequentemente, os métodos não paramétricos são em geral mais flexíveis. Eles podem descobrir a estrutura nos dados que, de outra forma, poderia ser perdida.

A exposição simples acima indica que a técnica de regressão não paramétrica pode ser aplicada a diferentes etapas da análise de regressão: exploração de dados, construção de modelos, teste de modelos paramétricos e diagnóstico. De fato, a suavização de *splines* é uma ferramenta poderosa e versátil para criar modelos estatísticos para explorar estruturas dos dados. Um *spline* é uma representação matemática para a qual é fácil construir uma interface que permitirá ao usuário projetar e controlar a forma de curvas e superfícies complexas. A abordagem geral é que o usuário insere uma sequência de pontos (nós) e é construída uma curva cuja forma segue de perto essa sequência. Os pontos são chamados de pontos de controle. Uma curva que realmente passa por cada ponto de controle é chamada de curva de interpolação; uma curva que passa perto dos pontos de controle, mas não necessariamente por eles, é chamada de curva de aproximação.

Considere, primeiro, o que aconteceria se tentássemos ajustar um modelo da forma (2.2) por mínimos quadrados, sem colocar restrições na curva f . E então, é claro, que a soma residual dos quadrados pode ser reduzida a zero escolhendo f para interpolar os dados fornecidos. A colocação de condições de suavidade em f não faz nenhuma diferença essencial. Um bom ajuste aos dados não é único, o importante, é o bom ajuste das curvas. Outro objetivo, muitas vezes conflitante, é obter uma estimativa de curva que não exiba flutuações súbitas. A ideia básica da abordagem da penalidade é quantificar a noção de uma curva que flutua rapidamente e, em seguida, colocar o problema de estimativa de uma maneira que explique o compromisso necessário entre os dois objetivos bastante diferentes na estimativa de curvas.

O modelo permite uma especificação flexível da dependência da resposta das covariáveis, mas especificando o modelo apenas em termos de ‘funções suaves’, em vez de relacionamentos paramétri-

cos detalhados, é possível evitar uma espécie de modelos pesados e complicados. Essa flexibilidade e conveniência costumam dois novos problemas teóricos. É necessário que ambas representem as funções suaves de alguma maneira e escolher o quão suaves elas devem ser.

Existem diversos métodos para representar f , algumas vezes requer-se que esta seja representada de tal forma que (2.2) torna-se um modelo linear, ou seja, representamos f em uma forma aditiva. De tal forma que estamos ajustando um modelo paramétrico em que f assume uma forma flexível paramétrica. Quando temos muitas variáveis explicativas contínuas é praticamente impossível ajustar uma função f para cada covariável. Assim, a principal aplicabilidade de usar f é controlar efeitos importantes tais como tempo, temperatura, umidade, e outros, em modelos com componentes paramétricos. Isso pode ser feito ao escolher uma *base*, definindo o espaço da função no qual f (ou uma aproximação fechada para isso) é um elemento. Escolher uma base significa escolher algumas funções base, as quais serão tratadas como completamente conhecidas: se $b_j(x)$ é a j -ésima função base, então f é assumida com a seguinte representação:

$$f(x) = \sum_{j=1}^r b_j(x)\beta_j \quad (2.3)$$

para valores dos parâmetros desconhecidos, β_j . Substituindo (2.3) em (2.2), produz-se claramente um modelo linear. A seguir, vamos especificar algumas bases para o modelo: polinomial, base linear por partes, *spline* cúbico natural e *spline* cúbico. Esta última será utilizada no decorrer da tese, uma vez que possui vantagens diante das demais, principalmente, menor custo computacional.

Polinomial: como um exemplo simples, suponha que f é comportada a ser polinomial de 4ª ordem, então o espaço dos polinômios é de ordem 4 e contém f . Uma base para esse espaço é $b_1(x) = 1$, $b_2(x) = x$, $b_3(x) = x^2$, $b_4(x) = x^3$ e $b_5(x) = x^4$, então de (2.3) tem-se que

$$f(x) = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4,$$

e o modelo (2.2) torna-se

$$y_i = \beta_0 + \beta_1x_i + \beta_2x_i^2 + \beta_3x_i^3 + \beta_4x_i^4 + \epsilon_i.$$

A abordagem clássica é usar para f um polinômio de ordem baixa em que os coeficientes são estimados por mínimos quadrados. Essa abordagem é amplamente usada na prática e facilmente implementada usando a abordagem de regressão múltipla.

Regressão polinomial é uma técnica popular, mas há diversas desvantagens. Um delas é que observações individuais podem exercer uma influência, de maneira inesperada, em partes remotas da curva. Outra dificuldade é que a elaboração implícita do modelo no aumento do grau polinomial ocorre em etapas discretas e não pode ser controlada continuamente. Um terceiro ponto, não necessariamente uma desvantagem, é que os polinômios de um determinado grau formam uma classe de modelos de dimensão finita ‘com arestas duras’ e pode haver alguma vantagem em permitir que os dados determinem o modelo ajustado de uma maneira um pouco mais flexível.

Base linear por partes: segundo Wood (2017), uma base para funções lineares por partes de uma variável univariada x é determinada inteiramente pelos locais das derivadas descontínuas da função, ou seja, pelos locais nos quais as partes lineares se juntam. Sejam esses nós denotados por $\{x_j^* : j = 1, \dots, r\}$, e suponha que $x_j^* > x_{j-1}^*$, então para $j = 2, \dots, r - 1$

$$b_j(x) = \begin{cases} (x - x_{j-1}^*)/(x_j^* - x_{j-1}^*), & x_{j-1}^* < x \leq x_j^*, \\ (x_{i+j}^* - x)/(x_{j+1}^* - x_j^*), & x_j^* < x < x_{j+1}^*, \\ 0, & \text{caso contrário,} \end{cases}$$

enquanto que

$$b_1(x) = \begin{cases} (x_2^* - x)/(x_2^* - x_1^*), & x < x_2^*, \\ 0, & \text{caso contrário,} \end{cases}$$

e

$$b_r(x) = \begin{cases} (x - x_{r-1}^*) / (x_r^* - x_{r-1}^*), & x > x_{r-1}^*, \\ 0, & \text{caso contrário.} \end{cases}$$

Portanto, $b_j(x)$ é zero em todos os lugares, exceto no intervalo entre os nós imediatamente para ambos os lados de x_j^* . Além disso, $b_j(x)$ aumenta linearmente de 0 em x_{j-1}^* a 1 em x_j^* e depois diminui linearmente a 0 em x_{j+1}^* . Funções base como essa, que são diferentes de zero apenas em alguns intervalos finitos, têm suporte compacto. Note que uma maneira exatamente equivalente de definir $b_j(x)$ é como a interpolação linear dos dados $\{x_i^*, \tau_i^j : i = 1, \dots, r\}$, em que $\tau_i^j = 1$ se $i = j$ e zero caso contrário. Usando esta base para representar $f(x)$, (2.2) torna-se o modelo linear $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, em que $X_{ij} = b_j(x_i)$.

A escolha do grau de suavidade do modelo, aqui controlado pela dimensão da base, r , pode ser arbitrária e levar a modelos e conclusões errôneas. Um modelo organizado em nós espaçados uniformemente nos quantis, $r - 1$, geralmente, não será aninhado dentro de um modelo baseado em r nós espaçados de modo uniforme (igualmente espaçados). É possível começar com uma grade de nós e simplesmente eliminar sequencialmente, como parte da seleção *backward*, mas o espaçamento desigual resultante entre nós pode levar a um desempenho ruim do modelo. Além disso, o ajuste de tais modelos de regressão tende a depender fortemente dos locais escolhidos para os nós.

Uma alternativa é manter a dimensão da base fixada em um tamanho um pouco maior do que se acredita ser razoavelmente necessário, mas controlar a suavidade do modelo adicionando uma penalidade “ondulada” ao objetivo de ajuste por mínimos quadrados. Para mais detalhes veja Wood (2017). Essa técnica oferece uma maneira perfeitamente razoável de representar as funções suaves em modelos aditivos, mas é possível uma melhora substancial. Em particular, se representarmos os termos do modelo suave usando bases de *spline*, é possível obter um erro de aproximação de função consideravelmente reduzido para uma dada dimensão da base de suavização.

Splines cúbicos: uma das primeiras técnicas são os *splines* cúbicos naturais os quais são interpoladores mais suaves. Veja em Green e Silverman (1994) mais detalhes e provas dessa técnica baseada no original de Schoenberg (1964). Além disso, Relvas e Paula (2016) desenvolveram sua teoria sob essa perspectiva. A propriedade de interpolação mais suave não é a única boa propriedade dos interpoladores de *splines* cúbicos. Em De Boor *et al.* (1978, Cap. 5) são apresentados vários resultados mostrando que a interpolação cúbica de *splines* é ótima, ou pelo menos muito boa, em diversos aspectos.

Essas propriedades dos interpolantes de *splines* sugerem que esses devem fornecer uma boa base para representar termos suaves em modelos estatísticos. Qualquer que seja a verdadeira função suave subjacente, um *spline* deve ser capaz de aproximá-la, e se quisermos construir modelos a partir de funções suaves de covariáveis, então representar essas funções com base em aproximações suaves é intuitivamente atraente.

Particularmente no contexto de regressão, é natural que qualquer medida de suavidade não seja afetada pela adição de uma função constante ou linear, de modo que se duas funções diferem apenas por uma função constante ou linear, sua suavidade deve ser idêntica. Isso leva naturalmente à ideia de uma suavidade funcional que depende da segunda derivada da curva em consideração. Claro, alguém poderia considerar o máximo de $|f''|$ ou o número de pontos de inflexão em f . Mas, dada uma curva f definida em um intervalo $[a, b]$, a maneira intuitivamente atraente de medir a suavidade de uma curva duas vezes diferenciável f é calcular sua segunda derivada quadrada integrada $\int_a^b \{f''(x)\}^2 dx$, a qual é uma medida global de suavidade que possui, como veremos, vantagens computacionais consideráveis.

Dessa forma, em trabalhos estatísticos, y_i é usualmente calculado com ruído, e isso geralmente é mais útil ao suavizador x_i , ao invés de interpolar os dados y_i . Para esse fim, em vez de tomar $f(x_i) = y_i$, pode ser melhor tratar $f(x_i)$ como n parâmetros livres do *spline* cúbico, e estimá-los em ordem a minimizar

$$\sum_{i=1}^n [y_i - f(x_i)]^2 + \lambda \int_a^b f''(x)^2 dx, \quad (2.4)$$

em que λ é um parâmetro ajustável (*tunable*) - ou parâmetro de penalização, usado para controlar o peso relativo a ser dado aos objetivos conflitantes de correspondência dos dados e produção de uma função f suave. A função $f(x)$ resultante é um *spline* de suavização (Reinsch, 1967).

Portanto, em vez de ser escolhida antecipadamente, a base do *spline* cúbico surge naturalmente da especificação do objetivo de suavização (2.4), no qual o que se entende por ajuste do modelo e suavidade é precisamente definido de maneira independente da base. Suavizar *splines* parece ser pouco ideal. O único problema substancial é o fato de eles terem tantos parâmetros livres quanto os dados a serem suavizados. Isso é um desperdício, já que, na prática, λ quase sempre será alto o suficiente para que o *spline* resultante seja mais suave do que os n graus de liberdade poderiam sugerir. De fato, muitos graus de liberdade de um *spline* são frequentemente suprimidos completamente pela penalidade. Para suavização univariada com *splines* cúbicos, o grande número de parâmetros acaba por não ser problemático (por exemplo, De Hoog e Hutchinson (1986) fornecem um algoritmo $O(n)$ estável), mas assim que tentamos lidar com mais covariáveis o custo computacional torna-se grave.

Um compromisso óbvio entre reter as boas propriedades de *splines* e a eficiência computacional é usar *splines* penalizados de regressão. Na sua forma mais simples, isso envolve a construção de uma base de *spline* (e penalidades associadas) para um conjunto de dados menor do que será analisado e, em seguida, o uso dessa base (mais penalidades) para modelar o conjunto de dados original. Os valores de covariáveis no conjunto de dados menor devem ser organizados para cobrir bem a distribuição dos valores de covariáveis no conjunto de dados original. Essa ideia de *splines* penalizados de regressão é apresentada em Wahba (1980) e Parker e Rice (1985), por exemplo.

Por fim, existem muitas bases equivalentes que podem ser usadas para representar *splines* cúbicos. Aqui iremos descrever os *splines* cúbicos penalizados de regressão, e sua versão cíclica, desenvolvidas por Wood (2017). Uma abordagem é parametrizar o *spline* em termos de valores nos nós. Considere definir uma função de *spline* cúbico, $f(x)$, com r nós fixos, x_1, \dots, x_r , seja $\beta_j = f(x_j)$ e $\zeta_j = f''(x_j)$. Dessa forma, o *spline* pode ser escrito como,

$$f(x) = a_j^-(x)\beta_j + a_j^+(x)\beta_{j+1} + c_j^-(x)\zeta_j + c_j^+(x)\zeta_{j+1}, \quad \text{se } x_j \leq x \leq x_{j+1}, \quad (2.5)$$

em que as funções base a_j^- , a_j^+ , c_j^- e c_j^+ estão definidas na Tabela 2.1. As condições em que o *spline* deve ser contínuo na segunda derivada, em x_j , e deve ter segunda derivada zero em x_1 e x_r , implica que

$$\mathbf{B}\boldsymbol{\zeta}^- = \mathbf{D}\boldsymbol{\beta},$$

em que $\boldsymbol{\zeta}^- = (\zeta_2, \dots, \zeta_{r-1})^\top$, $\zeta_1 = \zeta_r = 0$ e \mathbf{B} e \mathbf{D} são definidos na Tabela 2.1. Defina $\mathbf{F}^- = \mathbf{B}^{-1}\mathbf{D}$, e

$$\mathbf{F} = \begin{bmatrix} \mathbf{0} \\ \mathbf{F}^- \\ \mathbf{0} \end{bmatrix},$$

em que $\mathbf{0}$ é uma linha de zeros, temos que $\boldsymbol{\zeta} = \mathbf{F}\boldsymbol{\beta}$. Portanto, o *spline* pode ser reescrito na forma linear, em termos de $\boldsymbol{\beta}$ como

$$f(x) = a_j^-(x)\beta_j + a_j^+(x)\beta_{j+1} + c_j^-(x)\mathbf{F}_j\boldsymbol{\beta} + c_j^+(x)\mathbf{F}_{j+1}\boldsymbol{\beta}, \quad \text{se } x_j \leq x \leq x_{j+1},$$

a qual se resume em (2.3) da seguinte maneira,

$$f(x) = \sum_{j=1}^r b_j(x)\beta_j.$$

Portanto, dado um conjunto de valores x nos quais avalia-se o *spline*, é fácil obter um modelo de mapeamento da matriz $\boldsymbol{\beta}$ para estimá-lo. Também pode ser mostrado, por exemplo, em Lancaster e Salkauskas (1986) que

$$\int_{x_1}^{x_r} f''(x)^2 dx = \boldsymbol{\beta}^\top \mathbf{D}^\top \mathbf{B}^{-1} \mathbf{D} \boldsymbol{\beta},$$

em que $\mathbf{M} \equiv \mathbf{D}^\top \mathbf{B}^{-1} \mathbf{D}$ é a matriz penalidade para as bases. Essa base possui interpretação direta dos parâmetros, não requer algum tipo de reescala das variáveis preditoras, é preciso escolher os locais dos nós, x_j .

Tabela 2.1: Definição das funções base e matrizes usadas para definir o *spline* cúbico de regressão.
 $h_j = x_{j+1} - x_j$.

Funções base para o <i>spline</i> cúbico		
$a_j^-(x) = (x_{j+1} - x)/h_j$	$c_j^-(x) = [(x_{j+1} - x)^3/h_j - h_j(x_{j+1} - x)]/6$	
$a_j^+(x) = (x - x_j)/h_j$	$c_j^+(x) = [(x - x_j)^3/h_j - h_j(x - x_j)]/6$	
Elementos da matriz não-zero - <i>spline</i> não cíclico		
$D_{i,i} = 1/h_i$	$D_{i,i+1} = -1/h_i - 1/h_{i+1}$	$D_{i,i+2} = 1/h_{i+1}$
$B_{i,i} = (h_i + h_{i+1})/3$		$i = 1, \dots, k-2$
$B_{i,i+1} = h_{i+1}/6$	$B_{i+1,i} = h_{i+1}/6$	$i = 1, \dots, k-3$
Elementos da matriz não-zero - <i>spline</i> cíclico		
$\tilde{B}_{i-1,i} = \tilde{B}_{i,i-1} = h_{i-1}/6$	$\tilde{B}_{i,i} = (h_{i-1} + h_i)/3$	
$\tilde{D}_{i-1,i} = \tilde{D}_{i,i-1} = 1/h_{i-1}$	$\tilde{D}_{i,i} = -1/h_{i-1} - 1/h_i$	$i = 2, \dots, k-1$
$\tilde{B}_{1,1} = (h_{k-1} + h_1)/3$	$\tilde{B}_{1,k-1} = h_{k-1}/6$	$\tilde{B}_{k-1,1} = h_{k-1}/6$
$\tilde{D}_{1,1} = -1/h_1 - 1/h_{k-1}$	$\tilde{D}_{1,k-1} = 1/h_{k-1}$	$\tilde{D}_{k-1,1} = 1/h_{k-1}$

Fonte: extraído de Wood (2017, p. 201)

Muitas vezes, é apropriado que uma função suave do modelo seja “cíclica”, o que significa que a função tem o mesmo valor e as primeiras derivadas nos limites superior e inferior. Por exemplo, na maioria das aplicações, não seria apropriado que uma função suave no decorrer do ano fosse alterada descontinuamente no final do ano. O *spline* cúbico penalizado de regressão anterior pode ser modificado para produzir tal suavidade. O *spline* pode ser escrito da forma (2.5), mas temos que $\beta_1 = \beta_r$ e $\zeta_1 = \zeta_r$. Nesse caso, definimos o vetor $\boldsymbol{\beta}^\top = (\beta_1, \dots, \beta_{r-1})$ e $\boldsymbol{\zeta}^\top = (\zeta_1, \dots, \zeta_{r-1})$. As condições em que o *spline* deve ser contínuo na segunda derivada em cada nó, e que $f(x_1)$ deve corresponder a $f(x_r)$ até a segunda derivada, são equivalentes a

$$\tilde{\mathbf{B}}\boldsymbol{\zeta} = \tilde{\mathbf{D}}\boldsymbol{\beta},$$

em que $\tilde{\mathbf{B}}$ e $\tilde{\mathbf{D}}$ são definidos na Tabela 2.1. Por razões similares, pode-se escrever o *spline* da seguinte maneira:

$$f(x) = \sum_{j=1}^{r-1} \tilde{b}_j(x)\beta_j.$$

A segunda derivada penalizada é dada por

$$\int_{x_1}^{x_r} f''(x)^2 dx = \boldsymbol{\beta}^\top \tilde{\mathbf{D}}^\top \tilde{\mathbf{B}}^{-1} \tilde{\mathbf{D}}\boldsymbol{\beta},$$

em que $\tilde{\mathbf{M}} \equiv \tilde{\mathbf{D}}^\top \tilde{\mathbf{B}}^{-1} \tilde{\mathbf{D}}$ é a matriz penalidade para as bases. Essa base possui interpretação direta dos parâmetros, não requer um tipo de reescala das variáveis preditoras e é preciso escolher os locais dos nós, x_j .

Em relação ao *spline* penalizado de regressão tem-se que muitas escolhas aparentemente arbitrárias podem influenciar o ajuste. O *trade-off* típico está definido entre viés e variância. Deseja-se

fazer seleções de modo a evitar sobreajuste (*overfitting*) ou sub-ajuste (*underfitting*). Sabe-se que ao adicionar mais nós tem-se maior flexibilidade, mas isso pode potencialmente levar ao sobreajuste. O ajuste é influenciado pelo número e localização dos nós aos quais são incluídos. Além disso, há uma terceira variável, mais direta, que influencia o ajuste: a adição explícita da penalidade. Em suma, almeja-se selecionar as localizações dos nós e o parâmetro de suavização λ (penalidade) que melhor captura a forma subjacente de nossos dados. Segundo [Henderson e Souto \(2018\)](#) embora a seleção de todos os parâmetros: o número e a localização dos nós e o parâmetro de suavização, é fácil mostrar que as escolhas do grau e nós são muito menos cruciais do que a escolha de λ , o parâmetro de suavização. Ou seja, ao usar um número suficientemente alto de nós, a suavidade de nosso ajuste pode ser controlada por λ . Portanto, quando o número de nós é fixo, é preciso escolher λ de forma apropriada.

Existem diversas formas de selecionar os parâmetros de maneira orientada por dados, nos concentraremos em uma das abordagens mais comumente usada: validação cruzada generalizada (VCG), que será discutida nas próximas seções. De acordo com [Henderson e Souto \(2018\)](#) sobre os experimentos realizados em relação ao número de nós, a literatura indica que: (1) adicionar mais nós somente melhora o ajuste para um pequeno número de nós; (2) ao usar muitos nós, o mínimo da VCG para ajustes lineares e quadráticos torna-se indistinguível.

Embora existam critérios mais formais para selecionar o número e a localização dos nós, [Ruppert et al. \(2003\)](#) fornecem soluções simples que geralmente funcionam bem. Sua escolha padrão de r é

$$r = \min\{(1/4) \times \text{números únicos de } x_i; 35\},$$

em que r é o número de nós. Para os locais dos nós, eles sugerem

$$\kappa_r := \left(\frac{r' + 1}{r + 2} \right) - \text{ésimo quantil de amostra do único } x_i,$$

para $r' = 1, \dots, r$.

A proposta de [Eilers e Marx \(1996\)](#) mostra que quando f é um *B-spline* (base de De Boor) é possível apresentar que a penalização da segunda derivada ao quadrado pode ser aproximada por uma penalização discreta que é mais eficiente na estimação dos parâmetros. Esse procedimento é chamado de penalização por *P-splines*. Porém, esse não é objetivo do trabalho, pois estamos usando outros tipos de *splines*, assim mantemos a penalização do quadrado da segunda derivada.

[Eilers e Marx \(1996, 2010\)](#) argumentam que nós igualmente espaçados são sempre preferidos. [Eilers e Marx \(2010\)](#) apresentam um exemplo onde nós igualmente espaçados superam nós espaçados de quantis. O melhor tipo de espaçamento de nós ainda está em debate e ambos os métodos ainda são comumente usados. [Montoya et al. \(2014\)](#) testam o desempenho de diferentes métodos de seleção para determinar o número de nós através de simulações, de tal maneira em que os nós são dispostos de forma equidistantes em um modelo *P-spline*. Especificamente, eles comparam os métodos apresentados em [Ruppert et al. \(2003\)](#) com o método de seleção de nó do algoritmo *myopic* e o método de seleção de nó do algoritmo de busca completa. Seus resultados mostram que o método de escolha padrão possui um desempenho tão bom ou melhor do que os outros métodos ao usar diferentes métodos de seleção de parâmetro de suavização comumente usados.

De acordo com [Henderson e Souto \(2018\)](#), embora a localização dos nós tenha pouco efeito sobre o ajuste ao usar uma quantidade suficientemente grande de nós, eles podem se tornar importantes ao lidar com problemas mais complexos. Por exemplo, ao tentar suavizar funções de regressão com forte variabilidade local ou com dados esparsos. Nesses casos, usar um algoritmo mais complexo para fazer sua seleção pode ser mais apropriado.

2.2 Erros simétricos

A análise de regressão é uma das metodologias estatísticas mais utilizadas nos ramos da ciência e engenharia, as quais buscam relações funcionais entre as variáveis. O exemplo mais típico de análise de regressão é a modelagem de regressão linear múltipla, usada para explicar a relação entre variáveis explicativas e preditora bem como prever valores da variável resposta. Relatos abrangentes da teoria e aplicações do modelo de regressão linear são discutidos, por exemplo, em Weisberg (2005), Seber e Lee (2012) e Montgomery *et al.* (2012). Essa técnica geralmente é baseada em um modelo estatístico no qual os termos de erro são assumidos como variáveis aleatórias independentes e identicamente distribuídas de modo que a distribuição é considerada normal.

A estrutura das distribuições de erro desempenha um papel significativo na análise das equações de regressão. Se a distribuição de erros não for normal, exceto para grandes amostras, não será possível usar as técnicas de inferência para testar a adequação da equação de regressão, e a distribuição do coeficiente de regressão se desvia do procedimento padrão. Portanto, para se ter uma análise precisa, é necessário considerar a estrutura das distribuições dos erros.

A violação da suposição de normalidade pode afetar adversamente a eficiência dos estimadores, a falta de robustez e pode ser limitada para representar os dados com precisão. De fato, o impacto da violação da normalidade no desempenho dos estimadores é frequentemente negligenciado. Portanto, tem sido de enorme interesse desenvolver estimadores eficientes para distribuições de erros não normais.

Outra desvantagem dos erros Gaussianos é o fato da distribuição ser mesocúrtica. Em muitas ocasiões, os dados estatísticos mostram um pico ou achatamento, indicando algum tipo de curtose. Sendo assim, as variáveis em estudo são platicúrticas ou leptocúrticas. Para analisar esses tipos de variáveis, são necessários modelos com caudas mais leves ou mais pesadas que podem ser uma boa alternativa. Em situações como essa, mesmo que o formato da curva de frequência da amostra seja simétrico e em forma de sino, a aproximação normal pode se ajustar mal à distribuição (Rao *et al.*, 1997). Além disso, observações atípicas influenciam de forma desproporcional os ajustes sob normalidade. Por um tempo, uma alternativa era usar uma transformação na variável resposta para obter melhores resultados, ou até mesmo simetria dos dados. Contudo, mesmo com ajustes aos modelos com erros normais, as estimativas obtidas eram sensíveis a observações extremas, conhecidas como pontos aberrantes ou *outliers*. Por esses motivos, vários pesquisadores propuseram análises mais robustas. Dentre as soluções, destaca-se uma família de distribuição diferente para os termos do erro, nas quais se destacam as distribuições simétricas. Assim, é possível minimizar as influências de pontos aberrantes nas estimativas e inferência dos parâmetros, ou seja, assume para os erros distribuições com caudas mais pesadas ou mais leves que as da distribuição normal.

Essa classe de distribuições possui grandes avanços na literatura. Em Chmielewski (1981), Fang *et al.* (1990), Fang e Anderson (1990), Cysneiros (2004) e Gupta e Varga (2012) podem ser encontradas revisões das áreas em que são aplicadas as distribuições simétricas. Cambanis *et al.* (1981) observam que a classe de distribuições simétricas coincide com a família de distribuições elípticas univariadas. Em relação aos modelos com estimativas robustas, Lange *et al.* (1989) propõem modelos supondo erros com distribuição t -Student com ν graus de liberdade, Little (1988) e Yamaguchi (1990) propõem modelos baseados em erros com distribuição t -multivariada e normal contaminada, os quais adicionam-se parâmetros que permitem ajustar a curtose da distribuição dos dados. Taylor (1992) desenvolveu um modelo de regressão linear em que os erros têm distribuição exponencial potência com um parâmetro extra de forma. Arellano-Valle (1994) apresentou vários resultados sobre as propriedades, inferências e aplicações aos modelos de regressão em distribuições elípticas. Ferrari e Uribe-Opazo (2001) estendem esses resultados para modelos de regressão linear simétricos. Uribe-Opazo *et al.* (2008) desenvolveram teoria assintótica de segunda ordem para testes de escore na classe dos modelos de regressão lineares simétricos. Cysneiros (2004) desenvolveram métodos restritos e de validação de modelos de regressão simétricos e modelos de regressão lineares heterocedásticos simétricos, em que o trabalho propõe metodologias de influência local, bem como medidas de alavancagem. Posteriormente, Cysneiros e Paula (2005) apresentaram os principais resultados inferenciais e as técnicas de diagnóstico em modelos não lineares, lineares e lineares

heterocedásticos com erros simétricos e, por fim, discutiram algumas extensões de modelos elípticos para dados correlacionados. [Cysneiros et al. \(2007\)](#) discutiram sobre os modelos lineares heterocedásticos simétricos, bem como aspectos de robustez dos estimadores de máxima verossimilhança contra *outliers* e observações com variância grande, teste de escore e diagnósticos para heterocedasticidade. [Cysneiros \(2011\)](#) desenvolveu os modelos de regressão para os erros simétricos. [Cao et al. \(2010\)](#) estudaram conjuntamente heteroscedasticidade e autocorrelação em modelos de regressão não lineares com erros simétricos, e [Relvas e Paula \(2016\)](#) propuseram os modelos parcialmente lineares com erros simétricos autorregressivos de primeira ordem.

De [Cysneiros \(2004\)](#), formalmente, podemos definir uma distribuição sendo simétrica da seguinte maneira. Considere uma variável aleatória Y , a qual possui distribuição simétrica com parâmetro de posição μ , parâmetro de escala ϕ e função geradora de densidade $g(\cdot)$, sua função de densidade de probabilidade pode ser expressa como

$$h(y_i; \mu, \phi) = \frac{1}{\sqrt{\phi}} g(\delta_i), \quad i = 1, 2, \dots, n, \quad (2.6)$$

em que $\delta_i = \frac{(y_i - \mu)^2}{\phi}$, $\mu \in \mathcal{R}$, $\phi > 0$ e $g(\cdot)$ é uma função definida positiva \mathcal{R} , ou seja, $g : \mathcal{R} \rightarrow [0, \infty)$ tal que $\int_0^\infty u^{-\frac{1}{2}} g(u) du = 1$ ([Kelker, 1970](#)), sendo conhecida como uma geradora de densidades. O nome gerador de densidade deve-se ao fato de que se u é uma distribuição simétrica, não necessariamente possui densidade. Assim, se a densidade de u existe, ela deve assumir a forma $g(u^2)$ para alguma função não negativa $g(\cdot)$ de uma variável escalar, que satisfaça a condição $\int_0^\infty g(u^2) du = \int_0^\infty u^{-\frac{1}{2}} g(u) du = 1$ (veja, por exemplo, ([Fang et al., 1990](#), Seção 2.2.3)). Essa condição é necessária para que $h(y; \mu, \phi)$ seja uma função densidade de probabilidade. Dessa forma, denotamos $Y \sim S(\mu, \phi)$ a variável simétrica definida acima.

A função característica toma a forma $\exp(it\mu)\varphi(t^2\phi)$ para alguma função $\varphi(\cdot)$, em que $i = \sqrt{-1}$. [Kelker \(1970\)](#) mostrou que, se $\delta^{\frac{1}{2}(1-k)-1}g(\delta)$ é integrável, então existem os momentos de ordem k da variável aleatória Y . Pela definição acima, podemos observar que a função $g(\cdot)$ caracteriza a função de probabilidade simétrica. Além disso, se a esperança e a variância de Y existem, então elas são dadas por $E(Y) = \mu$ e $\text{Var}(Y) = \xi_g \phi$, respectivamente, em que $\xi_g = -2\varphi'(0)$ é uma constante positiva que depende da distribuição simétrica que está sendo utilizada, com $\varphi'(0) = d\varphi(\delta)/d\delta|_{\delta=0}$ e que não depende dos parâmetros μ e ϕ .

Se $Y \sim S(\mu, \phi)$, então $a + bY \sim S(a + b\mu, b^2\phi)$ em que $a, b \in \mathcal{R}$ com $b \neq 0$. Logo, a combinação linear de uma variável aleatória simétrica é também uma variável aleatória simétrica. Uma padronização que será utilizada no decorrer deste trabalho é a seguinte:

$$Z = \frac{(Y - \mu)}{\sqrt{\phi}} \sim S(0, 1),$$

ou seja, com função densidade $h(z) = h(z; 0, 1) = g(z^2)$ para $z \in \mathcal{R}$, de modo que Z é chamada de variável simétrica padrão.

As distribuições simétricas apresentam parâmetros de posição e escala e, em muitos casos, a função $g(\cdot)$ depende de um parâmetro adicional, chamado parâmetro de forma, que vai ser indicado como ν , controlando a curtose da distribuição. Esse parâmetro pode ser fixado ou estimado a partir dos dados. Nessa situação, tem-se uma maior flexibilização das caudas, ocasionando ajustes mais adequados. Com base nas distribuições com essas características, neste trabalho, a atenção estará voltada para as seguintes distribuições: normal, normal generalizada ([Nadarajah, 2005](#)), t -Student, t -Student generalizada, logísticas tipo I, tipo II e generalizada e exponencial potência. A Tabela 2.2 descreve as funções densidade de probabilidade das distribuições citadas anteriormente e o parâmetro ξ_g , para mais detalhes, veja [Cysneiros e Paula \(2005\)](#).

Tabela 2.2: Função geradora de densidade de probabilidade e ξ_g para algumas distribuições simétricas

Distribuição	$g(\delta)$	ξ_g
Normal	$\frac{1}{\sqrt{2\pi}} \exp(-\delta/2)$	1
Normal generalizada	$\frac{\nu}{2\Gamma(1/\nu)} \exp(-\sqrt{\delta}^\nu)$	$\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}$
t -Student	$\frac{\nu^{\nu/2}}{\text{Beta}(1/2, \nu/2)} (\nu + \delta)^{-\frac{\nu+1}{2}}$	$\frac{\nu}{\nu-2}, \quad \nu > 2$
t -Student generalizada	$\frac{\nu_1^{\nu_2/2}}{\text{Beta}(1/2, \nu_2/2)} (\nu_1 + \delta)^{-\frac{\nu_2+1}{2}}$	$\frac{\nu_1}{\nu_2-2}, \quad \nu_1 > 0, \nu_2 > 2$
Logística-I	$c \frac{\exp(-\delta)}{[1+\exp(\delta)]^2}$	0, 79569
Logística-II	$\frac{\exp(-\sqrt{\delta})}{[1+\exp(-\sqrt{\delta})]^2}$	$\frac{\pi^2}{3}$
Logística Generalizada	$\frac{\alpha}{\text{Beta}(\nu, \nu)} \left[\frac{\exp(-\alpha\sqrt{\delta})}{(1+\exp\{-\alpha\sqrt{\delta}\})^2} \right]^\nu$	$2\psi'(\nu)$
Exponencial Potência	$\frac{\exp(-\frac{1}{2}\delta^{1/(1+\nu)})}{\Gamma(1+\frac{1+\nu}{2}) 2^{1+(1+\nu)/2}}$	$\frac{2^{1+\nu}\Gamma\{3(\nu+1)/2\}}{\Gamma\{(\nu+1)/2\}}$

Nota: $\Gamma(\cdot)$ e $\text{Beta}(\cdot, \cdot)$ são as funções gama e beta, respectivamente, c é uma constante normalizadora, $c \approx 1.48430029$ e $\psi'(\cdot)$ é a função trigama

Fonte: extraída de [Cysneiros e Paula \(2005\)](#)

2.2.1 Modelo de regressão paramétrico linear com erros simétricos

Para definir a classe dos modelos de regressão simétricos, supomos primeiramente que Y_1, \dots, Y_n são variáveis aleatórias independentes, cada uma com distribuição pertencente à classe simétrica definida em (2.6). Com isso, o modelo linear simétrico é definido por

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n, \quad (2.7)$$

em que $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{r_0-1})^\top$ é o vetor de parâmetros desconhecidos de dimensão $(r_0 \times 1)$, $\mathbf{x}_i = (1, x_{i1}, \dots, x_{i(r_0-1)})^\top$ é o vetor que representa os valores das variáveis explicativas cuja dimensão é $(r_0 \times 1)$, tem-se que $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ com $\mathbf{X} = (\mathbf{1}_n^\top, \mathbf{x}_1^\top, \dots, \mathbf{x}_n^\top)^\top$ sendo uma matriz $(n \times r_0)$ de posto completo. Portanto, os modelos de regressão linear simétricos são definidos por (2.6) e (2.7). Quando a existência dos momentos for satisfeita para as distribuições da classe simétrica, temos que $E(y_i) = \mu_i$ e $\text{Var}(y_i) = \xi_g \phi$, alguns valores de ξ_g são apresentados na Tabela 2.2.

O logaritmo da função de verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ é dado por

$$L(\boldsymbol{\theta}) = -\frac{n}{2} \log \phi + \sum_{i=1}^n \log [g(\delta_i)],$$

em que $\delta_i = (y_i - \mu_i)^2 / \phi$ e $y_i \sim S(\mu, \phi)$.

Para obter a função escore e a matriz de informação de Fisher, precisamos derivar a função $L(\boldsymbol{\theta})$ em relação aos parâmetros desconhecidos. Para isso, supomos aqui que tais derivadas existem e que as distribuições simétricas utilizadas neste estudo satisfazem certas condições de regularidade necessárias para a validade dos resultados: (i) a existência dos momentos e (ii) a diferenciação da função log-verossimilhança em todo espaço paramétrico.

A função escore para $\boldsymbol{\beta}$ é dada por

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \left\{ \sum_{i=1}^n \log [g(\delta_i)] \right\}.$$

Em geral, no contexto simétrico é usual definir $W_g(\delta)$ e $W'_g(\delta)$ de tal forma que $W_g = \{d \log [g(\delta)]\} / d\delta = g'(\delta) / g(\delta)$ e $W'_g(\delta) = [dW_g(\delta)] / d\delta$. Então

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^n W_g(\delta_i) \frac{\partial \delta_i}{\partial \beta_j} = \sum_{i=1}^n \left\{ \frac{v_i}{\phi} (y_i - \mu_i) \frac{\partial \mu_i}{\partial \beta_j} \right\},$$

para $j = 0, \dots, r_0 - 1$. As expressões $W_g(\delta)$ e $W'_g(\delta)$ para as distribuições simétricas utilizadas nesta pesquisa são apresentadas na Tabela 2.3.

Tabela 2.3: Expressões para $W_g(\delta)$ e $W'_g(\delta)$ para algumas distribuições simétricas

Distribuição	$W_g(\delta)$	$W'_g(\delta)$
Normal	$-\frac{1}{2}$	0
Normal generalizada	$-\frac{\nu}{2} \delta^{(\nu/2)-1}$	$-\frac{(\nu^2-2\nu)}{4} \delta^{(\nu/2)-2}$
t -Student	$-\frac{\nu+1}{2(\nu+\delta)}$	$\frac{\nu+1}{2(\nu+\delta)^2}$
t -Student Generalizada	$-\frac{\nu_2+1}{2(\nu_1+\delta)}$	$\frac{\nu_2+1}{2(\nu_1+\delta)^2}$
Logística-I	$-\tan\left(\frac{\delta}{2}\right)$	$-\sec\left(\frac{\delta}{2}\right)/2$
Logística-II	$-\frac{\exp(-\sqrt{\delta})-1}{(-2\sqrt{\delta})[1+\exp(-\sqrt{\delta})]}$	$\frac{2\exp(-\sqrt{\delta})\sqrt{\delta}+\exp(-2\sqrt{\delta})-1}{(-4\delta^{3/2})[1+\exp(-\sqrt{\delta})]^2}$
Logística Generalizada	$\frac{-\alpha\nu[\exp(-\alpha\sqrt{\delta})-1]}{(-2\sqrt{\delta})[1+\exp(-\alpha\sqrt{\delta})]}$	$-\frac{\alpha\nu}{4} \frac{2\alpha\exp(-\alpha\sqrt{\delta})\sqrt{\delta}+\exp(-2\alpha\sqrt{\delta})-1}{(\delta^{3/2})[1+\exp(-\alpha\sqrt{\delta})]^2}$
Exponencial Potência	$-\frac{1}{2(1+\nu)\delta^{\nu/(1+\nu)}}$	$\frac{\nu}{(1+\nu)^2 2\delta^{(2\nu+1)/(1+\nu)}}$

Fonte: extraída de Cysneiros e Paula (2005)

Portanto, segundo Cysneiros e Paula (2005) a função escore para $\boldsymbol{\beta}$ na forma matricial é dada por

$$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{1}{\phi} \mathbf{X}^{\top} \mathbf{D}_v \boldsymbol{\epsilon},$$

em que $\mathbf{D}_v = \text{diag}\{v_1, \dots, v_n\}$ com $v_i = -2W_g(\delta_i)$, e $\boldsymbol{\epsilon} = (\mathbf{y} - \boldsymbol{\mu})$, para $i = 1, \dots, n$. Analogamente, a função escore para ϕ é obtida

$$\begin{aligned} \frac{\partial L(\boldsymbol{\theta})}{\partial \phi} &= \frac{d}{d\phi} \left\{ -\frac{n}{2} \log(\phi) + \sum_{i=1}^n \log[g(\delta_i)] \right\} \\ &= -\frac{n}{2\phi} + \sum_{i=1}^n \left\{ -\frac{W_g(\delta_i)}{\phi^2} (y_i - \mu_i)^2 \right\} \\ &= -\frac{n}{2\phi} + \sum_{i=1}^n \left\{ \frac{v_i}{2\phi^2} (y_i - \mu_i)^2 \right\}. \end{aligned}$$

Logo, a função escore para ϕ é expressa na forma

$$\mathbf{U}_{\phi}(\boldsymbol{\theta}) = \frac{1}{2\phi} \left\{ \frac{1}{\phi} \boldsymbol{\epsilon}^{\top} \mathbf{D}_v \boldsymbol{\epsilon} - n \right\}.$$

Para a estimação dos parâmetros nos modelos de regressão simétrica podemos utilizar o método de máxima verossimilhança como apresentado em Cysneiros e Paula (2005), o qual consiste em encontrar as estimativas do vetor de parâmetros $\boldsymbol{\beta}$ e do escalar ϕ através da solução das equações dadas por $\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = 0$ e $\mathbf{U}_{\phi}(\boldsymbol{\theta}) = 0$. A solução dessas equações não possui forma fechada, tornando-se necessário o uso de métodos iterativos para obter as estimativas de máxima verossimilhança. Sendo assim, essas estimativas podem ser obtidas por meio do algoritmo do método de escore de Fisher,

discutido em Cysneiros e Paula (2005), cujo processo iterativo conjunto é dado por

$$\boldsymbol{\beta}^{(u+1)} = \left[\mathbf{X}^\top \mathbf{D}_v^{(u)} \mathbf{X} \right]^{-1} \mathbf{X}^\top \mathbf{D}_v^{(u)} \mathbf{y} \quad (2.8)$$

$$\phi^{(u+1)} = \frac{1}{n} \boldsymbol{\epsilon}^{(u)\top} \mathbf{D}_v^{(u)} \boldsymbol{\epsilon}^{(u)}, \quad (2.9)$$

para $u = 0, 1, \dots$. O processo iterativo (2.8)-(2.9) garante solução positiva para a estimativa de máxima verossimilhança de ϕ . Deve-se iniciar o processo com valores $\boldsymbol{\beta}^{(0)}$ e $\phi^{(0)}$. O procedimento descrito acima é o mesmo descrito por Gourieroux e Monfort (1995, p. 170).

Por meio das funções escores, é possível obter a matriz Hessiana, de acordo com Cysneiros e Paula (2005) as submatrizes ficam dadas por

$$\begin{aligned} \ddot{\mathbf{L}}_{\beta\beta} &= \frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = -\frac{1}{\phi} \mathbf{X}^\top \mathbf{D}_a \mathbf{X}, \\ \ddot{\mathbf{L}}_{\phi\phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \phi^2} = \frac{1}{\phi^2} \left\{ \frac{n}{2} + \boldsymbol{\delta}^\top \mathbf{D}_c \boldsymbol{\delta} - \frac{1}{\phi} \boldsymbol{\epsilon}^\top \mathbf{D}_v \boldsymbol{\epsilon} \right\}, \\ \ddot{\mathbf{L}}_{\beta\phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \phi} = \frac{2}{\phi^2} \mathbf{X}^\top \mathbf{b}, \end{aligned}$$

em que $\mathbf{D}_a = \text{diag}\{a_1, \dots, a_n\}$, $\mathbf{D}_c = \text{diag}\{c_1, \dots, c_n\}$, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top$, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$ e $\mathbf{b} = (b_1, \dots, b_n)^\top$, sendo $a_i = -2 \{W_g(\delta_i) + 2\delta_i W'_g(\delta_i)\}$, $c_i = W'_g(\delta_i)$, $b_i = \{W_g(\delta_i) + \delta_i W'_g(\delta_i)\} \epsilon_i$ e $\epsilon_i = y_i - \mu_i$.

Dado que a função $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ é assumida regular com respeito ao vetor de parâmetros $\boldsymbol{\theta}$ segundo as condições apresentadas em Cox e Hinkley (1974), podemos assumir que o estimador de máxima verossimilhança de $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$ é um estimador consistente e segue distribuição assintoticamente normal, cuja matriz de variâncias e covariâncias depende da matriz de informação de Fisher $\mathcal{I}_{\theta\theta} = \mathbb{E} \left[-\ddot{\mathbf{L}}_{\theta\theta} \right]$, a qual pode ser expressa na forma

$$\mathcal{I}_{\theta\theta} = \begin{bmatrix} \mathcal{I}_{\beta\beta} & 0 \\ 0 & \mathcal{I}_{\phi\phi} \end{bmatrix}.$$

Para o modelo em questão, pode-se mostrar que as submatrizes da matriz de informação de Fisher são definidas por

$$\begin{aligned} \mathcal{I}_{\beta\beta} &= \mathbb{E} \left\{ -\frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right\} = \frac{4d_g}{\phi} \mathbf{X}^\top \mathbf{X}, \\ \mathcal{I}_{\phi\phi} &= \mathbb{E} \left\{ -\frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \phi^2} \right\} = \frac{n}{4\phi^2} (4f_g - 1), \\ \mathcal{I}_{\beta\phi} &= \mathbb{E} \left\{ -\frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \phi} \right\} = \mathbf{0}_{(r_0 \times 1)}, \end{aligned}$$

em que $f_g = \mathbb{E} [W_g^2(U^2)U^4]$ e $d_g = \mathbb{E} [W_g^2(U^2)U^2]$, com $U \sim S(0, 1)$. A Tabela 2.4, extraída de Cysneiros e Paula (2005), apresenta os valores de d_g e f_g para algumas distribuições simétricas.

Tabela 2.4: Valores de d_g e f_g para algumas distribuições simétricas.

Distribuição	d_g	f_g
Normal	$\frac{1}{4}$	$\frac{3}{4}$
Normal Generalizada	$\frac{(1+(-1)^{2\nu})\nu^2(1+\nu)}{8}$	$\frac{(1+(-1)^{2\nu})\nu^{3+(2/\nu)}\Gamma(2+\frac{3}{\nu})}{8\Gamma(1+\frac{1}{\nu})}$
t -Student	$\frac{(\nu+1)}{4(\nu+3)}$	$\frac{3(\nu+1)}{4(\nu+3)}$
t -Student Generalizada	$\frac{\nu_2(\nu_2+1)}{4\nu_1(\nu_2+3)}$	$\frac{3(\nu_2+1)}{4(\nu_2+3)}$
Logística-I	0,369310044	1,003445984
Logística-II	$\frac{1}{12}$	0,60749
Logística Generalizada	$\frac{\alpha^2\nu^2}{4(2\nu+1)}$	$\frac{2\nu(2+\nu^2\psi'(\nu))}{4(2\nu+1)}$
Exponencial Potência	$\frac{\Gamma\{(3-\nu)/2\}}{4(2^{\nu-1})(1+\nu)^2\Gamma\{(\nu+1)/2\}}$	$\frac{(\nu+3)}{4(\nu+1)}$

Nota: $\psi'(\cdot)$ é a função trigama.

Fonte: extraída de [Cysneiros e Paula \(2005\)](#)

Note que β e ϕ são ortogonais, ou seja, $E\{[\partial^2 L(\theta)]/\partial\beta\partial\phi\} = 0$. Logo, os estimadores $\hat{\beta}$ e $\hat{\phi}$ são assintoticamente independentes. Contudo, temos que $\hat{\mathcal{I}}_{\beta\beta}^{-1} = \frac{\hat{\phi}}{4d_g}(\mathbf{X}^\top \mathbf{X})^{-1}$ é um estimador consistente da matriz de variâncias e covariâncias assintóticas de $\hat{\beta}$. De forma similar, $\hat{\Gamma}_{\phi\phi}^{-1} = \frac{4\hat{\phi}^2}{n(4f_g-1)}$ é um estimador consistente da variância assintótica de $\hat{\phi}$. Os estimadores $\hat{\mathcal{I}}_{\beta\beta}^{-1}$ e $\hat{\Gamma}_{\phi\phi}^{-1}$ podem ser utilizados para encontrar regiões de confiança assintóticas para $\hat{\beta}$ e intervalos de confiança assintóticos para $\hat{\phi}$.

Podemos encontrar na literatura basicamente três métodos para a estimação do parâmetro de forma da distribuição normal generalizada: o método de máxima verossimilhança ([Nadarajah, 2005](#)), o método dos momentos ([Dominguez-Molina et al., 2003](#)) e ainda uma combinação de ambos ([Varanasi e Aazhang, 1989](#)). Da mesma forma, os graus de liberdade ν da distribuição t de Student podem ser estimados por meio do método de máxima verossimilhança e do método dos momentos ([Arellano-Valle, 1994](#)). No entanto, para obtermos estimativas mais estáveis para os parâmetros dos modelos, optamos por trabalhar com o parâmetro ν fixo no conjunto de valores inteiros, este parâmetro é selecionado por meio do procedimento de seleção que consiste em minimizar o critério de Akaike (AIC) e validação cruzada generalizada (VCG), os quais serão detalhados nas próximas seções.

2.2.2 Modelo parcialmente linear com erros simétricos

Os modelos paramétricos possuem vantagens definidas, quando apropriados. Os métodos inferenciais correspondentes geralmente possuem propriedades de eficiência. Além disso, os parâmetros podem ter algum significado físico que os tornam interpretáveis e de interesse do pesquisador. Infelizmente, por causa desses fatores, os modelos paramétricos são frequentemente usados quando há pouca informação disponível sobre a relação das variáveis em estudo. De fato, o uso de um modelo paramétrico inadequado pode produzir inferência enganosa ou incorreta sobre a curva de regressão.

A modelagem de regressão paramétrica é, sem dúvida, a abordagem mais prevalente na análise de regressão. No entanto, existem outros métodos para ajustar a relação dos dados. Com esse propósito temos as técnicas de regressão não paramétrica. Esses métodos permitem maior flexibilidade na forma possível da curva de regressão e, em particular, não fazem suposições sobre uma forma paramétrica ([Eubank, 1999](#)).

Técnicas de regressão não paramétricas superam a dificuldade inerente às técnicas paramétricas (a forma funcional de x é conhecida). No entanto, seu uso não é isento de custos. Geralmente, os estimadores não paramétricos são menos eficientes que a variedade paramétrica quando o modelo paramétrico é válido. Consequentemente, técnicas de regressão não paramétricas dependem mais

fortemente dos dados para obter informações sobre x do que suas contrapartes paramétricas.

Dessa forma, o que nos preocupa é simplesmente relaxar a suposição de linearidade na equação (2.7). A forma natural é considerar o modelo não paramétrico simples com apenas um preditor da seguinte maneira:

$$y_i = f(x_i) + \epsilon_i. \quad (2.10)$$

A regressão simples não paramétrica é frequentemente chamada de “suavização do gráfico de dispersão” porque uma aplicação importante permite traçar uma curva suave através de um gráfico de dispersão de y em relação a x . Quando há mais de dois ou três preditores, modelos mais restritivos são desenvolvidos. Um desses modelos é o modelo de regressão linear aditivo:

$$y_i = f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_k(x_{ki}) + \epsilon_i. \quad (2.11)$$

Nota-se que uma generalização desse modelo envolve uma parte paramétrica e outra não paramétrica, sendo assim, podemos construir os modelos semiparamétricos, ou modelos parcialmente lineares, como definidos em [Green e Silverman \(1994\)](#). O modelo parcialmente linear com erros simétricos é útil quando a relação da variável resposta com a variável explicativa não é linear nem tampouco conhecida, podendo ser usada de maneira não paramétrica. Além disso, podemos utilizar uma distribuição com caudas mais pesadas, diminuindo, assim, o peso de observações aberrantes nas estimativas do modelo, produzindo dessa forma ajustes mais robustos. Esses modelos foram desenvolvidas por [Ibache-Pulgar et al. \(2013\)](#) e emergem como uma ferramenta poderosa na modelagem estatística devido à sua flexibilidade para modelar efeitos das variáveis explicativas que podem contribuir de maneiras paramétrica ou não paramétrica. Tais modelos assumem a seguinte relação entre a resposta e os valores das variáveis explicativas:

$$y_i = f_0(x_{0i}) + f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_k(x_{ki}) + \epsilon_i, \quad (2.12)$$

com $i = 1, \dots, n$, y_i a i -ésima variável resposta, $f_0(x_{0i}) = \mathbf{x}_i^\top \boldsymbol{\beta}$, e \mathbf{x}_i^\top o vetor ($r_0 \times 1$) de valores das variáveis explicativas, $f(\cdot)$ uma função de uma variável explicativa que contribui de maneira não paramétrica e ϵ_i representa o termo do erro da i -ésima observação. Assume-se que ϵ_i segue uma distribuição simétrica, ou seja, $\epsilon_i \sim S(0, \phi)$. Este modelo, escrito em forma matricial, pode ser representado por

$$\mathbf{y} = \mathbf{N}_0 \boldsymbol{\gamma}_0 + \mathbf{N}_1 \boldsymbol{\gamma}_1 + \dots + \mathbf{N}_k \boldsymbol{\gamma}_k + \boldsymbol{\epsilon},$$

em que \mathbf{y} é um vetor ($n \times 1$), \mathbf{N}_l é uma matriz ($n \times r_l$) para $l = 0, \dots, k$ com cada linha igual a $\mathbf{n}_i^{(l)\top}$. Tem-se que $\mathbf{N}_0 = \mathbf{X}$ representa os dados da matriz de planejamento da parte paramétrica e \mathbf{N}_1 a \mathbf{N}_k referem-se às funções base da parte não paramétrica do experimento.

Nos trabalhos desenvolvidos por [Green e Silverman \(1994\)](#) e [Ibache-Pulgar et al. \(2013\)](#) temos que $f(\cdot)$ possui a forma de *spline* cúbico natural, nesse caso, o número de nós não é fixo, corresponde ao número de valores distintos da variável explicativa. Assim, a forma de $f(\cdot)$ não é conhecida a priori e tem-se apenas a suposição de que $f(\cdot)$ é pelo menos duplamente diferenciável e a solução através da penalização do quadrado da segunda derivada é de um *spline* cúbico natural.

A função de verossimilhança penalizada do modelo definido acima, a qual reduz a presença de super ajuste na presença de um número elevado de nós, pode ser escrita da seguinte forma:

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda}) = -\frac{n}{2} \log \phi + \sum_{i=1}^n \log [g(\delta_i)] - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k,$$

em que $\boldsymbol{\theta} = (\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_k, \phi)^\top$, $\boldsymbol{\lambda}$ é o vetor de parâmetros de suavização e \mathbf{M}_l é a matriz de penalização ($r_l \times r_l$) para $l = 1, \dots, k$ definidas na Seção 2.1. Nas próximas subseções, apresentaremos as funções escore e as matrizes de informação de Fisher deste modelo. Segundo [Ibache-Pulgar et al. \(2013\)](#) a função escore dos parâmetros do logaritmo da função de verossimilhança são dadas por

$$\begin{aligned} \mathbf{U}_p(\gamma_0) &= \frac{1}{\phi} \mathbf{N}_0^\top \mathbf{D}_v \boldsymbol{\epsilon}, \\ \mathbf{U}_p(\gamma_l) &= \frac{1}{\phi} \mathbf{N}_l^\top \mathbf{D}_v \boldsymbol{\epsilon} - \lambda_l \mathbf{M}_l \gamma_l, \quad l = 1, \dots, k, \quad \text{e} \\ \mathbf{U}_p(\phi) &= \frac{1}{2\phi} \left\{ \frac{1}{\phi} \boldsymbol{\epsilon}^\top \mathbf{D}_v \boldsymbol{\epsilon} - n \right\}. \end{aligned}$$

Usando as funções escores, também é possível obter a matriz Hessiana, sendo as submatrizes dadas por

$$\begin{aligned} \ddot{\mathbf{L}}_{\gamma_0 \gamma_0} &= \frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \gamma_0 \partial \gamma_0^\top} = -\frac{1}{\phi} \mathbf{N}_0^\top \mathbf{D}_a \mathbf{N}_0, \\ \ddot{\mathbf{L}}_{\gamma_l \gamma_l} &= \frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \gamma_l \partial \gamma_l^\top} = -\frac{1}{\phi} \mathbf{N}_l^\top \mathbf{D}_a \mathbf{N}_l - \lambda_l \mathbf{M}_l, \quad l = 1, \dots, k, \\ \ddot{\mathbf{L}}_{\phi \phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \phi^2} = \frac{1}{\phi^2} \left\{ \frac{n}{2} + \boldsymbol{\delta}^\top \mathbf{D}_c \boldsymbol{\delta} - \frac{1}{\phi} \boldsymbol{\epsilon}^\top \mathbf{D}_v \boldsymbol{\epsilon} \right\}, \\ \ddot{\mathbf{L}}_{\gamma_l \gamma_{l'}} &= \frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \gamma_l \partial \gamma_{l'}^\top} = -\frac{1}{\phi} \mathbf{N}_l^\top \mathbf{D}_a \mathbf{N}_{l'} \quad \text{e} \\ \ddot{\mathbf{L}}_{\gamma_l \phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \gamma_l \partial \phi} = \frac{2}{\phi^2} \mathbf{N}_l^\top \mathbf{b}, \end{aligned}$$

para $l, l' = 0, \dots, k$ e $l \neq l'$. Agora, com as matrizes obtidas acima, é possível obter a matriz de informação de Fisher, que é definida como $\mathcal{I}_{\theta\theta} = \mathbf{E} \left[-\ddot{\mathbf{L}}_{\theta\theta} \right]$. Para o modelo em questão, pode-se mostrar que as submatrizes da matriz de informação de Fisher são expressas nas formas

$$\begin{aligned} \mathcal{I}_{\gamma_0 \gamma_0} &= \mathbf{E} \left\{ -\frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \gamma_0 \partial \gamma_0^\top} \right\} = \frac{4d_g}{\phi} \mathbf{N}_0^\top \mathbf{N}_0, \\ \mathcal{I}_{\gamma_l \gamma_l} &= \mathbf{E} \left\{ -\frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \gamma_l \partial \gamma_l^\top} \right\} = \frac{4d_g}{\phi} \mathbf{N}_l^\top \mathbf{N}_l + \lambda_l \mathbf{M}_l, \quad l = 1, \dots, k, \\ \mathcal{I}_{\phi \phi} &= \mathbf{E} \left\{ -\frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \phi^2} \right\} = \frac{n}{4\phi^2} (4f_g - 1), \\ \mathcal{I}_{\gamma_l \gamma_{l'}} &= \mathbf{E} \left\{ -\frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \gamma_l \partial \gamma_{l'}^\top} \right\} = \frac{4d_g}{\phi} \mathbf{N}_l^\top \mathbf{N}_{l'} \quad \text{e} \\ \mathcal{I}_{\gamma_l \phi} &= \mathbf{E} \left\{ -\frac{\partial^2 L_p(\boldsymbol{\theta})}{\partial \gamma_l \partial \phi} \right\} = \mathbf{0}_{(r_l \times 1)}, \end{aligned}$$

para $l, l' = 0, \dots, k$ e $l \neq l'$.

Ibacache-Pulgar *et al.* (2013) propõem alguns procedimentos para a estimação dos parâmetros do modelo definido nesta Seção que combina procedimentos do método escore de Fisher e do algoritmo *backfitting*. O objetivo de tais métodos é maximizar o logaritmo da função de verossimilhança penalizada. Na literatura numérica, o algoritmo é geralmente chamado de método iterativo de Gauss-Seidel (Ralston e Rabinowitz (2001) e Golub e Van Loan (2012, Cap. 10)). Diversos trabalhos sobre esse princípio de redução de dimensão e modelos aditivos generalizados promoveram o uso do algoritmo de *backfitting*, tais como Friedman e Stuetzle (1981), Breiman e Friedman (1985), Hastie e Tibshirani (1986), Hastie e Tibshirani (1987), Hastie e Tibshirani (1990). Estabelecer convergência do algoritmo em casos de interesse estatístico mostrou-se bastante difícil (Buja *et al.*, 1989, Seção 3.4).

O algoritmo de ajuste *backfitting* percorre os termos individuais no modelo aditivo e atualiza

cada um usando um processo mais suave. Isso é feito suavizando os resíduos parciais adequadamente definidos até que a convergência seja alcançada. Os ciclos continuam até que nenhuma das funções mude de uma iteração para a seguinte. Normalmente, são necessárias três ou quatro iterações por variável. Isso pode ser uma tarefa computacionalmente intensiva, se a dimensão k for alta. Portanto, são necessárias bases rápidas de regressão unidimensional como blocos de construção elementares desse algoritmo. Nesse caso, o suavizador de *spline* cúbico de regressão penalizada com nós fixos desenvolvido em Wood (2017) irá apresentar maior eficiência computacional em relação aos *splines* cúbicos naturais.

2.3 Modelo com erros autorregressivos de primeira ordem

Os modelos lineares anteriores assumem que os erros são identicamente e independentemente distribuídos. A suposição de independência em alguns momentos não é satisfeita, mas geralmente é uma aproximação razoável. Quando as observações abrangem uma série temporal, em que um único indivíduo é rastreado por muitos períodos ou pontos no tempo as dependências entre os erros podem existir. Os períodos próximos um dos outros são prováveis de serem semelhantes aos períodos relativamente remotos. Essa semelhança pode muito bem se estender aos erros, que representam as causas omitidas da variável resposta. Embora a dependência de tempo entre os erros possa ser insignificante, não é aconselhável assumir *a priori* que esse seja o caso.

Considere que o processo mais usado na prática é o processo autorregressivo de primeira ordem, abreviado AR(1):

$$\epsilon_t = \rho\epsilon_{t-1} + e_t, \quad (2.13)$$

em que, $|\rho| < 1$, o erro no período t depende diretamente apenas do erro do período anterior, ϵ_{t-1} , e de um erro contemporâneo aleatório e_t . Diferentemente dos erros de regressão ϵ_t , assumiremos que os erros aleatórios e_t são independentes um do outro (e de ϵ de períodos anteriores) e que $e_t \sim N(0, \phi)$. A correlação serial nos erros de regressão, portanto, é totalmente gerada pela dependência parcial de cada erro do período anterior. Devido à sua importância nas aplicações e à sua simplicidade, vamos descrever o processo AR(1) com mais detalhes.

Suponha, no entanto, que o processo que gera os erros seja *estacionário*. Estacionariedade significa que todos os erros têm a mesma esperança (que, de fato, supomos ser 0), uma dispersão comum (ϕ) e que a covariância de dois erros depende apenas da separação entre eles no tempo. Denote ϵ_t como o erro para o instante de tempo t e ϵ_{t+s} o erro para o instante de tempo $t + s$ (em que s é um número inteiro - positivo, negativo ou 0), a estacionariedade implica que, para qualquer t , a covariância entre ϵ_t e ϵ_{t+s} é

$$\text{Cov}(\epsilon_t, \epsilon_{t+s}) = \text{E}(\epsilon_t \epsilon_{t+s}) = \phi \rho_s = \text{Cov}(\epsilon_t, \epsilon_{t-s}),$$

em que ρ_s , chamada autocorrelação (ou correlação serial) em defasagem (lag) s , é a correlação entre dois erros separados por instantes de tempo.

Algumas realizações de séries temporais geradas pela equação (2.13), com $e_t \sim N(0, 1)$, são mostradas na Figura 2.4. Na Figura 2.4(a), $\rho = 0$ e, conseqüentemente, os ϵ_t não estão correlacionados, um processo de série temporal às vezes denominado ruído branco. Na Figura 2.4(b), $\rho = 0,85$; observe como os valores da série próximos tendem a ser semelhantes. Na Figura 2.4(c), $\rho = -0,65$; veja aqui como a série tende a passar de valores negativos para positivos. Finalmente, a Figura 2.4(d) ilustra um processo não estacionário, com $\rho = 1,01$.

A Figura 2.4(b) também fornece algumas informações intuitivas sobre os problemas de estimativa apresentados por erros autocorrelacionados:

1. Como os erros que estão próximos no tempo provavelmente são semelhantes, há muito menos informações em uma série temporal altamente autocorrelacionada do que em uma amostra aleatória independente do mesmo tamanho. Isso provavelmente aumentará a autocorrelação dos erros. Por exemplo, muitas vezes é improdutivo gerar observações usando períodos de

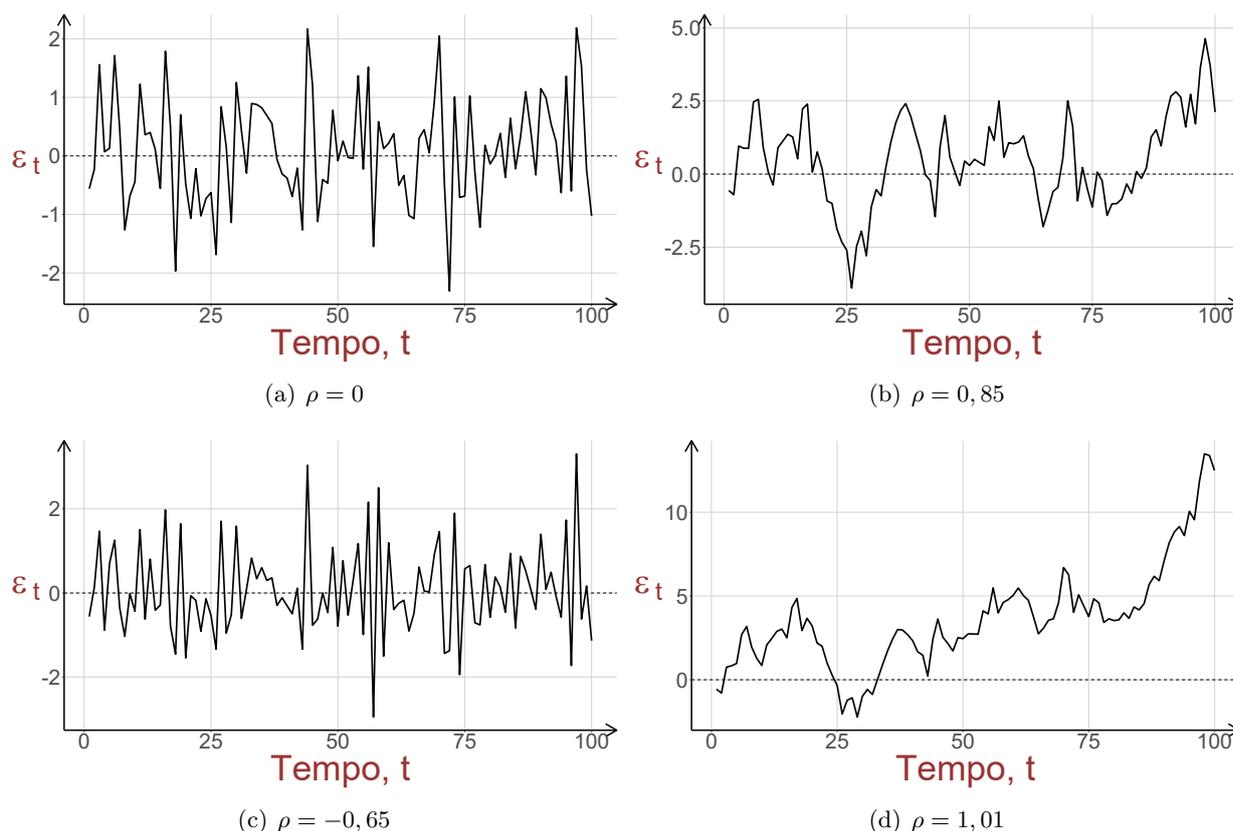


Figura 2.4: Ilustração de séries temporais para diferentes valores do termo de autocorrelação ρ .

tempo menos espaçados (dados mensais ou trimestrais em vez de anuais). Os dados mensais ou trimestrais também aumentam a possibilidade de efeitos “sazonais”. Uma abordagem simples para efeitos sazonais é incluir regressores fictícios por meses ou trimestres. Da mesma forma, regressores simulados para os dias da semana podem ser apropriados para algumas séries temporais diárias. Abordagens mais sofisticadas para efeitos sazonais são descritas na maioria dos textos sobre análise de séries temporais, como as de [Harvey \(1990\)](#) detalhadas na seção 7.6 e [George G. Judge \(1985\)](#) mencionadas nas seções 7.2.4 e 7.7.2.

2. Durante um período relativamente curto, é provável que uma série altamente autocorrelacionada suba ou desça, ou seja, mostre uma tendência positiva ou negativa. Isso é verdade mesmo que a série seja estacionária e, portanto, acabe retornando à sua expectativa de 0. Se, por exemplo, nossa amostra consistisse apenas nas 25 primeiras observações da Figura 2.4(b), haveria uma tendência negativa nos erros; se nossa amostra consistisse em observações de 75 a 100, haveria uma tendência positiva.
3. Como as variáveis explicativas em uma regressão de série temporal também manifestam tendências direcionais, um aumento ou queda nos erros de uma série temporal curta pode induzir uma correlação entre uma variável explicativa e os erros em uma amostra específica. É importante, no entanto, entender que não há implicações de que as estimativas por mínimos quadrados sejam tendenciosas devido à correlação entre as variáveis explicativas e os erros. Em muitas amostras, às vezes, haverá correlações negativas entre os erros e as variáveis explicativas, às vezes, correlações positivas e, às vezes, praticamente nenhuma correlação. Entretanto, as correlações – às vezes negativas, às vezes positivas - que ocorrem em amostras específicas podem aumentar acentuadamente a variação do estimador de mínimos quadrados.
4. Finalmente, como o estimador de mínimos quadrados força uma correlação nula da amostra entre as variáveis explicativas e os resíduos (em oposição aos erros não observados), as variações

de amostragem dos coeficientes de mínimos quadrados podem ser subestimadas. Lembre-se de que, em uma série curta, erros altamente correlacionados automaticamente manifestam uma tendência em uma amostra específica.

O processo AR(1) é o membro mais simples da família de processos autorregressivos. No processo autoregressivo de ordem p [abreviado AR(p)], ϵ_t depende diretamente dos erros p anteriores e de um erro aleatório e_t :

$$\epsilon_t = \rho_1\epsilon_{t-1} + \rho_2\epsilon_{t-2} + \dots + \rho_p\epsilon_{t-p} + e_t.$$

Conforme [Kutner et al. \(2005\)](#), uma das principais causas dos termos de erro autocorrelacionados positivamente em aplicações de regressão econômica e de negócios envolvendo dados de séries temporais é a omissão de uma ou várias variáveis-chave do modelo. Quando os efeitos ordenados pelo tempo dessas variáveis-chave “ausentes” são correlacionados positivamente, os termos de erro no modelo de regressão tendem a ser automaticamente correlacionados positivamente, pois incluem efeitos de variáveis ausentes.

Considere, por exemplo, a regressão das vendas anuais de um produto em relação ao preço médio anual do produto durante um período de 30 anos. Se o tamanho da população tiver um efeito importante nas vendas, sua omissão no modelo pode levar à correlação positiva dos termos de erro, porque o efeito do tamanho da população nas vendas provavelmente está positivamente correlacionado ao longo do tempo. Outra causa dos termos de erro autocorrelacionado positivamente nos dados econômicos é a presença de erros sistemáticos de cobertura nas séries temporais de variáveis de resposta, cujos erros geralmente tendem a se correlacionar positivamente no decorrer do tempo.

Quando os termos de erro no modelo de regressão são automaticamente correlacionados positivamente, o uso de procedimentos comuns de mínimos quadrados tem várias consequências importantes: (i) os coeficientes de regressão estimados ainda são imparciais, mas não possuem mais a propriedade de variação mínima e podem ser bastante ineficientes; (ii) o erro quadrático médio pode subestimar seriamente a variação dos termos do erro; (iii) o desvio padrão estimado calculado de acordo com procedimentos comuns de mínimos quadrados pode subestimar seriamente o verdadeiro desvio padrão do coeficiente de regressão estimado e (iv) intervalos de confiança e testes usando as distribuições t -Student e F de Snedecor não são mais estritamente aplicáveis.

Dessa forma, percebe-se a importância de incluir no modelo a presença da estrutura de correlação a fim de evitar os problemas. O modelo de regressão linear com erros autorregressivos de primeira ordem é definido como

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \epsilon_i, \\ \epsilon_i &= \rho\epsilon_{i-1} + e_t. \end{aligned}$$

Podemos expandir o modelo acima considerando um número maior de variáveis explicativas de forma similar. Este modelo de regressão apresenta algumas propriedades interessantes. Dentre elas, podemos destacar que a correlação imposta aos erros é função do parâmetro ρ . Mais precisamente, temos que a correlação entre erros defasados em s unidades é igual a ρ^s . Há diversos processos de estimação para os modelos lineares com erros autorregressivos, para mais detalhes veja [Fox \(2015\)](#) e [Kutner et al. \(2005\)](#).

Diante de tais circunstâncias, [Relvas \(2013\)](#) desenvolveu em sua dissertação a generalização de tal modelo para os modelos parcialmente lineares com erros simétricos autoregressivos de primeira ordem com o intuito de não restringir aos erros com distribuição normal. Em particular, a suposição de erros simétricos com caudas mais pesadas do que as caudas da distribuição normal leva a estimativas robustas contra observações aberrantes.

Capítulo 3

Modelos aditivos com erros simétricos condicionais AR(1)

Para desenvolver modelos precisos de forma consistente, um pesquisador deve utilizar um procedimento sistemático que possa ser aplicado rapidamente e modificado conforme necessário. Dessa forma, seguir um processo científico pode ajudar. O método científico é composto por instruções que são descrições de conjuntos de ações ou operações que podemos seguir com precisão. Referimos essas instruções como “definição operacional”, a qual deve conter uma especificação de todas as operações necessárias para alcançar os mesmos resultados quando repetidas.

O processo de modelagem pode ser simples ou complexo. Começa quando surge a necessidade de entender um determinado conjunto de dados que exige uma resposta. Perguntas como por quê a modelagem é necessária e como os resultados serão usados são muito importantes. Uma vez tomada a decisão de desenvolver uma previsão, é necessário elaborar um modelo teórico que responda às perguntas do pesquisador.

A construção teórica do modelo ajuda formar a relação que existe entre as inúmeras variáveis do modelo. Ao mesmo tempo, permite segregar influências em fatores internos ou externos. Aqueles que o pesquisador possui controle são os fatores internos. Já aqueles que ele opera fora do controle são os fatores externos. Por um lado, se estiver interessado em uma previsão de tendências a longo prazo, podemos escolher, por exemplo, uma tendência ou um modelo de regressão para esta tarefa. Por outro lado, se for do seu interesse uma previsão de curto prazo, ou seja, projeções semanais ou mensais, podemos escolher, por exemplo, a média móvel, suavização exponencial ou modelos Box-Jenkins para a análise.

Com os dados em mãos, estamos prontos para identificar o modelo teórico mais adequado, ou seja, aquele que suporta a estrutura da investigação. Os dados podem vir de documentos internos ou de fontes externas. Deve-se ter cuidado com o tipo e a qualidade do que está sendo coletado. A análise dos dados não deve ser tomada simplesmente como um processo mecânico. Antes de ajustar o modelo, devemos familiarizar-se com o problema em si. Qualquer informação que ajude no processo de ajuste, como natureza dos dados, variáveis externas e internas, tendência e sazonalidade, deve ser analisada cuidadosamente. Isso permite que o pesquisador veja o modelo ajustado de uma perspectiva dinâmica. Caso seja necessário reavaliar os resultados, é nesse estágio que as mudanças podem ser feitas. Os modelos devem ser testados quanto à sua confiabilidade e validade. Os resultados da previsão devem ser comparados com os reais para avaliar a precisão do modelo.

A avaliação nessa fase serve como um processo de controle. Às vezes, a confiabilidade e validade do modelo aumentam com a inclusão de outros fatores ou variáveis, alterações no período de tempo ou variações na periodicidade dos dados. O ciclo de *feedback* deve ser usado para ajustar a previsão adequadamente. A etapa final do processo da pesquisa é apresentar os resultados. É essencial perceber que, para o pesquisador, embora a tarefa de fornecer uma previsão tenha terminado, o processo de refiná-la ainda não foi finalizado.

Para enfatizar, uma boa previsão não é estática, mas dinâmica. Seria vantajoso estabelecer um procedimento para avaliar seu modelo continuamente e reconhecer que condições inesperadas no

fenômeno dos dados podem alterar as premissas do modelo previsto e, portanto, exigir uma nova estimativa.

O foco desta tese é examinar o problema de estimar a função $f(\cdot)$ na presença de correlação, não o de estimar a própria função de correlação. Nesse contexto, nossos objetivos são (1) explicar algumas das dificuldades associadas à presença de correlação na regressão não paramétrica, (2) fornecer uma visão geral da literatura de regressão não paramétrica que trata do caso de erros correlatos e (3) desenvolver métodos para contribuir neste processo. Grande parte da literatura em regressão não paramétrica se baseia em argumentos assintóticos para esclarecer o comportamento probabilístico dos métodos propostos. A mesma abordagem será usada aqui, mas tentamos fornecer intuição aos resultados também. Vários problemas, alguns bastante fundamentais, ocorrem quando se tenta a regressão não paramétrica na presença de erros correlatos. De fato, no cenário mais geral em que nenhuma forma paramétrica é assumida para a média nem a função de correlação, o modelo é essencialmente não identificável, de modo que é teoricamente impossível estimar qualquer uma das funções separadamente. Na maioria das aplicações práticas, no entanto, o pesquisador tem alguma ideia do que representa um tipo razoável de ajuste para os dados disponíveis, e ele usará essa expectativa para decidir o que é uma estimativa de função “aceitável” ou “inaceitável”. Para todas as técnicas de regressão não paramétrica, a forma e a suavidade da função estimada dependem em grande parte do valor específico escolhido para um “parâmetro de suavização”, definido diferentemente para cada técnica. Para evitar a necessidade de selecionar um valor para o parâmetro de suavização por tentativa e erro, vários tipos de métodos de seleção orientados a dados foram desenvolvidos para auxiliar os estudiosos nessa tarefa. No entanto, a presença de correlação entre os erros, se ignorada, faz com que os métodos de seleção de parâmetros de ajuste automático comumente usados, como validação cruzada, sejam interrompidos.

Para avaliar a utilidade de uma técnica de modelagem em uma situação específica, neste capítulo, introduziremos os modelos não paramétricos com erros simétricos autorregressivos de primeira ordem e, assim, prestaremos atenção ao padrão dos dados em relação à tendência e sazonalidade usando funções não paramétricas para descrevê-las. O modelo proposto é definido na Seção 3.1, enquanto a função de verossimilhança penalizada é considerada na Seção 3.2 para a estimação paramétrica. A função escore e a matriz de informação de Fisher para os parâmetros de interesse são derivados na Seção 3.3, assim como o processo iterativo de *backfitting* (Gauss-Seidel) para obter as estimativas de máxima verossimilhança penalizada (EMVPs). Procedimentos inferenciais são discutidos na Seção 3.4 e os graus de liberdade efetivos são derivados a partir de um suavizador apropriado na Seção 3.5.

3.1 O modelo

Baseado em Cleveland *et al.* (1990) e Wood (2017, Seção 7.7.2) propõem-se modelar a tendência e a sazonalidade de uma série temporal seguindo um modelo aditivo com erros simétricos autorregressivos de primeira ordem:

$$y_i = f_T(t_i) + f_S(s_i) + \epsilon_i, \quad (i = 1, \dots, n), \quad (3.1)$$

$$\epsilon_i = \rho\epsilon_{i-1} + e_i, \quad -1 < \rho < 1, \quad (3.2)$$

em que y_i denota a i -ésima resposta observada, $f_T(t_i)$ e $f_S(s_i)$ são as funções não paramétricas suavizadoras, t_i e s_i indicam, respectivamente, o i -ésimo tempo e o tempo do respectivo ano para o i -ésimo tempo, ρ é o coeficiente de autocorrelação, enquanto e_i são os erros simétricos independentes com média zero e parâmetro de dispersão ϕ , isto é, $e_i \stackrel{iid}{\sim} S(0, \phi)$ para $i = 1, \dots, n$. Há alguns atrativos ao considerar erros AR(1), como apontado por Greene (2012, p. 910), tem-se a conveniência e a aproximação razoável para modelar processos mais complexos. Todavia, extensões para erros AR(p) também podem ser consideradas no modelo (3.1)-(3.2) e será discutida com maiores detalhes no próximo capítulo.

As funções $f(\cdot)$ descritas pelos autores Green e Silverman (1994), Ibacache-Pulgar *et al.* (2013) e Relvas e Paula (2016), já discutidas anteriormente, possuem a forma de *spline* cúbico natural, ou seja, o número de nós não é fixo e corresponde ao número de valores distintos da variável explicativa. As funções $f(\cdot)$ aqui descritas assumem formas lineares flexíveis, conforme discutido em Wood (2017). Assume-se que $f(\cdot)$ pertence ao espaço funcional, e respeita algumas condições estabelecidas. Pretende-se encontrar uma combinação linear de funções nesse espaço capaz de explicar o comportamento e se ajuste à variável explicativa de forma não paramétrica. As estimativas do componente não paramétrico t podem ser obtidas pelos métodos de suavização por *splines*, um critério de penalização e com a subvenção do parâmetro de suavização. Dessa forma, para cada ponto t_i denominado nó, será obtida uma estimativa $f(t_i)$. É fundamental escolher um certo número de nós, mas, para isso, é necessário tomar cuidado em relação à escolha da quantidade de nós; se for um número muito grande, pode causar um ajuste próximo à interpolação dos dados, o que não é conveniente; se o número de nós for pequeno, pode-se acarretar um mau ajuste.

Sendo assim, as funções $f_T(t)$ e $f_S(s)$ são expressas na forma linear $f_T(t) = \sum_{j=1}^{r_T} \eta_{T_j}(t) \gamma_{T_j}$ e $f_S(s) = \sum_{l=1}^{r_S-1} \eta_{S_l}(s) \gamma_{S_l}$, respectivamente, com $\eta_{T_j}(t)$ e $\eta_{S_l}(s)$ denotando a j -ésima e l -ésima funções base e γ_{T_j} e γ_{S_l} são parâmetros desconhecidos a serem estimados no modelo. Em particular, $f_T(t)$ será representada por um *spline* cúbico penalizado de regressão com nós fixos nos valores t_j^0 ($j = 1, \dots, r_T$), enquanto que $f_S(s)$ será representado por um *spline* cúbico penalizado de regressão cíclico com nós fixos nos valores s_l^0 ($l = 1, \dots, r_S - 1$), já discutidos anteriormente na Seção 2.1 e descritos na Tabela 2.1 referente a Tabela 5.1 de Wood (2017). Note que as funções f_T e f_S não são estimadas, mas são expressas em formas lineares flexíveis.

Podemos escrever o modelo (3.1)-(3.2) na seguinte forma matricial:

$$\mathbf{y} = \mathbf{N}_T \boldsymbol{\gamma}_T + \mathbf{N}_S \boldsymbol{\gamma}_S + \boldsymbol{\epsilon}, \quad (3.3)$$

em que \mathbf{y} é um vetor ($n \times 1$) de respostas observadas, \mathbf{N}_T é uma matriz ($n \times r_T$) e função base com linhas $\mathbf{n}_T(t_i) = (\eta_{T_1}(t_i), \dots, \eta_{T_{r_T}}(t_i))^\top$ e o (i, j) -ésimo elemento igual à parametrização dos *splines* em termos de seus valores nos nós t_j^0 , \mathbf{N}_S é uma matriz ($n \times r_S - 1$) e função base com linhas $\mathbf{n}_S(t_i) = (\eta_{S_1}(s_i), \dots, \eta_{S_{r_S-1}}(s_i))^\top$ e o (i, l) -ésimo elemento igual à parametrização dos *splines* em termos de seus valores nos nós s_l^0 , para $i = 1, \dots, n$, $j = 1, \dots, r_T$ e $l = 1, \dots, r_S - 1$, e $\boldsymbol{\epsilon}$ é o vetor ($n \times 1$) de erros autoregressivos de primeira ordem dado por $(\epsilon_1, \dots, \epsilon_n)^\top$. Os parâmetros desconhecidos são $\boldsymbol{\gamma}_T = (\gamma_{T_1}, \dots, \gamma_{T_{r_T}})^\top$ e $\boldsymbol{\gamma}_S = (\gamma_{S_1}, \dots, \gamma_{S_{r_S-1}})^\top$. Expressões para $\eta_T(t)$ e $\eta_S(s)$ podem ser derivadas a partir da Tabela 5.1, elaborada com base em Wood (2017). Portanto, o modelo matricial em (3.3) possui as matrizes

$$\mathbf{N}_T = \begin{bmatrix} \eta_{T_1}(t_1) & \cdots & \eta_{T_{r_T}}(t_1) \\ \eta_{T_1}(t_2) & \cdots & \eta_{T_{r_T}}(t_2) \\ \vdots & \ddots & \vdots \\ \eta_{T_1}(t_n) & \cdots & \eta_{T_{r_T}}(t_n) \end{bmatrix} \quad \text{e} \quad \mathbf{N}_S = \begin{bmatrix} \eta_{S_1}(s_1) & \cdots & \eta_{S_{r_S-1}}(s_1) \\ \eta_{S_1}(s_2) & \cdots & \eta_{S_{r_S-1}}(s_2) \\ \vdots & \ddots & \vdots \\ \eta_{S_1}(s_{n-1}) & \cdots & \eta_{S_{r_S-1}}(s_{n-1}) \\ \eta_{S_1}(s_1) & \cdots & \eta_{S_{r_S-1}}(s_1) \end{bmatrix}.$$

Sabendo que e_i é uma variável aleatória simétrica independente, tem-se que a função densidade de probabilidade é dada por

$$h_e(e_i) = \frac{1}{\sqrt{\phi}} g(\delta_i), \quad e \in \mathcal{R},$$

em que $\delta_i = \phi^{-1} e_i^2$, com $g : \mathcal{R} \rightarrow [0, \infty)$ tal que $\int_0^\infty u^{-\frac{1}{2}} g(u) du = 1$, sendo conhecida como uma geradora de densidades (veja, por exemplo, Fang *et al.* (1990)). Quando a geradora existe, temos que $E(e_i) = 0$ e $\text{Var}(e_i) = \xi_g \phi$, em que $\xi_g > 0$ é uma constante que pode ser obtida a partir do valor esperado da variável radial ou a partir da derivação da função característica. Cysneiros e Paula (2005) fornecem informações adicionais sobre as expressões de ξ_g para algumas distribuições simétricas, como apresentado na Tabela 2.2. Assume-se que ξ_g é conhecido ou fixo, ou estimado separadamente.

Quando existem, a média e a variância de y_i , dado o conjunto de observações passadas y_{i-1} , similarmemente a [Relvas e Paula \(2016\)](#), pode-se derivar do modelo (3.1)-(3.2) da seguinte maneira:

1. Para a primeira observação, tem-se que $y_1 \sim S(\mu_1^*, \phi)$ com $\mu_1^* = E(y_1) = f_T(t_1) + f_S(s_1) = \boldsymbol{\eta}_T(t_1)^\top \boldsymbol{\gamma}_T + \boldsymbol{\eta}_S(t_1)^\top \boldsymbol{\gamma}_S$.
2. Para a distribuição condicional $y_i | y_{i-1} \stackrel{ind}{\sim} S(\mu_i^*, \phi)$, tem-se que $\mu_i^* = E(y_i | y_{i-1}) = f_T(t_i) + f_S(s_i) + \rho\{y_{i-1} - f_T(t_{i-1}) - f_S(s_{i-1})\}$, para $i = 2, \dots, n$.

Os primeiros dois momentos para y_i são derivados com mais detalhes em [Relvas e Paula \(2016\)](#), e nota-se que

$$E(y_i) = E[E(y_i | y_{i-1})] = \boldsymbol{\eta}_T(t_i)^\top \boldsymbol{\gamma}_T + \boldsymbol{\eta}_S(s_i)^\top \boldsymbol{\gamma}_S, \quad \forall i,$$

$$\text{Var}(y_i) = \text{Var}[E(y_i | y_{i-1})] + E[\text{Var}(y_i | y_{i-1})] = \phi\xi \sum_{j=0}^{i-1} \rho^{2j} = \phi\xi \frac{(1 - \rho^{2i})}{(1 - \rho^2)}.$$

Para obtermos as covariâncias entre diferentes observações de y_i e y_{i-j} , ou seja, $E(y_i, y_{i-j})$, tal que $j < i$, temos que,

$$\begin{aligned} E(y_i y_{i-j}) &= E \left\{ \left[(\boldsymbol{\eta}_T^\top(t_i) \boldsymbol{\gamma}_T + \boldsymbol{\eta}_S^\top(s_i) \boldsymbol{\gamma}_S + \rho(y_{i-1} - \boldsymbol{\eta}_T^\top(t_{i-1}) \boldsymbol{\gamma}_T - \boldsymbol{\eta}_S^\top(s_{i-1}) \boldsymbol{\gamma}_S) + e_i) \right. \right. \\ &\quad \left. \left. \times (\boldsymbol{\eta}_T^\top(t_{i-j}) \boldsymbol{\gamma}_T + \boldsymbol{\eta}_S^\top(s_{i-j}) \boldsymbol{\gamma}_S + \rho(y_{i-j-1} - \boldsymbol{\eta}_T^\top(t_{i-j-1}) \boldsymbol{\gamma}_T - \boldsymbol{\eta}_S^\top(s_{i-j-1}) \boldsymbol{\gamma}_S) + e_{i-j}) \right] \right\} \\ &= (\boldsymbol{\eta}_T^\top(t_i) \boldsymbol{\gamma}_T + \boldsymbol{\eta}_S^\top(s_i) \boldsymbol{\gamma}_S) (\boldsymbol{\eta}_T^\top(t_{i-j}) \boldsymbol{\gamma}_T + \boldsymbol{\eta}_S^\top(s_{i-j}) \boldsymbol{\gamma}_S) \\ &\quad + \rho^2 E \left\{ (y_{i-1} - \boldsymbol{\eta}_T^\top(t_{i-1}) \boldsymbol{\gamma}_T - \boldsymbol{\eta}_S^\top(s_{i-1}) \boldsymbol{\gamma}_S) (y_{i-j-1} - \boldsymbol{\eta}_T^\top(t_{i-j-1}) \boldsymbol{\gamma}_T - \boldsymbol{\eta}_S^\top(s_{i-j-1}) \boldsymbol{\gamma}_S) \right\} \\ &\quad + \rho (E(y_{i-1} e_{i-j}) + E(e_i y_{i-j-1})) \\ &= E(y_i) E(y_{i-j}) + \rho (E(y_{i-1} e_{i-j}) + E(e_i y_{i-j-1})) + \rho^2 \text{Cov}(y_{i-1}, y_{i-j-1}), \end{aligned}$$

mas, como $i - j - 1 < i$, temos que $E(e_i y_{i-j-1}) = 0$, portanto

$$\text{Cov}(y_i, y_{i-j}) = \rho E(y_{i-1} e_{i-j}) + \rho^2 \text{Cov}(y_{i-1}, y_{i-j-1}).$$

Uma expressão geral para $\text{Cov}(y_i, y_{i-j})$, $j < i$, com erros autorregressivos de primeira ordem podem ser expressas por dois casos $j = i - 1$ e $j < i - 1$.

Para o caso $j = i - 1$

$$\text{Cov}(y_i, y_{i-j}) = \text{Cov}(y_i, y_1) = \rho E(y_{i-1} e_1) = \phi\xi \rho^{i-1}.$$

Já no caso $j < i - 1$,

$$\begin{aligned} \text{Cov}(y_i, y_{i-j}) &= \phi\xi \rho^j + \rho^2 \text{Cov}(y_{i-1}, y_{i-j-1}) \\ &= \phi\xi \rho^j + \rho^2 \rho^j \text{Cov}(y_{i-2}, y_{i-j-2}) \\ &= \phi\xi \rho^j + \phi\xi \rho^{j+2} + \rho^2 \rho^{j+2} \text{Cov}(y_{i-3}, y_{i-j-3}) \\ &\quad \vdots \\ &= \phi\xi \sum_{l=0}^{i-2} \rho^{j+2l}. \end{aligned}$$

Podemos destacar que a maior covariância está entre observações consecutivas, para $|\rho| < 1$. Além disso, para pontos distantes, ou observações não consecutivas, a covariância torna-se cada vez mais próxima de 0.

3.2 Função de verossimilhança penalizada

Suponha que $y_i, i = 1, \dots, n$, é uma variável aleatória e assuma que condicionalmente ao conjunto de informações passadas, y_i segue uma distribuição simétrica, ou seja, $y_i | y_{i-1} \stackrel{ind}{\sim} S(\mu_i^*, \phi)$, logo podemos reescrever a função de distribuição conjunta do modelo da seguinte maneira:

$$h_Y(y_1, y_2, \dots, y_n) = h_n(y_n | y_{n-1}, \dots, y_1) h_{n-1}(y_{n-1} | y_{n-2}, \dots, y_1) \dots h_2(y_2 | y_1) h_1(y_1).$$

Logo, similarmente aos modelos simétricos, considere $\boldsymbol{\theta} = (\boldsymbol{\gamma}_T^\top, \boldsymbol{\gamma}_S^\top, \phi, \rho)^\top \in \boldsymbol{\Theta} \subseteq \mathcal{R}^r$, com $r = r_T + r_S + 1$ sendo o número de parâmetros a ser estimado. O logaritmo da função de verossimilhança associada a $\boldsymbol{\theta}$ é dado por

$$L(\boldsymbol{\theta}) = -\frac{n}{2} \log(\phi) + \sum_{i=1}^n \log\{g(\delta_i)\}. \quad (3.4)$$

A maximização de (3.4) sem impor restrições sobre as funções não paramétricas $f_T(t)$ e $f_S(s)$ pode levar a *overfitting* e a não identificabilidade dos parâmetros. Para resolver esse problema, deve-se incorporar uma função penalidade sobre cada componente não paramétrico. Analogamente a [Hastie e Tibshirani \(1990\)](#) (veja também [Green e Silverman \(1994\)](#)), assume-se que $f_T(t)$ e $f_S(s)$ são contínuas e integráveis na segunda derivada, então o logaritmo da função de verossimilhança penalizada fica dado por

$$L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = L(\boldsymbol{\theta}) - \frac{\lambda_T}{2} \int_{a_T}^{b_T} [f_T''(t)]^2 dt - \frac{\lambda_S}{2} \int_{a_S}^{b_S} [f_S''(s)]^2 ds, \quad (3.5)$$

isto é, tem-se que a penalização da segunda derivada de $f_T(t)$ e $f_S(s)$ para $a_T \leq t \leq b_T$ e $a_S \leq s \leq b_S$, em que $\lambda_T > 0$ e $\lambda_S > 0$, são os parâmetros suavizadores que serão estimados separadamente. De [Wood \(2017\)](#), tem-se que

$$\int_{a_T}^{b_T} [f_T''(t)]^2 dt = \boldsymbol{\gamma}_T^\top \mathbf{M}_T \boldsymbol{\gamma}_T \quad \text{e} \quad \int_{a_S}^{b_S} [f_S''(s)]^2 ds = \boldsymbol{\gamma}_S^\top \mathbf{M}_S \boldsymbol{\gamma}_S,$$

em que $\mathbf{M}_T = \mathbf{D}_T^\top \mathbf{B}_T^{-1} \mathbf{D}_T$ e $\mathbf{M}_S = \mathbf{D}_S^\top \mathbf{B}_S^{-1} \mathbf{D}_S$ são matrizes não negativas definidas de dimensões $(r_T \times r_T)$ e $(r_S - 1 \times r_S - 1)$, respectivamente, nomeadas de matrizes de penalidade. Essas matrizes dependem somente dos nós e são definidas na Tabela 2.1 para as versões *splines* cúbicas de regressão penalizada para a tendência e sua versão cíclica para a sazonalidade (veja Tabela 5.1 elaborada por [\(Wood, 2017\)](#)).

Assim, o logaritmo da função de verossimilhança penalizada reduz a

$$L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = L(\boldsymbol{\theta}) - \frac{\lambda_T}{2} \boldsymbol{\gamma}_T^\top \mathbf{M}_T \boldsymbol{\gamma}_T - \frac{\lambda_S}{2} \boldsymbol{\gamma}_S^\top \mathbf{M}_S \boldsymbol{\gamma}_S, \quad (3.6)$$

em que $\boldsymbol{\lambda} = (\lambda_T, \lambda_S)^\top$ denota o vetor paramétrico de suavização. Na próxima seção será derivado o processo iterativo *backfitting* (Gauss-Seidel) para maximizar (3.6) para $\boldsymbol{\lambda}$ fixo.

3.3 Estimação dos parâmetros

Nesta seção apresentamos a função score penalizada, a matriz Hessiana penalizada e a matriz informação de Fisher penalizada e o processo iterativo para estimação das componentes que envolvem a tendência, a sazonalidade e também os parâmetros da distribuição simétrica com erros autorregressivos de ordem um, os cálculos com maiores detalhes podem ser vistos no Apêndice A.1.

3.3.1 Função escore penalizada

A função escore penalizada de $\boldsymbol{\theta}$ pode ser expressa como $\mathbf{U}_p^\theta = \partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) / \partial \boldsymbol{\theta} = \mathbf{U}_\theta - \mathbf{M}(\boldsymbol{\lambda})\boldsymbol{\theta}$, em que \mathbf{U}_θ é a função escore regular dada por

$$\mathbf{U}_\theta = \begin{pmatrix} \mathbf{U}_{\gamma_T} \\ \mathbf{U}_{\gamma_S} \\ \dots \\ \mathbf{U}_\phi \\ \mathbf{U}_\rho \end{pmatrix} = \begin{pmatrix} \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_T} \\ \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_S} \\ \dots \\ \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi} \\ \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho} \end{pmatrix} = \begin{pmatrix} \frac{1}{\phi} (\mathbf{A}\mathbf{N}_T)^\top \mathbf{D}_v \mathbf{A} \boldsymbol{\epsilon} \\ \frac{1}{\phi} (\mathbf{A}\mathbf{N}_S)^\top \mathbf{D}_v \mathbf{A} \boldsymbol{\epsilon} \\ \dots \\ \frac{1}{2\phi} \mathbf{1}_n^\top (\mathbf{D}_m \mathbf{1}_n - \mathbf{1}_n) \\ -\frac{1}{\phi} (\mathbf{B}\boldsymbol{\epsilon})^\top \mathbf{D}_v \mathbf{A} \boldsymbol{\epsilon} \end{pmatrix},$$

em que $\mathbf{M}(\boldsymbol{\lambda}) = \text{blockdiag}\{\lambda_T \mathbf{M}_T, \lambda_S \mathbf{M}_S, \mathbf{0}_2\}$ com $\mathbf{0}_2$ sendo uma matriz de zeros (2×2), $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{N}\boldsymbol{\gamma}$ com $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_T^\top, \boldsymbol{\gamma}_S^\top)^\top$ e $\mathbf{N} = (\mathbf{N}_T, \mathbf{N}_S)$, as matrizes \mathbf{A} e \mathbf{B} de dimensão $(n \times n)$ foram usadas por [Cao et al. \(2010\)](#) e são definidas por

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ \vdots & & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & -\rho & 1 \end{pmatrix} \text{ e } \mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 0 & \dots & 0 & 0 \\ \vdots & & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & -1 & 0 \end{pmatrix},$$

em que $\mathbf{D}_v = \text{diag}\{v_1, \dots, v_n\}$ com $v_i = -2W_g(\delta_i)$ e $\mathbf{D}_m = \text{diag}\{m_1, \dots, m_n\}$ com $m_i = v_i \delta_i$, para $i = 1, \dots, n$, enquanto $W_g(\delta) = g'(\delta)/g(\delta)$ e $\mathbf{1}_n$ é um vetor de uns ($n \times 1$). As quantidades $W_g(\delta)$ e $W_g'(\delta)$ para algumas distribuições simétricas são discutidas em [Cysneiros e Paula \(2005\)](#) e podem ser conferidas na Tabela 2.3. A manipulação computacional a partir dessas matrizes tornam-se mais atrativas e facilita a estimação dos parâmetros do modelo.

3.3.2 Matriz Hessiana penalizada

A matriz Hessiana penalizada (segunda derivada) do logaritmo da função de verossimilhança é dada por $\ddot{\mathbf{L}}_p^{\theta\theta} = \partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top = \ddot{\mathbf{L}}_{\theta\theta} - \mathbf{M}(\boldsymbol{\lambda})$, a qual é detalhada no Apêndice A.2, e é dada por,

$$\ddot{\mathbf{L}}_{\theta\theta} = \begin{pmatrix} \ddot{\mathbf{L}}_{\gamma_T \gamma_T} & \ddot{\mathbf{L}}_{\gamma_T \gamma_S} & \ddot{\mathbf{L}}_{\gamma_T \phi} & \ddot{\mathbf{L}}_{\gamma_T \rho} \\ \ddot{\mathbf{L}}_{\gamma_S \gamma_T} & \ddot{\mathbf{L}}_{\gamma_S \gamma_S} & \ddot{\mathbf{L}}_{\gamma_S \phi} & \ddot{\mathbf{L}}_{\gamma_S \rho} \\ \ddot{\mathbf{L}}_{\phi \gamma_T} & \ddot{\mathbf{L}}_{\phi \gamma_S} & \ddot{\mathbf{L}}_{\phi \phi} & \ddot{\mathbf{L}}_{\phi \rho} \\ \ddot{\mathbf{L}}_{\rho \gamma_T} & \ddot{\mathbf{L}}_{\rho \gamma_S} & \ddot{\mathbf{L}}_{\rho \phi} & \ddot{\mathbf{L}}_{\rho \rho} \end{pmatrix}.$$

Sabe-se que

$$\begin{aligned} \ddot{\mathbf{L}}_{\gamma_T \gamma_S} &= \ddot{\mathbf{L}}_{\gamma_S \gamma_T}^\top, & \ddot{\mathbf{L}}_{\gamma_T \phi} &= \ddot{\mathbf{L}}_{\phi \gamma_T}^\top, & \ddot{\mathbf{L}}_{\gamma_T \rho} &= \ddot{\mathbf{L}}_{\rho \gamma_T}^\top, \\ \ddot{\mathbf{L}}_{\gamma_S \phi} &= \ddot{\mathbf{L}}_{\phi \gamma_S}^\top, & \ddot{\mathbf{L}}_{\gamma_S \rho} &= \ddot{\mathbf{L}}_{\rho \gamma_S}^\top \text{ e} & \ddot{\mathbf{L}}_{\phi \rho} &= \ddot{\mathbf{L}}_{\rho \phi}^\top. \end{aligned}$$

Além disso, as submatrizes da diagonal de $\ddot{\mathbf{L}}_{\theta\theta}$ podem ser escritas, em notação matricial, como

$$\begin{aligned}\ddot{\mathbf{L}}_{\gamma_T\gamma_T} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_T \partial \gamma_T^\top} = \frac{1}{\phi} \left\{ (\mathbf{AN}_T)^\top (-\mathbf{D}_v + 4\mathbf{D}_d) (\mathbf{AN}_T) \right\}, \\ \ddot{\mathbf{L}}_{\gamma_S\gamma_S} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_S \partial \gamma_S^\top} = \frac{1}{\phi} \left\{ (\mathbf{AN}_S)^\top (-\mathbf{D}_v + 4\mathbf{D}_d) (\mathbf{AN}_S) \right\}, \\ \ddot{\mathbf{L}}_{\phi\phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi^2} = \frac{1}{\phi^2} \left\{ \frac{n}{2} + \boldsymbol{\delta}^\top \mathbf{D}_c \boldsymbol{\delta} - \boldsymbol{\delta}^\top \mathbf{D}_v \mathbf{1}_n \right\} \quad e \\ \ddot{\mathbf{L}}_{\rho\rho} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho^2} = \frac{1}{\phi} \left\{ (\mathbf{B}\boldsymbol{\epsilon})^\top (-\mathbf{D}_v + 4\mathbf{D}_d) (\mathbf{B}\boldsymbol{\epsilon}) \right\}.\end{aligned}$$

Já as submatrizes cruzadas são expressas como

$$\begin{aligned}\ddot{\mathbf{L}}_{\gamma_T\gamma_S} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_T \partial \gamma_S^\top} = \frac{1}{\phi} \left\{ (\mathbf{AN}_T)^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{AN}_S) \right\}, \\ \ddot{\mathbf{L}}_{\gamma_T\phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_T \partial \phi^\top} = \frac{1}{\phi} \left\{ (\mathbf{AN}_T)^\top (2\mathbf{D}_d - \mathbf{D}_v) (\mathbf{A}\boldsymbol{\epsilon}) \right\}, \\ \ddot{\mathbf{L}}_{\gamma_T\rho} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_T \partial \rho^\top} = \frac{1}{\phi} \left\{ (\mathbf{B}\boldsymbol{\epsilon})^\top (\mathbf{D}_v - 4\mathbf{D}_d) (\mathbf{AN}_T) + (\mathbf{BN}_T)^\top \mathbf{D}_v \mathbf{A}\boldsymbol{\epsilon} \right\}, \\ \ddot{\mathbf{L}}_{\gamma_S\phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_S \partial \phi^\top} = \frac{1}{\phi} \left\{ (\mathbf{AN}_S)^\top (2\mathbf{D}_d - \mathbf{D}_v) (\mathbf{A}\boldsymbol{\epsilon}) \right\}, \\ \ddot{\mathbf{L}}_{\gamma_S\rho} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_S \partial \rho^\top} = \frac{1}{\phi} \left\{ (\mathbf{B}\boldsymbol{\epsilon})^\top (\mathbf{D}_v - 4\mathbf{D}_d) (\mathbf{AN}_S) + (\mathbf{BN}_S)^\top \mathbf{D}_v \mathbf{A}\boldsymbol{\epsilon} \right\} \quad e \\ \ddot{\mathbf{L}}_{\phi\rho} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi \partial \rho^\top} = \frac{1}{\phi^2} \left\{ (\mathbf{B}\boldsymbol{\epsilon})^\top (\mathbf{D}_v - 2\mathbf{D}_d) (\mathbf{A}\boldsymbol{\epsilon}) \right\}.\end{aligned}$$

Os elementos restantes da matriz são obtidos pela simetria. Com esses cálculos, completamos a matriz observada de Fisher. Entretanto, toda proposta de estimação neste trabalho será realizada com a matriz de informação de Fisher, que representa a matriz esperada, apresentada na próxima subseção.

3.3.3 Matriz de informação de Fisher penalizada

A matriz de informação forma uma ferramenta de interesse para verificar a eficiência, ou seja, a obtenção do limite de informações para a variação do estimador. A matriz de informação, \mathcal{I} , é o valor esperado negativo da matriz de derivadas parciais de segunda ordem do logaritmo da função de verossimilhança a ser maximizada, ou seja, $\mathcal{I}_p^{\theta\theta} = -E \left\{ \partial L_p^2(\boldsymbol{\theta}, \boldsymbol{\lambda}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top \right\} = \mathcal{I}_{\theta\theta} - \mathbf{M}(\boldsymbol{\lambda})$. Os cálculos com maiores detalhes são apresentados no Apêndice A.3 cujas submatrizes são dadas por $\mathcal{I}_{\gamma_T\gamma_T}$, $\mathcal{I}_{\gamma_S\gamma_S}$, $\mathcal{I}_{\phi\phi}$, $\mathcal{I}_{\rho\rho}$, $\mathcal{I}_{\gamma_T\gamma_S}$, $\mathcal{I}_{\gamma_T\phi}$, $\mathcal{I}_{\gamma_T\rho}$, $\mathcal{I}_{\gamma_S\phi}$, $\mathcal{I}_{\gamma_S\rho}$ e $\mathcal{I}_{\phi\rho}$. Além disso, a matriz é simétrica, portanto,

$$\begin{aligned}\mathcal{I}_{\gamma_T\gamma_S} &= \mathcal{I}_{\gamma_S\gamma_T}^\top, & \mathcal{I}_{\gamma_T\phi} &= \mathcal{I}_{\phi\gamma_T}^\top, & \mathcal{I}_{\gamma_T\rho} &= \mathcal{I}_{\rho\gamma_T}^\top, \\ \mathcal{I}_{\gamma_S\phi} &= \mathcal{I}_{\phi\gamma_S}^\top, & \mathcal{I}_{\gamma_S\rho} &= \mathcal{I}_{\rho\gamma_S}^\top \quad e & \mathcal{I}_{\phi\rho} &= \mathcal{I}_{\rho\phi}^\top.\end{aligned}$$

Com o auxílio dos cálculos no Apêndice A.3, a matriz de informação de Fisher, \mathcal{I} , pode ser considerada como uma matriz bloco diagonal da seguinte maneira:

$$\begin{aligned}\mathcal{I}_{\gamma\gamma} &= \frac{4dg}{\phi} \left(\mathbf{N}_A^\top \mathbf{N}_A \right), \\ \mathcal{I}_{\phi\phi} &= \frac{n}{4\phi^2} (4f_g - 1) \text{ e} \\ \mathcal{I}_{\rho\rho} &= \frac{4dg\xi}{1-\rho^2} \left\{ (n-1) - \frac{\rho^2(1-\rho^{2(n-1)})}{1-\rho^2} \right\},\end{aligned}$$

sendo que os demais elementos da matriz são $\mathbf{0}$, uma vez que $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_T^\top, \boldsymbol{\gamma}_S^\top)^\top$ é ortogonal aos parâmetros $(\phi, \rho)^\top$. Temos que $\mathbf{N}_A = (\mathbf{A}\mathbf{N}_T, \mathbf{A}\mathbf{N}_S)$, d_g e f_g são dadas por $E\{W_g^2(z^2)z^2\}$, $E\{W_g^2(z^2)z^4\}$, respectivamente, com $z \sim S(0, 1)$. Essas quantidades podem ser auferidas na Tabela 2.4. Podemos definir a matriz de informação de Fisher regular de $\boldsymbol{\theta}$ como

$$\mathcal{I}_{\theta\theta} = \begin{pmatrix} \mathcal{I}_{\gamma_T\gamma_T} & \mathcal{I}_{\gamma_T\gamma_S} & \mathbf{0} & \mathbf{0} \\ \mathcal{I}_{\gamma_S\gamma_T} & \mathcal{I}_{\gamma_S\gamma_S} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}_{\phi\phi} & 0 \\ \mathbf{0} & \mathbf{0} & 0 & \mathcal{I}_{\rho\rho} \end{pmatrix} = \text{blockdiag}\{\mathcal{I}_{\gamma\gamma}, \mathcal{I}_{\phi\phi}, \mathcal{I}_{\rho\rho}\}.$$

3.3.4 Processo iterativo

Analogamente a [Hastie e Tibshirani \(1990, Cap.5\)](#) e [Wood \(2017, Cap.6\)](#), deriva-se o processo iterativo de *backfitting* (Gauss-Seidel) para obter o EMVP de $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_T^\top, \hat{\boldsymbol{\gamma}}_S^\top)^\top$ alternado com o processo iterativo para obter os EMVP de ϕ e ρ . Assumindo $\boldsymbol{\lambda}$ fixo, tem-se que resolver as seguintes equações:

$$\left\{ (\mathbf{A}\mathbf{N}_T)^\top \mathbf{D}_v(\mathbf{A}\mathbf{N}_T) + \phi\lambda_T \mathbf{M}_T \right\} \boldsymbol{\gamma}_T = (\mathbf{A}\mathbf{N}_T)^\top \mathbf{D}_v \mathbf{A}(\mathbf{y} - \mathbf{N}_S \boldsymbol{\gamma}_S)$$

e

$$\left\{ (\mathbf{A}\mathbf{N}_S)^\top \mathbf{D}_v(\mathbf{A}\mathbf{N}_S) + \phi\lambda_S \mathbf{M}_S \right\} \boldsymbol{\gamma}_S = (\mathbf{A}\mathbf{N}_S)^\top \mathbf{D}_v \mathbf{A}(\mathbf{y} - \mathbf{N}_T \boldsymbol{\gamma}_T),$$

as quais podem ser simplificadas para

$$\boldsymbol{\gamma}_T = \mathbf{S}_T(\lambda_T) \mathbf{A} \{ \mathbf{y} - \mathbf{N}_S \boldsymbol{\gamma}_S \} \quad \text{e} \quad \boldsymbol{\gamma}_S = \mathbf{S}_S(\lambda_S) \mathbf{A} \{ \mathbf{y} - \mathbf{N}_T \boldsymbol{\gamma}_T \},$$

em que $\mathbf{S}_T(\lambda_T) = \left\{ (\mathbf{A}\mathbf{N}_T)^\top \mathbf{D}_v(\mathbf{A}\mathbf{N}_T) + \phi\lambda_T \mathbf{M}_T \right\}^{-1} (\mathbf{A}\mathbf{N}_T)^\top \mathbf{D}_v$ é a matriz de suavização do componente de tendência e $\mathbf{S}_S(\lambda_S) = \left\{ (\mathbf{A}\mathbf{N}_S)^\top \mathbf{D}_v(\mathbf{A}\mathbf{N}_S) + \phi\lambda_S \mathbf{M}_S \right\}^{-1} (\mathbf{A}\mathbf{N}_S)^\top \mathbf{D}_v$, a matriz de suavização da componente de sazonalidade.

Então, propõe-se, para ϕ, ρ e $\boldsymbol{\lambda}$ fixos, o seguinte processo iterativo para obter o EMVP de $\hat{\boldsymbol{\gamma}}$:

1. Defina os nós, e então calcule \mathbf{N}_T , \mathbf{N}_S , \mathbf{M}_T e \mathbf{M}_S .
2. Inicialize um contador como $u = 0$, determine um valor inicial para $\boldsymbol{\theta}^{(0)}$.
3. Baseado em $\boldsymbol{\theta}^{(u)}$, faça o seguinte:

- (a) A partir dos valores iniciais $\boldsymbol{\theta}^{(0)} = \left(\boldsymbol{\gamma}_T^{(0)\top}, \boldsymbol{\gamma}_S^{(0)\top}, \phi^{(0)}, \rho^{(0)} \right)^\top$, obtenha \mathbf{A} , os pesos $v_i^{(0)} = v_i |_{\theta^{(0)}}$ e $\mathbf{D}_v^{(0)} = \text{diag}_{1 \leq i \leq n} \{ v_i^{(0)} \}$. Então, calcule

$$\mathbf{S}_T^{(0)}(\lambda_T) = \left\{ (\mathbf{A}\mathbf{N}_T)^\top \mathbf{D}_v^{(0)}(\mathbf{A}\mathbf{N}_T) + \phi\lambda_T \mathbf{M}_T \right\}^{-1} (\mathbf{A}\mathbf{N}_T)^\top \mathbf{D}_v^{(0)}$$

e

$$\mathbf{S}_S^{(0)}(\lambda_S) = \left\{ (\mathbf{A}\mathbf{N}_S)^\top \mathbf{D}_v^{(0)}(\mathbf{A}\mathbf{N}_S) + \phi\lambda_S \mathbf{M}_S \right\}^{-1} (\mathbf{A}\mathbf{N}_S)^\top \mathbf{D}_v^{(0)}.$$

(b) Avalie a seguinte expressão para $\gamma_T^{(u+1)}$ e $\gamma_S^{(u+1)}$:

$$\gamma_T^{(u+1)} = \mathbf{S}_T^{(u)}(\lambda_T) \mathbf{A} \left\{ \mathbf{y} - \mathbf{N}_S \gamma_S^{(u)} \right\}$$

e

$$\gamma_S^{(u+1)} = \mathbf{S}_S^{(u)}(\lambda_S) \mathbf{A} \left\{ \mathbf{y} - \mathbf{N}_T \gamma_T^{(u+1)} \right\},$$

para $u = 0, 1, \dots$. Atualize $\mathbf{S}_T^{(u)}(\lambda_T)$ e $\mathbf{S}_S^{(u)}(\lambda_S)$. Repita (b) substituindo $\gamma^{(u)}$ por $\gamma^{(u+1)}$, respectivamente, até a convergência para algum critério. A condição de convergência utilizada em todos os passos do algoritmo foi que a diferença máxima entre os elementos dos vetores estimados da iteração (u) e da ($u + 1$) fossem menores do que 0,001.

A convergência do processo iterativo descrito acima e as condições de convergência seguem do Teorema 4.2 de [Green e Silverman \(1994, p. 68\)](#). Esse algoritmo conhecido como *backfitting* pode convergir em poucos passos, porém depende da escolha dos parâmetros de suavização λ_T e λ_S . Os quais são determinados por validação cruzada por minimizar a quantidade $VCG(\boldsymbol{\lambda})$, discutido nas próximas Seções, para uma grade selecionada de valores.

4. Repita o processo iterativo dado pelo passo (b) com o processo iterativo abaixo usando o método quase Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) ver [Davidon \(1991\)](#) e [Mittelhammer et al. \(2000\)](#), ou “L-BFGS-B” desenvolvido por [Byrd et al. \(1995\)](#) para obter o EMVP de $\zeta = (\phi, \rho)^\top$ dado por

$$\zeta^{(l+1)} = \arg \max_{(\phi, \rho)} L_p(\hat{\gamma}^{(u+1)}, \phi^{(l)}, \rho^{(l)}), \quad l = 0, 1, 2, \dots$$

Para a distribuição t -Student com ν graus de liberdade, o peso atual $v_i^{(u)} = (\nu + 1)/(\nu + \delta_i^{(u)})$, com $\delta_i^{(u)} = \delta_i |_{\hat{\theta}}^{(u)}$, é inversamente proporcional à distância entre o valor observado y_i e seu valor predito atual $\mu_i^{(u)}$ de modo que observações “aberrantes” (*outlier*) tendem a ter pesos menores no processo de estimação. Interpretações similares podem ser aplicadas a outras distribuições do erro com cauda pesada.

3.4 Procedimentos inferenciais

Modelos lineares sob erros normais e simétricos possuem condições de regularidade semelhantes para grandes amostras, o que garante consistência, eficiência e normalidade dos EMVPs. Como para o caso paramétrico, a matriz de variância-covariância aproximada $\hat{\boldsymbol{\theta}}$ pode ser obtida a partir da inversa da matriz de informação esperada (veja, por exemplo, [Wood \(2017, Cap.6\)](#)). Essencialmente, $\widehat{\text{Var}}(\hat{\boldsymbol{\theta}}) = (\mathcal{I}_p^{\theta\theta} |_{\hat{\theta}})^{-1}$, em que $\mathcal{I}_p^{\theta\theta}$ foi definida na seção 3.3.1. Essa abordagem tem suporte na abordagem Bayesiana para modelos lineares, conforme descrito em [Wood \(2017, p. 293\)](#). Se $\mathbf{y} = \mathbf{N}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$, em que $\boldsymbol{\epsilon}$ segue uma distribuição normal n -variada e uma priori imprópria é assumida para $\boldsymbol{\gamma}$, então a distribuição a posteriori $\boldsymbol{\gamma} | \mathbf{y}$ segue uma distribuição normal multivariada de média $\hat{\boldsymbol{\gamma}}$ e matriz de variância-covariância que corresponde a inversa da respectiva matriz de informação de Fisher penalizada para $\boldsymbol{\gamma}$. Assim, intervalos de credibilidade podem ser construídos para quaisquer quantidades derivadas de $\boldsymbol{\gamma}$. Esta abordagem pode ser estendida, para n grande, para a classe simétrica.

Considerando a ortogonalidade entre $\boldsymbol{\gamma}$ e $(\phi, \rho)^\top$, assim como entre ϕ e ρ , obtém-se

$$\widehat{\text{Var}}(\hat{\boldsymbol{\gamma}}) = (\mathcal{I}_p^{\gamma\gamma} |_{\hat{\theta}})^{-1},$$

em que $\mathcal{I}_p^{\gamma\gamma} = \mathcal{I}_{\gamma\gamma} + \mathbf{M}_\gamma(\boldsymbol{\lambda})$ com $\mathbf{M}_\gamma(\boldsymbol{\lambda}) = \text{blockdiag}\{\lambda_T \mathbf{M}_T, \lambda_S \mathbf{M}_S\}$. Além disso, $\widehat{\text{Var}}(\hat{\phi}) = \mathbf{K}_{\phi\phi}^{-1} |_{\hat{\theta}}$ e $\widehat{\text{Var}}(\hat{\rho}) = \mathbf{K}_{\rho\rho}^{-1} |_{\hat{\theta}}$. Desde que $\mathbf{f}_T = \mathbf{N}_T \boldsymbol{\gamma}_T$ e $\mathbf{f}_S = \mathbf{N}_S \boldsymbol{\gamma}_S$, em que $\mathbf{f}_T = (f_T(t_1), \dots, f_T(t_n))^\top$ e

$\mathbf{f}_S = (f_S(s_1), \dots, f_S(s_n))^\top$, pode-se obter a matriz de variância-covariância aproximada $\text{Var}(\widehat{\mathbf{f}}_T) = \mathbf{N}_T \text{Var}(\widehat{\boldsymbol{\gamma}}_T) \mathbf{N}_T^\top$ e $\text{Var}(\widehat{\mathbf{f}}_S) = \mathbf{N}_S \text{Var}(\widehat{\boldsymbol{\gamma}}_S) \mathbf{N}_S^\top$. Então, analogamente a [Vanegas e Paula \(2016\)](#), uma banda de confiança pode ser desenvolvida para $(f_T(t_1), \dots, f_T(t_n))^\top$ e $(f_S(s_1), \dots, f_S(s_n))^\top$.

3.5 Graus de liberdade efetivos

Desde que a penalização da função de suavização $f_T(t)$ e $f_S(s)$ leva a um encolhimento (*shrinkage*) dos EMVPs $\widehat{\boldsymbol{\gamma}}_T$ e $\widehat{\boldsymbol{\gamma}}_S$ com respeito aos EMV, é crucial para a seleção do modelo e dos testes de hipóteses a estimação dos graus de liberdade efetivos correspondente aos EMVPs. Analogamente a [Hastie e Tibshirani \(1990, Cap.5\)](#), [Green e Silverman \(1994, Cap.5\)](#) e mais recentemente [Wood \(2017, Cap.5\)](#), deriva-se abaixo os graus de liberdade efetivos correspondentes aos EMVPs $\widehat{\boldsymbol{\gamma}}_T$ e $\widehat{\boldsymbol{\gamma}}_S$ a partir de um suavizador apropriado.

Na solução das equações de estimação $\mathbf{U}_p^\theta = \mathbf{0}$, tem-se que

$$\begin{aligned} \widehat{\boldsymbol{\gamma}} &= \left\{ \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A + \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \right\}^{-1} \left\{ \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A \widehat{\boldsymbol{\gamma}} + \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{A}} (\mathbf{y} - \mathbf{N} \widehat{\boldsymbol{\gamma}}) \right\} \\ &= \left\{ \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A + \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \right\}^{-1} \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{A}} \mathbf{y}. \end{aligned}$$

Então, tem-se a relação $\widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{N}}_A \widehat{\boldsymbol{\gamma}} = \widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{A}} \widehat{\mathbf{N}} \widehat{\boldsymbol{\gamma}} = \widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{A}} \widehat{\boldsymbol{\mu}} = \widehat{\mathbf{H}}(\boldsymbol{\lambda}) \widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{A}} \mathbf{y}$, em que

$$\widehat{\mathbf{H}}(\boldsymbol{\lambda}) = \widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{N}}_A \left\{ \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A + \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \right\}^{-1} \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v^{\frac{1}{2}} \quad (3.7)$$

é nomeada suavizador linear para $\boldsymbol{\lambda}$ fixo e projeta o estimador da resposta transformada $\widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{A}} \mathbf{y}$ na média predita transformada $\widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{A}} \widehat{\boldsymbol{\mu}}$. Em particular, sob erros normais e $\rho = 0$, tem-se que $\widehat{\mathbf{H}}(\boldsymbol{\lambda}) = \mathbf{N} \{ \mathbf{N}^\top \mathbf{N} + \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \}^{-1} \mathbf{N}^\top$. Os graus de liberdade efetivos dos suavizadores são definidos como a soma dos autovalores de $\widehat{\mathbf{H}}(\boldsymbol{\lambda})$, a saber $df_s(\boldsymbol{\lambda}) = \text{tr}\{\widehat{\mathbf{H}}(\boldsymbol{\lambda})\}$. De [Eilers e Marx \(1996\)](#), tem-se que

$$\begin{aligned} df_s(\boldsymbol{\lambda}) &= \text{tr} \left\{ \widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{N}}_A \left(\widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A + \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \right)^{-1} \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v^{\frac{1}{2}} \right\} \\ &= \text{tr} \left\{ \left(\widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A + \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \right)^{-1} \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A \right\} \\ &= \text{tr} \left\{ \left(\mathbf{I}_{r_T+r_S-1} + \mathbf{Q}^{-\frac{1}{2}} \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \mathbf{Q}^{-\frac{1}{2}} \right)^{-1} \right\} \\ &= \sum_{i=1}^{r_T+r_S-1} \frac{1}{1 + \alpha_i(\boldsymbol{\lambda})}, \end{aligned}$$

em que $\alpha_i(\boldsymbol{\lambda}) \geq 0$ são os autovalores da matriz definida não negativa $\mathbf{Q}^{-\frac{1}{2}} \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \mathbf{Q}^{-\frac{1}{2}}$, os quais são dependentes dos parâmetros de suavização; $\mathbf{Q}^{1/2}$ é uma matriz positiva definida e $\mathbf{Q}^{\frac{1}{2}} \mathbf{Q}^{\frac{1}{2}} = \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A$, para $i = 1, \dots, r_T+r_S-1$. Os graus de liberdade efetivos $df_s(\boldsymbol{\lambda}_T)$ e $df_s(\boldsymbol{\lambda}_S)$ correspondem à soma dos primeiros r_T e os últimos r_S autovalores, respectivamente, do suavizador linear $\widehat{\mathbf{H}}(\boldsymbol{\lambda})$. Portanto, os graus de liberdade efetivos são dados por

$$df(\boldsymbol{\lambda}) = df_s(\boldsymbol{\lambda}) + 2$$

3.6 Comentários

Neste capítulo introduzimos os modelos aditivos com erros simétricos condicionais de ordem AR(1), em que o componente aditivo referente ao tempo é decomposto em um componente relacionado à tendência e outro relacionando à sazonalidade. Esses componentes são aproximados por *splines* cúbicos de regressão penalizados. Um processo iterativo do tipo *backfitting* para obter os coeficientes estimados dos *splines* cúbicos penalizados de regressão é alternado com um procedimento usual de máxima verossimilhança para obter $\hat{\phi}$ e $\hat{\rho}$. A função escore penalizada e a matriz de informação de Fisher penalizada são obtidas, bem como os graus de liberdade efetivos são derivados. No próximo capítulo discutiremos os modelos lineares parciais aditivos com erros simétricos condicionais AR(p).

Capítulo 4

Modelos lineares parciais aditivos com erros simétricos condicionais $AR(p)$

Os pesquisadores costumam ver uma série temporal como uma realização de um processo estocástico. Os principais objetivos em análise de séries temporais são entender as leis das probabilidades que as conduz e, assim, compreender a dinâmica subjacente, prever eventos futuros e controlá-los por meio de intervenções. Há uma infinidade de processos estocásticos que podem gerar os mesmos dados observados, pois o número de observações é sempre finito. No entanto, alguns desses processos são mais plausíveis e admitem melhor interpretação do que outros. Sem mais restrições no processo subjacente, é impossível identificar o processo a partir de um número finito de observações.

Uma abordagem popular é limitar as leis de probabilidades a uma família específica e, em seguida, selecionar um membro da família que seja mais plausível. O primeiro é chamado de modelagem e o último de estimação, ou inferência estatística. Quando a forma das leis de probabilidades em uma família é especificada, exceto por alguns parâmetros definidores de dimensão finita, esse modelo é conhecido como paramétrico. Quando os parâmetros de definição estão em um subconjunto de um espaço dimensional infinito ou as formas das leis de probabilidades não são completamente especificadas, esse modelo é frequentemente chamado de não paramétrico. A fronteira entre os dois tipos de modelos paramétricos e não paramétricos nem sempre é clara. Apesar disso, essa distinção nos ajuda a escolher um método de estimação apropriado.

Ao selecionar um modelo, a interpretabilidade, simplicidade e viabilidade desempenham papéis importantes. Um modelo selecionado deve refletir razoavelmente a lei física que governa os dados. Um modelo simples é geralmente preferível. A família de modelos de probabilidades deve ser razoavelmente grande para incluir as leis de probabilidades subjacentes que geraram os dados, mas não deve ser tão grande que os parâmetros de definição não possam mais serem estimados com uma precisão razoavelmente boa. Ao escolher um modelo de probabilidades, primeiro extrai-se características salientes dos dados observados e depois escolhe-se um modelo apropriado que possua essas características. Após estimar os parâmetros ou as funções no modelo, verifica-se se este ajusta razoavelmente bem aos dados e procura-se melhorias adicionais sempre que possível. Diferentes finalidades da análise também podem ditar o uso de outros modelos. Por exemplo, aquele que fornece um bom ajuste e admite uma boa interpretação não é necessariamente bom para a previsão.

Nos últimos anos, houve um interesse considerável na literatura estatística em relação a famílias flexíveis de distribuições capazes de modelar dados que apresentam caudas mais pesadas ou leves que a distribuição normal usual. Neste contexto, os modelos considerados que têm mostrado um comportamento promissor em cenários não padronizados são advindos das distribuições simétricas (vide, por exemplo, [Cysneiros e Paula \(2005\)](#)).

Neste capítulo propomos algumas extensões do modelo com erros simétricos autorregressivos de primeira ordem de [Relvas e Paula \(2016\)](#) para as situações em que os dados apresentam correlação de ordem p , com a possibilidade de incluir covariáveis no modelo, com características paramétricas ou não. Os novos modelos são particularmente atraentes devido ao fato de serem capazes de ajustar dados com maior ordem de correlação, caudas mais pesadas ou leves que a distribuição normal. As

vantagens desse modelo estão voltadas à simplicidade das funções *splines* cúbicas penalizadas de regressão apresentadas em Wood (2017), que diminuem o custo computacional dos *splines* cúbicos naturais discutidos ao longo do trabalho de Relvas (2013). Além disso, podemos decompor a série temporal em estudo nos componentes de tendência e sazonalidade, os quais podem necessitar de funções mais flexíveis para descrevê-los, e a sazonalidade pode exigir uma função cíclica. Outra vantagem está relacionada ao fato dos dados nem sempre serem autorregressivos de primeira ordem, sendo assim, a necessidade da ordem p leva a modelos mais flexíveis.

Neste capítulo introduzimos os modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p – MLPAS condicionais AR(p). Na Seção 4.1, apresentamos a classe e derivamos alguns resultados básicos. Na Seção 4.2 apresentamos a função de verossimilhança penalizada. Na Seção 4.3 derivamos a função escore penalizada e a matriz de informação de Fisher penalizada. Em seguida, desenvolvemos o processo iterativo para a estimação dos parâmetros de interesse do modelo. Na Seção 4.4 elaboramos os procedimentos inferências. Finalmente, na Seção 4.5, derivamos os graus de liberdade efetivos do modelo proposto.

4.1 O modelo

Considere um modelo em que as observações são o vetor de r_0 valores de covariáveis decorrentes de funções lineares paramétricas e a soma de k funções não paramétricas mais um erro aleatório, dado pela seguinte maneira

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + f_1(x_{1_i}) + f_2(x_{2_i}) + \dots + f_k(x_{k_i}) + \epsilon_i, \quad (i = 1, \dots, n). \quad (4.1)$$

Até agora, consideramos apenas a possibilidade de autocorrelação de primeira ordem. O termo de erro também pode ser correlacionado com valores próprios que ficaram mais de um período de tempo, de modo que, na regressão,

$$\epsilon_i = \rho_1 \epsilon_{i-1} + \rho_2 \epsilon_{i-2} + \dots + \rho_p \epsilon_{i-p} + e_i, \quad (4.2)$$

os vários ρ_p , para $p > 1$, também seriam significativos. É improvável, embora não impossível, que uma série inteira de ρ_p seja significativa, mas geralmente não é o que acontece. E ρ 's são os coeficientes de autocorrelação, enquanto e_i são os erros simétricos independentes com média zero e parâmetro de dispersão ϕ , isto é, $e_i \stackrel{iid}{\sim} S(0, \phi)$, para $i = 1, \dots, n$.

Em muitos casos o pesquisador está interessado em entender o desempenho de certas variáveis em relação a períodos anteriores. Dessa forma, o termo ρ_4 seria significativo para dados trimestrais. No caso mensal, o termo ρ_{12} seria significativo – mas não os ρ intermediários. Nessas situações, se a autocorrelação sazonal aparecer nos dados, geralmente é preferível ajustar os dados sazonalmente ou, se os dados forem suficientemente robustos, use as equações de teor da primeira diferença. Se o problema não desaparecer, adicione a variável defasada com o atraso de 4 (trimestre) ou 12 (mês), ou use o período atual menos 4 trimestres (ou 12 meses) atrás.

Dessa forma, podemos escrever o modelo (4.1) em notação matricial da seguinte forma:

$$\mathbf{y} = \mathbf{N}_0 \boldsymbol{\gamma}_0 + \mathbf{N}_1 \boldsymbol{\gamma}_1 + \mathbf{N}_2 \boldsymbol{\gamma}_2 + \dots + \mathbf{N}_k \boldsymbol{\gamma}_k + \boldsymbol{\epsilon}, \quad (4.3)$$

em que \mathbf{y} é um vetor ($n \times 1$) de respostas observadas, por conveniência, adotamos \mathbf{X} e $\boldsymbol{\beta}$ como \mathbf{N}_0 a matriz de experimento ($n \times r_0$) e $\boldsymbol{\gamma}_0$ o vetor paramétrico ($r_0 \times 1$), respectivamente. Aqui r_0 indica o número de variáveis explicativas resultantes da regressão linear mais o intercepto. \mathbf{N}_l é uma matriz ($n \times r_l$) e função base com linhas $\boldsymbol{\eta}_{l_i}(t_i) = (\eta_{l_1}(t_i^{(l)}), \dots, \eta_{l_{r_l}}(t_i^{(l)}))^\top$, e o (i, j) -ésimo elemento igual à parametrização dos *splines* em termos de seus valores nos nós $t_1^0, \dots, t_{r_l}^0$, sendo que $l = 1, \dots, k$, e k refere-se ao número de funções não paramétricas. Resumidamente temos as matrizes,

$$\mathbf{N}_0 = \begin{bmatrix} 1 & \eta_{12} & \dots & \eta_{1r_0} \\ 1 & \eta_{22} & \dots & \eta_{2r_0} \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \eta_{n2} & \dots & \eta_{nr_0} \end{bmatrix} \quad \text{e} \quad \mathbf{N}_l = \begin{bmatrix} \eta_{l1}(t_1^{(l)}) & \dots & \eta_{lr_l}(t_1^{(l)}) \\ \eta_{l2}(t_2^{(l)}) & \dots & \eta_{lr_l}(t_2^{(l)}) \\ \vdots & \ddots & \vdots \\ \eta_{ln}(t_n^{(l)}) & \dots & \eta_{lr_l}(t_n^{(l)}) \end{bmatrix}, \quad l = 1, 2, \dots, k.$$

Similarmente a [Relvas \(2013\)](#), pode-se derivar algumas propriedades para o modelo (4.1)-(4.2). Com p parâmetros autorregressivos, temos o seguinte:

1. Para a primeira observação, tem-se que $y_1 \sim S(\mu_1^*, \phi)$ com

$$\mu_1^* = \mathbb{E}(y_1) = \mathbf{x}_1^\top \boldsymbol{\beta} + f_1(t_{11}) + \dots + f_k(t_{k1}) = \boldsymbol{\eta}_{01}^\top \boldsymbol{\gamma}_0 + \boldsymbol{\eta}_{11}^\top(t_1^{(1)}) \boldsymbol{\gamma}_1 + \dots + \boldsymbol{\eta}_{k1}^\top(t_1^{(k)}) \boldsymbol{\gamma}_k.$$

2. Para a distribuição condicional $y_i \mid y_{i-1}, \dots, y_{\max(i-p,1)} \stackrel{ind}{\sim} S(\mu_i^*, \phi)$, tem-se que

$$\begin{aligned} \mu_i^* &= \mathbb{E}(y_i \mid y_{i-1}, \dots, y_{\max(i-p,1)}) = \\ &= \boldsymbol{\eta}_{0i}^\top \boldsymbol{\gamma}_0 + \boldsymbol{\eta}_{1i}^\top(t_i^{(1)}) \boldsymbol{\gamma}_1 + \dots + \boldsymbol{\eta}_{ki}^\top(t_i^{(k)}) \boldsymbol{\gamma}_k + \\ &\quad + \sum_{j=1}^{\min(i-1,p)} \rho_j \{y_{i-j} - \boldsymbol{\eta}_{0i-j}^\top \boldsymbol{\gamma}_0 - \boldsymbol{\eta}_{1i-j}^\top(t_{i-j}^{(1)}) \boldsymbol{\gamma}_1 - \dots - \boldsymbol{\eta}_{ki-j}^\top(t_{i-j}^{(k)}) \boldsymbol{\gamma}_j\}, \end{aligned}$$

para $i = 2, \dots, n$.

3. Em geral, temos que a média marginal é dada por

$$\mu_i = \mathbb{E}(y_i) = \boldsymbol{\eta}_{0i}^\top \boldsymbol{\gamma}_0 + \boldsymbol{\eta}_{1i}^\top(t_i^{(1)}) \boldsymbol{\gamma}_1 + \dots + \boldsymbol{\eta}_{ki}^\top(t_i^{(k)}) \boldsymbol{\gamma}_k, \quad i = 1, \dots, n;$$

4. Em relação à variância marginal do modelo, temos que levar em consideração a ordem de autocorrelação dos erros.

- (a) Para o modelo com erros AR(1):

$$\text{Var}(y_i) = \xi \phi \sum_{j=0}^{i-1} \rho_1^{2j} = \xi \phi \frac{1 - \rho_1^{2i}}{1 - \rho_1^2}.$$

- (b) Para o modelo com erros AR(p):

$$\text{Var}(y_i) = \xi \phi + \xi \phi \sum_{j=1}^{i-1} \rho_j^2 \text{Var}(y_{i-j}) + 2 \sum_{j=1}^{i-1} \sum_{\substack{j'=1 \\ j \neq j' < j'}}^{i-1} \rho_j \rho_{j'} \text{Cov}(y_{i-j}, y_{i-j'}).$$

5. Pode-se derivar as covariâncias entre duas observações, $\text{Cov}(y_i, y_{i-j})$, portanto

- (a) Para o modelo com erros AR(1):

- Para $j = i - 1$,

$$\text{Cov}(y_i, y_{i-j}) = \text{Cov}(y_i, y_1) = \rho \mathbb{E}(y_{i-1} e_1) = \phi \xi \rho^{i-1}.$$

- Para $j < i - 1$,

$$\begin{aligned}\text{Cov}(y_i, y_{i-j}) &= \phi\xi\rho^j + \rho^2\text{Cov}(y_{i-1}, y_{i-j-1}) \\ &= \phi\xi\rho^j + \rho^2\rho^j\text{Cov}(y_{i-2}, y_{i-j-2}) \\ &= \phi\xi\rho^j + \phi\xi\rho^{j+2} + \rho^2\rho^{j+2}\text{Cov}(y_{i-3}, y_{i-j-3}) \\ &\quad \vdots \\ &= \phi\xi \sum_{l=0}^{i-2} \rho^{j+2l}.\end{aligned}$$

(b) Para o modelo com erros AR(p):

- Para $j = i - 1$,

$$\text{Cov}(y_i, y_{i-j}) = \text{Cov}(y_i, y_1) = \sum_{l=1}^p \rho_l \mathbb{E}(y_{i-1}e_1).$$

- Para $j = i - 2$,

$$\begin{aligned}\text{Cov}(y_i, y_{i-j}) = \text{Cov}(y_i, y_2) &= \rho_1^2 \text{Cov}(y_{i-1}, y_{i-j-1}) + \sum_{l=1}^p \rho_l \mathbb{E}(y_{i-1}e_k) \\ &\quad + \rho_1 \sum_{l=2}^p \rho_l \text{Cov}(y_{i-1}, y_1).\end{aligned}$$

- Para $j = i - p$,

$$\begin{aligned}\text{Cov}(y_i, y_{i-j}) = \text{Cov}(y_i, y_p) &= \sum_{l=1}^{p-1} \rho_l^2 \text{Cov}(y_{i-l}, y_{k-l}) + \sum_{l=1}^p \rho_l \mathbb{E}(y_{i-l}e_k) \\ &\quad + \sum_{l=1}^p \sum_{\substack{l'=1 \\ l \neq l'}}^{p-1} \rho_l \rho_{l'} \text{Cov}(y_{i-l}, y_{k-l'}).\end{aligned}$$

- Para $j < i - p = k$,

$$\begin{aligned}\text{Cov}(y_i, y_{i-j}) = \text{Cov}(y_i, y_k) &= \sum_{l=1}^p \rho_l^2 \text{Cov}(y_{i-l}, y_{k-l}) + \sum_{l=1}^p \rho_l \mathbb{E}(y_{i-l}e_k) \\ &\quad + \sum_{l=1}^p \sum_{\substack{l'=1 \\ l \neq l'}}^p \rho_l \rho_{l'} \text{Cov}(y_{i-l}, y_{k-l'}).\end{aligned}$$

O modelo proposto é recomendado para estudar a relação da variável resposta e de uma ou mais variáveis explicativas, esta pode ser considerada paramétrica ou não paramétrica. No segundo caso, não há um certo interesse em estudar os efeitos, apenas controlar a variável resposta em estudo, como por exemplo, podemos controlar a tendência e a sazonalidade diante de outras variáveis. Pelo fato das observações serem correlacionadas ao longo do tempo, temos maior precisão no modelo em estudo. Além disso, a variável resposta possui caudas mais pesadas ou mais leves do que a distribuição normal, portanto, teremos ganhos na interpretação e resultados.

4.2 Função de verossimilhança penalizada

Seja $\boldsymbol{\theta} = (\boldsymbol{\gamma}_0^\top, \boldsymbol{\gamma}_1^\top, \dots, \boldsymbol{\gamma}_k^\top, \phi, \rho_1, \dots, \rho_p)^\top \in \boldsymbol{\Theta} \subseteq \mathcal{R}^r$, com $r = r_0 + r_1 + \dots + r_k + p + 1$ sendo o número de parâmetros a ser estimado. O logaritmo da função de verossimilhança regular associada com $\boldsymbol{\theta}$ é dado por

$$L(\boldsymbol{\theta}) = -\frac{n}{2} \log(\phi) + \sum_{i=1}^n \log\{g(\delta_i)\}. \quad (4.4)$$

A maximização de (4.4) sem impor restrições sobre as funções não paramétricas pode levar a *overfitting* e a não identificabilidade dos parâmetros. Para resolver esse problema, deve-se incorporar uma função penalidade sobre cada componente não paramétrico. Assume-se que essas funções são contínuas e integráveis na segunda derivada, então o logaritmo da função de verossimilhança penalizada é dado por

$$L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = L(\boldsymbol{\theta}) - \frac{\lambda_1}{2} \int_{a_1}^{b_1} [f_1''(t)]^2 dt - \dots - \frac{\lambda_k}{2} \int_{a_k}^{b_k} [f_k''(t)]^2 dt, \quad (4.5)$$

em que tem-se a penalização da segunda derivada de $f_l(t)$ para $a_l \leq t \leq b_l$, com $\lambda_l > 0$ para $l = 1, \dots, k$, os parâmetros de suavização que serão estimados separadamente. De Wood (2017), tem-se que

$$\int_{a_l}^{b_l} [f_l''(t)]^2 dt = \boldsymbol{\gamma}_l^\top \mathbf{M}_l \boldsymbol{\gamma}_l, \quad l = 1, \dots, k,$$

em que $\mathbf{M}_l = \mathbf{D}_l^\top \mathbf{B}_l^{-1} \mathbf{D}_l$, $l = 1, \dots, k$, são matrizes não negativas definidas de dimensões $(r_l \times r_l)$, nomeadas de matrizes de penalidade. Essas matrizes dependem somente dos nós e são definidas na Tabela 2.1. Assim, o logaritmo da função de verossimilhança penalizada reduz a

$$L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = L(\boldsymbol{\theta}) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k, \quad (4.6)$$

em que $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^\top$ denota o vetor paramétrico de suavização. Na próxima seção será derivado o processo iterativo *backfitting* (Gauss-Seidel) para maximizar (4.6) para $\boldsymbol{\lambda}$ fixo.

4.3 Estimação dos parâmetros

Nesta Seção apresentamos a função escore penalizada, a matriz Hessiana penalizada e a matriz de informação de Fisher penalizada bem como o processo iterativo para a estimação do componente não paramétrico e dos parâmetros que envolvem as variáveis independentes e, por fim, os parâmetros da distribuição simétrica com erros AR(p). Os cálculos com maiores detalhes podem ser vistos no Apêndice B.2.

4.3.1 Função escore penalizada

A função escore penalizada de $\boldsymbol{\theta}$ pode ser expressa como $\mathbf{U}_p^\theta = \partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) / \partial \boldsymbol{\theta} = \mathbf{U}_\theta - \mathbf{M}(\boldsymbol{\lambda}) \boldsymbol{\theta}$, em que \mathbf{U}_θ é a função escore regular dada por

Tem-se que a matriz $\ddot{\mathbf{L}}_{\theta\theta}$ é simétrica, portanto

$$\begin{aligned} \ddot{\mathbf{L}}_{\gamma_l\gamma_{l'}} &= \ddot{\mathbf{L}}_{\gamma_{l'}\gamma_l}^\top, & \ddot{\mathbf{L}}_{\gamma_l\phi} &= \ddot{\mathbf{L}}_{\phi\gamma_l}^\top, & \ddot{\mathbf{L}}_{\gamma_l\rho_j} &= \ddot{\mathbf{L}}_{\rho_j\gamma_l}^\top, & l \neq l' = 0, 1, \dots, k, \\ \ddot{\mathbf{L}}_{\rho_j\rho_{j'}} &= \ddot{\mathbf{L}}_{\rho_{j'}\rho_j}^\top \quad \text{e} & \ddot{\mathbf{L}}_{\phi\rho_j} &= \ddot{\mathbf{L}}_{\rho_j\phi}^\top, & & & j \neq j' = 1, \dots, p. \end{aligned}$$

Além disso, as submatrizes da diagonal de $\ddot{\mathbf{L}}_{\theta\theta}$ podem ser escritas, em notação matricial, como

$$\begin{aligned} \ddot{\mathbf{L}}_{\gamma_l\gamma_l} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_l \partial \gamma_l^\top} = \frac{1}{\phi} \left\{ (\mathbf{AN}_l)^\top (-\mathbf{D}_v + 4\mathbf{D}_d) (\mathbf{AN}_l) \right\}, \quad \text{para } l = 0, 1, \dots, k, \\ \ddot{\mathbf{L}}_{\phi\phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi^2} = \frac{1}{\phi^2} \left\{ \frac{n}{2} + \boldsymbol{\delta}^\top \mathbf{D}_c \boldsymbol{\delta} - \boldsymbol{\delta}^\top \mathbf{D}_v \mathbf{1}_n \right\} \quad \text{e} \\ \ddot{\mathbf{L}}_{\rho_j\rho_j} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho_j^2} = \frac{1}{\phi} \left\{ (\mathbf{B}_j \boldsymbol{\epsilon})^\top (-\mathbf{D}_v + 4\mathbf{D}_d) (\mathbf{B}_j \boldsymbol{\epsilon}) \right\}, \quad \text{para } j = 1, \dots, p. \end{aligned}$$

Já as submatrizes cruzadas são expressas como

$$\begin{aligned} \ddot{\mathbf{L}}_{\gamma_l\gamma_{l'}} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_l \partial \gamma_{l'}^\top} = \frac{1}{\phi} \left\{ (\mathbf{AN}_{l'})^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{AN}_l) \right\}, \quad l \neq l' = 0, 1, \dots, k, \\ \ddot{\mathbf{L}}_{\gamma_l\phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_l \partial \phi^\top} = \frac{1}{\phi^2} \left\{ (\mathbf{AN}_l)^\top (2\mathbf{D}_d - \mathbf{D}_v) (\mathbf{A}\boldsymbol{\epsilon}) \right\}, \\ \ddot{\mathbf{L}}_{\gamma_l\rho_j} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_l \partial \rho_j^\top} = \frac{1}{\phi} \left\{ (\mathbf{AN}_l)^\top (\mathbf{D}_v - 4\mathbf{D}_d) (\mathbf{B}_j \boldsymbol{\epsilon}) + (\mathbf{B}_j \mathbf{N}_l)^\top \mathbf{D}_v \mathbf{A}\boldsymbol{\epsilon} \right\}, \\ \ddot{\mathbf{L}}_{\phi\rho_j} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi \partial \rho_j^\top} = \frac{1}{\phi^2} \left\{ (\mathbf{B}_j \boldsymbol{\epsilon})^\top (\mathbf{D}_v - 2\mathbf{D}_d) (\mathbf{A}\boldsymbol{\epsilon}) \right\} \quad \text{e} \\ \ddot{\mathbf{L}}_{\rho_j\rho_{j'}} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho_j \partial \rho_{j'}^\top} = \frac{1}{\phi^2} \left\{ (\mathbf{B}_{j'} \boldsymbol{\epsilon})^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{B}_j \boldsymbol{\epsilon}) \right\}, \quad j \neq j' : 1, \dots, p. \end{aligned}$$

Os elementos restantes da matriz são obtidos pela simetria. Com esses cálculos, completamos a matriz de informação observada. Entretanto, toda proposta de estimação neste trabalho será realizada com a matriz de informação de Fisher, que representa a matriz esperada, apresentada na próxima subseção.

4.3.3 Matriz de informação de Fisher penalizada

A matriz de informação forma uma ferramenta de interesse para verificar a eficiência, ou seja, a obtenção do limite de informações para a variação do estimador. A matriz de informação, \mathcal{I} , é o valor esperado negativo da matriz de derivadas parciais de segunda ordem da log-verossimilhança a ser maximizada, ou seja, $\mathcal{I}_p^{\theta\theta} = -E \left\{ \partial L_p^2(\boldsymbol{\theta}, \boldsymbol{\lambda}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top \right\} = \mathcal{I}_{\theta\theta} - \mathbf{M}(\boldsymbol{\lambda})$. Os cálculos com maiores detalhes são apresentados no Apêndice B.3, em que submatrizes são dadas por $\mathcal{I}_{\gamma_l\gamma_{l'}}$, $\mathcal{I}_{\phi\phi}$, $\mathcal{I}_{\rho_j\rho_{j'}}$, $\mathcal{I}_{\gamma_l\phi}$, $\mathcal{I}_{\gamma_l\rho_m}$ e $\mathcal{I}_{\phi\rho_j}$. Além disso, a matriz é simétrica, portanto,

$$\begin{aligned} \mathcal{I}_{\gamma_l\gamma_{l'}} &= \mathcal{I}_{\gamma_{l'}\gamma_l}^\top, & \mathcal{I}_{\gamma_l\phi} &= \mathcal{I}_{\phi\gamma_l}^\top, & \mathcal{I}_{\gamma_l\rho_j} &= \mathcal{I}_{\rho_j\gamma_l}^\top, & l \neq l' = 0, 1, \dots, k, \\ \mathcal{I}_{\rho_j\rho_{j'}} &= \mathcal{I}_{\rho_{j'}\rho_j}^\top \quad \text{e} & \mathcal{I}_{\phi\rho_j} &= \mathcal{I}_{\rho_j\phi}^\top, & & & j \neq j' = 1, \dots, p. \end{aligned}$$

Considere algumas quantidades já discutidas anteriormente como $d_g = E [W_g^2(Z^2)Z^2]$ e $f_g = E [W_g^2(Z^2)Z^4]$ com $Z \sim S(0, 1)$. Duas condições de regularidade serão usadas para encontrar a matriz de informação de Fisher dos parâmetros $\boldsymbol{\theta}$ e são dadas por

$$(i) \quad E[\mathbf{U}\boldsymbol{\theta}] = \mathbf{0},$$

$$(ii) \quad E\left[-\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right] = E\left[\left(\frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta}}\right)^2\right].$$

Além disso, precisamos dos dois primeiros momentos e a matriz de covariâncias da variável ϵ_i no modelo. Sabe-se que $e_i \sim S(0, \phi)$.

No primeiro momento, vamos analisar dois casos para os modelos autorregressivos de ordem 1, ou AR(1),

- Considere $\epsilon_0 = 0$

Podemos reescrever o erro autorregressivo, $\epsilon_i = \rho_1 \epsilon_{i-1} + e_i$, usando a forma recursiva da seguinte maneira:

$$\begin{aligned} \epsilon_1 &= e_1, \\ \epsilon_2 &= \rho_1 e_1 + e_2, \\ \epsilon_3 &= \rho_1^2 e_1 + \rho_1 e_2 + e_3, \\ \epsilon_4 &= \rho_1^3 e_1 + \rho_1^2 e_2 + \rho_1 e_3 + e_4, \\ &\vdots \\ \epsilon_i &= \rho_1^{i-1} e_1 + \rho_1^{i-2} e_2 + \dots + \rho_1^2 e_{i-2} + \rho_1 e_{i-1} + e_i. \end{aligned} \tag{4.8}$$

Dessa forma, como escrevemos ϵ_i e pelo fato de e serem independentes, pode-se mostrar que

$$E(\epsilon_i) = 0,$$

$$\phi_\epsilon \equiv \text{Var}(\epsilon_i) = \left[\rho_1^{2(i-1)} + \rho_1^{2(i-2)} + \dots + \rho_1^2 + 1\right] \phi\xi = \phi\xi \sum_{j=0}^{i-1} \rho_1^{2j} = \phi\xi \frac{(1 - \rho_1^{2i})}{(1 - \rho_1^2)} = E(\epsilon_i^2), \tag{4.9}$$

$$\text{Cov}(\epsilon_i, \epsilon_{i-1}) = E(\epsilon_i \epsilon_{i-1}) = E[(\rho_1 \epsilon_{i-1} + e_i) \epsilon_{i-1}] = \rho_1 E[\epsilon_{i-1}^2] = \rho_1 \phi\xi \frac{(1 - \rho_1^{2(i-1)})}{(1 - \rho_1^2)}.$$

Para obter a covariância entre ϵ_i e ϵ_{i-s} para algum s , primeiro note que por aplicar repetidas substituições em (4.8) podemos reescrever como

$$\epsilon_i = \rho_1^s \epsilon_{i-s} + \sum_{j=1}^{s-1} \rho_1^j e_{i-j}.$$

Logo,

$$\text{Cov}(\epsilon_i, \epsilon_{i-s}) = E(\epsilon_i \epsilon_{i-s}) = E\left[\left(\rho_1^s \epsilon_{i-s} + \sum_{j=1}^{s-1} \rho_1^j e_{i-j}\right) \epsilon_{i-s}\right] = \rho_1^s \phi\xi \frac{(1 - \rho_1^{2(i-s)})}{(1 - \rho_1^2)}. \tag{4.10}$$

Para derivar a correlação entre s períodos de dois erros, dividimos (4.10) por (4.9) e obtemos

$$\frac{E(\epsilon_i \epsilon_{i-s})}{[E(\epsilon_i^2)E(\epsilon_{i-s}^2)]^{1/2}} = \frac{\rho_1^s \phi\xi (1 - \rho_1^{2(i-s)}) / (1 - \rho_1^2)}{\phi\xi / (1 - \rho_1^2) [(1 - \rho_1^{2i})(1 - \rho_1^{2(i-s)})]^{1/2}} = \rho_1^s \left[\frac{(1 - \rho_1^{2(i-s)})}{(1 - \rho_1^{2i})}\right]^{1/2}.$$

Assim, quanto maior for o intervalo de tempo entre duas perturbações, menor será a correlação entre elas. Agora estamos dispostos a escrever a matriz de covariância para ϵ . A partir de (4.9) e (4.10) obtemos

$$\mathbf{\Upsilon} = E(\epsilon\epsilon^\top) = E \begin{pmatrix} \epsilon_1^2 & \epsilon_1\epsilon_2 & \dots & \epsilon_1\epsilon_n \\ \epsilon_2\epsilon_1 & \epsilon_2^2 & \dots & \epsilon_2\epsilon_n \\ & & \dots & \\ \epsilon_n\epsilon_1 & \epsilon_n\epsilon_2 & \dots & \epsilon_n^2 \end{pmatrix} = \phi\xi\mathbf{V},$$

em que

$$\mathbf{V} = \frac{1}{1 - \rho_1^2} \begin{pmatrix} (1 - \rho_1^2) & \rho_1(1 - \rho_1^2) & \dots & \rho_1^{(n-1)}(1 - \rho_1^2) \\ \rho_1(1 - \rho_1^2) & (1 - \rho_1^4) & & \rho_1^{(n-2)}(1 - \rho_1^4) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_1^{(n-1)}(1 - \rho_1^2) & \rho_1^{(n-2)}(1 - \rho_1^4) & \dots & (1 - \rho_1^{2n}) \end{pmatrix}.$$

- Considere $\epsilon_0 \neq 0$

Segundo [Judge *et al.* \(1982\)](#) assume-se que a perturbação atual ϵ_i depende do erro no período anterior, ϵ_{i-1} , e de outro erro, e_i , que assume-se média zero e variância constante e não correlacionado ao longo do tempo. Isso é,

$$E(e_i) = 0, \quad E(e_i^2) = \phi\xi \quad \text{e} \quad E(e_i e_{i'}) = 0 \quad \text{para} \quad i \neq i'.$$

Essas propriedades podem ser escritas em notação matricial como

$$E(\mathbf{e}) = \mathbf{0} \quad \text{e} \quad E(\mathbf{e}\mathbf{e}^\top) = \phi\xi\mathbf{I}_n.$$

Essas suposições constituem o modelo estatístico linear geral com erros autorregressivo de primeira ordem. O parâmetro ρ_1 é desconhecido e, junto com γ , devemos estimá-los. Nesse momento vamos estudar as propriedades estatísticas de ϵ_i . Para investigá-los mais detalhadamente, substituímos recursivamente a expressão $\epsilon_i = \rho_1\epsilon_{i-1} + e_i$. Este processo produz,

$$\begin{aligned} \epsilon_i &= \rho_1\epsilon_{i-1} + e_i \\ &= \rho_1(\rho_1\epsilon_{i-2} + e_{i-1}) + e_i \\ &= \rho_1^2\epsilon_{i-2} + \rho_1 e_{i-1} + e_i \\ &= \rho_1^2(\rho_1\epsilon_{i-3} + e_{i-2}) + \rho_1 e_{i-1} + e_i \\ &= \rho_1^3\epsilon_{i-3} + \rho_1^2 e_{i-2} + \rho_1 e_{i-1} + e_i \\ &\vdots \\ &= \rho_1^j\epsilon_{i-j} + e_i + \rho_1 e_{i-1} + \rho_1^2 e_{i-2} + \dots + \rho_1^{j-1} e_{i-j+1}. \end{aligned}$$

Se assumimos que $-1 < \rho_1 < 1$ e seja $j \rightarrow \infty$, resumimos a expressão anterior como

$$\begin{aligned} \epsilon_i &= e_i + \rho_1 e_{i-1} + \rho_1^2 e_{i-2} + \dots \\ &= \sum_{j=0}^{\infty} \rho_1^j e_{i-j}. \end{aligned} \tag{4.11}$$

A suposição de que j vai ao infinito implica que o processo já está operando por um longo período no passado e a suposição de que $|\rho_1| < 1$ implica que os ϵ_i são estacionários. Esta última suposição também significa que as variâncias e covariâncias de ϵ_i são constantes ao longo do tempo. Se $|\rho_1| \geq 1$, então, conforme progredimos no tempo, a variância de ϵ_i aumenta,

tornando-se eventualmente infinita. A parte sistemática de y_i torna-se dominada pelo erro e o modelo, tal como está, é difícil de analisar. Assim, devemos supor que $|\rho_1| < 1$: esta é provavelmente uma suposição razoável, desde que \mathbf{N} tenha sido especificado corretamente.

Note pela equação (4.11) que o erro ϵ_i pode ser expresso como a soma ponderada de uma série de tempo de erros aleatórios não correlacionados e distribuídos de forma idêntica dados por $e_i, e_{i-1}, e_{i-2}, \dots$, com os pesos $1, \rho_1, \rho_1^2, \dots$ declinando geometricamente com o tempo. Esta situação indica que a influência dos fatores que afetam y e que não foram capturados por \mathbf{N} não é notada instantaneamente, mas é distribuída ao longo do tempo. O termo ϵ_i consiste na influência de eventos atuais (e_i), a influência de eventos que ocorreram um período atrás (e_{i-1}), a influência de eventos que ocorreram há dois períodos (e_{i-2}), e assim por diante. Além disso, se estendermos (4.11) para o futuro, temos

$$\begin{aligned}\epsilon_i &= e_i + \rho_1 e_{i-1} + \rho_1^2 e_{i-2} + \dots \\ \epsilon_{i+1} &= e_{i+1} + \rho_1 e_i + \rho_1^2 e_{i-1} + \dots \\ \epsilon_{i+2} &= e_{i+2} + \rho_1 e_{i+1} + \rho_1^2 e_i + \dots \\ &\vdots\end{aligned}$$

o que ilustra que, à medida que avançamos no futuro, a contribuição de um determinado termo, digamos, e_i , diminui. Se $0 < \rho_1 < 1$, sua contribuição terá sempre o mesmo sinal, e se $-1 < \rho_1 < 0$, sua contribuição oscilará no sinal.

Dessa forma, agora, podemos calcular a média e matriz de covariância para $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$. A partir de (4.11),

$$E(\epsilon_i) = \sum_{j=1}^{\infty} \rho_1^j E(e_{i-j}) = 0$$

e

$$\begin{aligned}\phi_\epsilon \equiv \text{Var}(\epsilon) = E(\epsilon_i^2) &= E[(e_i + \rho_1 e_{i-1} + \rho_1^2 e_{i-2} + \dots)^2] \\ &= E[e_i^2 + \rho_1^2 e_{i-1}^2 + \rho_1^4 e_{i-2}^2 + \dots + \rho_1 e_i e_{i-1} + \rho_1^2 e_i e_{i-2} + \dots] \\ &= E(e_i^2) + \rho_1^2 E(e_{i-1}^2) + \rho_1^4 E(e_{i-2}^2) + \dots \\ &= \phi_\xi (1 + \rho_1^2 + \rho_1^4 + \dots) \\ &= \frac{\phi_\xi}{(1 - \rho_1^2)}.\end{aligned}\tag{4.12}$$

Na derivação acima, observe que, como e_i 's não são correlacionados, a esperança de todos os termos do produto cruzado é zero, por exemplo, $E(e_i e_{i-1}) = 0$. Além disso, a última igualdade resulta da soma de uma progressão geométrica infinita, cuja razão comum é ρ_1^2 .

Também é uma questão simples encontrar a autocorrelação em lag s . Por exemplo, no lag 1, temos a autocovariância

$$\begin{aligned}\text{Cov}(\epsilon_i, \epsilon_{i-1}) &= E(\epsilon_i \epsilon_{i-1}) \\ &= E[(\rho_1 \epsilon_{i-1} + e_i) \epsilon_{i-1}] \\ &= \rho_1 \phi_\epsilon.\end{aligned}$$

Portanto, a autocorrelação no lag 1 é expressa por

$$\rho_1 = \frac{\text{Cov}(\epsilon_i, \epsilon_{i-1})}{\sqrt{\text{Var}(\epsilon_i) \text{Var}(\epsilon_{i-1})}} = \frac{\rho_1 \phi_\epsilon}{\phi_\epsilon} = \rho_1.$$

Da mesma forma, para o lag 2,

$$\begin{aligned}\text{Cov}(\epsilon_i, \epsilon_{i-2}) &= \text{E}(\epsilon_i \epsilon_{i-2}) \\ &= \text{E}\{[\rho_1(\rho_1 \epsilon_{i-2} + e_{i-1}) + e_i] \epsilon_{i-2}\} \\ &= \rho_1^2 \phi_\epsilon\end{aligned}$$

e, portanto, $\rho_2 = \rho_1^2$.

A covariância entre s períodos de dois erros é dada por

$$\text{Cov}(\epsilon_i, \epsilon_{i-s}) = \text{E}(\epsilon_i \epsilon_{i-s}) = \text{E}(\epsilon_{i+s} \epsilon_i) = \frac{\rho_1^s \phi_\xi}{1 - \rho_1^2}. \quad (4.13)$$

Para derivar a autocorrelação entre s períodos de dois erros, dividimos (4.13) por (4.12) e obtemos

$$\rho_s = \frac{\text{Cov}(\epsilon_i, \epsilon_{i-s})}{[\text{Var}(\epsilon_i) \text{Var}(\epsilon_{i-s})]^{1/2}} = \frac{\rho_1^s \phi_\xi / (1 - \rho_1^2)}{\phi_\xi / (1 - \rho_1^2)} = \rho_1^s.$$

Assim, quanto maior for o intervalo de tempo entre duas perturbações, menor será a correlação entre elas. Agora, estamos dispostos a escrever a matriz de covariância para ϵ . A partir de (4.12) e (4.13) obtemos

$$\mathbf{\Upsilon} = \text{E}(\epsilon \epsilon^\top) = \text{E} \begin{pmatrix} \epsilon_1^2 & \epsilon_1 \epsilon_2 & \dots & \epsilon_1 \epsilon_n \\ \epsilon_2 \epsilon_1 & \epsilon_2^2 & \dots & \epsilon_2 \epsilon_n \\ & & \dots & \\ \epsilon_n \epsilon_1 & \epsilon_n \epsilon_2 & \dots & \epsilon_n^2 \end{pmatrix} = \phi_\xi \mathbf{V},$$

em que

$$\mathbf{V} = \frac{1}{1 - \rho_1^2} \begin{pmatrix} 1 & \rho_1 & \rho_1^2 & \dots & \rho_1^{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_1^{n-2} \\ \rho_1^2 & \rho_1 & 1 & \dots & \rho_1^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_1^{n-1} & \rho_1^{n-2} & \rho_1^{n-3} & \dots & 1 \end{pmatrix}. \quad (4.14)$$

Além disso, temos que a inversa dessa matriz, é definida como \mathbf{V}^{-1} e dada por

$$\mathbf{V}^{-1} = \begin{pmatrix} 1 & -\rho_1 & 0 & \dots & 0 & 0 & 0 \\ -\rho_1 & 1 + \rho_1^2 & -\rho_1 & \dots & 0 & 0 & 0 \\ 0 & -\rho_1 & 1 + \rho_1^2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho_1 & 1 + \rho_1^2 & -\rho_1 \\ 0 & 0 & 0 & \dots & 0 & -\rho_1 & 1 \end{pmatrix}.$$

De maneira mais geral, para o processo autorregressivo de primeira ordem, $\rho_s = \rho_1^s$, e como $|\rho_1| < 1$, as autocorrelações dos erros decaem exponencialmente em direção a 0 conforme o lag s fica maior. Esse comportamento é aparente nos exemplos da Figura 4.1, que mostra as funções de autocorrelação AR(1) para $\rho = 0,95$ e $\rho = -0,75$. Observe que a autocorrelação no lag 0 é $\rho_0 = 1$.

Agora no segundo momento, vamos estudar os modelos autorregressivos de ordem 2, ou seja, AR(2), na primeira abordagem, tentamos reescrever de forma recursiva a expressão, $\epsilon_i = \rho_1 \epsilon_{i-1} +$

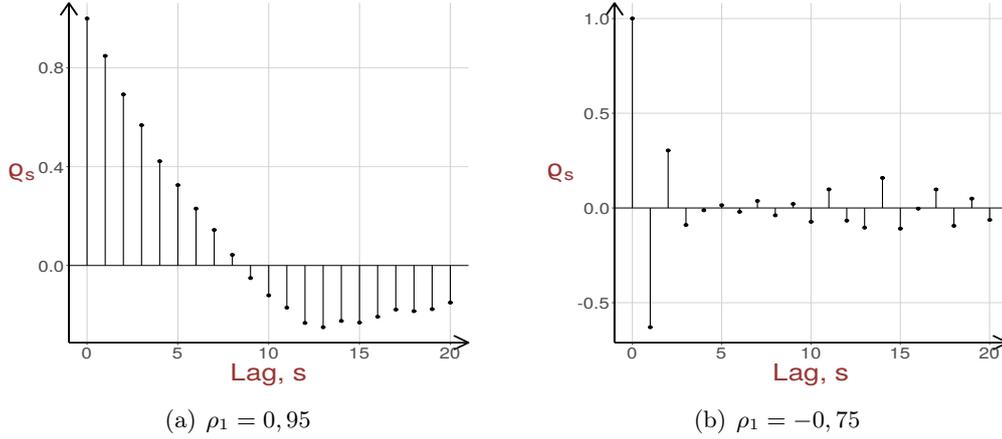


Figura 4.1: Autocorrelção teórica ρ_s para o processo $AR(2)$ com $\epsilon_i = \rho_1\epsilon_{i-1} + \rho_2\epsilon_{i-2} + e_i$.

$\rho_2\epsilon_{i-2} + e_i$, mas como pode ser visto abaixo, a medida que $i \rightarrow n$, torna-se custoso e trabalhoso.

$$\begin{aligned}
 \epsilon_i &= \rho_1\epsilon_{i-1} + \rho_2\epsilon_{i-2} + e_i \\
 &= \rho_1(\rho_1\epsilon_{i-2} + \rho_2\epsilon_{i-3} + e_{i-1}) + \rho_2(\rho_1\epsilon_{i-3} + \rho_2\epsilon_{i-4} + e_{i-2}) + e_i \\
 &= e_i + \rho_1e_{i-1} + \rho_2e_{i-2} + \dots \\
 &\vdots
 \end{aligned}
 \tag{4.15}$$

em que $E(e_i) = 0$, $E(e_ie_{i'}) = 0$ para $i \neq i'$, e $E(e_i^2) = \phi_\xi$. Esse processo será estacionário se $\rho_1 + \rho_2 < 1$, $\rho_2 - \rho_1 < 1$ e $-1 < \rho_2 < 1$. Segundo [Judge e Griffiths \(1985\)](#) e [Fox \(2015\)](#) os elementos da matriz de covariância $E(\epsilon\epsilon^\top) = \mathbf{\Upsilon} = \phi_\xi\mathbf{V}$ podem ser encontrados a partir da variância

$$\phi_\epsilon \equiv \text{Var}(\epsilon) = \phi_\xi \frac{(1 - \rho_2)}{(1 + \rho_2)[(1 - \rho_2)^2 - \rho_1^2]}.$$

Multiplicando (4.15) por ϵ_{i-1} e tomando as esperanças, temos que

$$\begin{aligned}
 \text{Cov}(\epsilon_i, \epsilon_{i-1}) &= \rho_1E(\epsilon_{i-1}^2) + \rho_2E(\epsilon_{i-1}\epsilon_{i-2}) \\
 &= \rho_1\phi_\epsilon + \rho_2\text{Cov}(\epsilon_i, \epsilon_{i-1}),
 \end{aligned}$$

já que $E(\epsilon_{i-1}^2) = \phi_\epsilon$ e $E(\epsilon_{i-1}\epsilon_{i-2}) = \text{Cov}(\epsilon_{i-1}, \epsilon_{i-2}) = \text{Cov}(\epsilon_i, \epsilon_{i-1})$. Resolvendo para a autocovariância,

$$\phi_1 \equiv \text{Cov}(\epsilon_i, \epsilon_{i-1}) = \frac{\rho_1}{1 - \rho_2}\phi_\epsilon.$$

Similarmente, para $s > 1$,

$$\begin{aligned}
 \phi_s \equiv \text{Cov}(\epsilon_i, \epsilon_{i-s}) &= \rho_1E(\epsilon_{i-1}\epsilon_{i-s}) + \rho_2E(\epsilon_{i-2}\epsilon_{i-s}) \\
 &= \rho_1\phi_{s-1} + \rho_2\phi_{s-2}
 \end{aligned}$$

e, assim, podemos encontrar as autocovariâncias recursivamente. Por exemplo, para $s = 2$,

$$\begin{aligned}
 \phi_2 &= \rho_1\phi_1 + \rho_2\phi_0 \\
 &= \rho_1\phi_1 + \rho_2\phi_\epsilon,
 \end{aligned}$$

em que $\phi_0 = \phi_\epsilon$, e para $s = 3$,

$$\phi_3 = \rho_1\phi_2 + \rho_2\phi_1.$$

Dessa forma, a inversa de \mathbf{V} é dada por

$$\mathbf{V}^{-1} = \begin{pmatrix} 1 & -\rho_1 & -\rho_2 & 0 & \dots & 0 & 0 \\ -\rho_1 & 1 + \rho_1^2 & -\rho_1 + \rho_1\rho_2 & -\rho_2 & \dots & 0 & 0 \\ -\rho_2 & -\rho_1 + \rho_1\rho_2 & 1 + \rho_1^2 + \rho_2^2 & -\rho_1 + \rho_1\rho_2 & \dots & 0 & 0 \\ 0 & -\rho_2 & -\rho_1 + \rho_1\rho_2 & 1 + \rho_1^2 + \rho_2^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 + \rho_1^2 & -\rho_1 \\ 0 & 0 & 0 & 0 & \dots & -\rho_1 & 1 \end{pmatrix}.$$

As autocorrelações para o modelo AR(2) seguem a divisão das autocovariância por σ_ϵ^2 :

$$\begin{aligned} \text{lag } 0 : \varrho_0 &= 1, \\ \text{lag } 1 : \varrho_1 &= \frac{\rho_1}{1 - \rho_2}, \\ \text{lag } 2 : \varrho_2 &= \frac{\rho_1^2}{1 - \rho_2} + \rho_2, \\ \text{lag } 3 : \varrho_3 &= \rho_1\varrho_2 + \rho_2\varrho_1 \text{ e} \\ \text{lag } s > 3 : \varrho_s &= \rho_1\varrho_{s-1} + \rho_2\varrho_{s-2}. \end{aligned}$$

Se o processo for estacionário, então essas autocorrelações decaem para 0, embora o padrão de decaimento possa ser mais ou menos complexo dependendo dos valores e sinais dos parâmetros autorregressivos ρ_1 e ρ_2 . Veja dois exemplo na Figura 4.2.

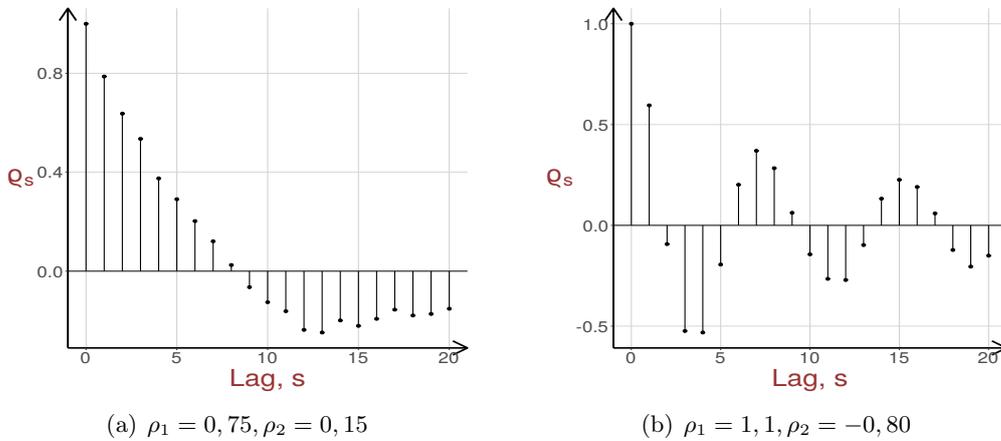


Figura 4.2: Autocorrelação teórica ϱ_s para o processo AR(2) com $\epsilon_i = \rho_1\epsilon_{i-1} + \rho_2\epsilon_{i-2} + e_i$.

No desenvolvimento anterior, descrevemos como o modelo linear geral com erros AR(1) pode ser estendido aos modelos com erros AR(2). Se tivermos observações suficientes, essa extensão prontamente generaliza para erros que seguem qualquer processo AR de ordem finita. No entanto, a expressão para a matriz variância-covariância Υ_p torna-se progressivamente mais complicada. Uma expressão geral para \mathbf{V} é dada por Wise (1955).

Thomas e Wallis (1971) sugerem que, quando obtemos dados trimestrais, um processo de quarta ordem pode ser apropriado. No entanto, em vez de um processo geral de quarta ordem, eles sugerem que apenas os distúrbios nos trimestres correspondentes de cada ano devem ser correlacionados. Isso leva à especificação

$$\epsilon_i = \rho_1 \epsilon_{i-4} + e_i,$$

cujas matriz de variância-covariância fica dada por

$$\phi \xi \mathbf{V}_4 = \phi \xi (\mathbf{V}_1 \otimes \mathbf{I}_4),$$

em que \mathbf{V}_1 é de dimensão $(n/4 \times n/4)$ e tem a mesma estrutura como \mathbf{V} obtida a partir de um processo AR(1), veja equação (4.14). A estrutura é tal que $E(\epsilon_i \epsilon_{i-s}) = 0$ a menos que s seja um múltiplo de 4, no caso em que temos $\phi_\epsilon \rho_1^{s/4}$. Também, $\phi_\epsilon = \phi \xi / (1 - \rho^2)$.

Com o auxílio dos cálculos no Apêndice B.3, a matriz de informação de Fisher, \mathcal{I} , pode ser considerada como uma matriz bloco diagonal do seguinte modo:

$$\begin{aligned} \mathcal{I}^{\gamma\gamma} &= \frac{4dg}{\phi} (\mathbf{N}_A^\top \mathbf{N}_A), \\ \mathcal{I}^{\phi\phi} &= \frac{n}{4\phi^2} (4f_g - 1) \quad e \\ \mathcal{I}^{\rho\rho} &= \frac{4dg}{\phi} \Upsilon_\rho, \end{aligned}$$

em que,

$$\Upsilon_\rho = \phi \xi \begin{pmatrix} \sum_{i=1}^{n-1} E(\epsilon_i^2) & \sum_{i=1}^{n-1} E(\epsilon_i \epsilon_{i+1}) & \dots & \sum_{i=1}^{n-p+1} E(\epsilon_i \epsilon_{i+p-1}) \\ \sum_{i=1}^{n-1} E(\epsilon_i \epsilon_{i+1}) & \sum_{i=1}^{n-2} E(\epsilon_i^2) & \dots & \sum_{i=1}^{n-p+2} E(\epsilon_i \epsilon_{i+p-2}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n-p+1} E(\epsilon_i \epsilon_{i+p-1}) & \sum_{i=1}^{n-p+2} E(\epsilon_i \epsilon_{i+p-2}) & \dots & \sum_{i=1}^{n-p} E(\epsilon_i^2) \end{pmatrix}.$$

é uma matriz $(p \times p)$ discutida no Apêndice B.3. Note que existe ortogonalidade entre $\gamma = (\gamma_0^\top, \gamma_1^\top, \dots, \gamma_k^\top)^\top$ e $(\phi, \rho^\top)^\top$ veja detalhes nos Apêndices B.2 e B.3. Portanto, os demais elementos da matriz são $\mathbf{0}$. Temos que $\mathbf{N}_A = (\mathbf{A}\mathbf{N}_0, \mathbf{A}\mathbf{N}_1, \dots, \mathbf{A}\mathbf{N}_k)$, d_g e f_g são dadas por $E\{W_g^2(z^2)z^2\}$, $E\{W_g^2(z^2)z^4\}$, respectivamente, com $z \sim S(0, 1)$. Essas quantidades podem ser auferidas na Tabela 2.4. Podemos definir a matriz de informação de Fisher regular de θ como

$$\mathcal{I}_{\theta\theta} = \begin{pmatrix} \mathcal{I}^{\gamma_0\gamma_0} & \mathcal{I}^{\gamma_0\gamma_1} & \dots & \mathcal{I}^{\gamma_0\gamma_k} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathcal{I}^{\gamma_1\gamma_0} & \mathcal{I}^{\gamma_1\gamma_1} & \dots & \mathcal{I}^{\gamma_1\gamma_k} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots & & & \vdots \\ \mathcal{I}^{\gamma_k\gamma_0} & \mathcal{I}^{\gamma_k\gamma_1} & \dots & \mathcal{I}^{\gamma_k\gamma_k} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathcal{I}^{\phi\phi} & 0 & \dots & 0 \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & 0 & \mathcal{I}^{\rho_1\rho_1} & \dots & \mathcal{I}^{\rho_1\rho_p} \\ \vdots & \vdots & & \vdots & \vdots & & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & 0 & \mathcal{I}^{\rho_p\rho_1} & \dots & \mathcal{I}^{\rho_p\rho_p} \end{pmatrix} = \text{blockdiag}\{\mathcal{I}^{\gamma\gamma}, \mathcal{I}^{\phi\phi}, \mathcal{I}^{\rho\rho}\}.$$

4.3.4 Processo iterativo

O processo iterativo de *backfitting* (Gauss-Seidel) para obter os EMVPs das partes paramétricas e as funções não paramétricas de $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_0^\top, \hat{\boldsymbol{\gamma}}_1^\top, \dots, \hat{\boldsymbol{\gamma}}_k^\top)^\top$ alternando com o processo iterativo para obter os EMVPs de ϕ e $\boldsymbol{\rho}$, assumindo $\boldsymbol{\lambda}$ fixo, é dado por resolver as seguintes equações:

$$\left\{ (\mathbf{AN}_0)^\top \mathbf{D}_v (\mathbf{AN}_0) \right\} \boldsymbol{\gamma}_0 = (\mathbf{AN}_0)^\top \mathbf{D}_v \mathbf{A} \left(\mathbf{y} - \sum_{j \neq 0} \mathbf{N}_j \boldsymbol{\gamma}_j \right), \quad (4.16)$$

$$\left\{ (\mathbf{AN}_l)^\top \mathbf{D}_v (\mathbf{AN}_l) + \phi \lambda_l \mathbf{M}_l \right\} \boldsymbol{\gamma}_l = (\mathbf{AN}_l)^\top \mathbf{D}_v \mathbf{A} \left(\mathbf{y} - \mathbf{N}_0 \boldsymbol{\gamma}_0 - \sum_{l' \neq l} \mathbf{N}_{l'} \boldsymbol{\gamma}_{l'} \right), \quad (l = 1, \dots, k), \quad (4.17)$$

as quais podem ser simplificadas para

$$\boldsymbol{\gamma}_0 = \mathbf{S}_0 \mathbf{A} \left\{ \mathbf{y} - \sum_{l \neq 0} \mathbf{N}_l \boldsymbol{\gamma}_l \right\}, \quad \boldsymbol{\gamma}_l = \mathbf{S}_l(\lambda_l) \mathbf{A} \left\{ \mathbf{y} - \mathbf{N}_0 \boldsymbol{\gamma}_0 - \sum_{l' \neq l} \mathbf{N}_{l'} \boldsymbol{\gamma}_{l'} \right\},$$

em que $\mathbf{S}_0 = \left\{ (\mathbf{AN}_0)^\top \mathbf{D}_v (\mathbf{AN}_0) \right\}^{-1} (\mathbf{AN}_0)^\top \mathbf{D}_v$ e $\mathbf{S}_l(\lambda_l) = \left\{ (\mathbf{AN}_l)^\top \mathbf{D}_v (\mathbf{AN}_l) + \phi \lambda_l \mathbf{M}_l \right\}^{-1} (\mathbf{AN}_l)^\top \mathbf{D}_v$. Então, propõe-se, para $\phi, \boldsymbol{\rho}$ e $\boldsymbol{\lambda}$ fixos, o seguinte processo iterativo para obter o EMVP de $\hat{\boldsymbol{\gamma}}$:

1. Defina os nós, e então calcule $\mathbf{N}_1, \dots, \mathbf{N}_k$ e $\mathbf{M}_1, \dots, \mathbf{M}_k$, \mathbf{N}_0 é conhecida.
2. Inicialize um contador como $u = 0$, determine um valor inicial para $\boldsymbol{\theta}^{(0)}$.
3. Baseado em $\boldsymbol{\theta}^{(u)}$, faça o seguinte:

- (a) A partir dos valores iniciais $\boldsymbol{\theta}^{(0)} = \left(\boldsymbol{\gamma}_0^{(0)\top}, \dots, \boldsymbol{\gamma}_k^{(0)\top}, \phi^{(0)}, \rho_1^{(0)}, \dots, \rho_p^{(0)} \right)^\top$, obtenha \mathbf{A} , os pesos $v_i^{(0)} = v_i |_{\boldsymbol{\theta}^{(0)}}$ e $\mathbf{D}_v^{(0)} = \text{diag}_{1 \leq i \leq n} \left\{ v_i^{(0)} \right\}$. Então, calcule

$$\begin{aligned} \mathbf{S}_0^{(0)} &= \left\{ (\mathbf{AN}_0)^\top \mathbf{D}_v^{(0)} (\mathbf{AN}_0) \right\}^{-1} (\mathbf{AN}_0)^\top \mathbf{D}_v^{(0)}, \\ \mathbf{S}_l^{(0)}(\lambda_l) &= \left\{ (\mathbf{AN}_l)^\top \mathbf{D}_v^{(0)} (\mathbf{AN}_l) + \phi \lambda_l \mathbf{M}_l \right\}^{-1} (\mathbf{AN}_l)^\top \mathbf{D}_v^{(0)}, \quad (l = 1, \dots, k). \end{aligned}$$

- (b) Avalie a seguinte expressão para $\boldsymbol{\gamma}_0^{(u+1)}, \boldsymbol{\gamma}_j^{(u+1)}$:

$$\begin{aligned} \boldsymbol{\gamma}_0^{(u+1)} &= \mathbf{S}_0^{(u)} \mathbf{A} \left\{ \mathbf{y} - \sum_{l \neq 0} \mathbf{N}_l \boldsymbol{\gamma}_l^{(u)} \right\}, \\ \boldsymbol{\gamma}_1^{(u+1)} &= \mathbf{S}_1^{(u)}(\lambda_1) \mathbf{A} \left\{ \mathbf{y} - \mathbf{N}_0 \boldsymbol{\gamma}_0^{(u+1)} - \sum_{l=2}^k \mathbf{N}_l \boldsymbol{\gamma}_l^{(u)} \right\}, \\ &\vdots \\ \boldsymbol{\gamma}_k^{(u+1)} &= \mathbf{S}_k^{(u)}(\lambda_k) \mathbf{A} \left\{ \mathbf{y} - \sum_{l=0}^{k-1} \mathbf{N}_l \boldsymbol{\gamma}_l^{(u+1)} \right\}, \end{aligned}$$

para $u = 0, 1, \dots$. Atualize $\mathbf{S}_0^{(u)}, \mathbf{S}_l^{(u)}(\lambda_l)$. Repita (b) substituindo $\boldsymbol{\gamma}_l^{(u)}$ por $\boldsymbol{\gamma}_l^{(u+1)}$, para $l = 1, \dots, k$ até a convergência para algum critério. A condição de convergência utilizada em todos os passos do algoritmo foi que a diferença máxima entre os elementos dos vetores estimados da iteração (u) e da ($u + 1$) fossem menores do que 0,001.

A convergência do processo iterativo descrito acima e as condições de convergência seguem do Teorema 4.2 de [Green e Silverman \(1994, p. 68\)](#). Esse algoritmo conhecido como

backfitting pode convergir em poucos passos, porém depende da escolha dos parâmetros de suavização λ_l . Os quais são determinados por validação cruzada por minimizar a quantidade $VCG(\boldsymbol{\lambda})$, discutido nas próximas Seções, para uma grade selecionada de valores.

4. Repita o processo iterativo do passo (b) com o processo iterativo abaixo usando o método quase Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) ver [Davidon \(1991\)](#) e [Mittelhammer et al. \(2000\)](#), ou “L-BFGS-B” desenvolvido por [Byrd et al. \(1995\)](#) para obter o EMVP de $\zeta = (\phi, \boldsymbol{\rho})^\top$ dado por

$$\zeta^{(s+1)} = \arg \max_{(\phi, \boldsymbol{\rho})} L_p(\hat{\boldsymbol{\gamma}}^{(u+1)}, \phi^{(s)}, \boldsymbol{\rho}^{(s)}), \quad s = 0, 1, 2, \dots$$

O cálculo direto dos EMVPs de $\hat{\boldsymbol{\gamma}}_l$, ($l = 0, \dots, k$) é raramente usado na prática, porque o algoritmo de *backfitting* é mais eficiente para obter $\hat{\boldsymbol{\gamma}}_l$; não requer matrizes de alta dimensão e sua inversa. Todavia, as funções escores obtidas em (4.7) e igualando-as a zero, conforme (4.16) e (4.17) podem ser úteis se desejarmos estudar algumas propriedades teóricas dos EMVPs. Para alguns casos especiais de MLPAS condicionais AR(p), considere $\boldsymbol{\lambda}$, ϕ e \mathbf{D}_v fixos, é possível manipular as equações algebricamente e obter formas fechadas para as estimativas de $\boldsymbol{\gamma}_l$ ($l = 0, \dots, k$). Por exemplo:

- (i) Para $\boldsymbol{\gamma}_0 = \mathbf{0}$ e $k = 2$, temos que

$$\begin{aligned} \hat{\boldsymbol{\gamma}}_1 &= \left\{ (\mathbf{AN}_1)^\top \mathbf{W}_2 (\mathbf{AN}_1) + \phi \lambda_1 \mathbf{M}_1 \right\}^{-1} (\mathbf{AN}_1)^\top \mathbf{W}_2 \mathbf{A} \mathbf{y}, \\ \hat{\boldsymbol{\gamma}}_2 &= \left\{ (\mathbf{AN}_2)^\top \mathbf{W}_1 (\mathbf{AN}_2) + \phi \lambda_2 \mathbf{M}_2 \right\}^{-1} (\mathbf{AN}_2)^\top \mathbf{W}_2 \mathbf{A} \mathbf{y}, \end{aligned}$$

em que

$$\begin{aligned} \mathbf{W}_1 &= \mathbf{D}_v - \mathbf{D}_v (\mathbf{AN}_1) \left\{ (\mathbf{AN}_1)^\top \mathbf{D}_v (\mathbf{AN}_1) + \phi \lambda_1 \mathbf{M}_1 \right\}^{-1} (\mathbf{AN}_1)^\top \mathbf{D}_v = \mathbf{D}_v [\mathbf{I}_n - (\mathbf{AN}_1) \mathbf{S}_1], \\ \mathbf{W}_2 &= \mathbf{D}_v - \mathbf{D}_v (\mathbf{AN}_2) \left\{ (\mathbf{AN}_2)^\top \mathbf{D}_v (\mathbf{AN}_2) + \phi \lambda_2 \mathbf{M}_2 \right\}^{-1} (\mathbf{AN}_2)^\top \mathbf{D}_v = \mathbf{D}_v [\mathbf{I}_n - (\mathbf{AN}_2) \mathbf{S}_2]. \end{aligned}$$

- (ii) Para $\boldsymbol{\gamma}_0 \neq \mathbf{0}$ e $k = 1$, temos que

$$\begin{aligned} \hat{\boldsymbol{\gamma}}_0 &= \left\{ (\mathbf{AN}_0)^\top \mathbf{W}_1 (\mathbf{AN}_0) \right\}^{-1} (\mathbf{AN}_0)^\top \mathbf{W}_1 \mathbf{A} \mathbf{y}, \\ \hat{\boldsymbol{\gamma}}_1 &= \left\{ (\mathbf{AN}_1)^\top \mathbf{W}_0 (\mathbf{AN}_1) + \phi \lambda_1 \mathbf{M}_1 \right\}^{-1} (\mathbf{AN}_1)^\top \mathbf{W}_0 \mathbf{A} \mathbf{y}, \end{aligned}$$

em que $\mathbf{W}_0 = \mathbf{D}_v [\mathbf{I}_n - (\mathbf{AN}_0) \mathbf{S}_0]$.

- (iii) Para $\boldsymbol{\gamma}_0 \neq \mathbf{0}$ e $k = 2$,

$$\begin{aligned} \hat{\boldsymbol{\gamma}}_0 &= \left\{ (\mathbf{AN}_0)^\top \mathbf{W}_{12} (\mathbf{AN}_0) \right\}^{-1} (\mathbf{AN}_0)^\top \mathbf{W}_{12} \mathbf{A} \mathbf{y}, \\ \hat{\boldsymbol{\gamma}}_1 &= \left\{ (\mathbf{AN}_1)^\top \mathbf{W}_{02} (\mathbf{AN}_1) + \phi \lambda_1 \mathbf{M}_1 \right\}^{-1} (\mathbf{AN}_1)^\top \mathbf{W}_{02} \mathbf{A} \mathbf{y}, \\ \hat{\boldsymbol{\gamma}}_2 &= \left\{ (\mathbf{AN}_2)^\top \mathbf{W}_{01} (\mathbf{AN}_2) + \phi \lambda_2 \mathbf{M}_2 \right\}^{-1} (\mathbf{AN}_2)^\top \mathbf{W}_{01} \mathbf{A} \mathbf{y}, \end{aligned}$$

em que

$$\begin{aligned} \mathbf{S}_{l'} &= \left\{ (\mathbf{AN}_l)^\top \mathbf{D}_v (\mathbf{AN}_l) - (\mathbf{AN}_l)^\top \mathbf{D}_v (\mathbf{AN}_{l'}) \mathbf{S}_{l'} (\mathbf{AN}_l) + \phi \lambda_l \mathbf{M}_l \right\}^{-1} (\mathbf{AN}_l)^\top \mathbf{D}_v \\ \mathbf{W}_{l'} &= \mathbf{D}_v [\mathbf{I}_n - \mathbf{A} \{ \mathbf{N}_l \mathbf{S}_{l'} - \mathbf{N}_l \mathbf{S}_{l'} (\mathbf{AN}_{l'}) \mathbf{S}_{l'} + \mathbf{N}_{l'} \mathbf{S}_{l'} - \mathbf{N}_{l'} \mathbf{S}_{l'} (\mathbf{AN}_l) \mathbf{S}_{l'} + \mathbf{N}_{l'} \mathbf{S}_{l'} (\mathbf{AN}_l) \mathbf{S}_{l'} (\mathbf{AN}_{l'}) \mathbf{S}_{l'} \}] \\ &\quad \forall l \neq l' : 0, 1, 2, \quad \text{e } \mathbf{M}_0 = \mathbf{0}. \end{aligned}$$

4.4 Procedimentos inferenciais

Nesta seção, discutiremos alguns aspectos inferenciais para os modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p . Primeiro, descrevemos um procedimento para obter os erros padrão aproximados para as estimativas dos parâmetros e derivamos faixas de erro padrão pontuais aproximadas para as funções não paramétricas. A seguir, são apresentadas discussões sobre a estimativa dos graus de liberdade efetivos.

4.4.1 Erro padrão aproximado

Modelos lineares sob erros normais e simétricos possuem condições de regularidade semelhantes para grandes amostras, o que garante consistência, eficiência e normalidade dos EMVPs. Como para o caso paramétrico, a matriz de variância-covariância aproximada $\widehat{\boldsymbol{\theta}}$ pode ser obtida a partir da inversa da matriz de informação esperada. Essencialmente, $\widehat{\text{Var}}(\widehat{\boldsymbol{\theta}}) = (\mathcal{I}_p^{\theta\theta} |_{\widehat{\boldsymbol{\theta}}})^{-1}$, em que $\mathcal{I}_p^{\theta\theta}$ foi definida na Seção 4.3.3 e com mais detalhes no Apêndice B.3. Essa abordagem tem suporte na abordagem Bayesiana para modelos lineares, conforme descrito em Wood (2017, p. 293). Se $\mathbf{y} = \mathbf{N}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$, em que $\boldsymbol{\epsilon}$ segue uma distribuição normal n -variada e uma priori imprópria é assumida para $\boldsymbol{\gamma}$, então a distribuição a posteriori $\boldsymbol{\gamma}|\mathbf{y}$ segue uma distribuição normal multivariada de média $\widehat{\boldsymbol{\gamma}}$ e matriz de variância-covariância que corresponde ao inverso da respectiva matriz de informação de Fisher penalizada para $\boldsymbol{\gamma}$. Assim, intervalos de credibilidade podem ser construídos para quaisquer quantidades derivadas de $\boldsymbol{\gamma}$. Esta abordagem pode ser estendida, para n grande, para a classe simétrica. Em particular, se estivermos interessados em extrair inferências para γ_0 , $\bar{\boldsymbol{\gamma}} = (\gamma_1, \dots, \gamma_k)$, ϕ e $\boldsymbol{\rho}$, as matrizes aproximadas de variância-covariância podem ser estimadas usando as matrizes bloco diagonal correspondentes obtidas de \mathcal{I}_p^{-1} , isto é,

$$\begin{aligned}\widehat{\text{Var}}(\widehat{\gamma}_0) &= \left(\mathcal{I}_p^{\gamma_0\gamma_0} - \mathcal{I}_p^{\gamma_0\bar{\boldsymbol{\gamma}}} \mathcal{I}_p^{\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\gamma}}}^{-1} \mathcal{I}_p^{\bar{\boldsymbol{\gamma}}\gamma_0} \right)^{-1} |_{\widehat{\boldsymbol{\theta}}} \quad \text{e} \\ \widehat{\text{Var}}(\widehat{\gamma}_1, \dots, \widehat{\gamma}_k) &= \left(\mathcal{I}_p^{\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\gamma}}} - \mathcal{I}_p^{\gamma_0\bar{\boldsymbol{\gamma}}\top} \mathcal{I}_p^{\gamma_0\gamma_0}^{-1} \mathcal{I}_p^{\gamma_0\bar{\boldsymbol{\gamma}}} \right)^{-1} |_{\widehat{\boldsymbol{\theta}}}.\end{aligned}$$

E considerando a ortogonalidade entre $\boldsymbol{\gamma}$ e $(\phi, \boldsymbol{\rho})^\top$, assim como entre ϕ e $\boldsymbol{\rho}$, obtém-se

$$\begin{aligned}\widehat{\text{Var}}(\widehat{\phi}) &= \mathbf{I}_{\phi\phi}^{-1} |_{\widehat{\boldsymbol{\theta}}} \quad \text{e} \\ \widehat{\text{Var}}(\widehat{\boldsymbol{\rho}}) &= \mathcal{I}_{\rho\rho}^{-1} |_{\widehat{\boldsymbol{\theta}}}.\end{aligned}$$

Uma banda de erro padrão pontual aproximada (BEP) para $\gamma_l(\cdot)$ (veja, por exemplo, Wahba (1983)) é dada por $\text{BEP}_{\text{approx}}(\gamma_l(t_{l_j}^0)) = \widehat{\gamma}_l(t_{l_j}^0) \pm 2\sqrt{\widehat{\text{Var}}_{\text{approx}}(\widehat{\gamma}_l(t_{l_j}^0))}$, em que $\widehat{\text{Var}}_{\text{approx}}(\widehat{\gamma}_l(t_{l_j}^0))$ denota a variância estimada aproximada de $\widehat{\gamma}_l(t_{l_j}^0)$, ($l = 1, \dots, k; j = 1, \dots, r_l$).

4.4.2 Intervalos de confiança simultâneos

Seja $100(1 - \alpha)\%$ um nível de confiança simultâneo desejado de $CI_{1,j}^{\alpha^{(*)}}, \dots, CI_{n,j}^{\alpha^{(*)}}$, em que $CI_{i,l}^{\alpha^{(*)}} = \left\{ f \in \mathcal{R} : |\widehat{f}_l(t_{l_j}^0) - f| \leq \Phi^{-1}(1 - \alpha^*/2) \left[\widehat{\text{Var}}(t_{l_j}^0) \right]^{1/2} \right\}$ é o intervalo de confiança baseado na normalidade de $100(1 - \alpha^*)\%$ de $f_l(t_{l_j}^0)$, com $\widehat{\text{Var}}(t_{l_j}^0)$ sendo a estimativa de $\text{Var}(t_{l_j}^0)$. A confiabilidade dos intervalos de confiança simultâneos de $CI_{1,j}^{\alpha^{(*)}}, \dots, CI_{n,j}^{\alpha^{(*)}}$ pode ser obtida tal que

$$\begin{aligned}P \left\{ \bigcap_{i=1}^n \left[f_l(t_{l_j}^0) \in CI_{ij}^{\alpha^{(*)}} \right] \right\} &= 1 - P \left\{ \bigcup_{i=1}^n \left[f_l(t_{l_j}^0) \notin CI_{ij}^{\alpha^{(*)}} \right] \right\} \\ &\geq 1 - \sum P \left[f_l(t_{l_j}^0) \notin CI_{ij}^{\alpha^{(*)}} \right] \\ &\geq 1 - r_l \alpha^*,\end{aligned}$$

em que r_l é o número de nós escolhidos. Assim, de acordo com o método de Bonferroni, α^* é um valor tal que de $\alpha^* = \alpha/r_l$ para que o nível de confiança simultâneo de $CI_{1,j}^{\alpha^*}, \dots, CI_{n,j}^{\alpha^*}$ seja pelo menos $100(1 - \alpha)\%$. Os intervalos de confiança simultâneos podem ser usados para testar informalmente o significado de uma variável explicativa cujo efeito é modelado não parametricamente. De fato, uma linha reta de inclinação zero que pode estar localizada dentro dos intervalos de confiança simultâneos é uma evidência informal de efeito nulo da covariável associada. Os métodos de Tukey e Scheffé são procedimentos alternativos ao método de Bonferroni para construir a banda de confiança simultânea para cada função não paramétrica.

4.4.3 Testes de hipóteses

Se tratarmos o logaritmo da função de verossimilhança penalizada como uma verossimilhança usual, podemos utilizar os testes de hipóteses já conhecidos na literatura. Por exemplo, para variâncias conhecidas podemos aplicar o teste da razão de verossimilhanças ou teste de Wald através da inversa da matriz de informação de Fisher para fazer inferência, para n grande, a respeito de γ e supondo normalidade assintótica $\hat{\gamma}_l$.

Vamos considerar as hipóteses $\mathcal{H}_0 : \gamma_l = \mathbf{0}$ contra $\mathcal{H}_1 : \gamma_l \neq \mathbf{0}$. O teste de Wald é dado por

$$\xi_W = \hat{\gamma}_l^\top \hat{\mathbf{V}}_{\gamma_l}^{-1} \hat{\gamma}_l,$$

em que $\hat{\mathbf{V}}_{\gamma_l}$ é a matriz de variância-covariância do l -ésimo componente não paramétrico para $l = 1, \dots, k$, e sob a hipótese nula, $\xi_W \sim \chi_{df_l}^2$, df são os graus de liberdade efetivos do modelo. No caso de \mathbf{V}_{γ_l} não ser de posto completo, a inversa de \mathbf{V}_{γ_l} pode ser substituída por uma inversa generalizada $\mathbf{V}_{\gamma_l}^-$. Segundo Wood (2017), quando o parâmetro de suavização é desconhecido, as estatísticas do teste sob a hipótese nula são geralmente menores que deveriam ser, e a hipótese nula acaba não sendo rejeitada muito facilmente. Em termos práticos, se essas estatísticas estiverem em torno de um limite para a aceitação ou rejeição de uma hipótese, devem ser analisadas com cautela.

4.5 Graus de liberdade efetivos

Desde que a penalização da função de suavização $f_1(t_1)$ a $f_k(t_k)$ leve a um encolhimento (*shrinkage*) dos EMVPs $\hat{\gamma}_1$ a $\hat{\gamma}_k$ com respeito aos EMVs, é crucial para a seleção do modelo e dos testes de hipóteses a estimação dos graus de liberdade efetivos correspondente aos EMVPs. Analogamente a Hastie e Tibshirani (1990, Cap. 5), Green e Silverman (1994, Cap. 5) e mais recentemente Wood (2017, Cap. 5), deriva-se abaixo os graus de liberdade efetivos correspondente aos EMVPs $\hat{\gamma}_0, \hat{\gamma}_1$ a $\hat{\gamma}_k$ a partir de um suavizador apropriado.

Na solução das equações de estimação $\mathbf{U}_p^\theta = \mathbf{0}$, tem-se que

$$\begin{aligned} \hat{\gamma} &= \left\{ \hat{\mathbf{N}}_A^\top \hat{\mathbf{D}}_v \hat{\mathbf{N}}_A + \hat{\phi} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \right\}^{-1} \left\{ \hat{\mathbf{N}}_A^\top \hat{\mathbf{D}}_v \hat{\mathbf{N}}_A \hat{\gamma} + \hat{\mathbf{N}}_A^\top \hat{\mathbf{D}}_v \hat{\mathbf{A}}(\mathbf{y} - \mathbf{N}\hat{\gamma}) \right\} \\ &= \left\{ \hat{\mathbf{N}}_A^\top \hat{\mathbf{D}}_v \hat{\mathbf{N}}_A + \hat{\phi} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \right\}^{-1} \hat{\mathbf{N}}_A^\top \hat{\mathbf{D}}_v^{\frac{1}{2}} \hat{\mathbf{D}}_v^{\frac{1}{2}} \hat{\mathbf{A}} \mathbf{y}, \end{aligned}$$

em que, $\mathbf{M}_\gamma(\boldsymbol{\lambda}) = \text{blockdiag}\{\mathbf{0}_{r_0}, \lambda_1 \mathbf{M}_1, \dots, \lambda_k \mathbf{M}_k\}$. Então, tem-se a relação $\hat{\mathbf{D}}_v^{\frac{1}{2}} \hat{\mathbf{N}}_A \hat{\gamma} = \hat{\mathbf{D}}_v^{\frac{1}{2}} \hat{\mathbf{A}} \hat{\mathbf{N}} \hat{\gamma} = \hat{\mathbf{D}}_v^{\frac{1}{2}} \hat{\mathbf{A}} \hat{\boldsymbol{\mu}} = \hat{\mathbf{H}}(\boldsymbol{\lambda}) \hat{\mathbf{D}}_v^{\frac{1}{2}} \hat{\mathbf{A}} \mathbf{y}$, em que

$$\hat{\mathbf{H}}(\boldsymbol{\lambda}) = \hat{\mathbf{D}}_v^{\frac{1}{2}} \hat{\mathbf{N}}_A \left\{ \hat{\mathbf{N}}_A^\top \hat{\mathbf{D}}_v \hat{\mathbf{N}}_A + \hat{\phi} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \right\}^{-1} \hat{\mathbf{N}}_A^\top \hat{\mathbf{D}}_v^{\frac{1}{2}} \quad (4.18)$$

é nomeada suavizador linear para $\boldsymbol{\lambda}$ fixo e projeta o estimador da resposta transformando $\hat{\mathbf{D}}_v^{\frac{1}{2}} \hat{\mathbf{A}} \mathbf{y}$ na média predita transformada $\hat{\mathbf{D}}_v^{\frac{1}{2}} \hat{\mathbf{A}} \hat{\boldsymbol{\mu}}$. Em particular, sob erros com distribuição normal e $\boldsymbol{\rho} = \mathbf{0}$, tem-se que $\hat{\mathbf{H}}(\boldsymbol{\lambda}) = \mathbf{N}\{\mathbf{N}^\top \mathbf{N} + \hat{\phi} \mathbf{M}_\gamma(\boldsymbol{\lambda})\}^{-1} \mathbf{N}^\top$.

Os graus de liberdade efetivos dos suavizadores são definidos como a soma dos autovalores de $\widehat{\mathbf{H}}(\boldsymbol{\lambda})$, ou seja, $\text{df}_s(\boldsymbol{\lambda}) = \text{tr}\{\widehat{\mathbf{H}}(\boldsymbol{\lambda})\}$. De Eilers e Marx (1996), tem-se que

$$\begin{aligned} \text{df}_s(\boldsymbol{\lambda}) &= \text{tr} \left\{ \widehat{\mathbf{D}}_v^{\frac{1}{2}} \widehat{\mathbf{N}}_A \left(\widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A + \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \right)^{-1} \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v^{\frac{1}{2}} \right\} \\ &= \text{tr} \left\{ \left(\widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A + \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \right)^{-1} \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A \right\} \\ &= \text{tr} \left\{ \left(\mathbf{I}_{r_T+r_S} + \mathbf{Q}^{-\frac{1}{2}} \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \mathbf{Q}^{-\frac{1}{2}} \right)^{-1} \right\} \\ &= \sum_{i=1}^{r_0+r_1+\dots+r_k} \frac{1}{1 + \alpha_i(\boldsymbol{\lambda})}, \end{aligned}$$

em que $\alpha_i(\boldsymbol{\lambda}) \geq 0$ são os autovalores da matriz definida não negativa $\mathbf{Q}^{-\frac{1}{2}} \widehat{\boldsymbol{\phi}} \mathbf{M}_\gamma(\boldsymbol{\lambda}) \mathbf{Q}^{-\frac{1}{2}}$ e $\mathbf{Q}^{\frac{1}{2}} \mathbf{Q}^{\frac{1}{2}} = \widehat{\mathbf{N}}_A^\top \widehat{\mathbf{D}}_v \widehat{\mathbf{N}}_A$, para $i = 1, \dots, r_0 + r_1 + \dots + r_k$. Os graus de liberdade efetivos $\text{df}_s(\lambda_0)$, $\text{df}_s(\lambda_1)$ a $\text{df}_s(\lambda_k)$ correspondem à soma dos primeiros r_0 , r_1 e os últimos r_k autovalores, respectivamente, do suavizador linear $\widehat{\mathbf{H}}(\boldsymbol{\lambda})$. Portanto, os graus de liberdade efetivos do modelo em estudo são dados por

$$\text{df}(\boldsymbol{\lambda}) = \text{df}_s(\boldsymbol{\lambda}) + p + 1.$$

4.6 Comentários

Neste capítulo introduzimos os modelos lineares parciais aditivos com erros simétricos condicionais $\text{AR}(p)$, em que o componente aditivo pode ser decomposto em um ou mais componentes, por exemplo, em um determinado estudo podemos estar interessados em saber se uma variável possui tendência não linear ao longo do tempo, mas também apresentar sazonalidade em relação aos meses do ano ou dias da semana. Além disso, alguns determinados horários podem estar diretamente relacionados ao aumento ou não dessa variável. Assim, aplicar componentes aditivos faz total sentido, visto que há possibilidade destas variáveis apresentarem comportamento flexível ao longo do tempo. E, logo são aproximados por *splines* cúbicos penalizados de regressão. Um processo iterativo do tipo *backfitting* para obter os coeficientes estimados dos *splines* cúbicos penalizados de regressão é alternado com um procedimento usual de máxima verossimilhança para obter $\widehat{\boldsymbol{\phi}}$ e $\widehat{\boldsymbol{\rho}}$. A função escore penalizada e a matriz de informação de Fisher penalizada são obtidas, bem como os graus de liberdade efetivos são derivados.

Capítulo 5

Análise de Diagnóstico

Após o processo de estimação dos parâmetros, é importante determinar se as premissas subjacentes ao modelo são atendidas e se a equação modela com precisão o processo em estudo. Sabe-se que nem todas as observações no conjunto de dados desempenham um papel harmônico ao ajustar um modelo de regressão. Algumas podem ter mais impacto, por exemplo, no processo de estimação do que outras, a escolha da distribuição para o erro pode não ser apropriada para o conjunto de dados investigados. Portanto, conclusões baseadas nesses modelos podem ser equivocadas. Diante de tal problema, a solução é examinar quanto o modelo considerado é adequado para os dados antes de fazer inferências. A identificação de observações que afetam as estimativas dos parâmetros é uma componente importante para validação de modelos. Observações que influenciam desproporcionalmente determinados resultados da análise de regressão são chamadas observações influentes. O estudo dos dados e como diferentes partes deles influenciam a inferência é chamado análise de influência ou análise de sensibilidade.

De maneira mais geral, deseja-se identificar subconjuntos de dados que têm influência desproporcional no modelo de regressão estimado. Como discutido por [Belsley et al. \(1980\)](#), esses subconjuntos influentes podem vir de várias fontes:

1. observações digitadas incorretamente (resultados de erros durante a coleta ou o processamento dos dados);
2. erros inerentes aos dados (ou conscientizar o pesquisador da possibilidade de que parte dos dados venha de outro regime ou subpopulação que possua características muito diferentes em comparação à população estudada);
3. valores extremos que são observações legitimamente extremas.

Sendo assim, [Andrews e Pregibon \(1978\)](#) destacaram que precisamos encontrar os valores extremos que importam. O que querem dizer com isso é que nem todos os discrepantes são prejudiciais, pois influenciam indevidamente, por exemplo, a estimativa dos parâmetros no modelo de regressão. Se nem todos os valores extremos importam, examinar apenas os resíduos pode não nos levar à detecção de observações aberrantes ou incomuns. Assim, são necessárias outras maneiras de encontrar observações influentes.

É preciso uma filosofia geral na identificação de modelos para permitir aos pesquisadores certa flexibilidade no exercício de seu julgamento subjetivo. É um fato que raramente existe um verdadeiro modelo na prática. Um bom mecanismo de identificação de modelos deve levar à seleção de um que seja estatisticamente sólido e significativo. Um modelo parcimonioso é sempre preferível quando dois modelos candidatos parecem igualmente bons. Como todo ajuste estatístico, é importante realizar uma verificação de diagnóstico para um modelo de série temporal ajustado.

Uma suposição básica do modelo de regressão linear é que os termos de erro seguem uma distribuição normal. Porém, sabe-se que alguns fenômenos nem sempre estão de acordo com essa suposição, produzindo dados de distribuição com caudas pesadas, assimetria ou multimodalidade. Essas características podem ser contornadas por transformações de dados (ou seja, Box-Cox, etc.), as

quais podem tornar a normalidade aproximada com resultados empíricos razoáveis. Apesar disso, algumas desvantagens possíveis desses métodos são: (i) transformações fornecem informações reduzidas sobre o esquema de geração dos dados subjacente; (ii) transformações das componentes podem não garantir a normalidade; (iii) parâmetros podem perder a interpretabilidade em uma escala transformada; e (iv) as transformações podem não ser universais e geralmente variam com o conjunto de dados. Portanto, de uma perspectiva prática, é necessário buscar um modelo teórico apropriado que evite a transformação dos dados. Na prática, quando nada se sabe sobre a verdadeira distribuição dos termos do erro, uma análise de regressão linear baseada em qualquer um dos modelos acima pode ser realizada usando um modelo especificado incorretamente. Além disso, pode haver situações em que uma única família paramétrica é incapaz de fornecer um modelo satisfatório para variações locais nos dados observados.

Para que a seleção do modelo seja feita de forma adequada, são propostos diversos métodos com intuito de verificar a qualidade do ajuste. Os objetivos dos procedimentos de diagnóstico são avaliar a existência de afastamento das suposições feitas inicialmente para o modelo, bem como detectar observações discrepantes, especialmente, pontos influentes e predições *outliers*. Essas observações podem exercer grande influência sobre as estimativas dos parâmetros e, em algumas situações, podem causar alterações inferenciais.

5.1 Análise de resíduos

Análise de resíduos é uma ferramenta importante para detectar observações aberrantes e desvio das suposições feitas para a distribuição dos erros nos modelos de regressão. Entre os vários estudos propostos, o resíduo quantílico (Dunn e Smyth, 1996) pode ser facilmente obtido quando a função de distribuição acumulada (fda) é conhecida. Tal resíduo para observações independentes $(y_1, \dots, y_n)^\top$ é definido como $r_{q_i} = \Phi^{-1}\{F_y(y_i; \boldsymbol{\theta})\}$, em que $F_y(y_i; \boldsymbol{\theta})$ e $\Phi(\cdot)$ denotam, respectivamente, a função de distribuição de probabilidade de y_i e a fda da $N(0, 1)$, para $i = 1, \dots, n$. Para n grande e sob o modelo postulado r_{q_1}, \dots, r_{q_n} seguem distribuição normal padrão.

No entanto, de acordo com Barros e Paula (2019), para observações correlacionadas deve-se considerar a distribuição condicional independente com respeito à respectiva função de distribuição acumulada condicional, a qual pode ser dada por $F_{y_1}(y_1; \boldsymbol{\theta}), F_{y_2|y_1}(y_2; \boldsymbol{\theta}), \dots, F_{y_n|(y_1, \dots, y_{n-1})}(y_n; \boldsymbol{\theta})$. Dessa forma, os resíduos quantílicos condicionais são dados por

$$r_{q_1}^c = \Phi^{-1}\{F_{y_1}(y_1; \hat{\boldsymbol{\theta}})\}, \dots, r_{q_n}^c = \Phi^{-1}\{F_{y_n|(y_1, \dots, y_{n-1})}(y_n; \hat{\boldsymbol{\theta}})\},$$

para $i = 1, \dots, n$. Para n grande e sob o modelo postulado (4.1)-(4.2) $r_{q_1}^c, \dots, r_{q_n}^c$ seguem distribuição normal padrão. O gráfico de probabilidade normal pode ser aplicado para avaliar desvios do modelo postulado, assim como detectar possíveis observações aberrantes.

A análise residual examina, assim, o erro para cada valor y previsto. Ao visualizar graficamente os resíduos contra o valores ajustados ou resíduos contra o tempo, é possível detectar se as suposições foram violadas. Se houver um padrão claro para os gráficos residuais, em vez de uma distribuição aleatória dos resíduos, isso indica que uma das suposições pode ser violada. Por exemplo, se os resíduos são inicialmente muito pequenos, mas se espalham e ficam cada vez maior, isso indicaria que a variação não é constante. Outra maneira de verificar graficamente as suposições é através de um gráfico Q-Q, que representa o quantil da amostra em relação ao quantil teórico para determinar se dois conjuntos de dados vêm de uma população com uma distribuição específica. Como estamos interessados em verificar se nossos resíduos seguem uma distribuição normal, queremos um gráfico Q-Q normal. Como o nome indica, estamos comparando os quantis do nosso primeiro conjunto de dados contra os quantis do nosso segundo conjunto, em que quantil se refere à quantidade de dados que fica abaixo desse ponto. Os erros em estudo são simétricos autoregressivos de ordem p . Contudo, foi possível encontrar um resíduo (resíduo quantílico condicional) que sob essas distribuições para os erros são assintoticamente iid $N(0, 1)$.

5.2 Influência local

Avaliar a sensibilidade dos parâmetros estimados sob perturbações no modelo/dados é outra ferramenta importante em modelos de regressão. Se essas perturbações causarem efeitos desproporcionais em determinados resultados do modelo, pode ser indício de que este está mal ajustado ou existem afastamentos importantes das suposições feitas para o mesmo. A identificação das observações responsáveis por essas discrepâncias pode ajudar na escolha de um modelo mais adequado aos dados. Excluir casos (veja, por exemplo, [Cook e Weisberg \(1982\)](#)) e influência local ([Cook, 1986](#)) são os procedimentos mais populares. Mas, apagar observações não é usual em modelos de regressão com erros em série temporal, então considera-se nesta seção somente o método de influência local, isto é, a avaliação de pequenas perturbações na estimação do modelo (4.1)-(4.2).

A medida de influência mais conhecida introduzida por [Cook \(1986\)](#) é avaliar a influência de pequenas perturbações nos dados ou no modelo utilizando o afastamento da função de verossimilhança definida por

$$LD(\omega) = 2 \left\{ L_p(\hat{\theta}, \lambda) - L_p(\hat{\theta}, \lambda \mid \omega) \right\},$$

em que $L_p(\hat{\theta}, \lambda)$ é o logaritmo da função de verossimilhança penalizada do modelo postulado, $L_p(\hat{\theta}, \lambda \mid \omega)$ é o logaritmo da função de verossimilhança penalizada do modelo perturbado pelo vetor de perturbações aplicadas no modelo. Tem-se que $L_p(\theta, \lambda \mid \omega_0) = L_p(\theta, \lambda)$ e conseqüentemente $LD(\omega_0) = 0$. Logo, desde que $LD(\omega) \geq 0$, ω_0 é um ponto de mínimo da função $LD(\omega)$. Em geral, o vetor ω é introduzido no modelo via $\omega = \omega_0 + a\ell$, em que ℓ , $\omega_0 \in \mathcal{R}^n$ e a medem a magnitude da perturbação na direção ℓ . Assumimos que existe um ω_0 tal que $L_p(\theta, \lambda \mid \omega_0) = L_p(\theta, \lambda)$ para todo θ . A proposta de [Cook \(1986\)](#) é estudar o comportamento de $LD(\omega)$, ou de alguma outra medida de influência, em torno do vetor de não-perturbação ω_0 e mostrar que a curvatura normal \mathbf{C}_ℓ de $LD(\omega)$ em ω_0 na direção de algum vetor unitário ℓ é dada por $\mathbf{C}_\ell = \mathbf{C}_\ell(\theta) = 2 \mid \ell^\top \Delta^\top (-\ddot{\mathbf{L}}_p)^{-1} \Delta \ell \mid$ com $\|\ell\| = 1$, em que $\ddot{\mathbf{L}}_p$ é a matriz de informação observada penalizada (veja os Apêndices A e B), Δ é uma matriz dada pelas derivadas de cada parâmetro ou função não paramétrica, ou seja, $\Delta = \partial^2 L_p(\theta, \lambda \mid \omega) / \partial \theta \partial \omega^\top$ avaliada em $\theta = \hat{\theta}$ e em $\omega = \omega_0$.

O gráfico dos elementos $\mid \ell_{max} \mid$ versus i (ordem dos dados) pode revelar que tipo de perturbação tem mais influência sobre $LD(\omega)$, na vizinhança de ω_0 ([Cook, 1986](#)). Ou, pode-se adotar o gráfico de índice de $C_i = 2 \mid c_{ii} \mid$ para avaliar a presença de observações influentes, em que c_{ii} forma a diagonal principal da matriz $\mathbf{C} = \Delta^\top (-\ddot{\mathbf{L}}_p)^{-1} \Delta$.

Vamos derivar a matriz Δ para avaliarmos a influência local das observações em cinco esquemas de perturbação diferentes. Sendo assim, é possível encontrar as matrizes de curvatura normal que serão diagonalizadas para encontrarmos o vetor ℓ_{max} o qual representa a maior curvatura da matriz $\mathbf{C}_\ell(\theta)$ em que os valores do vetor são interpretados conforme cada observação muda as inferências sobre pequenas alterações no modelo ou nos dados. Os esquemas de perturbação aplicadas ao modelo (4.1) e (4.2) podem ser resumidamente utilizados no modelo (3.1) e (3.2). O primeiro considera que cada observação possui um peso diferente no processo de estimação. O segundo pondera a perturbação no parâmetro de dispersão. O terceiro mensura que a variável resposta é deslocada de um valor determinado. O quarto avalia que a variável explicativa contínua na parte paramétrica seja deslocada por um valor. O quinto, e último, quantifica uma modificação na estrutura de autocorrelação dos dados.

Como \mathbf{C}_ℓ não é invariável sob mudança uniforme de escala, [Poon e Poon \(1999\)](#) propuseram a curvatura normal conformal para θ na direção unitária $\|\ell\| = 1$, a qual é definida por

$$\mathbf{C}_\ell(\theta) = \ell^\top \mathbf{C} \ell / \sqrt{\text{tr}(\mathbf{C}^2)},$$

em que $0 \leq \mathbf{C}_\ell(\theta) \leq 1$, $\mathbf{C} = \Delta^\top \{(-\ddot{\mathbf{L}}_p^{\hat{\theta}\hat{\theta}})^{-1}\} \Delta$ com $\ddot{\mathbf{L}}_p^{\hat{\theta}\hat{\theta}}$ sendo a matriz de informação observada (veja os apêndices A e B), $\text{tr}(\mathbf{C}^2) = \sum_{i=1}^n \alpha_i^2$, em que α são os autovalores de \mathbf{C} e $\Delta = (\Delta_0^\top, \Delta_1^\top, \dots, \Delta_k^\top, \Delta_{k+1}^\top, \Delta_{k+2}^\top, \dots, \Delta_{k+p+1}^\top)^\top$ é uma matriz $(r_0 + r_1 + \dots + r_k + p + 1 \times n)$ com suas quantidades calculadas no Apêndice C para cada tipo de esquema de perturbação. Tais

formas matriciais facilitam o desenvolvimento de códigos computacionais para avaliar as curvaturas normal conformal.

Vários gráficos de diagnóstico podem ser derivados. Por exemplo, denotando os autovalores normalizados como $\hat{\alpha}_{\max} = \hat{\alpha}_1 \geq \dots \geq \hat{\alpha}_k \geq q/\sqrt{n} > \hat{\alpha}_{k+1} \dots \hat{\alpha}_n \geq 0$, uma medida de influência agregada para todo q autovetor influente é definida como

$$m(q)_i = \sqrt{\sum_{j=1}^k \hat{\alpha}_j e_{ji}^2},$$

em que e_{ji} denota o i -ésimo elemento do j -ésimo autovetor \mathbf{e}_j , para $i = 1, \dots, n$, $j = 1, \dots, k$ e $q = 0, 1, 2, \dots$. O gráfico índice de $m(q)_i$ é sugerido para revelar aquelas observações que são q -influentes, isto é, influentes para todos os autovetores, de modo que $C_{e_j} \geq q/\sqrt{n}$.

Em particular, se o interesse é avaliar a curvatura normal conformal na direção da i -ésima observação, representada pelo vetor $(n \times 1)$ \mathbf{d}_i formada por zeros e com 1 na i -ésima posição, temos que

$$C_{d_i} = C_i = m^2(0)_i = \sum_{j=1}^n \hat{\alpha}_j e_{ji}^2,$$

que corresponde ao quadrado da contribuição total dos autovetores ortonormais. Um critério de corte sugerido por Lee e Xu (2004) considera como mais influentes as observações dadas por $\mathbf{C}_i > \bar{\mathbf{C}} + c\text{DP}(\mathbf{C})$, em que $\bar{\mathbf{C}}$ e $\text{DP}(\mathbf{C})$ denotam, respectivamente, a média e o desvio padrão de \mathbf{C} ($\mathbf{C}_i, i = 1, \dots, n$) e c é selecionado apropriadamente.

Para uma partição particular $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ com $\boldsymbol{\theta}_1$ definido, por exemplo, como $\boldsymbol{\theta}_1 = \boldsymbol{\gamma}$, $\theta_1 = \phi$ ou $\theta_1 = \rho_1$, pode-se avaliar a curvatura normal conformal por

$$\mathbf{C}_\ell(\boldsymbol{\theta}_1) = \boldsymbol{\ell}^\top \mathbf{C}_1 \boldsymbol{\ell} / \sqrt{\text{tr}(\mathbf{C}_1^2)},$$

em que $0 \leq \mathbf{C}_\ell(\boldsymbol{\theta}_1) \leq 1$, $\mathbf{C}_1 = \boldsymbol{\Delta}^\top \{(-\ddot{\mathbf{L}}_p^{\hat{\boldsymbol{\theta}}})^{-1} - \mathbf{G}_p^{\hat{\boldsymbol{\theta}}_2 \hat{\boldsymbol{\theta}}_2}\} \boldsymbol{\Delta}$ e $\mathbf{G}_p^{\hat{\boldsymbol{\theta}}_2 \hat{\boldsymbol{\theta}}_2} = \text{blockdiag} \{ \mathbf{0}, (\ddot{\mathbf{L}}_p^{\hat{\boldsymbol{\theta}}_2})^{-1} \}$. Novamente, pode-se representar o gráfico índice de $\mathbf{C}_i(\boldsymbol{\theta}_1)$ para avaliar a influência das observações em $\hat{\boldsymbol{\theta}}_1$.

Vários trabalhos têm sido desenvolvidos na derivação de curvaturas de influência local em modelos lineares com erros simétricos. Por exemplo, Galea *et al.* (2003) derivaram tais curvaturas em modelos lineares elípticos univariados, enquanto Osorio *et al.* (2007) e Ibacache-Pulgar *et al.* (2012) derivaram as curvaturas de influência local em modelos lineares elípticos paramétricos e semiparamétricos mistos, respectivamente. Galea *et al.* (2005) desenvolveram curvaturas para os modelos não lineares com erros simétricos. Da mesma forma, Cao *et al.* (2010) e Relvas e Paula (2016) obtiveram os gráficos de influência local para modelos com erros simétricos univariados AR(1). Recentemente, Borssoi *et al.* (2020) aplicaram a metodologia em modelos mistos elípticos com erros de medida em uma covariável.

5.2.1 Perturbação ponderação de casos

Considere que cada observação possui um peso diferente no processo de estimação, dessa forma, pode-se avaliar a contribuição individual de cada uma. O logaritmo da função de verossimilhança penalizada fica dada por

$$L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \boldsymbol{\omega}) = \sum_{i=1}^n \omega_i L_i(\boldsymbol{\theta}) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k$$

em que $L_i(\boldsymbol{\theta}) = -\frac{1}{2} \log(\phi) + \log(g(\delta_i))$ e $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)$ é o vetor de pesos, com $0 < \omega_i < 1$. Neste caso, o vetor de não perturbação é $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$. Com os cálculos do Apêndice C, tem-se os elementos da matriz $\boldsymbol{\Delta}$ avaliados em $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$ expressos da seguinte maneira:

$$\begin{aligned}\frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \gamma_l} &= \frac{1}{\phi} (\mathbf{A} \mathbf{N}_l)^\top \mathbf{D}(v) \mathbf{D}(\mathbf{A}_\epsilon), \quad l = 0, \dots, k, \\ \frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \phi} &= -\frac{\mathbf{1}_n^\top}{2\phi} - \frac{\mathbf{1}_n^\top \mathbf{D}(m)}{2\phi} \quad \text{e} \\ \frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \rho_j} &= -\frac{1}{\phi} (\mathbf{B}_j \boldsymbol{\epsilon})^\top \mathbf{D}(v) \mathbf{D}(\mathbf{A}_\epsilon), \quad j = 1, \dots, p,\end{aligned}$$

em que $\mathbf{1}_n = (1, \dots, 1)^\top$ e $\mathbf{D}(\hat{\mathbf{A}}_\epsilon)$ é uma matriz diagonal com elementos dados por $\hat{\mathbf{A}}_\epsilon$. As matrizes devem ser avaliadas em $\hat{\boldsymbol{\theta}}$.

5.2.2 Perturbação no parâmetro de dispersão

Considere como um esquema de perturbação, a mudança no parâmetro de dispersão de ϕ (modelo homoscedástico) para ϕ/ω_i (modelo heterocedástico), em que $0 < \omega_i \leq 1$. A função densidade de probabilidade Y_i assume, neste caso, a forma

$$h_{y_i}(y | \omega_i) = \sqrt{\frac{\omega_i}{\phi}} g(\omega_i \delta_i), \quad (5.1)$$

em que ω_i denota o peso correspondente ao i -ésimo caso, $i = 1, \dots, n$. Quando $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$, o modelo perturbado (5.1) reduz ao modelo postulado (2.6). A matriz $\boldsymbol{\Delta}$ avaliada em $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$ é dada por

$$\begin{aligned}\frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \gamma_l} &= \frac{1}{\phi} (\mathbf{A} \mathbf{N}_l)^\top [\mathbf{D}(v) - 2\mathbf{D}(d)] \mathbf{D}(\mathbf{A}_\epsilon), \quad l = 0, \dots, k, \\ \frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \phi} &= \frac{\boldsymbol{\delta}^\top}{2\phi} [\mathbf{D}(v) - 2\mathbf{D}(d)] \quad \text{e} \\ \frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \rho_j} &= \frac{1}{\phi} (\mathbf{B}_j \boldsymbol{\epsilon})^\top [2\mathbf{D}(d) - \mathbf{D}(v)] \mathbf{D}(\mathbf{A}_\epsilon), \quad j = 1, \dots, p.\end{aligned}$$

Por exemplo, os gráficos índices de $\mathbf{C}_\ell(\boldsymbol{\gamma}_0)$ e $\mathbf{C}_\ell(\phi)$ podem ser úteis na escolha de um modelo de regressão linear simétrico robusto contra estruturas heterocedásticas.

5.2.3 Perturbação na variável resposta

Considere a i -ésima resposta perturbada como $y_{\omega_i} = y_i + \omega_i$, em que $\omega_i \in \mathcal{R}$ e $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$. Dessa maneira, é possível avaliar a sensibilidade das estimativas quando há mudanças nas respostas das observações. Esse esquema de perturbação leva a

$$\begin{aligned}\epsilon_{\omega_i} &= y_i + \omega_i - \eta_{0_i}^\top \gamma_0 - \eta_{1_i}^\top \gamma_1 - \dots - \eta_{k_i}^\top \gamma_k, \\ \delta_{\omega_i} &= \frac{(\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}})^2}{\phi}.\end{aligned}$$

Os cálculos para a obtenção da matriz $\boldsymbol{\Delta}$ estão presentes no Apêndice C. Essa pode ser decomposta em

$$\left[\left(\frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \gamma_0} \right), \left(\frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \gamma_1} \right), \dots, \left(\frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \gamma_k} \right), \left(\frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \phi} \right), \left(\frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \rho_1} \right), \dots, \left(\frac{\partial^2 \mathbf{L}_p}{\partial \boldsymbol{\omega} \partial \rho_p} \right), \right]^\top,$$

em que essas medidas avaliadas em $\widehat{\boldsymbol{\theta}}$ e sob o vetor de não perturbação $\boldsymbol{\omega}_0$ são dadas por

$$\begin{aligned}\frac{\partial^2 L_p}{\partial \boldsymbol{\omega} \partial \gamma_l} &= \frac{1}{\phi} (\mathbf{A} \mathbf{N}_l)^\top [\mathbf{D}(v) - 4\mathbf{D}(d)] \mathbf{A}, \quad l = 0, \dots, k, \\ \frac{\partial^2 L_p}{\partial \boldsymbol{\omega} \partial \phi} &= \frac{1}{\phi} (\mathbf{A} \boldsymbol{\epsilon})^\top [\mathbf{D}(v) - 2\mathbf{D}(d)] \mathbf{A} \quad \text{e} \\ \frac{\partial^2 L_p}{\partial \boldsymbol{\omega} \partial \rho_j} &= \frac{1}{\phi} (\mathbf{B}_j \boldsymbol{\epsilon})^\top [4\mathbf{D}(d) - \mathbf{D}(v)] \mathbf{A} - \frac{1}{\phi} (\mathbf{A} \boldsymbol{\epsilon})^\top \mathbf{D}(v) \mathbf{B}_j, \quad j = 1, \dots, p.\end{aligned}$$

5.2.4 Perturbação na variável explicativa

Neste esquema de perturbação, considera-se que a r -ésima variável explicativa contínua seja acrescida de uma constante. Portanto, supõe-se que a r -ésima variável explicativa fique expressa por $\eta_{\omega 0_i}^{(r)} = \eta_{0_i}^{(r)} + \omega_i, \forall i = 1, \dots, n$. Para tal caso, temos as seguintes quantidades:

$$\begin{aligned}\epsilon_{\omega_i} &= y_i - \eta_{0_i}^\top \gamma_0 - \eta_{1_i}^\top \gamma_1 - \dots - \eta_{k_i}^\top \gamma_k - \omega_i \gamma_0^{(r)}, \\ \delta_{\omega_i} &= \frac{(\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}})^2}{\phi}.\end{aligned}$$

De acordo com os cálculos do Apêndice C, tem-se se a matriz $\boldsymbol{\Delta}$ avaliada em $\widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$ com os seguintes elementos:

$$\begin{aligned}\frac{\partial^2 L_p}{\partial \boldsymbol{\omega} \partial \gamma_l} &= \frac{\gamma_l^{(r)}}{\phi} (\mathbf{A} \mathbf{N}_l)^\top [4\mathbf{D}(d) - \mathbf{D}(v)] \mathbf{A}, \quad l = 0, \dots, k, \\ \frac{\partial^2 L_p}{\partial \boldsymbol{\omega} \partial \phi} &= \frac{\gamma_0^{(r)}}{\phi} (\mathbf{A} \boldsymbol{\epsilon})^\top [2\mathbf{D}(d) - \mathbf{D}(v)] \mathbf{A} \quad \text{e} \\ \frac{\partial^2 L_p}{\partial \boldsymbol{\omega} \partial \rho_j} &= \frac{\gamma_0^{(r)}}{\phi} (\mathbf{B}_j \boldsymbol{\epsilon})^\top [\mathbf{D}(v) - 4\mathbf{D}(d)] \mathbf{A} - \frac{\gamma_0^{(r)}}{\phi} (\mathbf{A} \boldsymbol{\epsilon})^\top \mathbf{D}(v) \mathbf{B}_j, \quad j = 1, \dots, p.\end{aligned}$$

5.2.5 Perturbação no coeficiente de autocorrelação

Sun (2017) desenvolveu uma abordagem diferente aos esquemas de perturbação que aplicada ao coeficiente de autocorrelação. Neste esquema de perturbação, considere o r -ésimo coeficiente acrescido de uma constante apropriada. Portanto, supõe-se que $\rho_{\omega r_i} = \rho_r + \omega_i, \forall i = 1, \dots, n$ e $r = 1, \dots, p$. Para tal caso tem-se as seguintes quantidades:

$$\begin{aligned}\epsilon_i &= y_i - \eta_{0_i}^\top \gamma_0 - \eta_{1_i}^\top \gamma_1 - \dots - \eta_{k_i}^\top \gamma_k, \\ \delta_{\omega_i} &= \frac{(\epsilon_i - \rho_1 \epsilon_{(i-1)} - \dots - \rho_p \epsilon_{(i-p)} - \omega_i \epsilon_{(i-r)})^2}{\phi}.\end{aligned}$$

De acordo com os cálculos do Apêndice C, tem-se se a matriz $\boldsymbol{\Delta}$ avaliada em $\widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$ com os seguintes elementos:

$$\begin{aligned}\frac{\partial^2 L_p}{\partial \omega \partial \gamma_l} &= \frac{1}{\phi} (\mathbf{A} \mathbf{N}_l)^\top [\mathbf{D}(v) - 4\mathbf{D}(d)] \mathbf{D}(\mathbf{B}_{r\epsilon}) + (\mathbf{B}_r \mathbf{N}_l)^\top \mathbf{D}(v) \mathbf{D}(\mathbf{A}_\epsilon), \quad l = 0, \dots, k, \\ \frac{\partial^2 L_p}{\partial \omega \partial \phi} &= \frac{1}{\phi^2} (\mathbf{A} \epsilon)^\top [\mathbf{D}(v) - 2\mathbf{D}(d)] \mathbf{D}(\mathbf{B}_{r\epsilon}) \quad \text{e} \\ \frac{\partial^2 L_p}{\partial \omega \partial \rho_j} &= \frac{1}{\phi} (\mathbf{B}_j \epsilon)^\top [4\mathbf{D}(d) - \mathbf{D}(v)] \mathbf{D}(\mathbf{B}_{r\epsilon}), \quad j = 1, \dots, p.\end{aligned}$$

em que $\mathbf{D}(\mathbf{B}_{r\epsilon})$ é uma matriz diagonal com elementos dados por $\mathbf{B}_r \epsilon$.

5.3 Seleção de modelos

A seleção de modelos é um problema fundamental em estatística. Suponha que realizamos uma coleção de modelos que refletem uma variedade de estruturas e informações nos dados e a tarefa é selecionar um dentre eles que melhor explique ou ajuste os dados. Nesse sentido, o objetivo da seleção de modelos equivale a selecionar um que se adapte bem aos dados, mas não seja muito complexo. Outro objetivo importante é selecionar modelos que tenham uma boa capacidade preditiva (fora da amostra), ou seja, com uma motivação preditiva explícita.

A classe de modelos simétricos é bastante ampla, contendo distribuições de caudas pesadas e caudas mais leves do que a distribuição normal. Além disso, a classe admite modelos em que todos os parâmetros da distribuição da variável resposta são ajustados em função de variáveis predictoras, e os modelos dessa classe podem ter termos fixos, aleatórios, de alisamento e os coeficientes de autocorrelação. Para a comparação de modelos encaixados e não encaixados nessa classe, pode-se usar o critério de informação de Akaike (AIC) (Akaike, 1973) ou o critério de informação Bayesiano (BIC) (Schwarz, 1978). Dessa forma, podemos selecionar a distribuição para o erro simétrico apropriado no modelo, assim como o parâmetro de suavização $\lambda_1, \dots, \lambda_k$ no processo iterativo descrito nas Seções 3.3.4 e 4.3.4. Ambos os critérios consistem em minimizar as funções

$$\text{AIC}(\boldsymbol{\lambda}) = -2L_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + 2\text{df}(\boldsymbol{\lambda}) \quad \text{e} \quad \text{BIC}(\boldsymbol{\lambda}) = -2L_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + \log(n)\text{df}(\boldsymbol{\lambda}).$$

Além disso, tem-se o critério GAIC, que está disponível no pacote `gam1ss`, para mais detalhes veja Stasinopoulos *et al.* (2015), permite a utilização de penalizações para valores $2 \leq k \leq \log(n)$. Assim, o critério consiste em minimizar a quantidade

$$\text{GAIC}(\boldsymbol{\lambda}) = -2L_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + k\text{df}(\boldsymbol{\lambda}).$$

5.4 Estimação do parâmetro de suavização

Nas seções anteriores, as derivações e inferências para as funções $f(\cdot)$ consideravam o parâmetro de suavização $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^\top$ como fixos. Contudo, esse parâmetro também pode ser estimado. De acordo com Green e Silverman (1994), pode-se escolher esse parâmetro após teste de vários valores para o mesmo conjunto de dados e analisar as curvas obtidas, ou, seguir um método automático de tal forma que os parâmetros de suavização serão obtidos através dos dados.

Wood (2017) mostra que se $\boldsymbol{\lambda}$ for muito alto, os dados serão sobre-suavizados demais e, se for muito baixo, os dados serão sub-suavizados: nos dois casos, isso significa que a estimativa $\hat{f}(\cdot)$ não será próxima da verdadeira função $f(\cdot)$. Idealmente, seria bom escolher $\boldsymbol{\lambda}$ para que $\hat{f}(\cdot)$ fique o mais próximo possível de $f(\cdot)$. Dessa forma, o método de validação cruzada comum é uma abordagem razoável. Se os modelos são julgados apenas por sua capacidade de ajustar os dados a partir dos quais foram estimados, modelos complicados são sempre selecionados em relação aos mais simples.

O método de validação cruzada tem como objetivo estimar o erro de predição de um modelo ajustado aos dados. Esse método utiliza parte dos dados para a construção do modelo e a outra parte

para a sua validação, ou seja, parte dos dados são considerados na avaliação da adequabilidade do ajuste. Segundo [Green e Silverman \(1994\)](#), quando a estimativa da função $f(\cdot)$ é um *spline* cúbico, a expressão de cálculo da validação cruzada é dada por

$$\text{VC}(\boldsymbol{\lambda}) = n^{-1} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{1 - \hat{h}_{ii}(\boldsymbol{\lambda})} \right)^2,$$

em que \hat{y}_i são as estimativas de \mathbf{y} , dado $\boldsymbol{\lambda}$ fixo, e $\hat{h}_{ii}(\boldsymbol{\lambda})$ é o i -ésimo elemento da matriz $\hat{\mathbf{H}}(\boldsymbol{\lambda})$ definida em (3.7) e (4.18). Deve-se escolher $\boldsymbol{\lambda}$ que minimize $\text{VC}(\boldsymbol{\lambda})$.

A validação cruzada generalizada é uma forma modificada da validação cruzada bastante utilizada para a escolha do parâmetro de suavização diferente de VC, basta substituir o respectivo valor h_{ii} de $\mathbf{H}(\boldsymbol{\lambda})$ por sua média $n^{-1} \text{tr} [\mathbf{H}(\boldsymbol{\lambda})]$. É computacionalmente caro calcular $\text{VC}(\boldsymbol{\lambda})$, uma vez que é possível encontrar o traço de $\mathbf{H}(\boldsymbol{\lambda})$ sem necessariamente encontrar todos os valores de cada elemento de h_{ii} :

$$\text{VCG}(\boldsymbol{\lambda}) = \frac{n \|\sqrt{\hat{\mathbf{D}}_v} \hat{\mathbf{A}}(\mathbf{y} - \hat{\boldsymbol{\mu}})\|^2}{\left[n - \text{tr} \{ \hat{\mathbf{H}}(\boldsymbol{\lambda}) \} \right]^2}.$$

Da mesma forma que na validação cruzada, define-se uma grade de valores dos parâmetros de suavização do vetor $\boldsymbol{\lambda}$ e aqueles valores que minimizam a quantidade $\text{VCG}(\boldsymbol{\lambda})$ são escolhidos.

Capítulo 6

O Pacote - APLMS

Esta seção apresenta uma visão geral de como o pacote APLMS, sigla em inglês que se refere a *additive partial linear models with symmetric errors*, está estruturado no *software* R. Esse pacote permite que saídas numéricas e gráficas sejam exibidas para estimação, inferência, simulação e técnicas de diagnóstico para os modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p e *splines* cúbicos penalizados de regressão já descritos anteriormente. Este software deve ser usado com o ambiente R para computação e gráficos estatísticos. Nosso pacote é composto por 7 funções que permitem aos usuários obter informações dos métodos propostos.

O R (Thaka e Gentleman, 1996) é uma linguagem e um ambiente estatístico que possui facilidades para a manipulação de dados, geração de gráficos e modelagem estatística. É um *software* livre e pode ser copiado e distribuído entre usuários, bem como pode ser instalado em diversos computadores. Os códigos fontes do R estão disponíveis no site <http://r-project.org> onde pode ser baixado e executado, e atualmente são gerenciados por um grupo chamado Core Development Team.

O pacote aqui desenvolvido será disponibilizado no *The Comprehensive R Archive Network* (CRAN) e pode ser facilmente baixado e utilizado nas bases de dados dos pesquisadores. Com a ideia de difundir a modelagem estatística para dados correlacionados ao longo do tempo, desenvolvemos as funções `aplms`, `summary.aplms`, `residuals.aplms`, `influence.aplms` e `simulation.aplms`. A função `aplms` consiste nas rotinas computacionais que permitem a estimação dos parâmetros do modelo para os diferentes tipos de erros simétricos e o ajuste dos parâmetros autorregressivos e da dispersão do modelo MLPAScarp ou em inglês - *APLMScarp* pelo método *backfitting*. A função `summary.aplms` permite gerar o resumo de estatísticas, inferências e teste de hipóteses do modelo ajustado. A função `residuals.aplms` retorna os resíduos quantílicos do modelo, além de realizar o gráfico normal de probabilidades com envelope. A função `influence.aplms` permite acessar a influência local do modelo ajustado. E, por último, a função `simulation.aplms` gera erros simétricos autorregressivos de ordem p .

A seguir, a sintaxe do comando para o ajuste de um modelo APLMS sob uma distribuição simétrica qualquer.

Tabela 6.1: *Resumo das funções no pacote APLMS*

Função	Descrição
<code>aplms</code>	Estimação dos modelos lineares parciais aditivos com erros simétricos condicionais $AR(p)$
<code>family.aplms</code>	Famílias de distribuição em <code>aplms</code>
<code>summary.aplms</code>	Resumo para o objeto <code>aplms</code> ajustado
<code>residuals.aplms</code>	Resíduos do objeto <code>aplms</code>
<code>influence.aplms</code>	Influência local do objeto <code>aplms</code>
<code>simulation.aplms</code>	Simulação de erros simétricos $AR(p)$
<code>confidence.band</code>	Bandas de confiança

Fonte: Elaborada pelo autor (2021)

Detalhes sobre o uso das funções (descritas na Tabela 6.1) podem ser obtidos nas páginas de ajuda correspondentes. Uma chamada da rotina da função `aplms` procede como

aplms	<i>Estimação dos modelos lineares parciais aditivos com erros simétricos condicionais AR(p)</i>
--------------	---

Descrição

`aplms` é usado para ajustar modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p , fornece uma descrição simbólica do preditor aditivo e uma descrição da distribuição dos erros. O comando usa o algoritmo de *backfitting* para combinar diferentes métodos de suavização ou ajuste. Os métodos atualmente suportados são regressão local e *splines* de suavização.

Uso

```
aplms(formula, family=fam(), data, p=1, control, lam)
```

Argumentos

<code>formula</code>	uma expressão da fórmula como em outros modelos de regressão, da forma <code>resposta ~ preditoras</code> . Para mais detalhes, veja a documentação de <code>lm</code> e <code>formula</code> . Os termos de suavização não paramétricos incorporados são indicados por <code>s</code> para suavizar os <i>splines</i> . Veja a documentação para <code>s</code> para seus argumentos. Suavizadores adicionais podem ser adicionados por criar as funções de interface apropriadas. Interações com termos de suavização não paramétricas não são totalmente suportadas, mas não produzirão erros; simplesmente produzirão a interação paramétrica usual.
<code>family</code>	uma descrição da distribuição simétrica do erro a ser usada no modelo. Isso pode ser uma <i>string</i> de caracteres que nomeia uma função, ou o resultado de uma chamada de uma função de família (consulte <code>family</code> para mais detalhes ou funções de família.).
<code>data</code>	um <i>data frame</i> opcional que contém as variáveis no modelo. Se não encontrado em <code>data</code> , as variáveis são obtidas do ambiente <code>formula</code> , geralmente o ambiente do qual o <code>aplms</code> é chamado.
<code>p</code>	indica o número de coeficientes de autocorrelação a serem estimados no processo de ajuste dos dados.
<code>control</code>	uma lista de parâmetros para controlar o processo de ajuste. Consulte a documentação para <code>aplms.control</code> para mais detalhes. Eles também podem ser definidos com um conjunto de argumentos para o próprio <code>aplms</code> .
<code>lam</code>	uma lista de parâmetros de suavização λ caso esteja modelando alguma função de suavização no modelo aditivo. Se haver alguma função de suavização, o padrão será considerado 100. Para três funções, por exemplo, use <code>lam=c(200,10,5)</code> .

Detalhes

O modelo `aplms` é ajustado usando o algoritmo de escore local, com ajustes iterativos do modelo semiparamétrico pelo *backfitting*. O algoritmo *backfitting* é o método Gauss-Seidel para ajustar modelos semiparamétricos, suavizando iterativamente os resíduos parciais.

Valores

`aplms` retorna um objeto que pode ser examinado por `print`, `summary` e `plot`. Componentes podem ser extraídos usando as funções `extractor` `predict`, `fitted`, `residuals`, `deviance`, `formula`, `family`. Pode ser modificada usando `update`.

Os `residuals`, `fitted`, `values`, `coeficientes` e `effects` podem ser extraídos por funções genéricas de mesmo nome, ao invés do operador "\$". A função `family` retorna o objeto da família inteiro usado no ajuste.

<code>AIC</code>	AIC do modelo ajustado: lembre-se de que os graus de liberdade usados nesses cálculos são os graus de liberdade efetivo do modelo, e a verossimilhança é avaliada no máximo da verossimilhança penalizada.
<code>coefficients</code>	os coeficientes da parte paramétrica do <code>additive.pred</code> multiplica as colunas da matriz do modelo. Os nomes dos coeficientes são os nomes dos efeitos de grau único de liberdade (as colunas da matriz do modelo).
<code>additive.pred</code>	o ajuste aditivo dado pelo produto da matriz modelo e os coeficientes, mais as colunas dos componentes <code>\$smooth</code> .
<code>fitted.values</code>	os valores médios ajustados, obtidos por transformar os componentes <code>additive.pred</code> usando a inversa da função de ligação.
<code>smooth, edf, var</code>	esses três caracterizam os aspectos não paramétricos do ajuste. <code>smooth</code> é uma matriz dos termos de suavizamento, com as colunas correspondente a cada termo do suavizador no modelo; se não há termos no suavizador, todas esses componentes estarão ausentes. Cada coluna corresponde a uma parte não paramétrica estrita do termo, enquanto a parte paramétrica é obtida a partir da matriz do modelo. <code>edf</code> é um vetor dado os graus de liberdade efetivo aproximados para cada coluna do suavizador <code>smooth</code> . <code>var</code> é a matriz contendo a variância aproximada para os suavizadores.
<code>residuals</code>	os resíduos quantílicos a partir do ajuste semiparamétrico do modelo.
<code>iter</code>	o número de iterações locais usadas para obter as estimativas.
<code>control</code>	a lista de controle <code>aplms</code> usada no ajuste.
<code>converged</code>	indica se o método de ajuste iterativo convergiu ou não.
<code>family</code>	um vetor característico com três elementos que fornece o nome da família, principalmente para fins de impressão.

Autores

Escrito por Rodrigo Oliveira e Gilberto Paula, seguindo exclusivamente o trabalho realizado nesta tese, e a filosofia em [Wood \(2017\)](#) do pacote `mgcv`.

Exemplos

```
library(aplms)
library(mgcv)
data(SFranc)
fit<-aplms(temp~s(day.of.year,bs="cc",k=20)+s(time,bs="cr"),
           data=SFranc,p=2)
summary(fit)
plot(fit)
```

family.aplms *Famílias de distribuição em aplms*

Descrição

Referem-se às famílias das distribuições simétricas estudadas no Capítulo 2 que podem ser usadas na função `aplms`. Também é possível acrescentar outras distribuições simétricas, desde que as mesmas quantidades das tabelas 2.2 e 2.3, e as funções, densidade de probabilidade e a acumulada, sejam fornecidas.

Detalhes

As seguintes famílias pertencem à família simétrica de distribuições. Elas são úteis em todas as funções de modelagem.

- **Normal**: `family=Normal()`;
- **Normal Generalizada**: `family=GNormal(ν_1)`;
- **t-Student**: `family=Stud(ν_1)`;
- **t-Student Generalizada**: `family=GStud(ν_1, ν_2)`;
- **Logística I**: `family=LogisI()`;
- **Logística II**: `family=LogisII()`;
- **Logística Generalizada**: `family=GLogis(α, ν_1)`;
- **Exponencial Potência**: `family=Powerexp(ν_1)`.

summary.aplms *Resumo para o objeto aplms ajustado*

Descrição

Dado um objeto ajustado por `aplms()` e produz vários resultados a partir dele.

Uso

```
summary.aplms(model, test=TRUE)
```

Argumentos

<code>model</code>	um objeto ajustado produzido por <code>aplms()</code>
<code>test</code>	por padrão os testes para os termos dos parâmetros estimados são calculados e apresentados. Caso <code>test=FALSE</code> , esses resultados são omitidos.

Detalhes

Os graus de liberdade efetivos do modelo são tomados como o traço discutido nos Capítulos 3 e 4. Outra alternativa de encontrar os graus de liberdade efetivos do modelo são: para o caso paramétrico, basta calcular o traço da matriz de planejamento $\text{tr}(\mathbf{ANS}_0)$, espera-se que seja aproximadamente o número de covariáveis no modelo r_0 , já para o caso não paramétrico calculamos $\text{tr}(\mathbf{ANS}_k)$.

Por padrão os valores- P para os termos paramétricos do modelo são baseados no teste de Wald usando a matriz de covariância dos coeficientes.

Valores

<code>summary.aplms</code>	produz uma lista de informações resultante de um objeto <code>aplms</code> ajustado.
<code>p.coeff</code>	é um array das estimativas dos coeficientes estritamente paramétricos do modelo.
<code>p.err</code>	é um array dos <code>p.coeff</code> dividido pelos seus erros padrão.
<code>p.pv</code>	é um array dos valores- P para a hipótese nula de que o parâmetro correspondente é zero. Calculado com referência à distribuição t , com os graus de liberdade efetivos estimados para o ajuste do modelo, se o parâmetro de dispersão tiver sido estimado, e a distribuição normal padrão, se não.
<code>k</code>	o número de termos de suavização no modelo.
<code>s.coeff</code>	um array das estimativas das funções não paramétricas suavizadas.
<code>v.err</code>	é um array dos <code>v.coeff</code> dividido pelos seus erros padrão.
<code>p.pv</code>	é um array dos valores- p aproximados para a hipótese nula de que cada termo do suavizador é zero. Essas quantidades são apenas aproximações.
<code>se</code>	é um array dos erros padrão estimados para todos parâmetros estimados.
<code>edf</code>	graus de liberdade efetivos do modelo.
<code>residual.edf</code>	estimação dos graus de liberdade dos resíduos.
<code>n</code>	número de dados.
<code>np</code>	número de coeficientes do modelo (coeficientes de regressão, sem os parâmetros de suavização ou outros parâmetros da verossimilhança).
<code>sp.criterion</code>	o valor minimizado do critério de seleção de suavidade.
<code>family</code>	a família em uso.
<code>formula</code>	a fórmula original <code>aplms</code> .
<code>dispersion</code>	o parâmetro de escala estimado.
<code>data</code>	os dados originais fornecidos no ajuste do modelo.

Exemplos

```
library(aplms)
library(mgcv)
data(SFranc)
fit<-aplms(temp~s(day.of.year,bs="cc",k=20)+s(time,bs="cr"),
            data=SFranc,p=2)
summary(fit)
plot(fit)
```

residuals.aplms *Resíduos do objeto aplms*

Descrição

Retorna os resíduos para o objeto do modelo `aplms` ajustado. Está disponível os resíduos Quantílicos e de Pearson.

Uso

```
residuals.aplms(model,type=c("Quant","Pearson"),plot=NULL)
```

Argumentos

<code>model</code>	um objeto ajustado produzido por <code>aplms()</code> .
<code>type</code>	o tipo de resíduos desejados. Usualmente um dos "Quant", "Pearson".
<code>plot</code>	o tipo de gráfico desejado. O valor padrão é <code>NULL</code> , ou seja, nenhum gráfico será mostrado. Usualmente um dos seguintes comandos pode ser escolhido, "index", "QQ".

Detalhes

Os resíduos de Pearson são os resíduos brutos divididos pelo desvio padrão dos dados, de acordo com a relação de variância média do modelo e o parâmetro de escala estimado. Os resíduos quantílicos é uma transformação que segue distribuição normal padrão. Caso deseja-se imprimir os gráficos dos resíduos, é possível obter os gráficos índices aos quais se resumem no *plot* dos resíduos ao longo do tempo, além de verificar se há heteroscedasticidade dos dados. Ou, pode-se realizar o gráfico Q-Q normal, em que é possível verificar se as suposições do modelo são atendidas, simetria, ou se é necessário usar distribuições com cauda mais pesadas ou mais leves.

Valores

Um vetor de resíduos.

Exemplos

```
library(aplms)
library(mgcv)
data(SFranc)
fit<-aplms(temp~s(day.of.year,bs="cc",k=20)+s(time,bs="cr"),
           data=SFranc,p=2)
residuals.aplms(fit,type="Quant",plot="QQ")
```

influence.aplms *Influência local do objeto aplms*

Descrição

Toma um objeto `aplms` ajustado e produz algumas informações de diagnóstico sobre o procedimento e resultados de ajuste. Retorna a curvatura conformal normal do objeto do modelo `aplms` ajustado. Estão disponíveis as perturbações `Weight-Case`, `Dispersion`, `Response`, `Explanatory` e `CorAR`.

Uso

```
influence.aplms(model,
               pert=c("Weight", "Dispersion", "Response",
                     "Explanatory", "CorAR"),
               param=NULL, ident=NULL, p, plot=FALSE)
```

Argumentos

<code>model</code>	um objeto ajustado produzido por <code>aplms()</code> .
<code>pert</code>	o tipo de perturbação desejada. Usualmente pode-se decidir em escolher as seguintes perturbações: "Weight", "Dispersion", "Response", "Explanatory" e "CorAR".

<code>param</code>	caso a perturbação escolhida seja "Explanatory", ou seja, envolva uma das variáveis explicativas paramétricas, é necessário identificar qual delas estará sob análise. Nesse caso indicar o nome da variável conforme está no banco de dados.
<code>ident</code>	indica a quantidade de pontos influentes a serem considerados ao apresentar o gráfico.
<code>p</code>	indicar qual coeficiente de autocorrelação está sob análise.
<code>plot</code>	por padrão <code>plot</code> é <code>FALSE</code> , caso deseje-se retorno do gráfico para a influência local coloque <code>plot=TRUE</code> .

Detalhes

Essa função fornece os pontos influentes do conjunto de dados em análise conforme o tipo de perturbação desejada. O gráfico dos pontos influentes pode ser solicitado desde que o argumento `plot` seja `TRUE`.

Valores

Um vetor relacionado à influência local do modelo ajustado aos dados.

Exemplos

```
library(aplms)
library(mgcv)
data(SFranc)
fit<-aplms(temp~s(day.of.year,bs="cc",k=20)+s(time,bs="cr"),
            data=SFranc,p=2)
influence.aplms(fit,perturbation="Weight",ident=7)
```

simulation.aplms *Simulação de erros simétricos AR(p)*

Descrição

Um simples conjunto de dados com erros simétricos autorregressivos de ordem p .

Uso

```
simulation.aplms(n,family=Normal(),dispersion=1,corAR)
```

Argumentos

<code>n</code>	o tamanho do vetor a ser simulado.
<code>family</code>	a família desejada para a distribuição do erro simétrico. Para mais detalhes, veja family.aplms .
<code>dispersion</code>	o parâmetro de dispersão para a distribuição do erro.
<code>corAR</code>	o vetor dos coeficientes de autocorrelação.

Detalhes

Esse conjunto de dados é artificial, e é usado para estudos de comportamento dos modelos ajustados.

Valores

Um vetor dos erros da família de distribuições simétricas e autorregressivos de ordem p .

Exemplos

```
simulation.aplms(n=1000, family=Powerexp(0.3), corAR=c(0.635, -0.32))
```

confidence.band *Bandas de confiança para o modelo ajustado*

Descrição

Função para extrair as bandas de confiança das variáveis do modelo `aplms()` ajustado.

Uso

```
confidence.band(model, variable=NULL, all.data=FALSE, points=FALSE)
```

Argumentos

<code>model</code>	um modelo ajustado do objeto de classe <code>aplms</code> .
<code>variable</code>	por padrão, esse argumento é NULL , nesse caso, a função retornará o gráfico das bandas de confiança de todos as variáveis no modelo. Caso, deseja-se uma variável no modelo ajustado, basta indicá-la.
<code>all.data</code>	por padrão, esse argumento é FALSE , nesse caso, a função retornará o gráfico das bandas de confiança de todos as variáveis no modelo. Se, all.data=TRUE , tem-se a banda de confiança para todo o conjunto de dados. Gráfico de superfície está disponível nesse caso. E, variable será considerada como NULL .
<code>points</code>	por padrão, esse argumento é FALSE , assim os gráficos não retornarão os pontos dos dados. Se points=TRUE tem-se a inserção dos pontos no gráfico.

Detalhes

A função retorna um ou mais gráficos para as bandas de confiança do modelo `aplms` ajustado. Quando tem-se mais de uma variável paramétrica, ou uma paramétrica e uma função de suavização, é possível obter uma superfície de resposta para o modelo ajustado.

Exemplos

```
library(aplms)
library(mgcv)
data(SFranc)
fit<-aplms(temp~s(day.of.year, bs="cc", k=20)+s(time, bs="cr"),
           data=SFranc, p=2)
confidence.band(fit, all.data=TRUE, points=FALSE)
```

Capítulo 7

Simulações

A simulação de dados é uma técnica fundamental na programação e pesquisas estatísticas, além de eficaz para desenvolver o conhecimento e as habilidades em torno dos métodos propostos. Sempre que um novo método estatístico é desenvolvido ou usado, há premissas que precisam ser testadas e confirmadas, portanto, gera-se dados artificiais para testá-las. As vantagens dos dados simulados se devem ao baixo custo em gerar valores aleatórios em vez de obter dados coletados. Ao simular cenários que se alinham ou até mesmo se destoam com os métodos propostos em estudo, é possível detectar quais extraem informações mais robustas e eficientes no conjunto de dados reais.

Há situações em que um conjunto de dados é apenas aproximadamente normal, nesse caso, é bem provável que as estimações de parâmetros, inferência estatística e métodos de diagnóstico não sejam adequados. Uma maneira de entender esses fenômenos é simular dados de populações não normais, ou melhor, simular diversos cenários com distribuições diferentes e adversidades que podem acometer os métodos propostos. A simulação fornece controle completo sobre as características do modelo da população a partir do qual os dados (simulados) são obtidos.

Neste capítulo será apresentado um estudo comparativo entre os modelos propostos por meio de simulações de Monte Carlo. O objetivo é avaliar o desempenho e o comportamento dos estimadores de máxima verossimilhança penalizada para os parâmetros autorregressivos (ρ 's), a dispersão ϕ , a parte paramétrica e as funções de suavização, no processo iterativo desenvolvido nos Capítulos 3 e 4, nos modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p . Esses, são baseados em amostras de tamanho finito e sem observações atípicas.

Estimativas de máxima verossimilhança de modelos simétricos com distribuição dos erros de caudas mais pesadas que a normal tendem a ser menos sensíveis a observações extremas. Dessa forma, estamos interessados em investigar a robustez dessas estimativas nesses casos. A fim de esclarecer este ponto para alguns modelos simétricos, realizamos 1000 réplicas de Monte Carlo com conjunto de dados de tamanho $n = 100, 300$ e 500 .

Todas as rotinas computacionais foram desenvolvidas durante esta pesquisa e organizadas em uma biblioteca denominada `aplms` utilizando a linguagem de programação R em sua versão 4.0.3. Detalhes da sintaxe de comandos para ajuste de um modelo MLPAS condicionais AR(p)-APLMS utilizando o pacote `aplms` foram apresentados no capítulo 6. Além disso, todas as rotinas escritas estão disponíveis mediante solicitação.

Os critérios abaixo foram utilizados para avaliar e comparar o desempenho dos estimadores de máxima verossimilhança penalizada dos parâmetros γ_0 , ϕ e ρ 's. Com o intuito de generalizar e facilitar a notação, consideremos $\hat{\theta}_i$ como substituto para $\hat{\gamma}_{0,i}$, $\hat{\phi}_i$ e $\hat{\rho}_i$, ou seja, $\hat{\theta}_i$ é o valor estimado do respectivo parâmetro θ na i -ésima réplica de Monte Carlo.

1. Média dos valores estimados:

$$\bar{\hat{\theta}} = \frac{\sum_{i=1}^{1000} \hat{\theta}_i}{1000}.$$

2. Viés médio dos valores estimados de $\hat{\theta}$:

$$\widehat{\text{Viés}}(\hat{\theta}) = \frac{\sum_{i=1}^{1000} (\hat{\theta}_i - \theta)}{1000}.$$

3. Erro quadrático médio (EQM) estimado de $\hat{\theta}$:

$$\widehat{\text{EQM}}(\hat{\theta}) = \frac{\sum_{i=1}^{1000} (\hat{\theta}_i - \theta)^2}{1000}.$$

7.1 Simulação AR(1)

No primeiro momento, o principal objetivo desse estudo de simulação é analisar o comportamento das estimações sob diversos tamanhos amostrais e diferentes distribuições simétricas com erros autorregressivos de primeira ordem AR(1). Na simulação, consideramos duas funções não paramétricas que representam a tendência f_1 e a sazonalidade f_2 e o erro simétrico AR(1) da seguinte maneira:

$$\begin{aligned} y_i &= f_1(t_i) + f_2(s_i) + \epsilon_i, \quad (i = 1, \dots, n), \\ \epsilon_i &= \rho\epsilon_{i-1} + e_i, \quad -1 < \rho < 1, \end{aligned} \quad (7.1)$$

em que os parâmetros assumem os valores $\phi = 1$ e $\rho = -0,75, -0,25, 0,25$ e $0,75$. Para ilustrar e_i , consideramos as distribuições normal, t -Student com $\nu = 3$ graus de liberdade, exponencial potência com parâmetros $\nu = -0,3$ e $\nu = 0,5$, normal generalizada com $\nu = 3$ e Logística-II, nomeadas como NO, t_3 , EP₁, EP₂, NG e L₂, respectivamente, veja a Figura 7.1. Pode-se notar que as distribuições normal generalizada e exponencial potência com parâmetro de forma negativo apresentam caudas mais leves que a distribuição normal, enquanto que as distribuições t -Student, Logística-II e exponencial potência com parâmetro de forma positivo apresenta caudas mais pesadas que a distribuição normal.

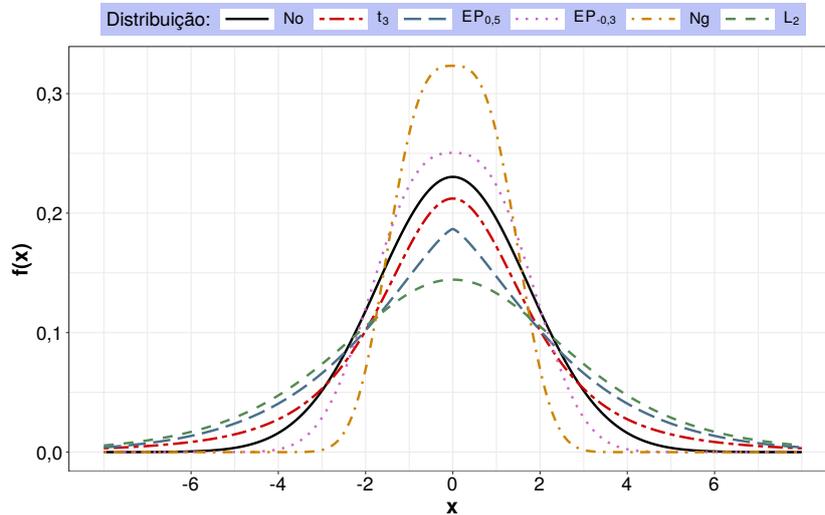


Figura 7.1: Ilustração da distribuição dos erros e_i

Além disso, a função $f_1(t_i) = 2(t_i^3 + \cos(\pi t_i^3))$, em que $t_i = i/n$ e $f_2(i) = 2\text{sen}(i/(\pi n/100))$. As funções f_1 e f_2 são fixas para todas as simulações em cada cenário, veja Figura 7.2.

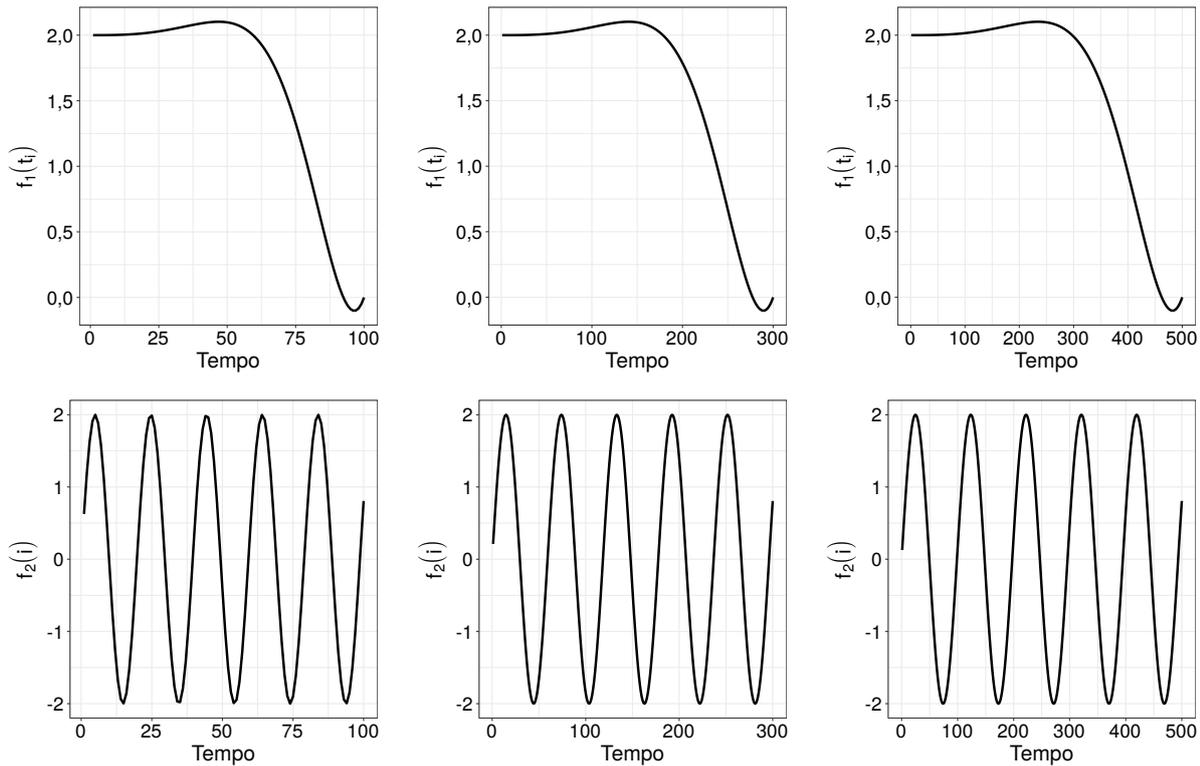


Figura 7.2: Disposição das funções não paramétricas teóricas para a tendência (cima) e sazonalidade (baixo), para os tamanhos amostrais iguais a 100, 300 e 500.

As Tabelas D.1, D.2, D.3, D.4, D.5 e D.6 apresentam os resultados das simulações para cada distribuição do erro que são: normal, t -Student ($\nu = 3$), exponencial potência ($\nu = -0,3$ e $\nu = 0,5$), normal generalizada ($\nu = 3$) e Logística-II, respectivamente. Nas seis tabelas é possível observar que o viés das estimativas dos parâmetros é menor quando adota-se a verdadeira distribuição do erro da qual os dados foram gerados. Além disso, pode-se notar que a precisão das estimativas está diretamente relacionada ao tamanho amostral, quando $n \rightarrow \infty$ tem-se que o viés de $\hat{\rho}$ e $\hat{\phi}$ diminui consideravelmente em média. Além disso, a consistência de $\hat{\phi}$ só é alcançada quando estamos sob a verdadeira distribuição dos erros.

Note que ao aumentarmos o tamanho amostral, o critério AIC tende a selecionar o modelo correto para as verdadeiras distribuições dos erros. Isso não ocorre para as distribuições exponencial potência com $\nu = 0,5$, uma vez que, tendemos a selecionar o modelo com erros Logística-II. E, o modelo com distribuição normal generalizada, já que tendemos a selecionar o modelo exponencial potência com $\nu = -0,3$. Esse erro na seleção dos modelos em relação à verdadeira distribuição dos erros se deve ao fato das proximidade entre as distribuições como pode ser vista pela Figura 7.1.

As Figuras D.1 e D.2 referem-se à média das estimativas para as funções de tendência e sazonalidade, respectivamente, sob a verdadeira distribuição dos erros simulados. Constata-se que de acordo que aumentamos o tamanho amostral a média das estimativas das funções tende à verdadeira função.

7.2 Simulação AR(2)

No segundo estudo de simulações, o principal objetivo é analisar o comportamento das estimações dos modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p maior do que 1. Nessa simulação, consideramos $p = 2$, ou seja, AR(2). Diferentes tamanhos amostrais e distribuições simétricas são analisados. No momento da simulação, consideramos a parte paramétrica f_0 e duas funções não paramétricas representadas por f_1 e f_2 e o erro simétrico AR(2) da seguinte forma:

$$\begin{aligned} y_i &= f_0(x_i) + f_1(t_i) + f_2(s_i) + \epsilon_i, \quad (i = 1, \dots, n), \\ \epsilon_i &= \rho_1 \epsilon_{i-1} + \rho_2 \epsilon_{i-2} + e_i, \end{aligned} \quad (7.2)$$

em que $e_i \sim S(0, \phi)$ e o parâmetro ϕ assume os valores 1 e 3, consideramos as distribuições normal, t -Student com $\nu = 3$ graus de liberdade, exponencial potência com parâmetros $\nu = -0, 3$ e $\nu = 0, 5$, normal generalizada com $\nu = 3$ e logística-II, nomeadas, respectivamente, como NO, t_3 , PE₁, PE₂, NG e L₂. Os coeficientes de correlação são dados por $\rho_1 = 0,656$ e $\rho_2 = -0,309$. Além disso, as funções são definidas da seguinte maneira

$$\begin{aligned} f_0(x_i) &= \beta_0 + \beta_1 x_i, \\ f_1(t_i) &= -\frac{(2t_i)^2}{10} + 1,2 \quad \text{e} \\ f_2(s_i) &= 3\sqrt{(s_i(1-s_i))} \operatorname{sen} \left(\frac{2\pi(1+2^{-3/5})}{s_i+2^{-3/5}} \right) - 0,2, \end{aligned} \quad (7.3)$$

em que $\beta_0 = 2$, $\beta_1 = 7$, $x_i \sim U[-1; 1]$, $t_i = -3 + i6/(n-1)$ com $i = 0, \dots, n-1$ e $s_i = \operatorname{rep}(i, \text{times} = 4)$ com $i = \operatorname{seq}(0, 1, \text{length} = (n/4))$. As funções f_0 , f_1 e f_2 são fixas para todas as simulações em cada cenário, veja a representação dos componentes não paramétricas dadas pela Figura 7.3.

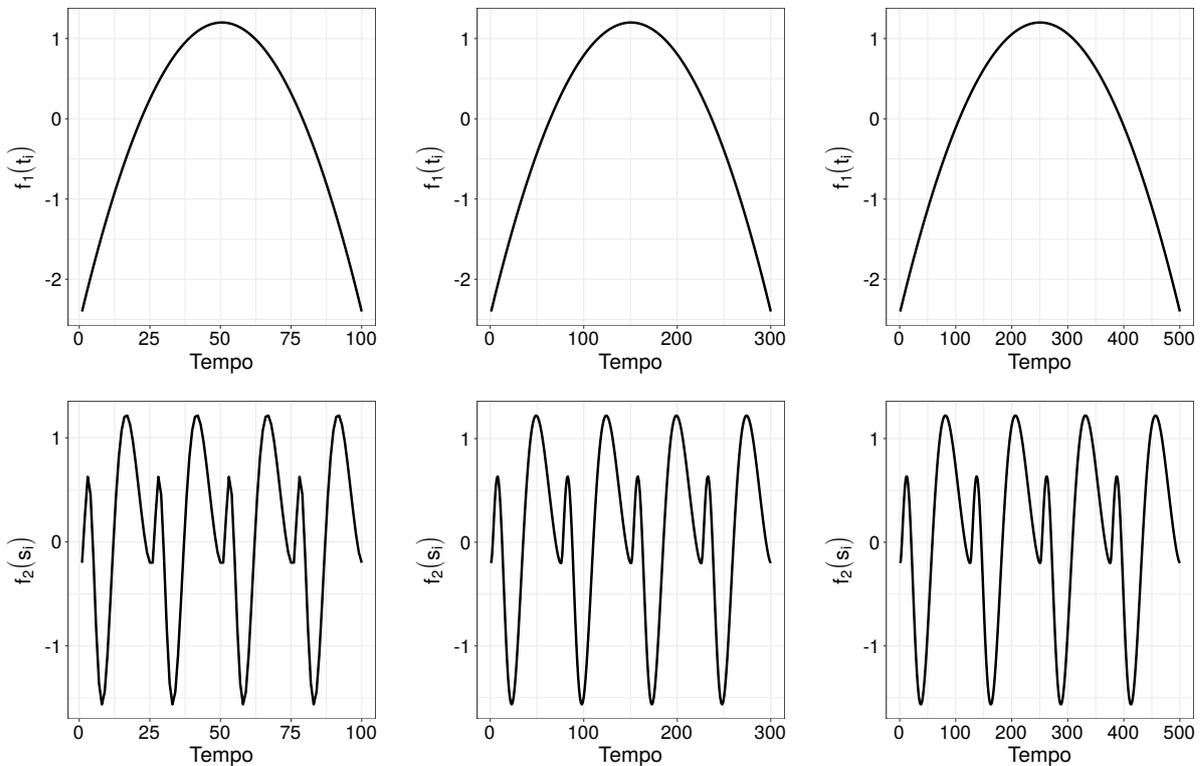


Figura 7.3: Disposição das funções não paramétricas teóricas para os componentes f_1 (cima) e f_2 (baixo), para os tamanhos amostrais iguais a 100, 300 e 500.

A Tabela D.7 representa o percentual de acerto, dado pelo critério AIC, referente a cada distribuição do erro simulado. Note que à medida que aumentamos o tamanho amostral também aumentamos o nosso percentual de acertos para a verdadeira distribuição dos erros. As Tabelas D.8, D.9, D.10, D.11, D.12 e D.13 apresentam os resultados das simulações para cada distribuição do erro que são normal, t -Student ($\nu = 3$), exponencial potência ($\nu = -0,3$ e $\nu = 0,5$), normal generalizada ($\nu = 3$) e logística-II, respectivamente. Nas seis tabelas é possível observar que o Erro Quadrático Médio, entre parênteses, das estimativas dos parâmetros é menor quando adota-se a verdadeira distribuição do erro da qual os dados foram gerados. Além disso, pode-se notar que a precisão das estimativas está diretamente relacionada ao tamanho amostral, quando $n \rightarrow \infty$, tem-se que o EQM de $\hat{\beta}_0$, $\hat{\rho}_1$, $\hat{\rho}_2$ e $\hat{\phi}$ diminuem consideravelmente em média. Quando incrementamos o tamanho amostral de $n = 100$ para $n = 500$, a redução do EQM da estimativa de ρ_1 , por exemplo, para a distribuição do erro normal, passou de 0,0119 para 0,0023, redução de 80,67%, o EQM para $\hat{\rho}_2$ passou de 0,0219 para 0,0023, redução de 89,50%, e o EQM para $\hat{\phi}$ passou de 0,0439 para 0,0053, redução de 87,93%. Em todos os cenários considerados, os parâmetros ρ_1 , ρ_2 , em média, foram subestimados, ou seja, as estimativas fornecidas foram menor do que o verdadeiro valor do parâmetro. Os erros quadráticos médios tendem a zero para $n \rightarrow \infty$ e quando escolhemos a mesma fonte de distribuição dos erros geradora dos dados.

As Figuras D.3 e D.4 referem-se à média das estimativas para as funções f_1 e f_2 , respectivamente, sob a verdadeira distribuição dos erros simulados. Constata-se que de acordo que aumentamos o tamanho amostral a média das estimativas das funções tende à verdadeira função simulada.

Capítulo 8

Aplicações

8.1 Anomalia da temperatura média global na superfície terra-mar

O aquecimento global designa o aumento das temperaturas médias do planeta ao longo dos últimos tempos e é um dos principais problemas que a Terra enfrenta atualmente. Em tese, é causado pelas práticas humanas, embora existam discordâncias quanto a isso no campo científico. Uma das causas desse problema climático que afeta todo o planeta é a intensificação do efeito estufa, fenômeno natural responsável pela manutenção do calor na Terra e que vem apresentando uma maior intensidade em razão da poluição do ar resultante das práticas humanas. A emissão de gases de efeito estufa antrópico é proveniente, principalmente, pelo dióxido de carbono devido à queima de combustíveis fósseis. Além desse, há o metano, compostos orgânicos voláteis, material particulado (MP_{2,5} e MP₁₀, por exemplo) e ozônio.

Devido à mudança da temperatura média global da superfície terrestre, oceânica e na atmosfera, o padrão climático global também sofre alterações. O que resulta em chuvas e secas incomuns em diversos lugares. A consequência mais alarmante dessa tendência de aumento da temperatura global é o derretimento do gelo, que inclui tanto o gelo marinho quanto as camadas de gelo continentais e geleiras. O derretimento do gelo é geralmente considerado o resultado do aumento do nível do mar global, vale salientar que o efeito do derretimento no aumento do nível do mar global não é uniforme.

A tendência de aumento da temperatura foi observada como aumento da temperatura do oceano e da terra. A Figura 8.1 ilustra o aumento da anomalia da temperatura média global na superfície terra-mar de 1880 até recentemente por meio de anomalias de temperatura a partir de medições do ar de superfície em estações meteorológicas e medições de temperatura da superfície do mar em navios e satélites. Este valor foi calculado a partir da anomalia da temperatura média global anual referenciada aos 14°C que é a média estimada de 1951-1980 Hansen *et al.* (2006). Hansen *et al.* (1981) observaram um aumento na temperatura da Terra em cerca de 0,4°C entre 1880 e 1970.

Anomalias positivas mostram que a temperatura observada estava mais quente do que a linha de base, enquanto uma anomalia negativa indica que a temperatura observada estava abaixo da linha de base. Desde a década de 1980, a variação anual da temperatura do século 20 tem sido consistentemente positiva. Em 2020, a anomalia da temperatura da superfície terra-mar era de 0,98°C. As anomalias de temperatura são geralmente mais importantes no estudo das mudanças climáticas do que a temperatura absoluta. Isso ocorre porque, ao calcular as temperaturas absolutas médias, fatores como localização e elevação da estação podem ter impactos críticos nas temperaturas absolutas, mas ser menos significativos nos cálculos de anomalias. A Figura 8.1 ilustra a mudança na temperatura global da superfície em relação às temperaturas médias de 1951-1980. Dezenove dos anos mais quentes ocorreram desde 2000, com exceção de 1998. O ano de 2020 empatou com 2016 para o ano mais quente já registrado desde o início da manutenção de registros em 1880 (fonte: NASA/GISS¹).

¹https://data.giss.nasa.gov/gistemp/graphs/graph_data/Global_Mean_Estimates_based_on_Land_and_Ocean_Data/graph.txt

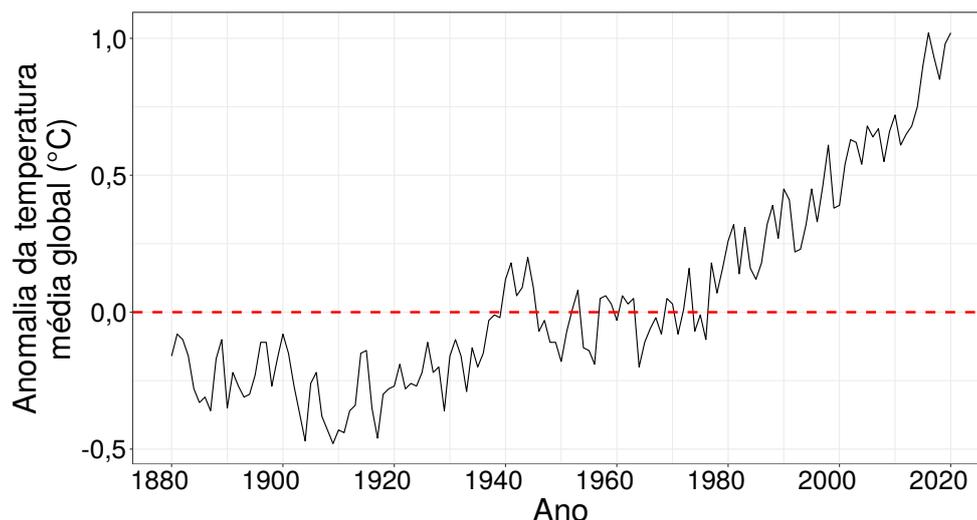


Figura 8.1: Ilustração da série temporal da anomalia da temperatura média global ($^{\circ}\text{C}$) anual na superfície terra-mar.

8.1.1 Modelagem

Estudar a anomalia da temperatura média global anual é importante uma vez que podemos entender as causas e os efeitos catastróficos aos quais os seres humanos podem enfrentar. Dessa forma, para entender melhor o problema, podemos desenvolver um modelo com erros simétricos condicionais autorregressivos e penalização com *splines* cúbico dado da seguinte maneira:

$$\overline{\text{anom}}_i = \beta_0 + f_T(\text{tempo}_i) + \epsilon_i, \quad i = 1, 2, \dots, 141, \quad (8.1)$$

em que β_0 é o intercepto do modelo, $f_T(\cdot)$ é uma função suave aproximada por *spline* de regressão cúbica, enquanto que $\overline{\text{anom}}_i$ indica a anomalia da temperatura média global da superfície terra-mar do i -ésimo ano, tempo_i é o i -ésimo ano e ϵ_i são os erros simétricos autorregressivos de ordem 1, dados por $\epsilon_i = \rho_1 \epsilon_{i-1} + e_i$ com $e_i \sim S(0, \phi)$. Nessa aplicação, consideramos os erros com distribuição normal, normal generalizada e o exponencial potência. Definimos uma série de valores para o parâmetro de suavização e $\lambda_1 = 350$ foi escolhido por minimizar o critério de validação cruzada generalizada. Consideramos 7 nós fixos para o tempo, o qual foi obtido a partir dos quantis da variável tempo.

A Tabela 8.1 nos mostra os resultados da variável anomalia da temperatura média global anual da superfície terra-mar ao longo dos anos seguindo o modelo (8.1) autorregressivo de primeira ordem. Pelo critério AIC, temos que o melhor modelo é dado pela distribuição exponencial potência com $\nu = -0,33$, ou seja, uma distribuição com cauda mais curta que a normal.

Tabela 8.1: Qualidade de ajuste para o modelo (8.1) sob erros com distribuição normal, normal generalizada e exponencial potência ajustados aos dados de anomalia da temperatura média global anual da superfície terra-mar.

Modelo	AIC(λ)	VCG(λ)	df(λ)	df(λ_1)
Normal	-252,8843	0,0097	3,8291	2,8291
Normal generalizada($\nu=2$)	-259,9224	0,0186	5,8381	4,8381
EP($\nu=-0,33$)	-267,5862	0,0151	5,4348	4,4348

O pacote `aplms` desenvolvido em linguagem R e discutido no Capítulo 6 é usado para obter as informações definidas nos Capítulos 3, 4 e 5 a fim de estudar melhor a variável dependente ao longo do tempo. Dessa maneira, o modelo (8.1) pode ser ajustado no R, basta carregar o pacote `aplms` e dispor os comandos da seguinte maneira:

```

> library(aplms)
> ajustel <- aplms(anom~s(tempo,bs="cr",k=7),
                  family=Powerexp(-0.33),p=1,lam=c(350))
> summary(ajustel)

Family: PowerExponential(-0.33)

Link function: identity

Formula:
aplms(anom~s(tempo,bs="cr",k=7),family=Powerexp(-0.33),p=1,lam=c(350))

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.0495    0.0124    4.0020 0.000102 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf    Chisq    p-value
s(tempo)    4.4348 10.25787 0.00098766 ***
--
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure:
              Estimate Std. Error t value Pr(>|t|)
rho1 0.4191    0.0723    5.7993 3.32816e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC = -267.5862 BIC = -251.5602 GCV = 0.0151
Dispersion est. = 0.0140 n= 141

```

Após ajustar o modelo e definir que o erro com distribuição simétrica exponencial potência apresenta melhores resultados, vamos à análise de resíduos e sensibilidade do modelo. A Figura 8.2 descreve o histograma, os gráficos de índices, probabilidade normal, a função de autocorrelação, função de autocorrelação parcial dos resíduos quantílicos condicionais e os pesos contra os resíduos quantílicos condicionais do modelo (8.1) sob erros com distribuição exponencial potência com $\nu = -0,33$ ajustados aos dados. Os gráficos de índices não apresentam tendência nem variabilidade ao longo do tempo, o que indica que a variação do erro parece constante e que a tendência foi controlada. A partir do gráfico de probabilidade normal, é possível identificar que o modelo de erro de cauda mais leve é adequado para ajustar os dados. As funções de autocorrelação (FACs) e funções de autocorrelação parcial (FACPs) dos resíduos condicionais parecem confirmar a suposição de independência. E, por fim, nota-se que a distribuição exponencial potência adota pesos diferentes para as observações.

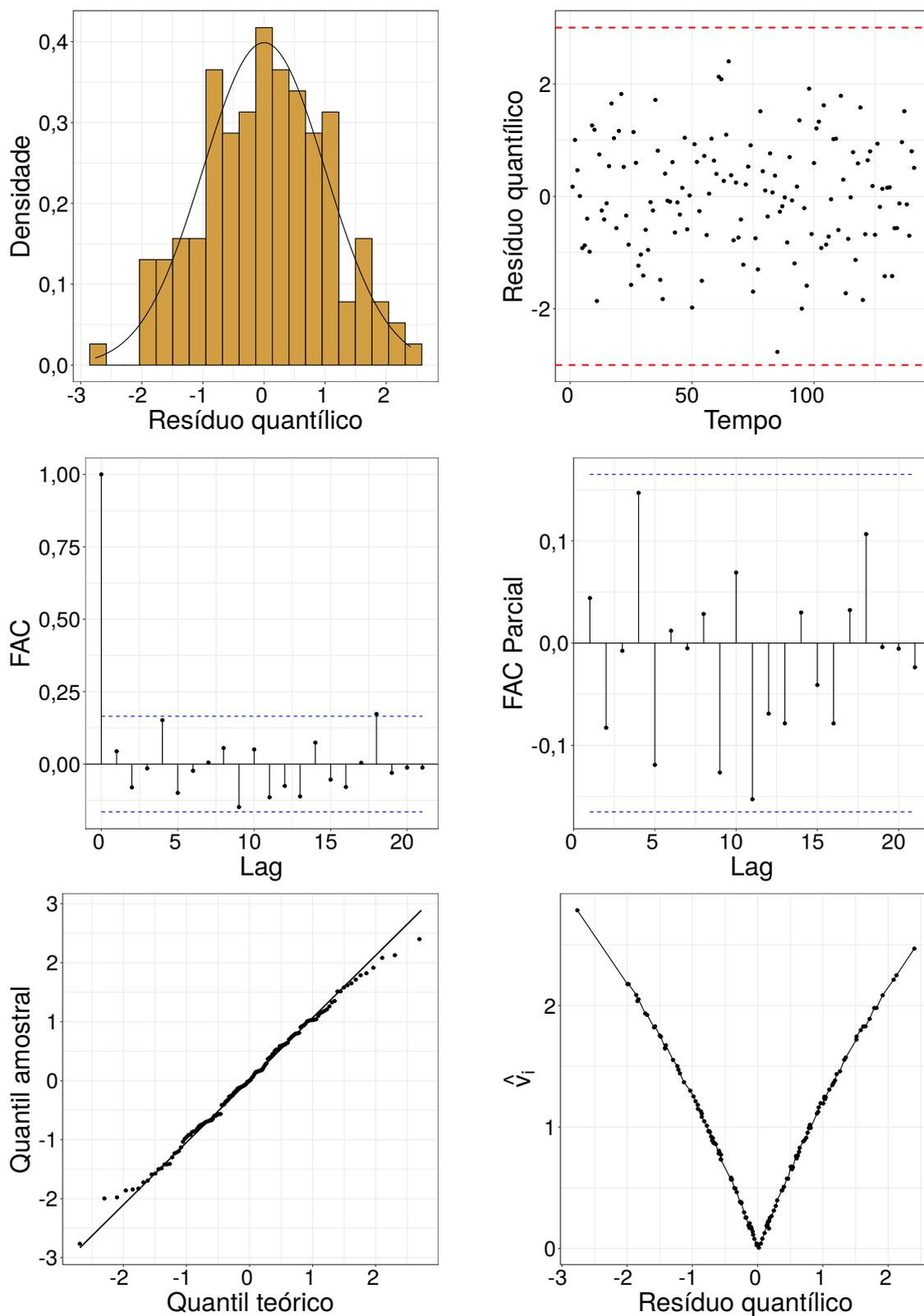


Figura 8.2: Histograma (primeiro-esquerdo), gráfico de índices dos resíduos (primeiro-direito), função de autocorrelação (segundo-esquerdo), função de autocorrelação parcial (segundo-direito), probabilidade normal (terceiro-esquerdo), e peso \hat{v}_i contra os resíduos (terceiro-direito) do modelo ajustado (8.1) para a anomalia da temperatura média global na superfície terrestre-oceânica sob erros com distribuição exponencial potência com $\nu = -0,33$.

E, por último, podemos analisar a curvatura conformal das observações para diferentes esquemas de perturbação. A Figura 8.3 representa os esquemas de perturbação ponderação de casos, do parâmetro de dispersão, da variável resposta e o coeficiente de autocorrelação dado por ρ_1 para o modelo (8.1) definido anteriormente para a anomalia da temperatura média global na superfície terra-mar. A variável em estudo apresenta poucos pontos que possivelmente são influentes e podem alterar as estimativas do estudo. O método de perturbação pelo coeficiente de autocorrelação apresenta o maior grau de influência. Na próxima análise, os valores entre parênteses indicam, o primeiro reflete ao valor observado e o segundo o valor esperado naquele período. O ano de 1941 (0,18; 0,0016) aparenta ser influente com relação aos casos: ponderado e coeficiente de autocorrelação. O ano de 1944 (0,2; 0,0009) demonstrou ser influente em relação aos casos de perturbação: ponderado, parâmetro de dispersão e coeficiente de autocorrelação. E, o ano de 1964 (-0,2; 0,0215) mostrou ser influente quanto aos casos: ponderado, parâmetro de dispersão e na variável resposta.

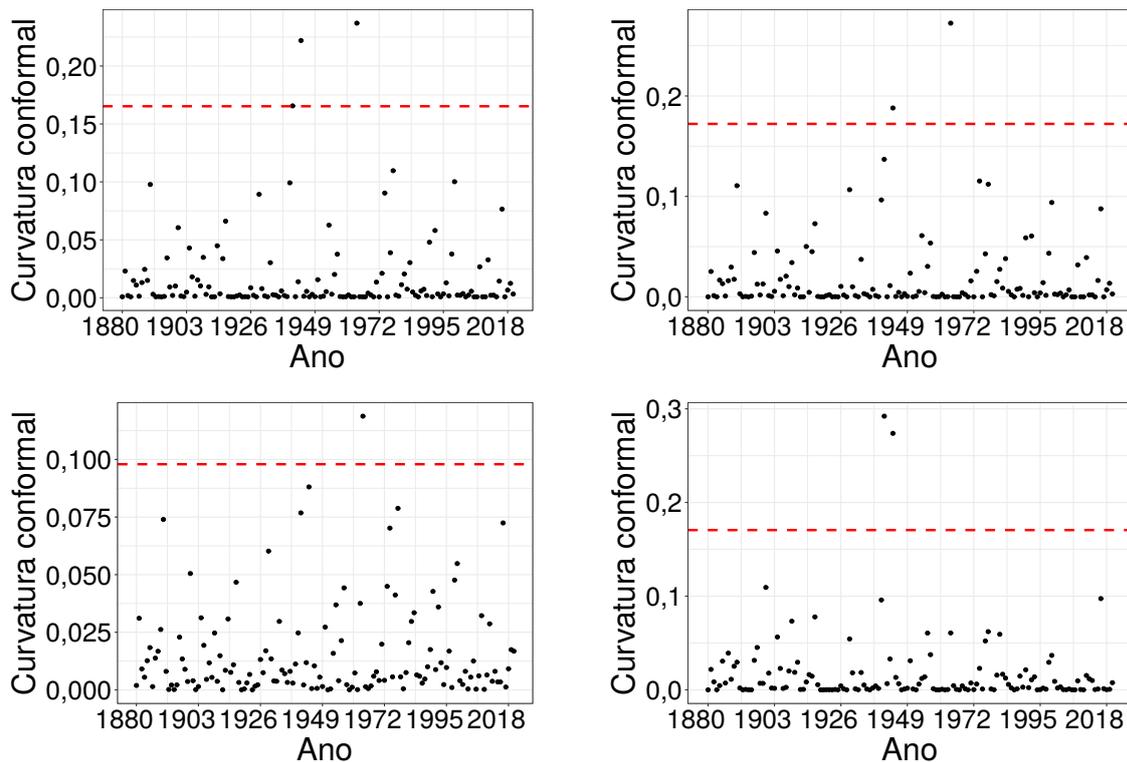


Figura 8.3: Gráficos de índices C_i nos esquemas de perturbação, ponderação de casos (primeiro-esquerdo), parâmetro de dispersão (primeiro-direito), variável resposta (segundo-esquerdo) e coeficiente de autocorrelação (ρ_1) (segundo-direito) do modelo (8.1) para a anomalia da temperatura média global na superfície terra-mar sob erros com distribuição dos erros exponencial potência com $\nu = -0,33$.

A Figura 8.4 mostra a tendência positiva da série histórica em estudo, a anomalia da temperatura média global anual da superfície terra-mar. Esse aumento detectado pode trazer consequências diversas e complexas para o planeta, além de danos irreversíveis à humanidade. Alguns efeitos já são percebidos, como ondas de calor, elevação dos níveis oceânicos. O calor intenso aumenta os riscos de doenças cardio-respiratórias, alergias, infecções, crises de asma e estresse térmico. Com o aquecimento global, fenômenos cíclicos e anomalias climáticas vêm tornando-se cada vez mais frequentes, tais como o *El Niño*, que, entre outras consequências, proporciona secas severas em muitas regiões do globo. O que dificulta o acesso a água, uma vez que há maior ocorrência de secas, e, em períodos mais prolongados. Essa situação prejudica a oferta de recursos naturais para os seres vivos e a manutenção da cadeia alimentar. O que pode levar muitas espécies à extinção. De toda forma, ainda não existe um consenso específico sobre todos os fenômenos causados pelo aquecimento global, que pode gerar ainda a maior incidência de tufões e furacões ou a presença deles em áreas onde não eram comuns. Os desequilíbrios climáticos provocam uma maior incidência

de tempestades em certas áreas, que passam a sofrer com esse problema.

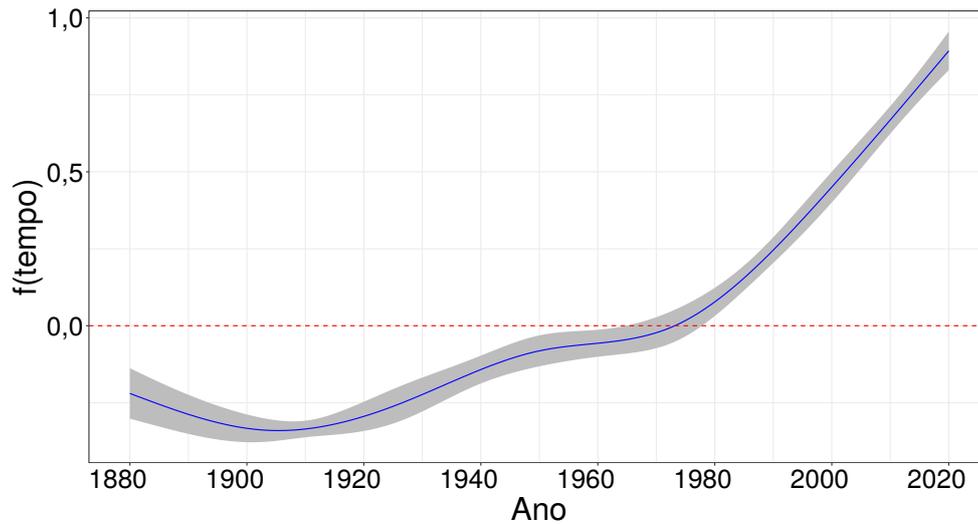


Figura 8.4: Bandas de confiança pontuais para a tendência do modelo (8.1) sob erros com distribuição exponencial potência com $\nu = -0,33$ ajustado à anomalia da temperatura média global anual na superfície terra-mar ($^{\circ}\text{C}$).

8.2 Temperatura em São Francisco

Uma outra ilustração das metodologias postas nesta tese consiste na análise da temperatura média (em $^{\circ}\text{C}$) diária da cidade de São Francisco no período de Janeiro de 1995 a Abril de 2020, um total de 9252 observações. Algumas observações ausentes foram imputadas pela média dos três dias anteriores e posteriores à data. Os dados podem ser obtidos a partir do arquivo de temperatura média diária da Universidade de Dayton - Agência de Proteção Ambiental ², (Kissock, 1999). As temperaturas médias diárias publicadas no site são obtidas do conjunto de dados *Global Summary of the Day (GSOD)* e são calculadas a partir da leitura da temperatura por 24 horas.

São Francisco está localizada na costa oeste dos Estados Unidos, no extremo norte da Península de São Francisco e inclui trechos significativos do Oceano Pacífico e da Baía de São Francisco dentro de seus limites. Cercada nos três lados por água, o clima é bastante influenciado pelas correntes frias do Oceano Pacífico. O tempo da cidade é ameno o ano inteiro, caracterizado por um clima temperado mediterrânico, o verão é longo, agradável, árido e de céu quase sem nuvens; o inverno é curto, fresco, com precipitação e de céu parcialmente encoberto. A temperatura média no inverno é de 10 a 12 $^{\circ}\text{C}$, raramente caindo abaixo de zero, e a temperatura média no verão é de 15 a 21 $^{\circ}\text{C}$.

Da mesma forma que Wood (2017, Seção 7.7.2), propomos o seguinte modelo aditivo para explicar a temperatura média diária de São Francisco:

$$\overline{\text{temp}}_i = \beta_0 + f_T(\text{tempo}_i) + f_S(\text{dia.do.ano}_i) + \epsilon_i, \quad (8.2)$$

em que $f_T(\cdot)$ e $f_S(\cdot)$ são funções suaves aproximadas por splines de regressão cúbica e de regressão cúbica cíclica, respectivamente, enquanto $\overline{\text{temp}}_i$ indica a temperatura média diária do i -ésimo dia, tempo_i é o i -ésimo dia, dia.do.ano_i indica o dia do ano respectivo ao i -ésimo dia, e ϵ_i são erros simétricos autorregressivos de ordem p , conforme discutido no Capítulo 4, para $i = 1, \dots, 9252$. Como alguns anos são bissextos, o padrão é dia.do.ano variar de 1 a 366. Para o nosso modelo, consideramos 80 e 15 nós fixos para o tempo e dia do ano, respectivamente, aos quais foram obtidos a partir dos quantis da variável tempo. A homogeneidade da sazonalidade também é assumida ao longo dos anos.

²<http://academic.udayton.edu/kissock/http/Weather/default.htm>

É importante observar que nas análises apresentadas por Wood (2017, Seção 7.2.2) para explicar a temperatura média da cidade do Cairo, foi aplicado o mesmo modelo (8.2) com $\epsilon_i = \rho\epsilon_{i-1} + e_i$, estrutura AR(1), e_i iid $N(0, \sigma^2)$, mas com efeito aleatório do ano e independência entre os anos.

8.2.1 Análise descritiva

Sabemos que a tendência é o aumento ou diminuição a longo prazo dos dados. Nos dados em estudo é possível observar uma leve tendência crescente. Com o apoio computacional poderemos perceber se essa tendência é estatisticamente significativa. Por outro lado, o padrão sazonal ocorre quando uma série temporal é afetada por fatores sazonais, como a época do ano ou o dia da semana. A Figura 8.5 descreve a série temporal e a distribuição empírica, para cada mês, da temperatura média diária de São Francisco. Os dados mensais da temperatura média na Figura 8.5 mostram a sazonalidade provavelmente induzida pela mudança das estações na cidade de São Francisco. O ciclo ocorre quando a exibição de dados aumenta e diminui sem um período fixo. Essas flutuações geralmente são devidas, nesse caso, às condições de fenômenos naturais (furacões, La Niña, El Niño) ou desastres provocados pelo homem (queimadas). Esses ciclos não podem ser diretamente observados pela Figura 8.5.

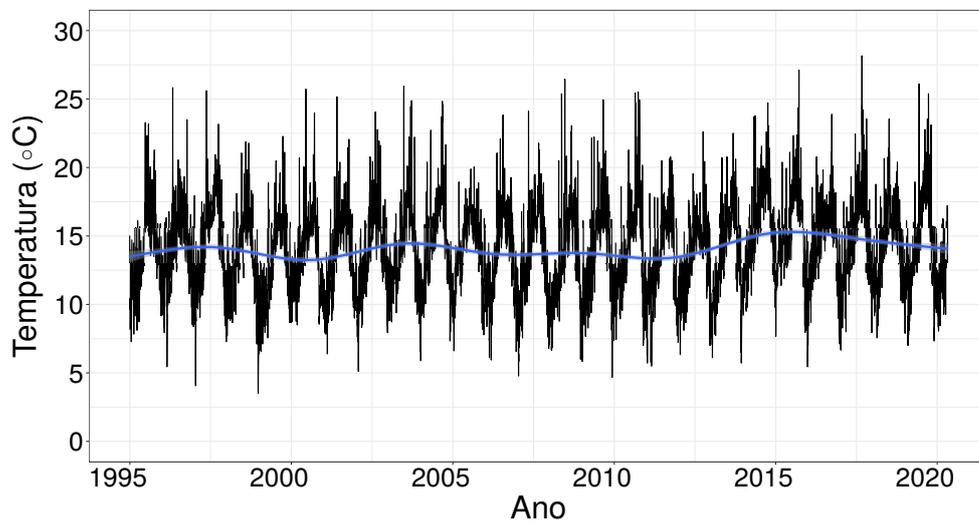


Figura 8.5: Série temporal e distribuição empírica da temperatura média ($^{\circ}C$) diária em São Francisco de Janeiro de 1995 a Abril de 2020.

A sazonalidade parece ser bastante semelhante ao longo de todos os anos. A fim de entender melhor como o fenômeno da sazonalidade afeta a série, dividimos o ano pelos meses. A Figura 8.6 representa o *Boxplot* e o gráfico de densidade para cada mês ao longo dos anos. Pode-se perceber que os meses de dezembro (21) a março (20) são os meses mais frios que remetem ao inverno no hemisfério norte. E, da mesma forma, temos os dias mais quentes no período de junho(21) a setembro(23).

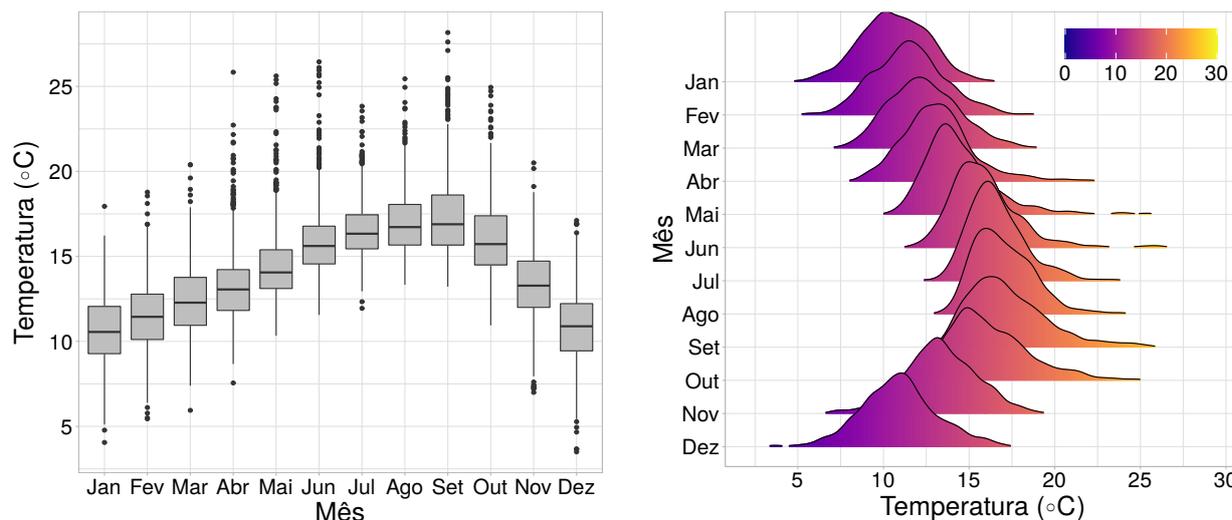


Figura 8.6: *Boxplot e gráfico de densidades por mês da série temporal da temperatura média ($^{\circ}$ C) diária em São Francisco de Janeiro de 1995 a Abril de 2020.*

A Figura 8.6 nos remete a uma sazonalidade mensal. Sendo assim, calculamos a média mensal dos dados diários ao longo dos anos, veja a primeira imagem da Figura 8.7. Esse gráfico enfatiza o padrão sazonal mensal esperado ao longo do tempo. Primeiro, o gráfico sazonal é semelhante a um gráfico de tempo, exceto que os dados são plotados em relação aos meses individuais em que os dados foram observados. Um segundo gráfico alternativo que enfatiza os padrões sazonais é descrever a média dos dados mensais ao longo dos anos, veja o segundo gráfico da Figura 8.7. Esse gráfico nos revela uma certa homogeneidade na sazonalidade da série. O terceiro gráfico da Figura 8.7 revela mini gráficos de tempo para cada mês. Aqui, a média para cada estação é mostrada como uma linha horizontal azul. Essa forma de plotagem permite que o padrão sazonal subjacente seja visto claramente e também mostra as mudanças na sazonalidade ao longo do tempo. É especialmente útil na identificação de mudanças em determinadas estações. Outro ponto interessante com este gráfico é perceber uma leve tendência positiva da temperatura ao longo dos anos. Durante todos os meses, o relacionamento é positivo, com o atraso 12 mostrando o efeito sazonal mais forte.

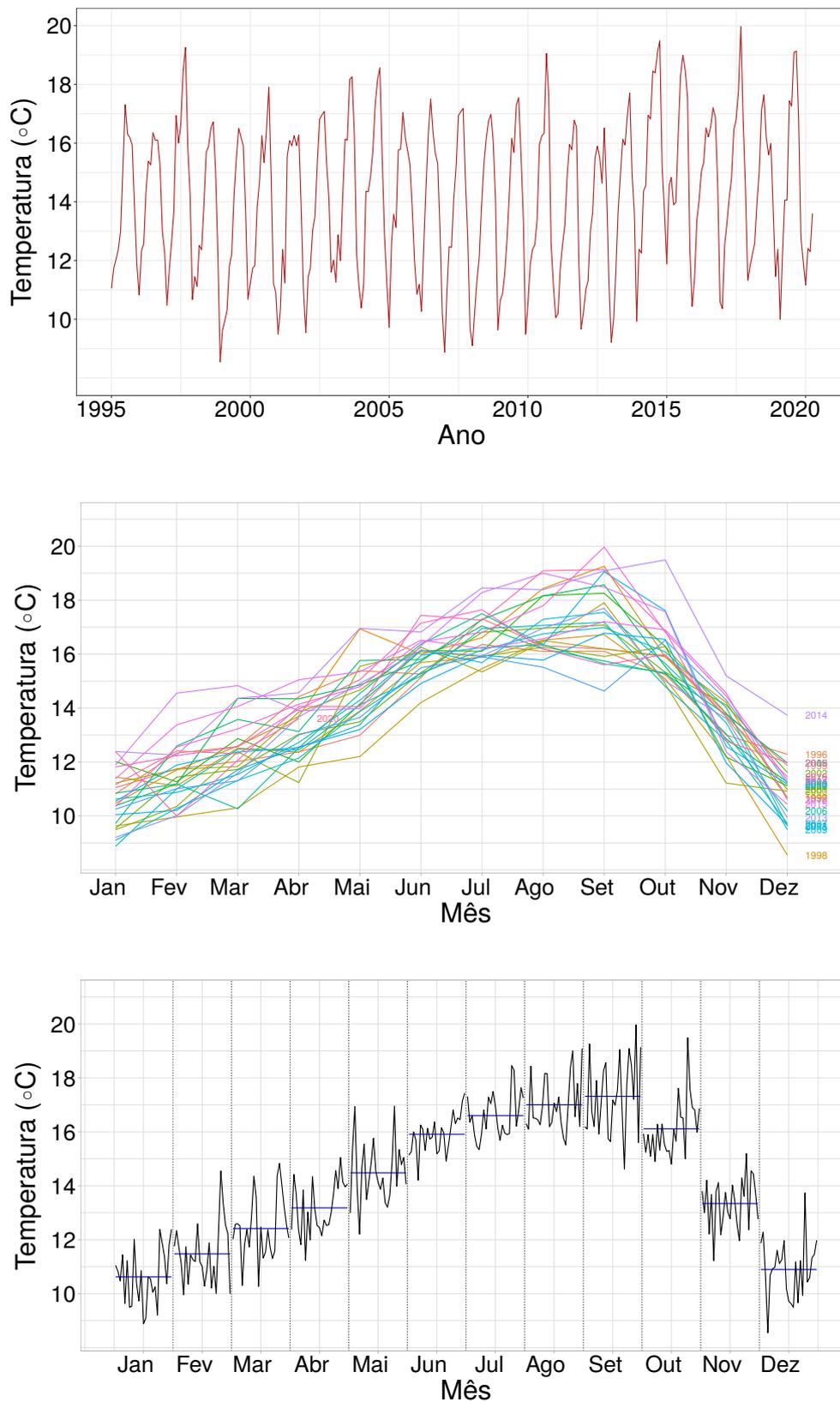


Figura 8.7: Média mensal da série de temperatura média diária de São Francisco, ao longo dos anos (cima), gráfico dos meses em relação aos anos individuais (meio) e dados de cada estação coletados juntos em mini gráficos de tempos separados (baixo).

Outra maneira de analisar os dados das séries temporais é plotar cada observação em relação a outra observação que ocorreu algum tempo antes. Por exemplo, pode-se plotar y_i e y_{i-1} . Isso é chamado de gráfico de defasagem, porque se está plotando as séries temporais contra defasagem em si. As correlações associadas aos gráficos de defasagem formam o que é chamado de “função de autocorrelação”. A autocorrelação é quase a mesma que a correlação. No entanto, a autocorrelação é a correlação de uma série temporal com uma cópia defasada de si mesma. A Figura 8.8 mostra o gráfico de y_i contra y_{i-s} para diferentes valores de s , em que o eixo horizontal mostra valores defasados da série temporal. A relação é positivamente forte na defasagem 12, refletindo forte sazonalidade nos dados.

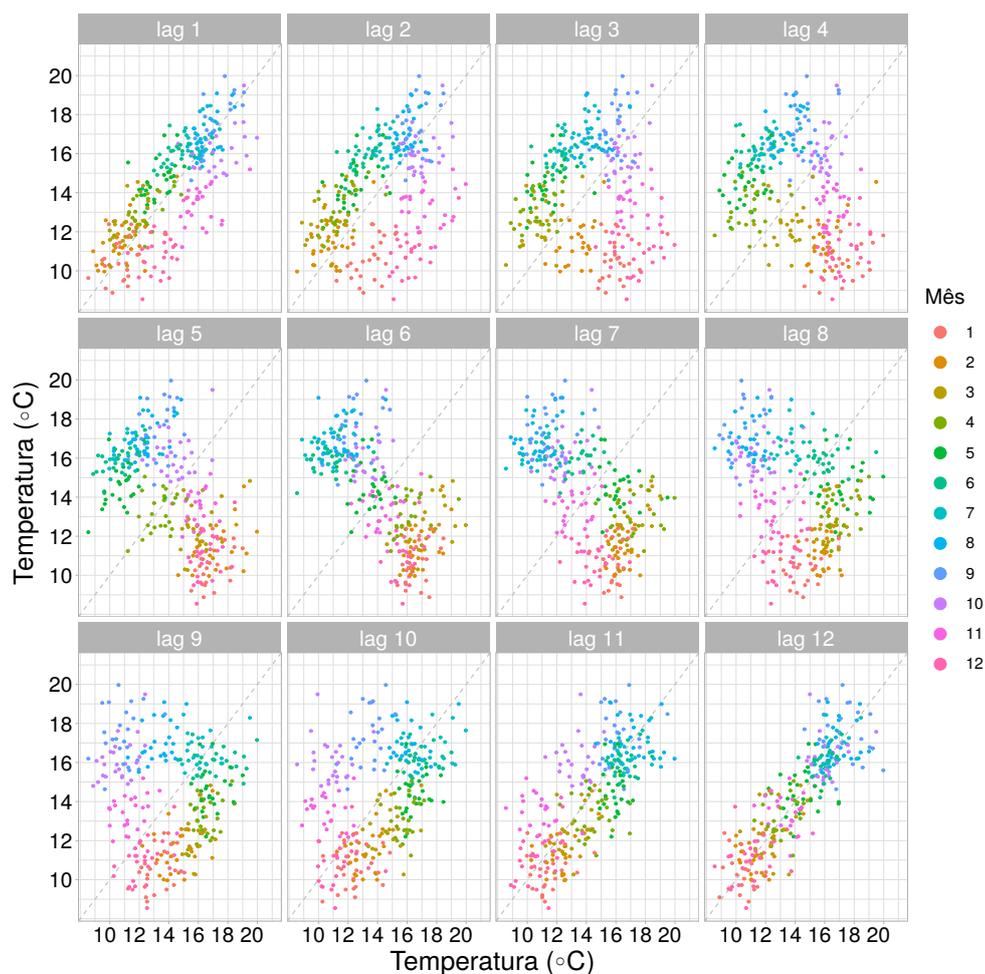


Figura 8.8: Gráficos de dispersão defasados para a média mensal da temperatura média ($^{\circ}\text{C}$) diária em São Francisco de Janeiro de 1995 a Abril de 2020.

8.2.2 Modelagem

Diante a investigação preliminar, pode-se observar que nossa série possui sazonalidade e possivelmente uma tendência positiva ao longo do tempo. A fim de entender melhor o processo que define os dados em estudo, iremos aplicar os modelos estudados nos Capítulos 3 e 4. Os resultados podem auxiliar na tomada de decisão de órgãos públicos e empresas privadas. Estes aos quais dependem e envolvam a temperatura média diária de São Francisco.

Sob o modelo (8.2), esperamos controlar a sazonalidade e depois detectar a tendência da temperatura média diária ao longo de décadas. Para ter uma flexibilidade de curtose, ajustaremos o modelo (8.2) aos dados na distribuição de erros normal, bem como na distribuição de erros t–Student com ν graus de liberdade (distribuição de caudas mais pesadas) e distribuição de erro exponencial

potência com o parâmetro de forma ν (distribuição de caudas mais pesadas para $0 < \nu < 1$ e distribuição de caudas mais curtas para $-1 < \nu < 0$).

Aplicamos o algoritmo *backfitting* descrito nas Seções 3.3.4 e 4.3.4 para uma grade de valores (λ_T, λ_S) para escolher os valores adequados dos parâmetros de suavização. Em seguida, obtemos as medidas de AIC e VCG sob as distribuições com erros normais, t -Student e exponencial potência com parâmetros de forma ν . Esses parâmetros foram estimados separadamente em cada ajuste, juntamente com os parâmetros de suavização. A Tabela 8.2 descreve os valores de AIC e VCG, além dos graus de liberdade efetivos $df(\lambda_T)$ e $df(\lambda_S)$ do modelo (8.2), sob distribuição dos erros normais, t -Student (com $\nu = 5$ graus de liberdade) e exponencial potência (com $\nu = 0,6$) ajustadas à temperatura média diária de São Francisco, com estruturas AR(1), AR(2) e AR(3). Os valores selecionados para os parâmetros de suavização foram $(\lambda_T, \lambda_S) = (500, 5)$.

Tabela 8.2: Qualidade de ajuste para o modelo (8.2) sob erros com distribuição normal, t -Student e Exponencial Potência ajustado à temperatura média diária de São Francisco.

AR	Model	AIC(λ)	VCG(λ)	df(λ)	df(λ_T)	df(λ_S)
1	N	32381,33	1,9379	33,2762	18,9216	11,3545
	t_5	31749,37	1,1910	36,0735	21,2246	11,8489
	EP _{0,6}	31773,41	0,5901	42,6532	27,1344	12,5187
2	N	32180,79	1,8960	36,0015	20,3660	11,6355
	t_5	31590,79	1,1735	38,5008	22,4844	12,0164
	EP _{0,6}	31610,23	0,5844	44,6756	28,1178	12,5578
3	N	32155,33	1,8904	36,3490	19,8111	11,5379
	t_5	31560,70	1,1692	38,8441	21,8986	11,9455
	EP _{0,6}	31583,22	0,5825	45,6914	28,1353	12,5560

A Tabela 8.2 nos revela que o modelo com erros t -Student autorregressivos de ordem três apresenta o menor valor para o critério de Akaike(AIC), enquanto o modelo normal tem o menor custo para estimar $\hat{\gamma}$, o modelo t -Student é o intermediário e o modelo exponencial potência possui o maior custo. Os modelos ajustados podem ser obtidos através do pacote `aplms` desenvolvido na linguagem R e discutida no Capítulo 6 e é dado pelo seguinte comando

```
library(aplms)
ajuste <- aplms(temperatura~s(tempo,bs="cr",k=80)+
  s(dia.do.ano,bs="cc",k=15),
  data=SFranc,p=3,family=Stud(5),lam=c(500,5))
```

As estimativas dos parâmetros dos modelos ajustados estão descritas na Tabela 8.3. Note que os interceptos dos modelos são próximos. Podemos notar que as estimativas do coeficiente de autocorrelação e seus erros padrão aproximados são muito parecidos, mas as estimativas dos parâmetros de dispersão não são comparáveis, pois estão em escalas diferentes.

Tabela 8.3: Estimativas dos parâmetros (erros padrão aproximados) do modelo (8.2) sob erros com distribuições normais, t -Student e exponencial potência com estruturas $AR(1)$, $AR(2)$ e $AR(3)$ ajustado à temperatura média diária de São Francisco.

AR	Model	$\hat{\beta}_0$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\phi}$
1	N	14,0999(0,0485)	0,7029(0,0074)	—	—	1,9249(0,0283)
	t_5	13,9079(0,0443)	0,7055(0,0066)	—	—	1,1823(0,0220)
	EP _{0,6}	13,9149(0,0438)	0,7140(0,0063)	—	—	0,5897(0,0110)
2	N	14,1003(0,0417)	0,8051(0,0103)	-0,1468(0,0103)	—	1,8825(0,0277)
	t_5	13,9450(0,0393)	0,7949(0,0102)	-0,1241(0,0097)	—	1,1645(0,0217)
	EP _{0,6}	13,9403(0,0389)	0,8022(0,0100)	-0,1211(0,0094)	—	0,5791(0,0108)
3	N	14,1002(0,0442)	0,8135(0,0104)	-0,1912(0,0132)	0,0554(0,0104)	1,8774(0,0276)
	t_5	13,9345(0,0415)	0,8023(0,0103)	-0,1691(0,0125)	0,0556(0,0096)	1,1604(0,0216)
	EP _{0,6}	13,9243(0,0410)	0,8092(0,0095)	-0,1652(0,0112)	0,0540(0,0096)	0,5773(0,0107)

De acordo com as análises, podemos perceber que o modelo com erro t -Student com $\nu = 5$ graus de liberdade apresenta melhores resultados. Para acessar esses resultados use o comando `summary` da seguinte maneira:

```
> summary(ajuste)

Family: Student(5)
Link function: identity

Formula:
temperatura~s(tempo,bs="cr",80)+s(dia.do.ano,bs="cc",k=15)

Parametric coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.93453 0.041542 335.43567 < 2.2e-16 ***
----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
edf      Chisq    p-value
s(tempo)    21.8986 252.6028 < 2.2e-16 ***
s(dia.do.ano) 11.9455 2957.924 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure:
Estimate Std. Error t value Pr(>|t|)
rho1    0.8023 0.0103    77.9831 < 2.2e-16 ***
rho2   -0.1691 0.0125   -13.5707 < 2.2e-16 ***
rho3    0.0556 0.0096    5.8062 6.6e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC=31560.6996      BIC=31837.7587      GCV=1.1692
Dispersion est. = 1.1604  n=9262
```

Na sequência, comparamos os modelos ajustados de acordo com a análises de resíduos e estudos de sensibilidade. A Figura 8.9 descreve os gráficos de índices, probabilidade normal, a função de autocorrelação, função de autocorrelação parcial e os pesos contra os resíduos quantílicos condicionais do modelo (8.2) sob erros com distribuição normal, t -Student com $\nu = 5$ e exponencial potência com $\nu = 0,6$ ajustadas aos dados. Os gráficos de índices não apresentam tendência nem variabilidade ao longo do tempo, o que indica que a variação do erro parece constante e que a tendência e a sazonalidade foram controladas. A partir dos gráficos de probabilidade normais, é possível identificar que os modelos de erros de caudas mais pesadas são mais adequados para ajustar os dados do que o normal. As funções de autocorrelação (FACs) e funções de autocorrelação parcial (FACPs) das três séries dos resíduos condicionais parecem confirmar a suposição de independência. E, por fim, nota-se que a distribuição t -Student e exponencial potência adotam pesos diferentes para as observações.

Com relação aos estudos de sensibilidade, a Figura 8.10 descreve os gráficos de índices de B_i , sob o esquema de perturbação de ponderação de casos, para avaliar a influência local da curvatura normal conformal, e também pela partição particular dado por $\hat{\mathbf{f}}$, $\hat{\phi}$ e $\hat{\rho}$ do modelo (8.2), sob erros com distribuições normais, t -Student e exponencial potência ajustadas à temperatura média diária de São Francisco, considerando o valor de corte de $c = 4$, semelhante a [Ibacache-Pulgar et al. \(2013\)](#). Os gráficos indicam que as estimativas dos parâmetros são menos sensíveis no modelo de erros com distribuição t -Student. O que corrobora com a escolha do modelo de acordo com o critério de Akaike, e também pela análise preliminar dos resíduos.

Finalmente, a Figura 8.11 descreve as bandas de confiança pontuais para a sazonalidade (ciclo anual) e para a tendência da temperatura média diária de São Francisco a partir do modelo escolhido, t -Student com $\nu = 5$ graus de liberdade e estrutura de correlação dada por AR(3). O painel esquerdo indica um crescimento lento da temperatura média diária de janeiro a setembro onde alcança seu pico, e em seguida, uma diminuição mais rápida com menor variabilidade. O painel da direita nos revela a tendência observada ao longo do tempo através do modelo selecionado. Note que o intervalo de confiança ao longo dos anos é diferente de zero, o que nos leva a acreditar que há alteração climática. A tendência de longo prazo parece estar estacionária até 2013, com um aumento após este ano até meados de 2019, onde houve uma queda, mas se sustentou em uma média diária na temperatura maior do que era esperado nos anos anteriores. Podemos observar que a década atual de estudo apresenta uma temperatura média superior de $0,6218$ °C com intervalo de $(0,6198; 0,6239)$ em relação aos outros anos dentro do estudo. Note também que, a maior diferença média na temperatura diária na última década em relação ao período anterior em estudo foi de $2,8495$ °C.

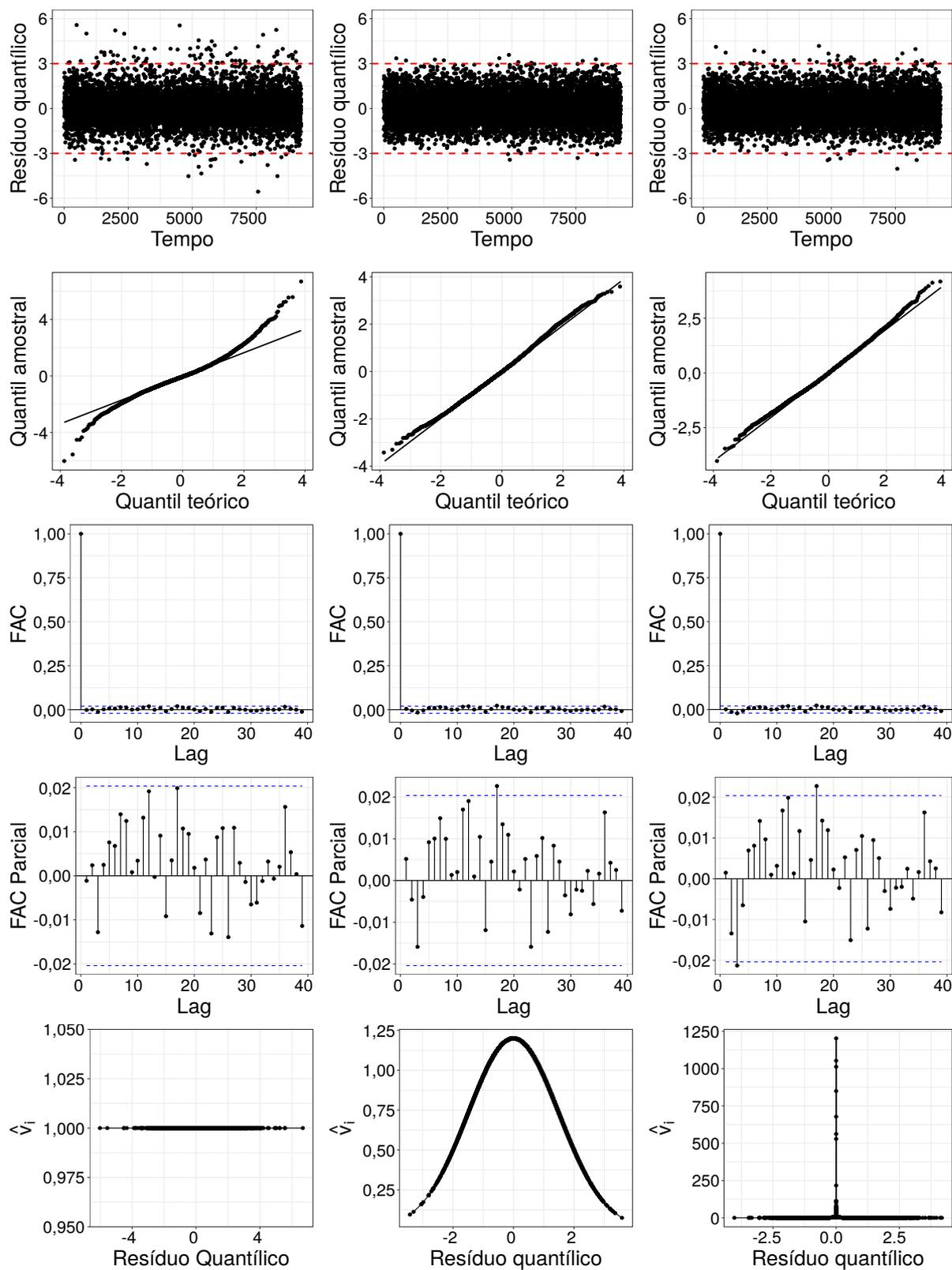


Figura 8.9: Gráficos de índices dos resíduos (primeiro), gráficos de probabilidade normal (segundo), função de autocorrelação - FACs (terceiro), função de autocorrelação parcial - FACPs (quarto) e peso das observações contra os resíduos quantílicos condicionais do modelo (8.2) sob distribuições de erro normal (esquerda), *t*-Student (centro) e exponencial potência (direita) ajustadas à temperatura média diária de São Francisco.

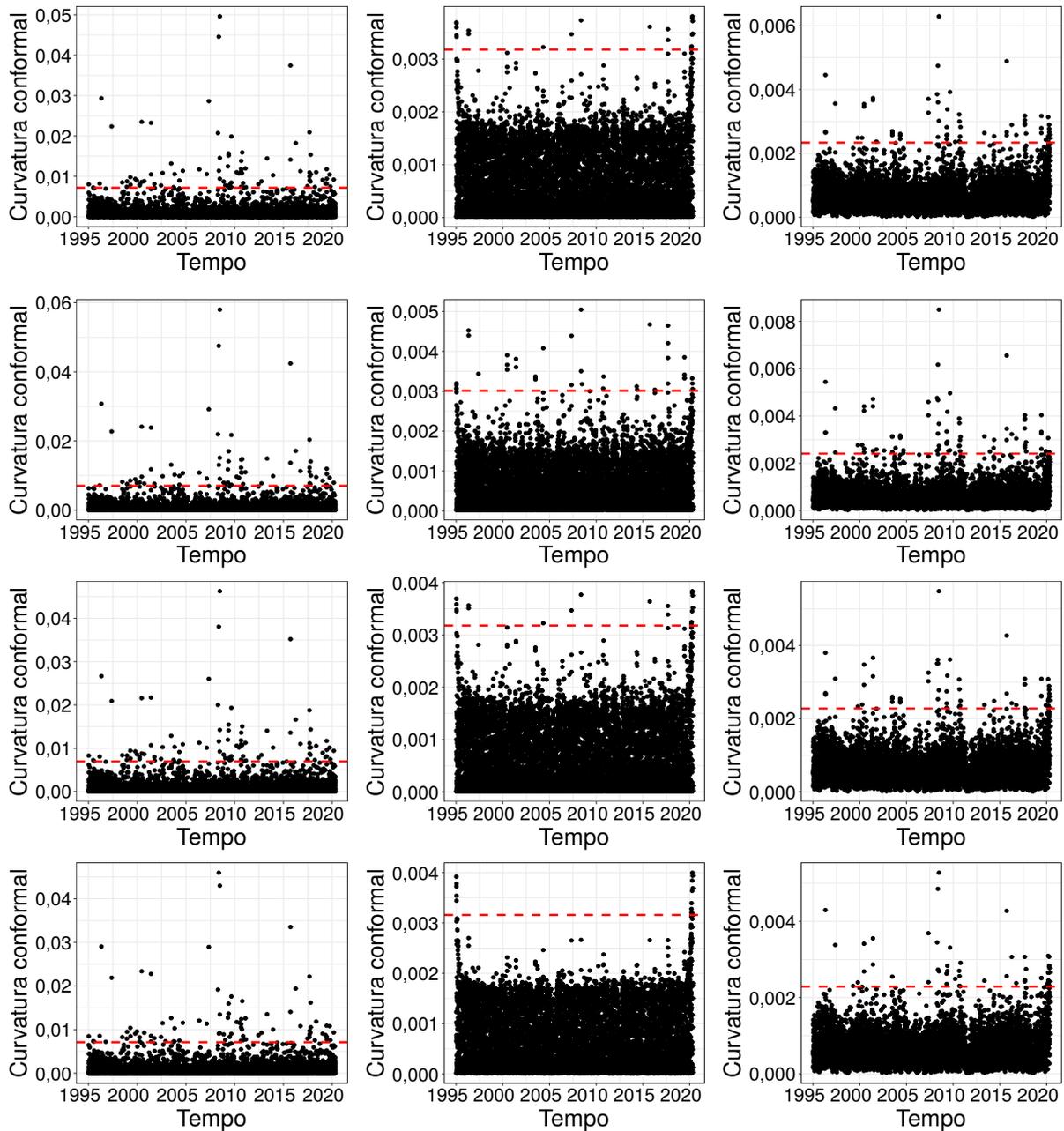


Figura 8.10: Gráficos de índices de C_i no esquema de perturbação ponderação de casos para o modelo (8.2) e também pela partição particular de $\hat{\mathbf{I}}$, $\hat{\phi}$ e $\hat{\rho}$ sob erros com distribuições normal (esquerda), t -Student (centro) e exponencial potência (direita) ajustadas à temperatura diária média de São Francisco.

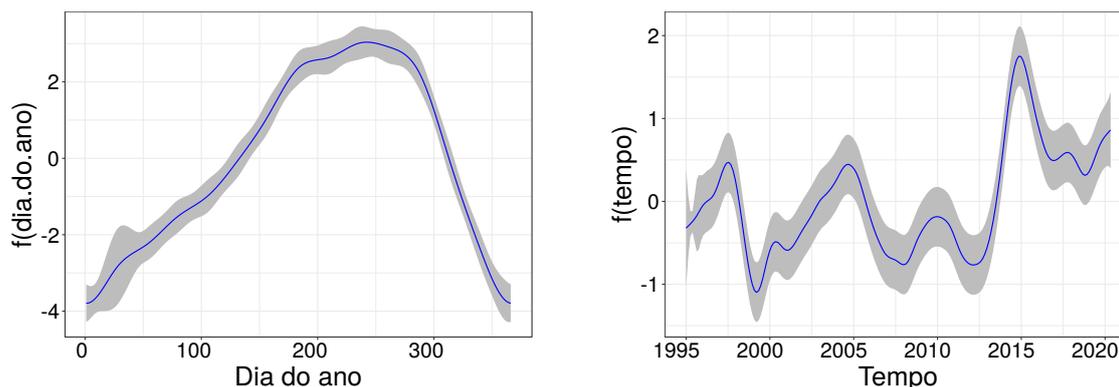


Figura 8.11: Bandas de confiança pontuais para os componentes de sazonalidade (esquerda) e de tendência (direita) do modelo (8.2) sob erros com distribuição t -Student com $\nu = 5$ graus de liberdade ajustado à temperatura média diária de São Francisco.

8.3 Partículas de poluição no município de São Paulo

As partículas de poluição são um problema recorrente. Entende-se de poluente como alguma substância presente no ar, a qual pode tornar impróprio e prejudicial a saúde humana, danos materiais, a fauna e a flora. O nível de poluição atmosférica é medido pela quantidade de substâncias poluentes presentes no ar. Existe uma vasta quantidade de substâncias que podem ser encontradas na atmosfera. Sendo assim, é difícil identificar e classificar os poluentes. As interações entre as diferentes fontes de poluição e a atmosfera definem o nível da qualidade do ar.

Segundo a Companhia Ambiental do Estado de São Paulo - Cetesb³, “a medição sistemática da qualidade do ar é restrita a um número de poluentes, definidos em razão de sua importância e dos recursos disponíveis para seu acompanhamento”. Um desses grupos de poluentes que indicam a qualidade do ar são os materiais particulados (MP).

De acordo com a Agência de Proteção Ambiental dos EUA (EPA), o material particulado é uma mistura de partículas de diversos materiais, sendo todas cerca de cinco vezes mais finas que um fio de cabelo ou mesmo menores que gotículas de substâncias líquidas. O material particulado pode ser gerado de processos como compostos químicos orgânicos, ácidos, como sulfatos e nitratos, metais, poeira, agricultura, incêndios, processos industriais, queima de combustíveis fósseis (resíduo de extrema toxicidade) a qual também é responsável pela liberação de dióxido de carbono (CO₂) - gás de efeito estufa que contribui para o aquecimento global. O material particulado pode também se formar na atmosfera a partir de gases como dióxido de enxofre (SO₂), óxidos de nitrogênio (NO_x) e compostos orgânicos voláteis (COVs), transformando-se em partículas como resultado de reações químicas no ar.

Em complemento a EPA, o material particulado (MP) pode ser dividido em duas categorias. O MP_{2,5} é formado por partículas cujo tamanho chega até 2,5 micrômetros (μm) e pode ser encontrado em nevoeiros e na fumaça. Já o MP₁₀, com partículas de tamanho entre 2,5 e 10 micrômetros (μm), pode ser encontrado em regiões próximas a indústrias. Essas partículas podem ficar retidas na parte superior do sistema respiratório ou penetrar mais profundamente, alcançando os alvéolos pulmonares.

O tamanho das partículas está associado ao seu potencial para causar problemas à saúde, quanto menores maiores os efeitos provocados. Os efeitos negativos da poluição atmosférica nos aparelhos respiratório e cardiovascular são bastante conhecidos, por serem temas recorrentes de estudos, mas os danos à saúde provocados pela exposição aos poluentes vão muito além, veja Pope III e Dockery (2006) e Brunekreef e Holgate (2002). Estudos indicam diversas doenças motivadas pelo material particulado, como doenças cardiovasculares, ataques cardíacos e arritmia cardíaca Brook *et al.* (2010), infarto do miocárdio Chen *et al.* (2020), relatos sobre o desenvolvimento de asma em crianças

³<https://cetesb.sp.gov.br/ar/poluentes/>

Nastos *et al.* (2010), problemas respiratórios Kesavachandran *et al.* (2013). Outras complicações podem estar relacionadas ao aumento na densidade das nuvens, o que dificulta a entrada da luz solar na atmosfera, isso causaria impacto na diminuição da frequência das precipitações e chuva ácida, o que implica incerteza nas previsões climáticas.

Após listar esses problemas relacionados ao material particulado é de suma importância entender o fenômeno e assim, auxiliar os órgãos fiscalizadores, o estado e as empresas a tomarem decisões assertivas ao real problema que acomete os seres humanos, a fauna e a flora. A legislação brasileira de controle da poluição do ar encontra-se defasada em relação à de outros países. Muitos países já criaram leis que regulam os níveis de emissões de material particulado e, para se adequarem aos padrões exigidos, indústrias se utilizam da tecnologia para diminuir suas emissões. Em algumas cidades há o sistema de inspeção veicular, que fiscaliza os níveis de emissões de monóxido de carbono (CO), dióxido de carbono (CO₂), hidrocarbonetos (HC) e material particulado dos carros registrados na cidade.

8.3.1 Análise descritiva

Neste estudo, estudamos as concentrações de MP_{2,5} e MP₁₀ sob efeitos de outras variáveis CO, NO, NO₂, NO_x e SO₂, e variáveis meteorológicas (temperatura do ar, umidade relativa e velocidade do vento). As concentrações de cada poluente foram medidas em estações de monitoramento da CETESB. Os dados estão disponíveis na plataforma e foram extraídos no período de janeiro de 2014 a dezembro de 2020 para a estação Marginal Tietê na ponte dos Remédios, resultando em 2557 observações. O critério para a escolha foi a disponibilidade de dados para os poluentes escolhidos e o perfil do tráfego de veículos na região da estação. Os dados foram obtidos através do sistema Qualar, a partir do pacote `koffing`, na linguagem R, para mais detalhes veja Amorim (2019). Foram empregadas para prever a concentração interna anual e sazonal de MP_{2,5} e MP₁₀ usando os métodos estatísticos desenvolvidos nas seções anteriores desta tese. Um dos principais objetivos deste estudo foi determinar se as concentrações de partículas diminuíram nos últimos 7 anos como resultado das estratégias de controle de emissão de fontes implementadas em São Paulo.

Os dados estão dispostos por hora, ou seja, ao longo do dia há 24 observações. Em nossa análise consideramos a média diária para cada variável. Dessa forma, a Tabela 8.4 mostra as medidas descritivas dos dados em si, consideramos o número de dados faltantes, a média, o desvio padrão, mínimo e máximo, os quartis de 25%, 50% e 75%, intervalo interquartil (IIQ), além das medidas de assimetria e curtose.

Tabela 8.4: Estatísticas resumidas dos poluentes e meteorológicos do ar na estação Marginal Tietê na Ponte dos Remédios, São Paulo, de janeiro de 2014 a dezembro de 2020.

Variável	Faltantes	Média ± DP	Min	P25	P50	P75	Max	IIQ	Assimetria	Curtose
MP _{2,5}	1,53%	20,42 ± 11,71	1,67	11,83	17,58	25,92	81,67	14,08	1,32	4,96
log(MP _{2,5})		2,86 ± 0,56	0,51	2,47	2,87	3,25	4,40	0,78	-0,07	2,77
MP ₁₀	1,41%	33,58 ± 18,81	4,54	20,04	29,54	42,12	130,54	22,08	1,34	5,25
log(MP ₁₀)		3,37 ± 0,55	1,51	3,00	3,39	3,74	4,87	0,74	-0,12	2,80
CO	0,59%	0,75 ± 0,36	0,14	0,47	0,69	0,94	2,71	0,47	1,20	5,26
NO	2,46%	62,18 ± 56,93	0,26	18,00	45,43	90,37	349,43	72,37	1,49	5,58
NO ₂	2,46%	57,31 ± 24,42	8,91	38,07	54,77	73,18	153,52	35,11	0,55	2,98
NO _x	2,46%	80,78 ± 57,37	6,04	35,52	66,76	112,24	357,00	76,72	1,26	4,80
SO ₂	1,60%	2,59 ± 1,47	0,00	1,52	2,26	3,35	9,57	1,83	1,18	4,66
TEMP	0,55%	21,45 ± 3,40	10,34	19,10	21,61	23,94	30,48	4,84	-0,25	2,74
UR	0,55%	69,34 ± 10,00	32,06	63,90	69,88	76,15	93,50	12,25	-0,46	3,33
VV	0,55%	2,12 ± 0,48	0,90	1,79	2,10	2,42	4,29	0,63	0,37	3,34

Nota-se que a média da variável MP_{2,5} é 20,42µg/m³ com um desvio padrão de 11,71µg/m³, no período de janeiro de 2014 a dezembro de 2020 o mínimo dessa variável foi atingido com a concentração de 1,67µg/m³ e máximo de 81,67µg/m³. Pode-se avaliar uma assimetria positiva de 1,32 e curtose de 4,96. Em relação à variável MP₁₀ tem-se uma média de 33,58µg/m³ com um desvio padrão de 18,81µg/m³, no período de janeiro de 2014 a dezembro de 2020 o mínimo dessa variável

foi atingido com a concentração de $4,54\mu\text{g}/\text{m}^3$ e máximo de $130,54\mu\text{g}/\text{m}^3$. Pode-se avaliar uma assimetria positiva de 1,34 e curtose de 5,25. Para fins de estudo decidimos aplicar a transformação logarítmica nessas variáveis com o intuito de conseguir simetria nos dados e, portanto, aplicar os métodos desenvolvidos. Ao realizar a transformação logarítmica temos que a assimetria diminui consideravelmente, veja que para a variável $\text{MP}_{2,5}$ é -0,05; já para a variável MP_{10} é -0,12. E, em relação à curtose temos uma diminuição de 4,96 para 2,77 para a variável $\text{MP}_{2,5}$ e de 5,25 para 2,80 para a variável MP_{10} .

A Figura 8.12 representa o histograma das variáveis $\text{MP}_{2,5}$ e MP_{10} , bem como, a transformação logarítmica das respectivas variáveis. Repare pelas Figuras 8.12(b) e 8.12(d) que a distribuição dos dados após a transformação apresentam uma leve assimetria negativa. Esses dados serão analisados em conjunto com as demais variáveis com o propósito de entender quais delas influenciam de forma paramétrica e não paramétrica os materiais particulados, os quais trazem grandes prejuízos à saúde das pessoas, às atividades agropecuárias, principalmente às incertezas nas previsões climáticas, e também problemas relacionados à fauna e à flora.

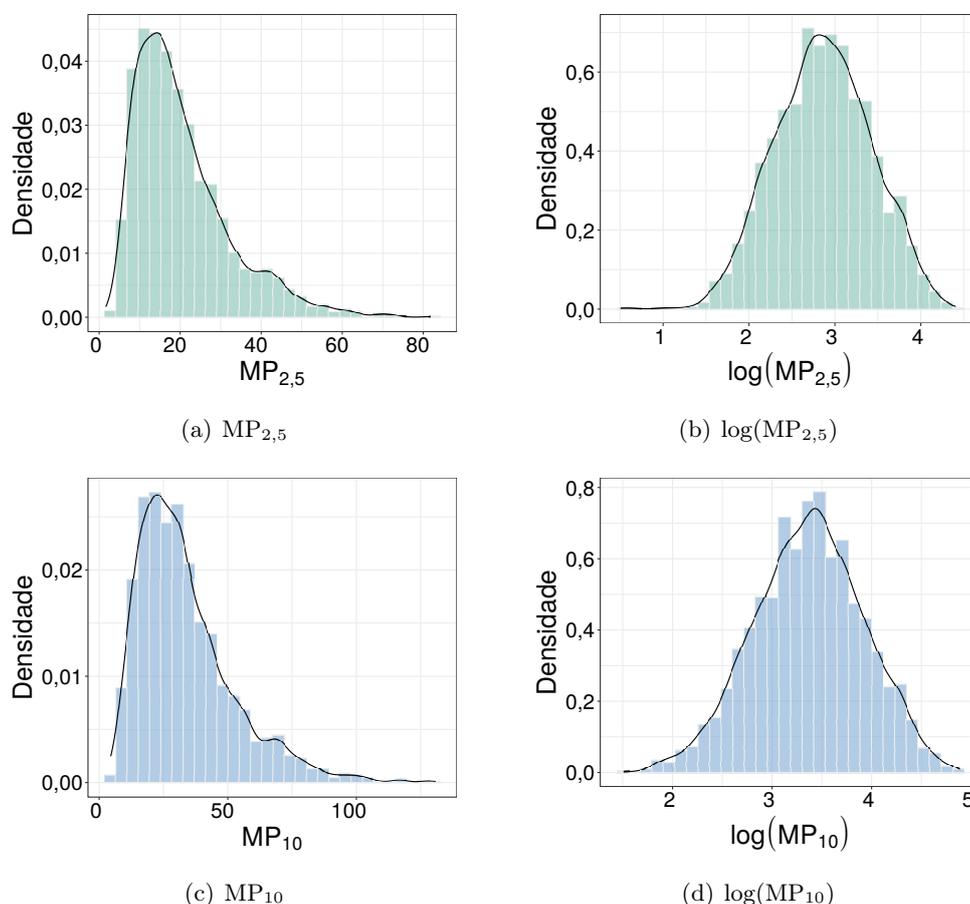


Figura 8.12: Histograma das variáveis $\text{MP}_{2,5}$ (cima) e MP_{10} (baixo) e respectiva transformação \log (direita).

A Figura 8.13 mostra o fator médio de impacto das variáveis de material particulado em relação aos anos, ao mês do ano e também ao dia da semana. Conforme mostrado nas Figuras 8.13(a) e 8.13(d), os níveis de $\text{MP}_{2,5}$ e MP_{10} diminuíram durante o período do estudo. As Figuras 8.13(b) e 8.13(e) sugerem tendências sazonais em relação às concentrações mensais. Os fatores médio de impacto da concentração do material particulado são mais prevalentes nos meses de junho a setembro, meses relacionados ao inverno na cidade de São Paulo. As Figuras 8.13(c) e 8.13(f) sugerem uma sazonalidade nos dias da semana, observe que segunda a sexta há uma maior concentração de ambos materiais particulados. Isso pode estar diretamente relacionado à maior incidência de veículos perto da estação, uma vez que esta está situada em uma região de tráfego intenso.

Além disso, uma simples comparação das médias ou medianas anuais pode não ser suficiente

para examinar as tendências de concentração e, portanto, uma análise estatística mais sofisticada foi empregada. Usando modelagem de regressão, a variável ano expressa a tendência de concentração enquanto controlamos outros parâmetros. As concentrações dependem não apenas das fontes de emissão, mas também das condições meteorológicas que podem variar de ano para ano. Assim, é necessária a inclusão de parâmetros meteorológicos no modelo para distinguir tendências devido a mudanças nas emissões de fonte daquelas relacionadas ao clima. Isso é especialmente importante para o estudo de tendências em um período relativamente curto, em que alguns anos atípicos podem dificultar a comparação das concentrações médias anuais.

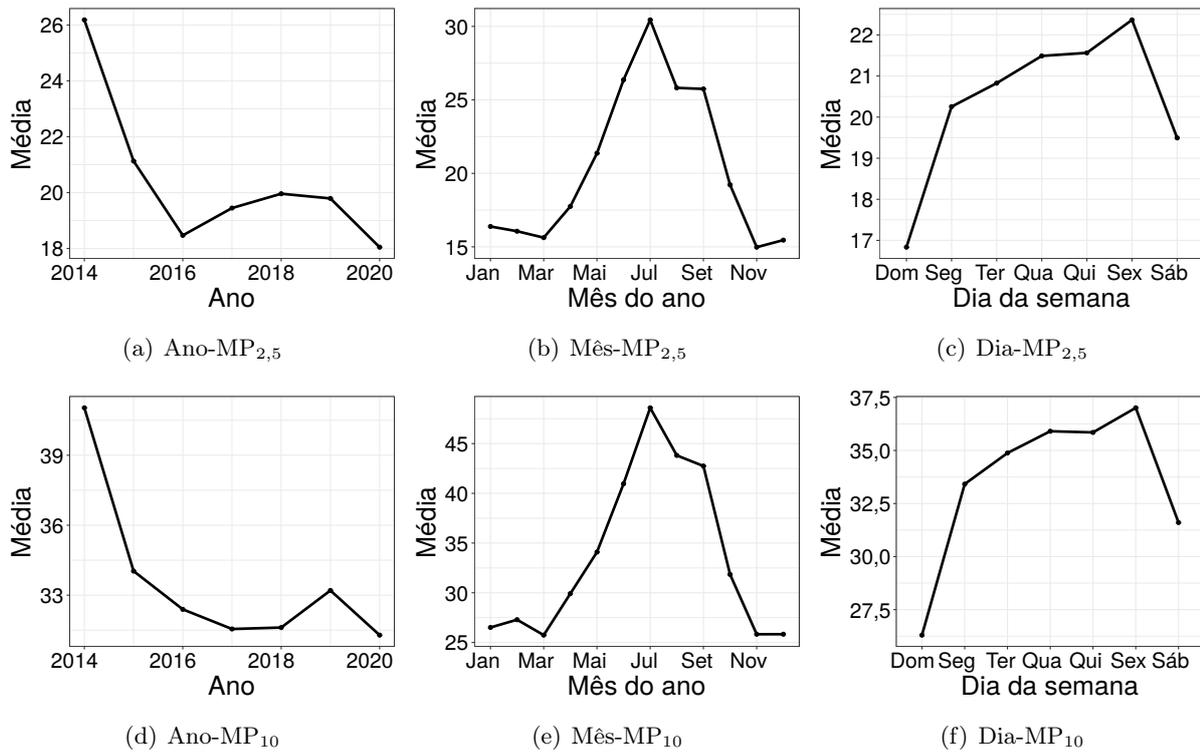


Figura 8.13: Fator médio de impacto das variáveis $MP_{2,5}$ (cima) e MP_{10} (baixo) por ano (esquerda), mês (meio) e dia da semana (direita).

A Figura 8.14 mostra a matriz de correlação das variáveis coletadas na estação da Marginal Tietê na Ponte dos Remédios. As variáveis materiais particulados $MP_{2,5}$ e MP_{10} apresentam uma correlação positiva forte de 0,95. Já em relação às outras variáveis, há uma correlação positiva forte com as variáveis CO, NO, NO_2 , NO_x e SO_2 , uma baixa correlação com a variável temperatura, e por fim, apresentam uma correlação negativa com as variáveis meteorológicas umidade relativa e velocidade do vento.

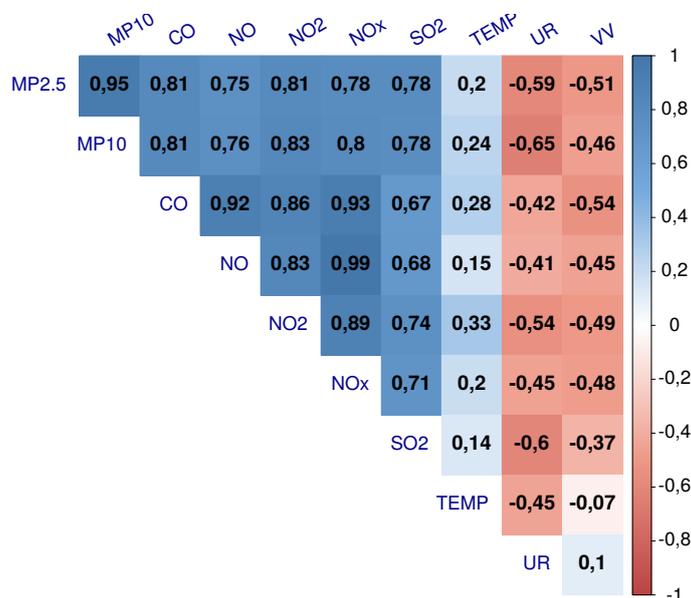


Figura 8.14: Matriz de correlação das variáveis de poluição e meteorológicas na estação da Marginal Tietê na Ponte dos Remédios.

8.3.2 Modelagem

As análises descritivas são fundamentais para nos alertar sobre possíveis relações entre as variáveis, uma possível tendência negativa da variável em estudo, a sazonalidade decorrente dos dias da semana que refletem ao maior número de veículos nas rodovias e a plena produção industrial nos dias comerciais, e por fim, a sazonalidade mensal que pode estar relacionada com as estações do ano. Como já percebemos os meses de inverno apresentam maior concentração de materiais particulados. Uma outra questão que possa ser interessante para os pesquisadores é a influência das horas nas variáveis. Mas este trabalho não irá lidar com esses dados. Horários de pico podem refletir uma maior concentração de materiais particulados, já que haverá mais automóveis em trânsito, assim como o horário comercial o qual envolve a produção industrial. Essas hipóteses podem ser implementadas no modelo em estudo, e assim, verificar a possível influência das horas na concentração de poluição do ar, o que pode auxiliar pessoas a evitarem circular em certos locais e horários para prevenir problemas de saúde, como já discutido anteriormente.

Os materiais particulados $MP_{2,5}$ e MP_{10} são assimétricos, dessa forma, realizamos a transformação log para torná-los simétricos, a fim de implementar e estudar os modelos propostos. Desenvolvemos diversas análises para encontrar as variáveis discutidas na Tabela 8.4, que melhor explicam os materiais particulados. Para tanto, os modelos escolhidos são dados por

$$\begin{aligned} \log(MP_{2,5}) = & CO + NO + SO_2 + TEMP + UR + VV \\ & + s(tempo) + s(dia.do.ano) + s(dia.da.semana) + \epsilon_i \end{aligned} \quad (8.3)$$

$$\begin{aligned} \log(MP_{10}) = & CO + NO_2 + SO_2 + TEMP + UR + VV \\ & + s(tempo) + s(dia.do.ano) + s(dia.da.semana) + \epsilon_i \end{aligned} \quad (8.4)$$

em que $\epsilon_i = \rho_1\epsilon_{i-1} + \dots + \rho_p\epsilon_{i-p} + e_i$ é autocorrelacionado de ordem p e e_i possui distribuição simétrica com média 0 e parâmetro de dispersão ϕ , a qual, pode ser expressa da seguinte maneira $e_i \stackrel{\text{iid}}{\sim} S(0, \phi)$. Nessa aplicação, consideramos os erros com distribuição normal, t -Student e exponencial potência.

A Tabela 8.5 nos mostra os resultados das variáveis $MP_{2,5}$ e MP_{10} , transformadas, seguindo os modelos dados por (8.3) e (8.4). Note que para a variável $\log(MP_{2,5})$ pelo critério AIC, temos que o melhor modelo é dado pelo esquema de autocorrelação de ordem 3 e com a distribuição t -Student com $\nu = 5$ graus de liberdade. Já para a variável $\log(MP_{10})$, pelo critério AIC, temos que o melhor modelo segue distribuição t -Student com $\nu = 5$ graus de liberdade e o esquema de autocorrelação de ordem um.

Tabela 8.5: Qualidade de ajuste para os modelos (8.3) e (8.4) sob erros com distribuição normal, t -Student e Exponencial Potência ajustados aos dados de materiais particulados na estação marginal Tietê, Ponte dos Remédios.

Variável	AR(p)	Modelo	AIC(λ)	VCG(λ)	df(λ)	df(λ_1)	df(λ_2)	df(λ_3)
$\log(MP_{2,5})$	1	Normal	58,1805	0,0598	22,7752	3,9089	7,8682	1,9980
		t -Student($\nu=5$)	-189,7868	0,0412	23,2211	3,9357	8,2867	1,9987
		EP($\nu=0,6$)	-150,3678	0,0220	24,0428	3,9738	9,0696	1,9994
	2	Normal	55,1131	0,0597	23,6569	3,9008	7,7579	1,9982
		t -Student($\nu=5$)	-201,9630	0,0360	24,2236	3,9358	8,2888	1,9989
		EP($\nu=0,6$)	-154,0940	0,0219	25,1159	3,9770	9,1393	1,9996
	3	Normal	41,1630	0,0593	24,4582	3,8857	7,5741	1,9984
		t -Student($\nu=5$)	-214,0122	0,0358	25,0857	3,9278	8,1589	1,9990
		EP($\nu=0,6$)	-164,8815	0,0218	25,8816	3,9675	8,9146	1,9996
$\log(MP_{10})$	1	Normal	-283,3535	0,0523	27,9979	7,3152	9,6845	1,9982
		t -Student($\nu=5$)	-443,2088	0,0324	28,5762	7,7756	9,8017	1,9989
		EP($\nu=0,6$)	-438,0337	0,0162	29,3373	8,4056	9,9321	1,9996
	2	Normal	-281,5642	0,0523	29,0178	7,3307	9,6889	1,9982
		t -Student($\nu=5$)	-441,5150	0,0324	29,5556	7,7587	9,7980	1,9989
		EP($\nu=0,6$)	-436,2317	0,0162	30,4014	8,4639	9,9379	1,9996
	3	Normal	-281,5289	0,0523	29,9447	7,2735	9,6729	1,9983
		t -Student($\nu=5$)	-435,4319	0,0298	30,5963	7,7916	9,8057	1,9990
		EP($\nu=0,6$)	-435,1216	0,0162	31,3502	8,4227	9,9278	1,9996

O pacote `aplms` desenvolvido em linguagem R e discutido no Capítulo 6 é usado para obter as informações definidas no Capítulo 4 e 5 a fim de estudar melhor a influência das variáveis independentes nas variáveis dependentes associadas à poluição, material particulado, que são definidas por $MP_{2,5}$ e MP_{10} . Dessa maneira, os modelos (8.3) e (8.4) podem ser ajustados no R, basta carregar o pacote `aplms` e dispor os comandos da seguinte maneira:

```
> library(aplms)
> ajuste1 <- aplms(MP2.51~CO+NO+SO2+TEMP+s(tempo,bs="cr",k=5)+
s(dia.do.ano,bs="cc",k=12)+s(dia.da.semana,bs="cc",k=4),
family=Stud(5),p=3,lam=c(100,500,5))
> ajuste2 <- aplms(MP101~CO+NO2+SO2+TEMP+s(tempo,bs="cr",k=10)+
s(dia.do.ano,bs="cc",k=12)+s(dia.da.semana,bs="cc",k=4),
family=Stud(5),p=1,lam=c(700,50,5))
```

```
> summary(ajuste1)
```

```
Family: Student(5)
```

Link function: identity

Formula:

MP2.5l~CO+NO+SO2+TEMP+s(tempo,bs="cr",k=5)+s(dia.do.ano,bs="cc",k=12)+
s(dia.da.semana,bs="cc",k=4)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.732069	0.120091	14.423008	< 2.2e-16 ***
CO	0.676363	0.038185	17.712942	< 2.2e-16 ***
NO	-0.001279	0.000217	-5.909113	3.899e-09 ***
SO2	0.092382	0.005387	17.149490	< 2.2e-16 ***
UR	-0.003676	0.000807	-4.557596	5.418e-06 ***
TEMP	0.048981	0.003261	15.019805	< 2.2e-16 ***
VV	-0.145356	0.012612	-11.524968	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

edf	Chisq	p-value
s(tempo)	3.9278	35.46413 3.405e-07 ***
s(dia.do.ano)	8.1589	308.8652 < 2.2e-16 ***
s(dia.da.semana)	1.9990	33.84365 4.469e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure:

	Estimate	Std. Error	t value	Pr(> t)
rho1	0.3651	0.0195	18.73908	< 2.2e-16 ***
rho2	0.0283	0.0200	1.41607	0.1569
rho3	0.0684	0.0184	3.72254	0.0002 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC = -214.0122 BIC = -67.3465 GCV = 0.0358

Dispersion est. = 0.0352 n= 2557

> summary(ajuste2)

Family: Student(5)

Link function: identity

Formula:

MP10l~CO+NO2+SO2+TEMP+s(tempo,bs="cr",k=10)+s(dia.do.ano,bs="cc",k=12)+
s(dia.da.semana,bs="cc",k=4)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.336285	0.112970	20.680496	< 2.2e-16 ***
CO	0.333092	0.027579	12.077871	< 2.2e-16 ***
NO2	0.004401	0.000455	9.676395	< 2.2e-16 ***
SO2	0.073570	0.005313	13.845889	< 2.2e-16 ***

```

UR          -0.005840 0.000758   -7.705904  1.848e-14 ***
TEMP        0.044286 0.003063   14.456604 < 2.2e-16 ***
VV          -0.088762 0.012046   -7.368848  2.314e-13 ***
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
edf      Chisq      p-value
s(tempo)      7.7756      73.4719  7.625e-13 ***
s(dia.do.ano)  9.8017      315.3062 < 2.2e-16 ***
s(dia.da.semana) 1.9989      18.8783  7.942e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure:
Estimate Std. Error t value Pr(>|t|)
rho1  0.3652      0.0165   22.177 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC = -443.2088      BIC = -276.1357      GCV = 0.0324
Dispersion est. = 0.0317  n = 2557

```

O ajuste do modelo (8.3) apresenta resultados satisfatórios em relação às variáveis paramétricas, então podemos notar que monóxido de carbono (CO), dióxido de enxofre (SO₂) e temperatura (TEMP) possuem influência positiva quanto à variável $\log(\text{MP}_{2,5})$. Enquanto que óxido de nitrogênio (NO), umidade relativa (UR) e velocidade do vento (VV) salienta influência negativa em relação ao material particulado com diâmetro de $2,5\mu\text{m}$. Além disso, note que a parte aditiva do modelo é significativa, logo, temos que o material particulado foi influenciado ao longo dos anos, veja Figura 8.15(a) em que houve uma diminuição na concentração de $\text{MP}_{2,5}$ no período de 2014 a 2018. Logo em seguida, em 2019, há um leve aumento, mas tende a diminuir em 2020, o que pode estar relacionado as restrições impostas durante a pandemia. Quanto aos meses do ano, veja a Figura 8.15(c) em que o período de junho a setembro apresenta os maiores picos de concentração de material particulado com diâmetro $2,5\mu\text{m}$, o que pode ser influência do inverno. E os dias da semana, veja na Figura 8.15(e) que terça, quarta e quinta são os dias com maior concentração, isso pode estar diretamente relacionado com a maior movimentação de carros na região, dias comerciais. Outro ponto a ser observado é que o coeficiente de autocorrelação ρ_2 não é significativo, mas, ρ_3 é significativo, sendo assim, escolhemos o modelo com erros de ordem 3.

Da mesma forma, analisamos o modelo (8.4) e os resultados são aceitáveis em relação as variáveis paramétricas ajustadas. As variáveis monóxido de carbono (CO), dióxido de nitrogênio (NO₂), dióxido de enxofre (SO₂) e temperatura (TEMP) apresentam influência positiva quanto a variável $\log(\text{MP}_{10})$. Já as variáveis umidade relativa (UR) e velocidade do vento (VV) possuem influência negativa. Ademais, a parte aditiva do modelo é expressamente significativo, sendo assim, podemos concluir que o material particulado com diâmetro de $10\mu\text{m}$ sofreu alteração ao longo dos anos, para mais detalhes veja a Figura 8.15(b), note que há uma estabilização nos anos de 2014 a 2017. Logo em seguida, há uma diminuição na concentração do material particulado até 2018, mas logo, há uma tendência positiva, no período de 2018 a 2020. Mas, neste último ano, houve um decréscimo, como discutido anteriormente, há indícios de que a pandemia decretada no início de 2020 possa ter influenciado na diminuição da concentração de MP_{10} . Em relação aos dias do ano, note na Figura 8.15(d) que julho a setembro, meses de inverno no hemisfério sul, possuem uma maior concentração do material. E, por fim, os dias de terça, quarta e quinta apresentam maior concentração de partículas com diâmetro de $10\mu\text{m}$ como pode ser visto na Figura 8.15(f). Além do mais, pelo critério AIC, escolhemos o modelo com erros autorregressivos de ordem um, e pode-se observar

pelos resultados da saída do R que o coeficiente $\rho_1 = 0,3652$ é significativo.

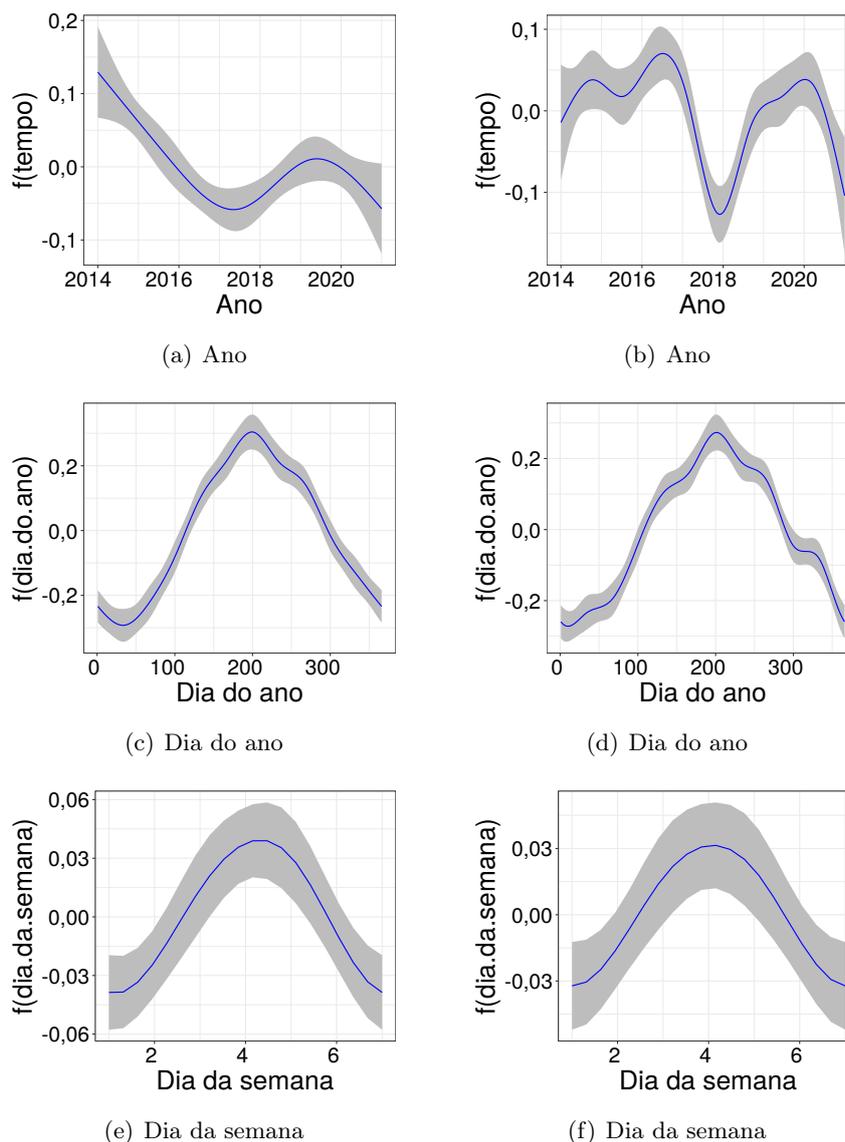


Figura 8.15: Impacto médio do ano, meses do ano e dias da semana em relação aos materiais particulados $MP_{2,5}$ (esquerda) e MP_{10} (direita) sob erros com distribuição t -Student com $\nu = 5$ graus de liberdade.

Para que tais conclusões sejam discutidas, é necessário avaliar se o modelo está bem ajustado. Portanto, cabe ao pesquisador fazer análise dos resíduos dos modelos ajustados. Note que para as análises aqui propostas, usamos os resíduos quantílicos condicionais, conforme discutimos na Seção 5.1. A Figura 8.16 apresenta o índice do resíduo ao longo do tempo em que percebe-se que não existe um padrão, e, portanto há homocedasticidade dos resíduos, para a variável $\log(MP_{2,5})$ há alguns pontos que podem ser considerados como *outliers*. O gráfico quantil nos indica que existem alguns pontos fora da normalidade. Os gráficos da Função de Autocorrelação (FAC) e a Função de Autocorrelação Parcial nos mostram que os resíduos quantílicos não são autocorrelacionados e, por fim, o gráfico dos pesos dados por v_i contra os resíduos, nos mostram que as observações apresentam diferentes pesos no processo de estimação do modelo, o que difere da distribuição normal a qual possui peso um para todas as observações.

E, por fim, podemos analisar a curvatura conformal das observações para diferentes esquemas de perturbação. A Figura 8.17 representa os esquemas de perturbação ponderação de casos, para o parâmetro de dispersão, variável resposta, variável explicativa (nesse caso selecionamos a variável dióxido de carbono - CO), e por fim, o coeficiente de autocorrelação dado por ρ_1 para os modelos

(8.3) e (8.4) definidos anteriormente para os materiais particulados $\log(\text{MP}_{2,5})$ e $\log(\text{MP}_{10})$, respectivamente. A variável $\log(\text{MP}_{2,5})$ apresenta poucos pontos que possivelmente são influentes e podem alterar as estimativas do estudo. O método de perturbação pelo coeficiente de autocorrelação apresenta o maior grau de influência. A variável $\log(\text{MP}_{10})$ apresenta poucos pontos influentes e com pouca intensidade para os quatro primeiros esquemas de perturbação. E, novamente, a perturbação pelo coeficiente de autocorrelação apresenta 3 pontos com alto índice influência local.

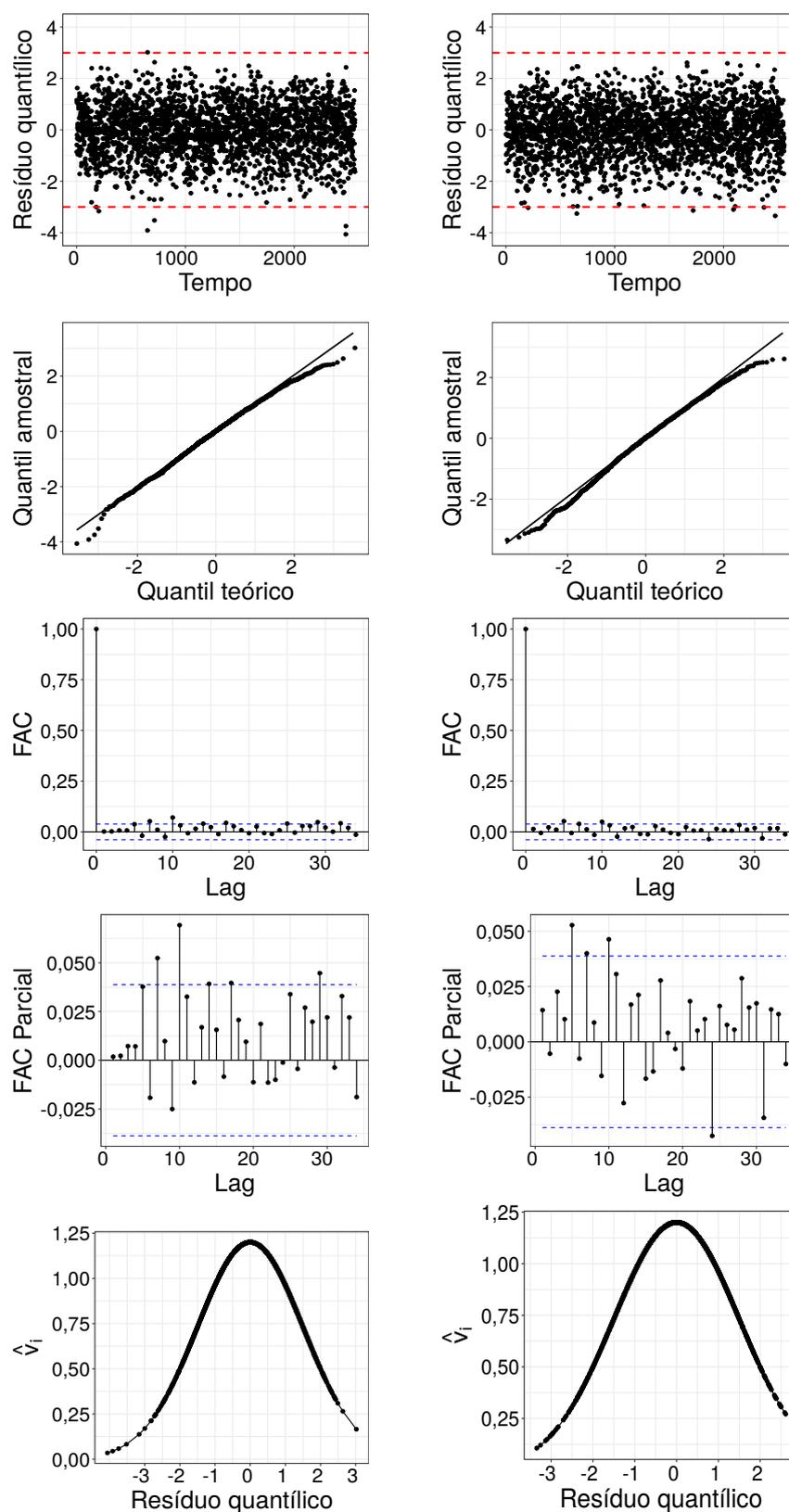


Figura 8.16: Gráficos de índices dos resíduos (primeiro), probabilidade normal (segundo), função de autocorrelação (terceiro), função de autocorrelação parcial (quarto) e peso das observações contra os resíduos quantílicos condicionais dos modelos ajustados (8.3) e (8.4) para os materiais particulados $\log(MP_{2,5})$ (esquerda) e $\log(MP_{10})$ (direita) sob erros com distribuição t -Student com $\nu = 5$.

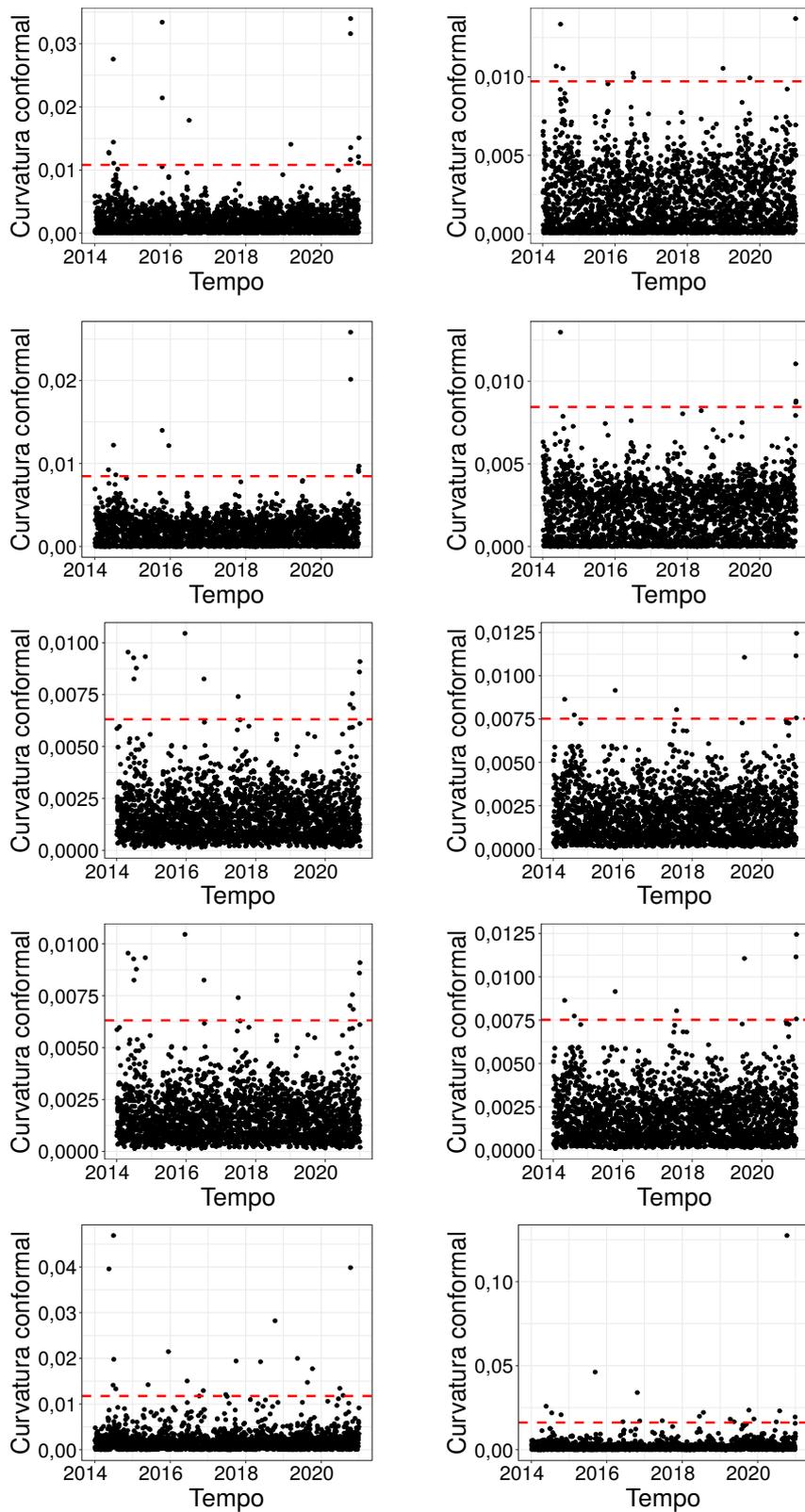


Figura 8.17: Gráficos de índices C_i nos esquemas de perturbação, ponderação de casos (primeiro), parâmetro de dispersão (segundo), variável resposta (terceiro), variável explicativa (CO) (quarto) e coeficiente de autocorrelação (ρ_1) (quinto) dos modelos (8.3) e (8.4) com distribuição dos erros t -Student com $\nu = 5$ graus de liberdade para $\log(MP_{2,5})$ (esquerda) e $\log(MP_{10})$ (direita).

8.4 Mortalidade cardiovascular

Segundo a OMS – "As doenças cardiovasculares (DCVs) são a causa número um de morte no mundo, ceifando cerca de 17,9 milhões de vidas a cada ano. As DCVs são um grupo de doenças do coração e dos vasos sanguíneos e incluem doença coronariana, doença cerebrovascular, doença cardíaca reumática e outras condições. Quatro em cada 5 mortes por DCV são devido a ataques cardíacos e derrames, e um terço dessas mortes ocorrem prematuramente em pessoas com menos de 70 anos de idade."⁴(tradução livre)

Dessa forma, é importante estudar as possíveis variáveis que podem afetar o avanço ou declínio do nível de mortalidade cardiovascular ao longo dos anos. Nos países desenvolvidos, observa-se redução da força da mortalidade por doenças cardiovasculares desde a metade do século XX, (Levi *et al.*, 2002). Mensah *et al.* (2017) descrevem possíveis causas e implicações do declínio da mortalidade por doenças cardiovasculares ao longo do século XX.

A possibilidade de relações entre os níveis de poluição do ar de curto ou longo prazo e mortalidade por doenças cardiovasculares é motivo de preocupação para a população e órgãos públicos responsáveis por estabelecer padrões de saúde devido a poluição do ar. As agências ambientais definem os padrões para proteger a população em geral, incluindo subgrupos sensíveis, contra os efeitos adversos à saúde. Há algumas preocupações em relação aos possíveis efeitos da poluição do ar e a mortalidade ao longo do tempo. Então, é importante identificar os fatores associados da mortalidade quanto a poluição na presença de fatores ambientais. E, identificar padrões que permita prever os níveis de mortalidade em função da poluição e dos efeitos do clima.

8.4.1 Análise descritiva

Nesse sentido, iremos analisar os possíveis efeitos da temperatura e poluição na mortalidade semanal no condado de Los Angeles no período de 10 anos, de 1970 a 1979. Os dados podem ser carregados diretamente no R através do pacote `astsa` os quais são extraídos de um estudo de Shumway *et al.* (1988), e também discutido por Shumway e Stoffer (2016). Dessa forma, a Tabela 8.6 mostra as medidas descritivas dos dados em si, consideramos a média, o desvio padrão, mínimo e máximo, os quartis de 25%, 50% e 75%, intervalo interquartil (IIQ), além das medidas de assimetria e curtose.

Tabela 8.6: Estatísticas resumidas da mortalidade cardiovascular média semanal, temperatura e partículas de poluição no condado de Los Angeles no período de 1970 a 1979.

Variável	Média ± DP	Min	P25	P50	P75	Max	IIQ	Assimetria	Curtose
Mortalidade	88,70 ± 10,00	68,11	81,90	87,33	94,36	132,04	12,46	0,81	4,00
Temperatura	74,26 ± 9,01	50,91	67,23	74,06	81,49	99,88	14,25	0,09	2,55
Partículas	47,41 ± 15,14	20,25	35,85	44,25	57,54	97,94	21,69	0,57	2,54

Nota-se que a média da variável mortalidade cardiovascular é 88,70 com um desvio padrão de 10, no período de 1970 a 1979 o mínimo dessa variável foi atingido com 68,11 e máximo de 132,04. Pode-se avaliar uma assimetria positiva de 0,81 e curtose de 4,00. Em relação à variável temperatura tem-se uma média de 74,26°F com um desvio padrão de 9,01 °F, no período de 1970 a 1979 o mínimo dessa variável foi atingido com 50,91°F e máximo de 99,88°F. Pode-se avaliar uma assimetria positiva de 0,09 e curtose de 2,55. E, por fim, temos a variável partículas de poluição com média de concentração, 47,41, e desvio padrão de 15,14, registrou o mínimo no período de 20,25, e máximo de concentração com 97,94.

A Figura 8.18 representa as três séries ao longo dos anos dos possíveis efeitos da temperatura e poluição na mortalidade cardiovascular média semanal. Note-se a forte componente sazonal em todas as séries, correspondendo às variações inverno-verão e à tendência de redução da mortalidade cardiovascular no período de 10 anos, como já discutido anteriormente.

⁴https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1

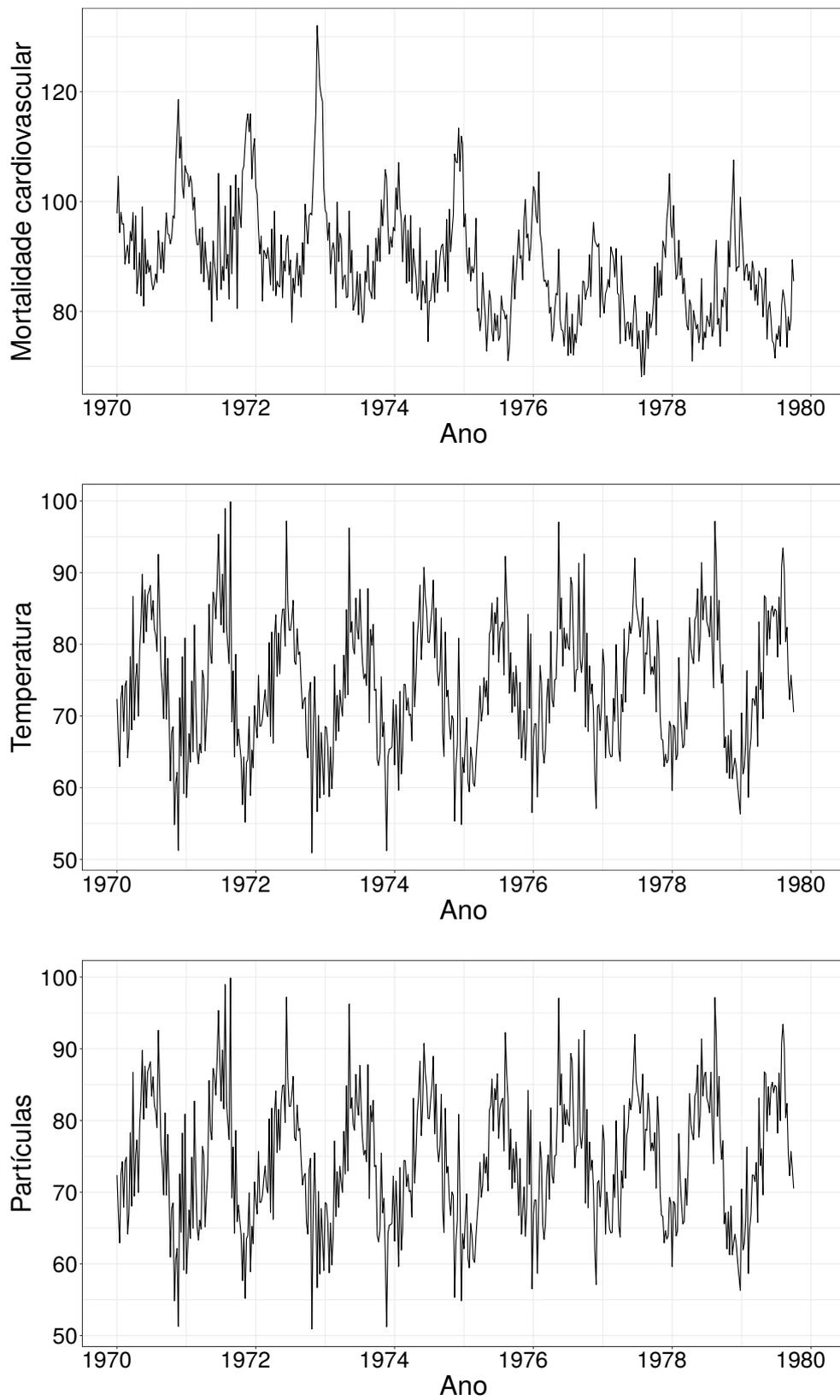


Figura 8.18: Mortalidade cardiovascular média semanal (parte superior), temperatura (parte central) e poluição por partículas de poluição (parte inferior) no Condado de Los Angeles. Existem 508 médias suavizadas de seis dias obtidas pela filtragem de valores diários ao longo do período de 10 anos 1970-1979.

O gráfico de dispersão na Figura 8.19 indica uma possível relação linear entre a mortalidade e as partículas de poluição e uma possível relação quadrática com a temperatura. Observe a forma curvilínea da curva de mortalidade por temperatura, indicando que temperaturas mais altas, bem como temperaturas mais baixas, estão associadas a aumentos na mortalidade cardiovascular. Tem-se que o coeficiente de correlação entre a variável mortalidade, temperatura e partículas é $-0,4386$ e $0,4439$, respectivamente, enquanto que temperatura e partículas de poluição é $-0,0172$.

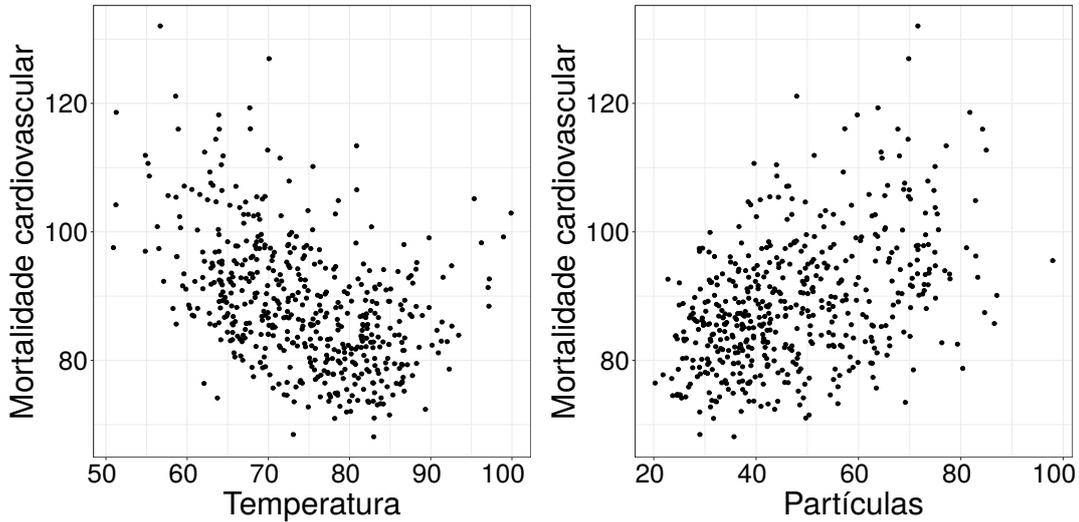


Figura 8.19: Diagrama de dispersão mostrando a relação entre mortalidade cardiovascular média semanal e as variáveis temperatura (primeiro) e partículas de poluição (segundo) no Condado de Los Angeles.

8.4.2 Modelagem

Baseado nos diagramas de dispersão, consideramos alguns modelos em que M_i denota a mortalidade cardiovascular, T_i denota a temperatura e P_i denota os níveis de partículas, representados da seguinte maneira:

$$\begin{aligned}
 M_i &= \beta_0 + \beta_1 T_i + \beta_2 P_i + f_T(\text{tempo}_i) + f_S(\text{semana.do.ano}_i) + \epsilon_i, \\
 M_i &= \beta_0 + \beta_1 T_i + \beta_2 P_i + \beta_3 \text{tempo}_i + f_S(\text{semana.do.ano}_i) + \epsilon_i, \\
 M_i &= \beta_0 + \beta_1 T_i + \beta_2 T_i^2 + \beta_3 P_i + f_T(\text{tempo}_i) + f_S(\text{semana.do.ano}_i) + \epsilon_i, \\
 M_i &= \beta_0 + \beta_1 T_i + \beta_2 T_i^2 + \beta_3 P_i + \beta_4 \text{tempo}_i + f_S(\text{semana.do.ano}_i) + \epsilon_i, \\
 M_i &= \beta_0 + f(T_i) + \beta_1 P_i + f_T(\text{tempo}_i) + f_S(\text{semana.do.ano}_i) + \epsilon_i, \\
 M_i &= \beta_0 + f(T_i) + \beta_1 P_i + \beta_2 \text{tempo}_i + f_S(\text{semana.do.ano}_i) + \epsilon_i,
 \end{aligned} \tag{8.5}$$

em que β_0 é o intercepto do modelo, β_1, β_2 e β_3 são os parâmetros a serem estimados, enquanto que $f(\cdot)$ representa uma função não paramétrica para T_i , $f_T(\cdot)$ e $f_S(\cdot)$ são funções suaves aproximadas por *spline* de regressão cúbica e de regressão cúbica cíclica, respectivamente, que representam a tendência e sazonalidade na série em estudo. O termo tempo_i é o vetor de tempos do i -ésimo ano ao longo das semanas, semana.do.ano_i indica a semana do ano respectivo ao i -ésimo ano, e ϵ_i são erros simétricos autorregressivos de ordem p , conforme discutido no Capítulo 4, para $i = 1, \dots, 508$. A variável semana.do.ano varia de 1 a 52. Para o nosso modelo, consideramos 5, 8 e 9 nós fixos para a temperatura, o tempo e a semana do ano, respectivamente, aos quais foram obtidos a partir dos quantis da variável tempo. A homogeneidade da sazonalidade também é assumida ao longo dos anos. Os valores selecionados para os parâmetros de suavização foram $(\lambda; \lambda_T; \lambda_S) = (10; 50; 10)$.

Após diversas análises chegamos na conclusão que o melhor modelo é dado por

$$M_i = \beta_0 + \beta_1 T_i + \beta_2 T_i^2 + \beta_3 P_i + \beta_4 \text{tempo}_i + f_S(\text{semana.do.ana}_i) + \epsilon_i, \quad (8.6)$$

em que várias distribuições foram ajustadas com erros autorregressivos de ordem um e dois. Os resultados desses modelos podem ser vistos na Tabela 8.7. Note que o modelo com o melhor ajuste foi realizado com a distribuição dos erros t -Student e autorregressivo de ordem dois.

Tabela 8.7: Qualidade de ajuste para o modelo (8.6) sob erros com distribuição normal, t -Student, exponencial potência e Logística-II ajustados à mortalidade cardiovascular média semanal no condado de Los Angeles, entre 1970 e 1979.

AR	Model	AIC(λ)	VCG(λ)	df(λ)	df(λ_S)
1	N	3125,7155	27,5306	6,9731	1,9731
	t_{16}	3122,0541	24,0499	7,1217	2,1217
	EP _{0,2}	3122,3114	19,1752	7,3712	2,3712
	L ₂	3123,9089	8,6638	7,2906	2,2906
2	N	3063,9311	24,3774	6,6315	1,6315
	t_{16}	3060,8482	21,3520	6,7758	1,7758
	EP _{0,2}	3061,9853	17,0290	7,0195	2,0195
	L ₂	3064,1966	7,7275	6,9253	1,9253

A Tabela 8.8 representa a estimativa dos parâmetros e o erro padrão aproximado (em parênteses) do modelo (8.6) sob erros com distribuições normal, t -Student, exponencial potência e Logística - II com estrutura AR(2) ajustado à mortalidade cardiovascular média semanal no condado de Los Angeles, entre 1970 e 1979. Note que as estimativas são bem próximas para todos os parâmetros, exceto para o parâmetro de dispersão.

Tabela 8.8: Estimativas dos parâmetros (erros padrão aproximados) do modelo (8.6) sob erros com distribuições normal, t -Student, exponencial potência e Logística-II com estrutura AR(2) ajustado à mortalidade cardiovascular média semanal no condado de Los Angeles, entre 1970 e 1979.

Variável	N	t_{16}	EP _{0,2}	L ₂
$\hat{\beta}_0$	3042,3397 (436,9003)	3019,3835 (418,5490)	3094,4664 (426,3001)	3021,2603 (409,8690)
tempo	-1,4551 (0,2213)	-1,4445 (0,2120)	-1,4826 (0,2159)	-1,4458 (0,2076)
part	0,1471 (0,0248)	0,1464 (0,0245)	0,1459 (0,0244)	0,1477 (0,0242)
temp	-2,3686 (0,3053)	-2,3207 (0,3029)	-2,3230 (0,3009)	-2,3099 (0,2987)
temp ²	0,0159 (0,0020)	0,0156 (0,0020)	0,0157 (0,0020)	0,0155 (0,0020)
$\hat{\rho}_1$	0,2921 (0,0412)	0,2868 (0,0419)	0,2984 (0,0412)	0,2892 (0,0418)
$\hat{\rho}_2$	0,3705 (0,0413)	0,3640 (0,0409)	0,3610 (0,0406)	0,3592 (0,0406)
$\hat{\phi}$	23,7560 (1,4906)	20,7943 (1,4218)	16,5682 (1,1388)	7,5204 (0,5581)

O pacote `aplms` desenvolvido em linguagem R e discutido no Capítulo 6 é usado para obter as informações definidas nos Capítulos 3, 4 e 5 a fim de estudar melhor a variável dependente ao longo do tempo. Dessa maneira, o modelo (8.6) pode ser ajustado no R, basta carregar o pacote `aplms` e dispor os comandos da seguinte maneira:

```
> library(aplms)
> library(astsa)
> ajuste <- aplms(cmort~tempr+I(tempr^2)+part+tempo+
  s(semana.do.ano, bs="cc", k=9), p=2,
  family=Stud(16), lam=c(10))
> summary(ajuste)
```

Family: Student(16)

Link function: identity

Formula:

cmort~tempr+I(tempr^2)+part+tempo+s(semana.do.ano, bs="cc",k=9)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3019.3835	418.5490	7.2139	2.057e-12	***
tempr	-2.3207	0.3029	-7.6623	9.732e-14	***
I(tempr^2)	0.0156	0.0020	7.8117	3.414e-14	***
part	0.1464	0.0245	5.9645	4.68e-09	***
tempo	-1.4445	0.2120	-6.8139	2.774e-11	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Chisq	p-value	
s(semana.do.ano)	1.7758	228.8835	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure:

	Estimate	Std. Error	t value	Pr(> t)	
rho1	0.2868	0.0419	6.84810	2.231e-11	***
rho2	0.3640	0.0409	8.91074	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC = 3060.8482 BIC = 3089.5130 GCV = 21.3520

Dispersion est. = 20.7943 n= 508

Após estimar o modelo e definir que o erro com distribuição simétrica t -Student apresenta melhores resultados, vamos à análise de resíduo e sensibilidade do modelo. A Figura 8.20 descreve o histograma, os gráficos de índices, probabilidade normal, a função de autocorrelação, função de autocorrelação parcial dos resíduos quantílicos condicionais e os pesos contra os resíduos quantílicos condicionais do modelo (8.6) sob erros com distribuição t -Student com $\nu = 16$ graus de liberdade ajustados aos dados. Os gráficos de índices não apresentam tendência nem variabilidade ao longo do tempo, o que indica que a variação do erro parece constante e que a tendência e a sazonalidade foram controladas. A partir do gráfico de probabilidade normal, é possível identificar que o modelo de erro de cauda mais pesada é adequado para ajustar os dados. As funções de autocorrelação (FACs) e funções de autocorrelação parcial (FACPs) dos resíduos condicionais parecem confirmar a suposição de independência. E, por fim, nota-se que a distribuição t -Student adota pesos diferentes para as observações.

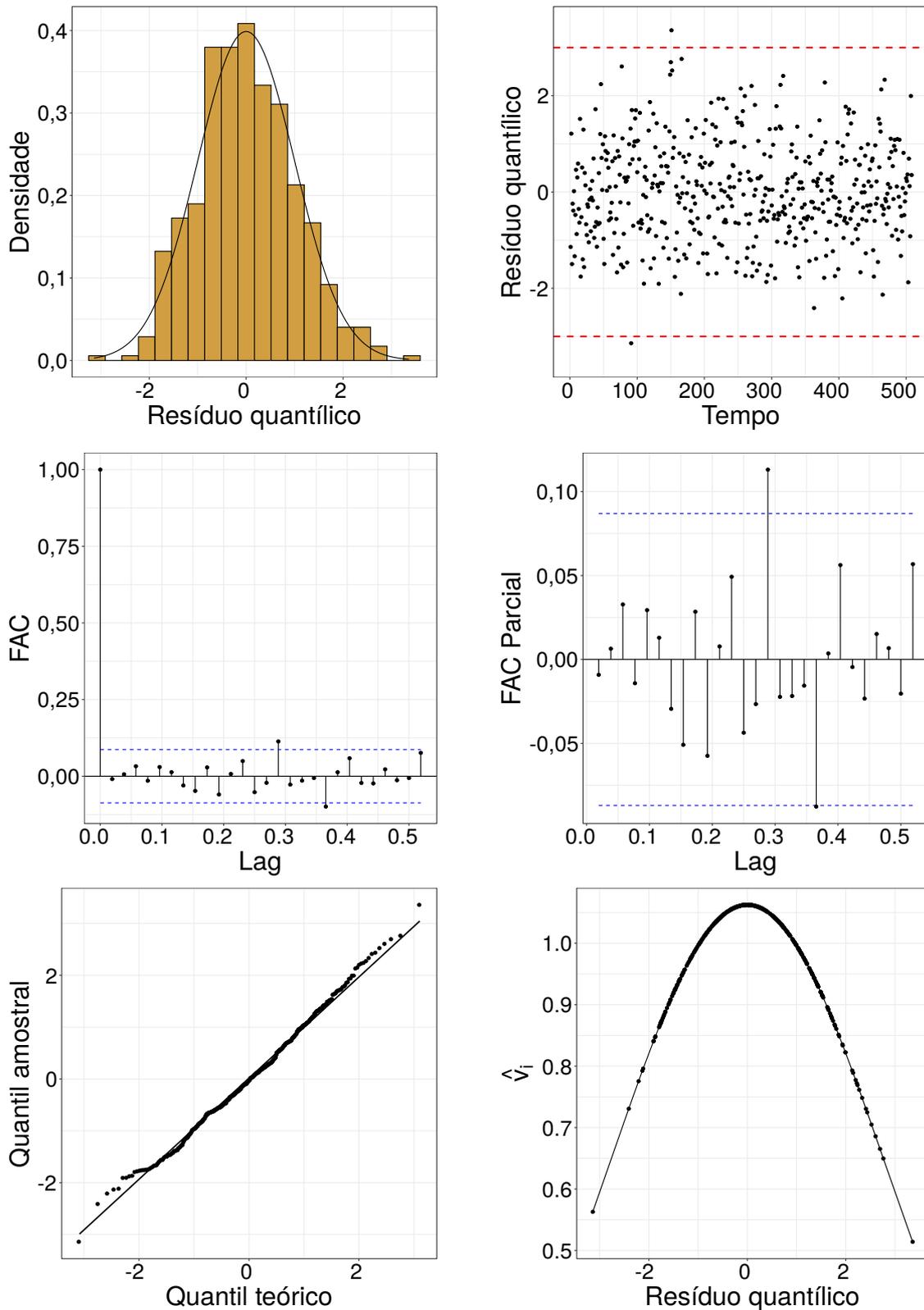


Figura 8.20: Histograma (primeiro-esquerdo), gráficos de índices dos resíduos (primeiro-direito), função de autocorrelação (segundo-esquerdo), função de autocorrelação parcial (segundo-direito), probabilidade normal (terceiro-esquerdo), e peso das observações contra os resíduos quantílicos condicionais do modelo ajustado (8.6) para a mortalidade cardiovascular média semanal sob erros com distribuição t -Student com $\nu = 16$ graus de liberdade.

E, por último, podemos analisar a curvatura conformal das observações para diferentes esquemas de perturbação. A Figura 8.21 representa os esquemas de perturbação ponderação de casos, do parâmetro de dispersão, da variável resposta, das variáveis explicativas (tempo, temperatura e partículas) e os coeficientes de autocorrelação dados por ρ_1 e ρ_2 para o modelo (8.6) definido anteriormente para a mortalidade cardiovascular média semanal. A variável em estudo apresenta poucos pontos que possivelmente são influentes e podem alterar as estimativas do estudo. Os métodos de perturbação dados pelo tempo e pelo coeficiente de autocorrelação ρ_2 apresentam o maior grau de influência.

A Figura 8.22 mostra a sazonalidade da série histórica em estudo, mortalidade cardiovascular média semanal. Note que o período das primeiras semanas (1-7) e últimas semanas do ano (45-52) apresentam maiores índices de mortalidade cardiovascular no condado de Los Angeles.

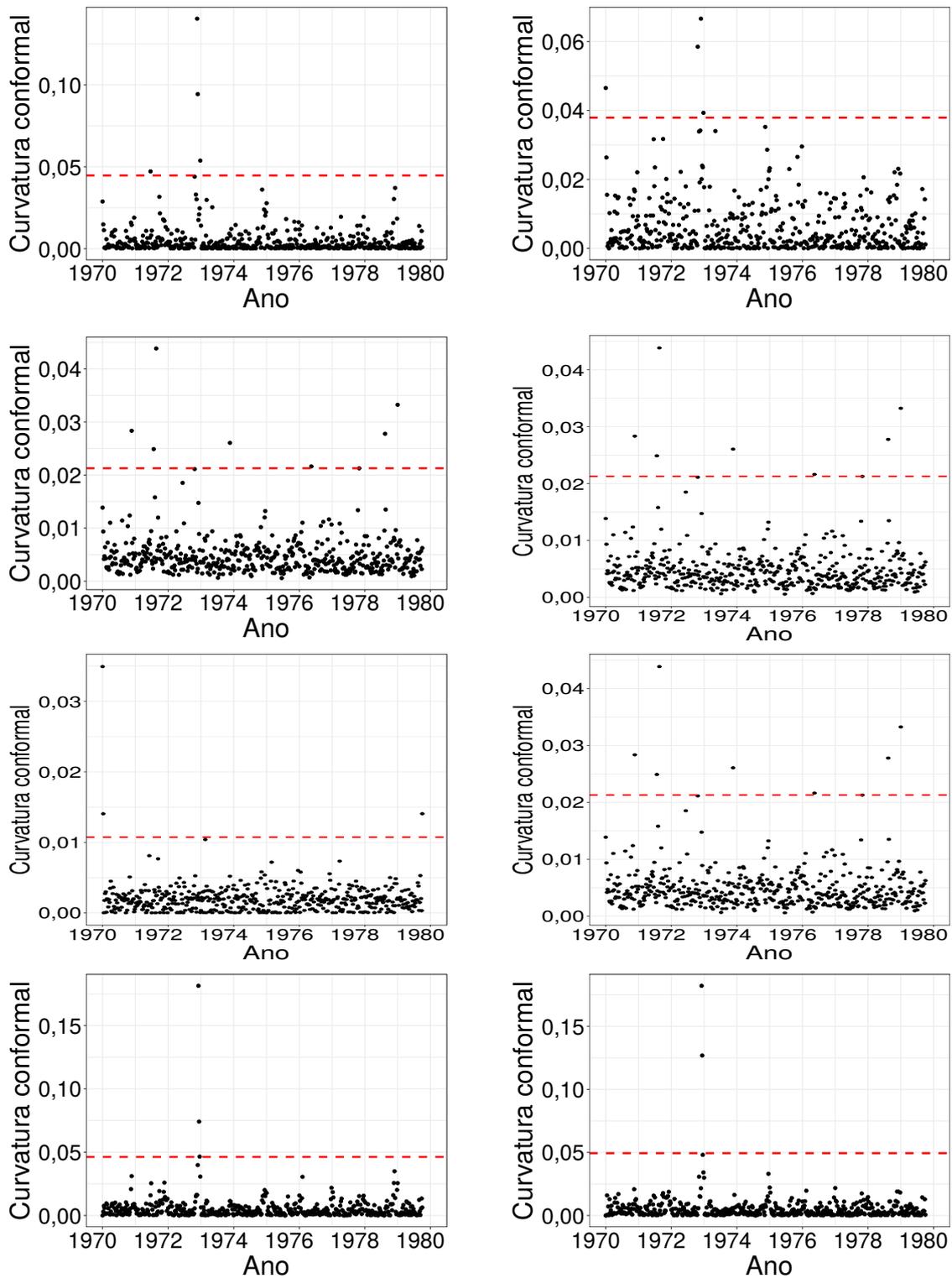


Figura 8.21: Gráficos de índices C_i nos esquemas de perturbação, ponderação de casos (primeiro-esquerdo), parâmetro de dispersão (segundo-esquerdo), variável resposta (terceiro-esquerdo), variável explicativa tempo (quarto-esquerdo), variável explicativa temperatura (primeiro-direito), variável explicativa partículas (segundo-direito), coeficiente de autocorrelação (ρ_1) (terceiro-direito) e (ρ_2) (quarto-direito) do modelo (8.6) para a mortalidade cardiovascular sob erros com distribuição t -Student com $\nu = 16$ graus de liberdade.

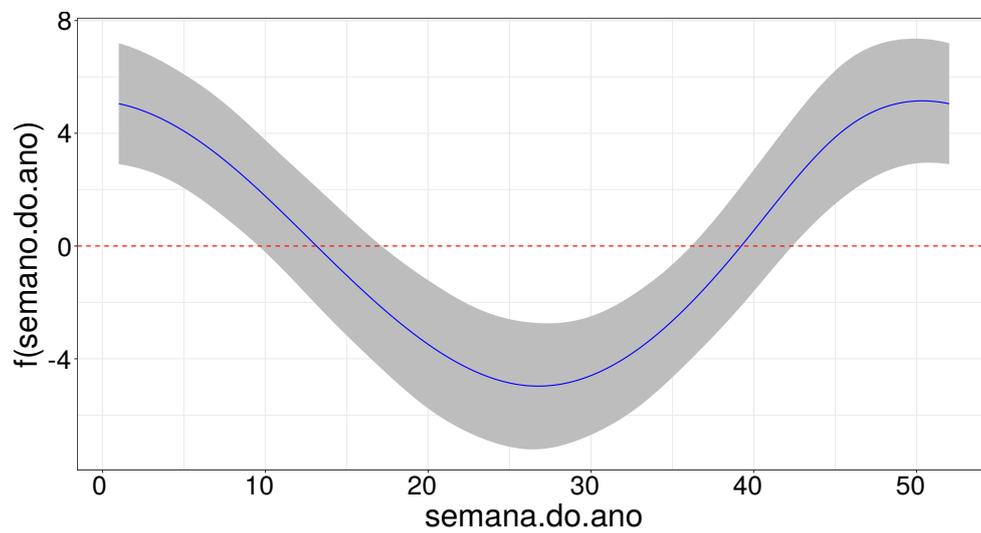


Figura 8.22: Bandas de confiança pontuais para a sazonalidade do modelo (8.6) sob erros com distribuição t -Student com $\nu = 16$ graus de liberdade ajustado à mortalidade cardiovascular média semanal.

Capítulo 9

Conclusões

9.1 Considerações finais

Nesta tese apresentamos inicialmente uma revisão dos modelos paramétricos e semiparamétricos, em consonância aos erros simétricos sem correlação e autocorrelacionados de ordem um. Em seguida, propomos os modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p , em que o componente não paramétrico pode ser expresso em uma combinação linear. Para tanto, usamos os *splines* cúbicos penalizados de regressão, muitas vezes, é apropriado que uma função suave do modelo seja “cíclica”, portanto construímos a versão cíclica do *spline* cúbico penalizado de regressão. Essas funções de suavização exigem uma forma diferente de penalização no logaritmo da função de verossimilhança, a qual foi considerada. Apresentamos a função escore penalizada, a matriz Hessiana penalizada e a matriz de informação de Fisher penalizada do modelo. A combinação dos processos iterativos *backfitting* e escore de Fisher foram implementados para estimarem os parâmetros do modelo. Alguns resultados assintóticos estão disponíveis para derivar um estimador consistente da matriz de covariância-variância para os estimadores paramétricos e não paramétricos.

Procedimentos usuais de diagnóstico tais como resíduos quantílicos são descritos, e a influência local das curvaturas conformais são derivados para cinco esquemas de perturbação, caso ponderado, dispersão, variável resposta, variável explicativa e coeficiente de autocorrelação. Estudos de simulação foram desenvolvidos para avaliar as distribuições empíricas dos estimadores paramétrico e não paramétrico e quatro conjuntos de dados foram analisados pelos procedimentos desenvolvidos neste trabalho. Finalmente, um pacote denominado como `aplms` foi produzido na linguagem R para ajustar os modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p .

9.2 Sugestões para pesquisas futuras

Existem diversas possibilidades de extensão para este trabalho. Seguem algumas delas:

- Nesse trabalho consideramos o modelo com erros autorregressivos de ordem p , podemos estender e desenvolver o modelo com erros autorregressivos e de médias móveis, ou seja, $ARMA(p, q)$.
- Assumir que os erros são heterocedásticos, ou seja, considere o parâmetro de dispersão ϕ_i variando ao longo do tempo, dessa forma, podemos desenvolver os modelos $ARCH(p)$.
- Explorar mais propriedades dos modelos lineares parciais aditivos com erros simétricos.
- Desenvolver testes para verificar se a sazonalidade é constante ao longo da série histórica.
- Verificar se os esquemas de perturbação são apropriados.

Artigo publicado

Oliveira, R.A., Paula, G.A. Additive models with autoregressive symmetric errors based on penalized regression splines. *Comput Stat* (2021). <https://doi.org/10.1007/s00180-021-01106-2>

Artigos em desenvolvimento

Oliveira, R. A. and Paula, G. A. (2021). The `aplms` package.

Oliveira, R. A. and Paula, G. A. (2021). Modeling and forecast global land-ocean temperature anomaly by light tailed symmetric distributions.

Apêndice A

Função escore, matriz Hessiana e informação de Fisher: caso AR(1)

A.1 Função escore penalizada

Seja $L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ o logaritmo da função de verossimilhança penalizada do vetor paramétrico $\boldsymbol{\theta} = (\boldsymbol{\gamma}_T^\top, \boldsymbol{\gamma}_S^\top, \phi, \rho)^\top$ denotado por,

$$L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = -\frac{n}{2} \log(\phi) + \sum_{i=1}^n \log\{g(\delta_i)\} - \frac{\lambda_T}{2} \boldsymbol{\gamma}_T^\top \mathbf{M}_T \boldsymbol{\gamma}_T - \frac{\lambda_S}{2} \boldsymbol{\gamma}_S^\top \mathbf{M}_S \boldsymbol{\gamma}_S,$$

em que $\delta_i = \frac{(\epsilon_i - \rho\epsilon_{i-1})^2}{\phi}$, $\epsilon_i = y_i - \mathbf{n}_T(t_i)^\top \boldsymbol{\gamma}_T - \mathbf{n}_S(s_i)^\top \boldsymbol{\gamma}_S$ e $\epsilon_0 = 0$, para $i = 1, \dots, n$.

A função escore penalizada para ϕ e ρ são, respectivamente, dadas por

$$U_p^\phi = \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi} = -\frac{1}{2\phi} \left\{ n + \sum_{i=1}^n 2W_g(\delta_i)\delta_i \right\}$$

e

$$U_p^\rho = \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho} = -\frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i)(\epsilon_i - \rho\epsilon_{i-1})(\epsilon_{i-1}).$$

Além disso, as derivadas de $L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ com respeito a γ_{T_j} e γ_{S_l} produz em

$$\begin{aligned} U_p^{\gamma_{T_j}} &= \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{T_j}} \\ &= -\frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i)(\epsilon_i - \rho\epsilon_{i-1})(n_{ij_T} - \rho n_{(i-1)j_T}) - \lambda_T [\mathbf{M}_T \boldsymbol{\gamma}_T]_j, \quad j = (1, \dots, r_T) \end{aligned}$$

e

$$\begin{aligned} U_p^{\gamma_{S_l}} &= \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{S_l}} \\ &= -\frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i)(\epsilon_i - \rho\epsilon_{i-1})(n_{iS_l} - \rho n_{(i-1)S_l}) - \lambda_S [\mathbf{M}_S \boldsymbol{\gamma}_S]_l, \quad l = (1, \dots, r_S), \end{aligned}$$

em que $[\mathbf{M}_T \boldsymbol{\gamma}_T]_j$ e $[\mathbf{M}_S \boldsymbol{\gamma}_S]_l$ denota a j -ésima e l -ésima posição dos vetores $\mathbf{M}_T \boldsymbol{\gamma}_T$ e $\mathbf{M}_S \boldsymbol{\gamma}_S$, respectivamente.

Na notação matricial obtém-se

$$\begin{aligned} \mathbf{U}_p^{\gamma_T} &= \frac{\partial \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_T} = \frac{1}{\phi} (\mathbf{A}\mathbf{N}_T)^\top \mathbf{D}_v \mathbf{A} \boldsymbol{\epsilon} - \lambda_T \mathbf{M}_T \boldsymbol{\gamma}_T, \\ \mathbf{U}_p^{\gamma_S} &= \frac{\partial \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_S} = \frac{1}{\phi} (\mathbf{A}\mathbf{N}_S)^\top \mathbf{D}_v \mathbf{A} \boldsymbol{\epsilon} - \lambda_S \mathbf{M}_S \boldsymbol{\gamma}_S, \\ \mathbf{U}_p^\phi &= \frac{\partial \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi} = \frac{1}{2\phi} \mathbf{1}_n^\top (\mathbf{D}_m \mathbf{1}_n - \mathbf{1}_n) \quad \text{e} \\ \mathbf{U}_p^\rho &= \frac{\partial \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho} = -\frac{1}{\phi} (\mathbf{B}\boldsymbol{\epsilon})^\top \mathbf{D}_v \mathbf{A} \boldsymbol{\epsilon}, \end{aligned}$$

em que as quantidades $\mathbf{N}_T, \mathbf{N}_S, \boldsymbol{\epsilon}, \mathbf{D}_v, \mathbf{D}_m, \mathbf{A}$ e \mathbf{B} foram definidas na Seção 3.

A.2 Matriz Hessiana penalizada

Por simplicidade de notação, considera-se $n_{ij_T} = n_{T_j}(t_i)$ e $n_{i_S} = n_{S_l}(t_i)$, ($i = 1, \dots, n, j = 1, \dots, r_T$ e $l = 1, \dots, r_S$).

Considere os parâmetros $(\gamma_{T_j}, \gamma_{T_{j'}})$ para qual obtém-se as derivadas

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\gamma_{T_j}, \gamma_{T_{j'}}} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{T_j} \partial \gamma_{T_{j'}}} = 4\phi^{-1} \sum_{i=1}^n W'_g(\delta_i) \delta_i \left\{ (n_{ij_T} - \rho n_{(i-1)j_T})(n_{ij'_T} - \rho n_{(i-1)j'_T}) \right\} \\ &+ 2\phi^{-1} \sum_{i=1}^n W_g(\delta_i) (n_{ij_T} - \rho n_{(i-1)j_T})(n_{ij'_T} - \rho n_{(i-1)j'_T}) - \lambda_T [\mathbf{M}_T]_{jj'}, \\ &(j, j' = 1, \dots, r_T). \end{aligned}$$

Em notação matricial, obtém-se

$$\ddot{\mathbf{L}}_p^{\gamma_T \gamma_T} = \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_T \partial \boldsymbol{\gamma}_T^\top} = \frac{1}{\phi} \left\{ (\mathbf{A}\mathbf{N}_T)^\top (-\mathbf{D}_v + 4\mathbf{D}_d) (\mathbf{A}\mathbf{N}_T) \right\} - \lambda_T \mathbf{M}_T.$$

Analogamente para o vetor paramétrico $\boldsymbol{\gamma}_S$ tem-se

$$\ddot{\mathbf{L}}_p^{\gamma_S \gamma_S} = \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_S \partial \boldsymbol{\gamma}_S^\top} = \frac{1}{\phi} \left\{ (\mathbf{A}\mathbf{N}_S)^\top (-\mathbf{D}_v + 4\mathbf{D}_d) (\mathbf{A}\mathbf{N}_S) \right\} - \lambda_S \mathbf{M}_S.$$

A segunda derivada de $\mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ com respeito a ϕ e ρ produz

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\phi\phi} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi^2} = \frac{n}{2\phi^2} + \frac{2}{\phi^2} \sum_{i=1}^n W_g(\delta_i) \delta_i + \frac{1}{\phi^2} \sum_{i=1}^n W'_g(\delta_i) \delta_i^2 \\ &= \frac{1}{\phi^2} \left\{ \frac{n}{2} + \boldsymbol{\delta}^\top \mathbf{D}_c \boldsymbol{\delta} - \boldsymbol{\delta}^\top \mathbf{D}_v \mathbf{1}_n \right\} \end{aligned}$$

e

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\rho\rho} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho^2} = \frac{4}{\phi} \sum_{i=1}^n W'_g(\delta_i) \delta_i \epsilon_{i-1}^2 + \frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i) \epsilon_{i-1}^2 \\ &= \frac{1}{\phi} \left\{ (\mathbf{B}\boldsymbol{\epsilon})^\top (-\mathbf{D}_v + 4\mathbf{D}_d) (\mathbf{B}\boldsymbol{\epsilon}) \right\}, \end{aligned}$$

em que $\mathbf{D}_c = \text{diag}\{c_1, \dots, c_n\}$ com $c_i = W'_g(\delta_i)$ e $\mathbf{D}_d = \text{diag}\{d_1, \dots, d_n\}$ com $d_i = \delta_i W'_g(\delta_i)$.

As derivadas de $L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ com respeito a $(\gamma_{T_j}, \gamma_{S_l})$ produz em

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\gamma_{T_j} \gamma_{S_l}} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{T_j} \partial \gamma_{S_l}^\top} = \frac{1}{\phi} \sum_{i=1}^n (n_{ij_{T_j}} - \rho n_{(i-1)j_{T_j}})(n_{il_{S_l}} - \rho n_{(i-1)l_{S_l}}) \times \\ &\times \{4W'_g(\delta_i)\delta_i + 2W_g(\delta_i)\}, \quad (j = 1, \dots, r_T) \quad \text{e} \quad (l = 1, \dots, r_S). \end{aligned}$$

Em notação matricial, obtém-se

$$\ddot{\mathbf{L}}_p^{\gamma_T \gamma_S} = \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_T \partial \boldsymbol{\gamma}_S^\top} = \frac{1}{\phi} \left\{ (\mathbf{AN}_T)^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{AN}_S) \right\}.$$

Para as derivadas de $L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ com respeito a (γ_{T_j}, ϕ) and (γ_{T_j}, ρ) , obtém-se

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\gamma_{T_j} \phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{T_j} \partial \phi} \\ &= \phi^{-2} \left[\sum_{i=1}^n (n_{ij_{T_j}} - \rho n_{(i-1)j_{T_j}}) \{2W_g(\delta_i) + 2W'_g(\delta_i)\delta_i\} (\epsilon_i - \rho \epsilon_{i-1}) \right] \end{aligned}$$

e

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\gamma_{T_j} \rho} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{T_j} \partial \rho} \\ &= \frac{1}{\phi} \left[\sum_{i=1}^n \epsilon_{i-1} \{4W'_g(\delta_i)\delta_i + 2W_g(\delta_i)\} (n_{ij_{T_j}} - \rho n_{(i-1)j_{T_j}}) \right] \\ &\quad + \frac{1}{\phi} \left\{ 2 \sum_{i=1}^n n_{(i-1)j_{T_j}} W_g(\delta_i) (\epsilon_i - \rho \epsilon_{i-1}) \right\}, \end{aligned}$$

que na forma matricial podem ser expressas como

$$\ddot{\mathbf{L}}_p^{\gamma_T \phi} = \frac{1}{\phi} \left\{ (\mathbf{AN}_T)^\top (2\mathbf{D}_d - \mathbf{D}_v) (\mathbf{A}\boldsymbol{\epsilon}) \right\}$$

e

$$\ddot{\mathbf{L}}_p^{\gamma_T \rho} = \frac{1}{\phi} \left\{ (\mathbf{AN}_T)^\top (\mathbf{D}_v - 4\mathbf{D}_d) (\mathbf{B}\boldsymbol{\epsilon}) + (\mathbf{BN}_T)^\top \mathbf{D}_v \mathbf{A}\boldsymbol{\epsilon} \right\}.$$

Similarmente, as derivadas de $L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ com respeito a (γ_{S_l}, ϕ) e (γ_{S_l}, ρ) produz em

$$\ddot{\mathbf{L}}_p^{\gamma_{S_l} \phi} = \frac{1}{\phi} \left\{ (\mathbf{AN}_S)^\top (2\mathbf{D}_d - \mathbf{D}_v) (\mathbf{A}\boldsymbol{\epsilon}) \right\}$$

e

$$\ddot{\mathbf{L}}_p^{\gamma_{S_l} \rho} = \frac{1}{\phi} \left\{ (\mathbf{AN}_S)^\top (\mathbf{D}_v - 4\mathbf{D}_d) (\mathbf{B}\boldsymbol{\epsilon}) + (\mathbf{BN}_S)^\top \mathbf{D}_v \mathbf{A}\boldsymbol{\epsilon} \right\}.$$

Finalmente, para a derivada $L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ com respeito a (ϕ, ρ) obtém-se

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\phi \rho} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi \partial \rho} = \frac{1}{\phi^2} \left[\sum_{i=1}^n \epsilon_{i-1} \{2W'_g(\delta_i)\delta_i + 2W_g(\delta_i)\} (\epsilon_i - \rho \epsilon_{i-1}) \right] \\ &= \frac{1}{\phi^2} \left\{ (\mathbf{B}\boldsymbol{\epsilon})^\top (\mathbf{D}_v - 2\mathbf{D}_d) (\mathbf{A}\boldsymbol{\epsilon}) \right\}. \end{aligned}$$

A.3 Matriz informação de Fisher penalizada

Aqui, deriva-se a matriz de informação de Fisher penalizada. Similarmente a [Relvas e Paula \(2016\)](#), usa-se as condições de regularidade aplicadas para cada função log-verossimilhança regular $L(\boldsymbol{\theta})$, nomeadamente $E\{\partial L(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\} = \mathbf{0}$ e $E\{\partial^2 L(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top\} = -E\{[\partial L(\boldsymbol{\theta})/\partial \boldsymbol{\theta}] [\partial L(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\top]\}$, e os resultados $f_g = E\{W_g^2(z^2)z^4\}$ e $d_g = E\{W_g^2(z^2)z^2\}$ com $z \sim S(0, 1)$, $E(\epsilon_i) = 0$ e $\text{Var}(\epsilon_i) = \phi \xi(1 - \rho^{2i})/(1 - \rho^2)$, para $i = 1, \dots, n$.

Para os parâmetros $\boldsymbol{\gamma}_T$ segue-se que

$$\mathcal{I}_p^{\boldsymbol{\gamma}_T \boldsymbol{\gamma}_T} = -E \left\{ \frac{1}{\phi} (\mathbf{AN}_T)^\top (-\mathbf{D}_v + 4\mathbf{D}_d) (\mathbf{AN}_T) - \lambda_T \mathbf{M}_T \right\}.$$

Pode-se mostrar que $E(-\mathbf{D}_v + 4\mathbf{D}_d) = -4d_g$ e conseqüentemente a matriz de informação de Fisher penalizada para $\boldsymbol{\gamma}_T$ fica dada por

$$\mathcal{I}_p^{\boldsymbol{\gamma}_T \boldsymbol{\gamma}_T} = \frac{4d_g}{\phi} (\mathbf{AN}_T)^\top (\mathbf{AN}_T) + \lambda_T \mathbf{M}_T.$$

Similarmente, obtém-se

$$\mathcal{I}_p^{\boldsymbol{\gamma}_S \boldsymbol{\gamma}_S} = \frac{4d_g}{\phi} (\mathbf{AN}_S)^\top (\mathbf{AN}_S) + \lambda_S \mathbf{M}_S.$$

Pelas condições de regularidade $E(\mathbf{U}_\theta^\phi) = 0$ obtém-se $E\{W_g(\delta_i)\delta_i\} = -\frac{1}{2}, \forall i$. Então,

$$\begin{aligned} E \left\{ \left(\frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi} \right)^2 \right\} &= E \left\{ \left(\frac{1}{2\phi} + \frac{1}{\phi} W_g(\delta_i)\delta_i \right)^2 \right\} \\ &= \frac{1}{\phi^2} \left(f_g - \frac{1}{4} \right), \quad \forall i, \end{aligned}$$

em que $L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ denota o i -ésimo elemento da função de log-verossimilhança penalizada. Portanto, obtém-se $\mathbf{I}_p^{\phi\phi} = \frac{n}{4\phi^2}(4f_g - 1)$.

Para o parâmetro ρ tem-se que

$$\left\{ \left(\frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho} \right)^2 \right\} = \left\{ \left(\frac{2}{\phi} W_g(\delta_i)(\epsilon_i - \rho\epsilon_{i-1})(\epsilon_{i-1}) \right)^2 \right\} = \frac{4}{\phi} W_g^2(\delta_i)\delta_i\epsilon_{i-1}^2, \quad \forall i.$$

Assim,

$$E \left(\frac{4}{\phi} W_g^2(\delta_i)\delta_i\epsilon_{i-1}^2 \right) = E \left[E \left\{ \frac{4}{\phi} (W_g^2(\delta_i)\delta_i\epsilon_{i-1}^2) \middle| y_{i-1} \right\} \right] = E \left(\frac{4}{\phi} d_g \epsilon_{i-1}^2 \right),$$

e por usar os momentos de ϵ_i , obtém-se

$$\begin{aligned} \mathbf{I}_p^{\rho\rho} &= \frac{4}{\phi} d_g \sum_{i=2}^n E(\epsilon_{i-1}^2) = \frac{4}{\phi} d_g \sum_{i=1}^{n-1} E(\epsilon_i^2) = 4d_g \xi \sum_{i=1}^{n-1} \frac{(1 - \rho^{2i})}{(1 - \rho^2)} \\ &= \frac{4d_g \xi}{1 - \rho^2} \left\{ (n-1) - \frac{\rho^2(1 - \rho^{2(n-1)})}{1 - \rho^2} \right\}. \end{aligned}$$

Conseqüentemente, a matriz informação de Fisher penalizada para $(\boldsymbol{\gamma}_T^\top, \boldsymbol{\gamma}_S^\top)^\top$ fica dado por

$$\mathcal{I}_p^{\boldsymbol{\gamma}_T \boldsymbol{\gamma}_S} = -E \left\{ \frac{1}{\phi} (\mathbf{AN}_T)^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{AN}_S) \right\} = \frac{4d_g}{\phi} \left\{ (\mathbf{AN}_T)^\top (\mathbf{AN}_S) \right\},$$

e pode-se ver que γ_T e γ_S não são ortogonais.

A partir das propriedades da distribuição simétrica tem-se que

$$E \left\{ (W_g(\delta_i) = W'_g(\delta_i)\delta_i) (\epsilon_i - \rho\epsilon_{i-1}) \mid y_{i-1} \right\} = 0.$$

Então, pode-se obter as seguintes matrizes informação de Fisher penalizada:

$$\begin{aligned} \mathcal{I}_p^{\phi\gamma_T} &= -E \left[\frac{1}{\phi^2} \left\{ (\mathbf{AN}_T)^\top (-\mathbf{D}_v + 2\mathbf{D}_d)(\mathbf{A}\epsilon) \right\} \right] = \mathbf{0}, \\ \mathcal{I}_p^{\phi\gamma_S} &= -E \left[\frac{1}{\phi^2} \left\{ (\mathbf{AN}_S)^\top (-\mathbf{D}_v + 2\mathbf{D}_d)(\mathbf{A}\epsilon) \right\} \right] = \mathbf{0} \text{ e} \\ \mathcal{I}_p^{\phi\rho} &= -E \left[\frac{1}{\phi^2} \left\{ (\mathbf{B}\epsilon)^\top (\mathbf{D}_v - 2\mathbf{D}_d)(\mathbf{A}\epsilon) \right\} \right] = 0. \end{aligned}$$

A partir $E(\mathbf{U}_p^{\gamma_T}) = \mathbf{0}$ segue-se que

$$E \left\{ \frac{1}{\phi} (\mathbf{AN}_T)^\top \mathbf{D}_v \mathbf{A}\epsilon \right\} = \mathbf{0},$$

então $E(\mathbf{D}_v \mathbf{A}\epsilon) = \mathbf{0}$, deste modo obtém-se

$$\begin{aligned} \mathcal{I}_p^{\rho\gamma_T} &= -\frac{1}{\phi} E \left\{ (\mathbf{B}\epsilon)^\top (\mathbf{D}_v - 4\mathbf{D}_d)(\mathbf{AN}_T) + (\mathbf{BN}_T)^\top \mathbf{D}_v \mathbf{A}\epsilon \right\} \\ &= -\frac{1}{\phi} E \left\{ (\mathbf{B}\epsilon)^\top (\mathbf{D}_v - 4\mathbf{D}_d)(\mathbf{AN}_T) \right\}, \end{aligned}$$

e desde que $E\{(\mathbf{B}\epsilon)^\top\} = \mathbf{0}$ tem-se que

$$\begin{aligned} \mathcal{I}_p^{\rho\gamma_T} &= -E \left[E \left\{ \frac{1}{\phi} \left((\mathbf{B}\epsilon)^\top (\mathbf{D}_v - 4\mathbf{D}_d) \mathbf{AN}_T \right) \mid y_{i-1} \right\} \right] \\ &= E \left\{ \frac{1}{\phi} (\mathbf{B}\epsilon)^\top 4d_g(\mathbf{AN}_T) \right\} = \mathbf{0}. \end{aligned}$$

Similarmente, pode-se mostrar que $\mathcal{I}_p^{\rho\gamma_S} = \mathbf{0}$.

Apêndice B

Função escore, Matriz Hessiana e informação de Fisher: caso AR(p)

Neste Apêndice derivamos alguns resultados para os modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p discutidos na Seção 4, a função escore, a matriz Hessiana e a matriz informação de Fisher penalizadas .

B.1 Função escore penalizada

Na estatística, a função escore, sob condições de regularidade adequadas, é o gradiente (o vetor de derivadas parciais) do logaritmo da função de verossimilhança em relação ao vetor de parâmetros. Avaliada em um ponto específico do vetor de parâmetro, a função escore indica a inclinação do logaritmo da função de verossimilhança e, portanto, a sensibilidade de pequenas alterações nos valores dos parâmetros. Se o logaritmo da função de verossimilhança for contínua no espaço paramétrico, o escore desaparecerá no máximo ou no máximo local; esse fato é usado na estimativa de máxima verossimilhança para encontrar os valores dos parâmetros que maximizam a função de verossimilhança.

O logaritmo da função de verossimilhança penalizada dos modelos lineares parciais aditivos com erros simétricos condicionais autorregressivos de ordem p com r_0 variáveis paramétricas e k variáveis não paramétricas discutido na Seção 4 é denotado por

$$L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = -\frac{n}{2}\log(\phi) + \sum_{i=1}^n \log[g(\delta_i)] - \frac{\lambda_1}{2}\boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2}\boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k,$$

em que,

$$\delta_i = \frac{(\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p})^2}{\phi}, \quad \epsilon_i = y_i - \eta_{0_i}^\top \boldsymbol{\gamma}_0 - \eta_{1_i}^\top \boldsymbol{\gamma}_1 - \dots - \eta_{k_i}^\top \boldsymbol{\gamma}_k.$$

Considere $\epsilon_0 = 0$, podemos derivar o logaritmo da função de verossimilhança penalizada com respeito aos parâmetros $\boldsymbol{\theta} = (\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_k, \phi, \rho_1, \dots, \rho_p)^\top$ para obter a função escore penalizada da seguinte maneira:

$$\begin{aligned}
\mathbf{U}_p^{\gamma_{0j}} &= \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{0j}} \\
&= -\frac{2}{\phi} \sum_{i=1}^n \frac{g'(\delta_i)}{g(\delta_i)} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{0i_j} - \rho_1 \eta_{0(i-1)_j} - \dots - \rho_p \eta_{0(i-p)_j}) \\
&= -\frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i) (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{0i_j} - \rho_1 \eta_{0(i-1)_j} - \dots - \rho_p \eta_{0(i-p)_j}), \\
\mathbf{U}_p^{\gamma_{lj}} &= \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{lj}} \\
&= -\frac{2}{\phi} \sum_{i=1}^n \frac{g'(\delta_i)}{g(\delta_i)} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{li_j} - \rho_1 \eta_{l(i-1)_j} - \dots - \rho_p \eta_{l(i-p)_j}) - [\lambda_l \mathbf{M}_l \boldsymbol{\gamma}_l]_j \\
&= -\frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i) (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{li_j} - \rho_1 \eta_{l(i-1)_j} - \dots - \rho_p \eta_{l(i-p)_j}) - [\lambda_l \mathbf{M}_l \boldsymbol{\gamma}_l]_j \\
&\quad \forall l = 1, \dots, k \quad \text{e} \quad j = 1, 2, \dots, r_l, \\
\mathbf{U}_p^\phi &= \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi} \\
&= -\frac{n}{2\phi} - \frac{1}{\phi^2} \sum_{i=1}^n \frac{g'(\delta_i)}{g(\delta_i)} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p})^2 = -\frac{1}{2\phi} \left[n + \sum_{i=1}^n 2W_g(\delta_i) \delta_i \right], \\
\mathbf{U}_p^{\rho_j} &= \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho_j} \\
&= -\frac{2}{\phi} \sum_{i=1}^n \frac{g'(\delta_i)}{g(\delta_i)} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\epsilon_{i-j}) \\
&= -\frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i) (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\epsilon_{i-j}), \quad \forall j = 1, \dots, p,
\end{aligned}$$

em que $[\lambda_l \mathbf{M}_l \boldsymbol{\gamma}_l]_j$ representa a j -ésima posição do vetor $[\lambda_l \mathbf{M}_l \boldsymbol{\gamma}_l]$. Sabe-se que $g(\delta)$ identifica a distribuição simétrica a qual está sob estudo e $W_g(\delta) = \frac{g'(\delta)}{g(\delta)}$, para mais detalhes veja o Capítulo 2 e a Tabela 2.3. Para cada parâmetro a função escore pode ser escrita em forma matricial a qual facilita nos cálculos computacionais. Sendo assim, obtemos:

$$\begin{aligned}
\mathbf{U}_p^{\gamma_0} &= \frac{1}{\phi} (\mathbf{A} \mathbf{N}_0)^\top \mathbf{D}_v \mathbf{A} \boldsymbol{\epsilon}, \\
\mathbf{U}_p^{\gamma_l} &= \frac{1}{\phi} (\mathbf{A} \mathbf{N}_l)^\top \mathbf{D}_v \mathbf{A} \boldsymbol{\epsilon} - \lambda_l \mathbf{M}_l \boldsymbol{\gamma}_l, \quad \forall l = 1, \dots, k, \\
\mathbf{U}_p^\phi &= \frac{\mathbf{1}_n^\top}{2\phi} [\mathbf{D}_m \mathbf{1}_n - \mathbf{1}_n], \\
\mathbf{U}_p^{\rho_j} &= -\frac{1}{\phi} (\mathbf{B}_j \boldsymbol{\epsilon})^\top \mathbf{D}_v \mathbf{A} \boldsymbol{\epsilon}, \quad \forall j = 1, \dots, p.
\end{aligned}$$

B.2 Matriz Hessiana penalizada

A matriz Hessiana é uma matriz quadrada de derivadas parciais de segunda ordem de uma função com valor escalar ou campo escalar. Descreve a curvatura local de uma função de muitas variáveis. Suponha que h é uma função com valor real de n variáveis $\delta_1, \delta_2, \dots, \delta_n$. A matriz Hessiana de h é uma função da matriz $(n \times n)$ com valor no domínio de um subconjunto de domínio de h , definido da seguinte forma: a matriz Hessiana em qualquer ponto do domínio é a matriz Jacobiana

do vetor gradiente de h no ponto.

Se todas as derivadas parciais de segunda ordem (puras e mistas) de h são funções contínuas em torno de um ponto específico do domínio, a matriz Hessiana existe e é dada pela expressão da matriz em termos das parciais de segunda ordem. Em particular, se todas as parciais de segunda ordem de h existem e são contínuas em todos os lugares, a matriz Hessiana de h existe em toda parte e é dada pela expressão da matriz em toda parte. O papel principal da continuidade é nos ajudar que, de fato, a matriz Hessiana existe.

No primeiro momento vamos calcular a diagonal principal da matriz Hessiana, ou seja, a segunda derivada pura do logaritmo da função de verossimilhança para cada parâmetro. Dessa forma, temos

$$\begin{aligned}
 \ddot{\mathbf{L}}_p^{\gamma_0 \gamma_{0j'}} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{0j} \partial \gamma_{0j'}} \\
 &= \frac{4}{\phi} \sum_{i=1}^n W'_g(\delta_i) \delta_i (\eta_{0i_j} - \rho_1 \eta_{0(i-1)_j} - \dots - \rho_p \eta_{0(i-p)_j}) (\eta_{0i_{j'}} - \rho_1 \eta_{0(i-1)_{j'}} - \dots - \rho_p \eta_{0(i-p)_{j'}}) \\
 &\quad + \frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i) (\eta_{0i_j} - \rho_1 \eta_{0(i-1)_j} - \dots - \rho_p \eta_{0(i-p)_j}) (\eta_{0i_{j'}} - \rho_1 \eta_{0(i-1)_{j'}} - \dots - \rho_p \eta_{0(i-p)_{j'}}), \\
 &\quad j, j' = 1, 2, \dots, r_0, \\
 \ddot{\mathbf{L}}_p^{\eta_j \eta_{j'}} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \eta_j \partial \eta_{j'}} \\
 &= \frac{4}{\phi} \sum_{i=1}^n W'_g(\delta_i) \delta_i (\eta_{li_j} - \rho_1 \eta_{l(i-1)_j} - \dots - \rho_p \eta_{l(i-p)_j}) (\eta_{li_{j'}} - \rho_1 \eta_{l(i-1)_{j'}} - \dots - \rho_p \eta_{l(i-p)_{j'}}) \\
 &\quad + \frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i) (\eta_{li_j} - \rho_1 \eta_{l(i-1)_j} - \dots - \rho_p \eta_{l(i-p)_j}) (\eta_{li_{j'}} - \rho_1 \eta_{l(i-1)_{j'}} - \dots - \rho_p \eta_{l(i-p)_{j'}}) \\
 &\quad - \lambda_l [\mathbf{M}_l]_{jj'}, \quad \forall l = 1, \dots, k \quad \text{e} \quad j, j' = 1, 2, \dots, r_l, \\
 \ddot{\mathbf{L}}_p^{\phi \phi} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi^2} = \frac{n}{2\phi^2} + \frac{2}{\phi^2} \sum_{i=1}^n W_g(\delta_i) \delta_i + \frac{1}{\phi^2} \sum_{i=1}^n W'_g(\delta_i) \delta_i^2, \\
 \ddot{\mathbf{L}}_p^{\rho_j \rho_j} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho_j^2} = \frac{4}{\phi} \sum_{i=1}^n W'_g(\delta_i) \delta_i (-\epsilon_{i-j})(-\epsilon_{i-j}) + \frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i) (-\epsilon_{i-j})(-\epsilon_{i-j}), \\
 &\quad \forall j = 1, \dots, p.
 \end{aligned}$$

A forma matricial das expressões anteriores se resumem a

$$\begin{aligned}
 \ddot{\mathbf{L}}_p^{\gamma_0 \gamma_0} &= \frac{1}{\phi} \left[(\mathbf{AN}_0)^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{AN}_0) \right], \\
 \ddot{\mathbf{L}}_p^{\gamma_l \gamma_l} &= \frac{1}{\phi} \left[(\mathbf{AN}_l)^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{AN}_l) \right] - \lambda_l \mathbf{M}_l, \quad \forall l = 1, \dots, k, \\
 \ddot{\mathbf{L}}_p^{\phi \phi} &= \frac{1}{\phi^2} \left[\frac{n}{2} + \boldsymbol{\delta}^\top \mathbf{D}_c \boldsymbol{\delta} - \boldsymbol{\delta}^\top \mathbf{D}_v \mathbf{1}_n \right], \\
 \ddot{\mathbf{L}}_p^{\rho_j \rho_j} &= \frac{1}{\phi} \left[(\mathbf{B}_j \boldsymbol{\epsilon})^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{B}_j \boldsymbol{\epsilon}) \right], \quad \forall j = 1, \dots, p,
 \end{aligned}$$

em que $\mathbf{D}_c = \text{diag} \{c_1, \dots, c_n\}$ com $c_i = W'_g(\delta_i)$ e $\mathbf{D}_d = \text{diag} \{d_1, \dots, d_n\}$ com $d_i = \delta_i W'_g(\delta_i)$.

A derivada de $L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ com respeito a $(\gamma_{l_j}, \gamma_{l'_j})$ para todo $l \neq l' : 0, 1, 2, \dots, k$, é dada por

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\gamma_{l_j} \gamma_{l'_j}} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{l_j} \partial \gamma_{l'_j}} \\ &= \frac{1}{\phi} \sum_{i=1}^n (\eta_{li_j} - \rho_1 \eta_{l(i-1)_j} - \dots - \rho_p \eta_{l(i-p)_j}) (\eta_{l'_j} - \rho_1 \eta_{l'(i-1)_j} - \dots - \rho_p \eta_{l'(i-p)_j}) \\ &\quad \times \{4W'_g(\delta_i) \delta_i + 2W_g(\delta_i)\}, \quad l \neq l' : 0, \dots, k, \quad j = 1, \dots, r_l \quad \text{e} \quad j' = 1, \dots, r_{l'}, \end{aligned}$$

o qual, em notação matricial, torna-se

$$\ddot{\mathbf{L}}_p^{\gamma_l \gamma_{l'}} = \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_l \partial \gamma_{l'}^\top} = \frac{1}{\phi} \left[(\mathbf{AN}_l)^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{AN}_{l'}) \right].$$

As derivadas de $L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ com respeito a (γ_{l_j}, ϕ) e $(\gamma_{l_j}, \rho_{j'})$ são dadas por

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\gamma_{l_j} \phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{l_j} \partial \phi} \\ &= \frac{2}{\phi^2} \sum_{i=1}^n (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{li_j} - \rho_1 \eta_{l(i-1)_j} - \dots - \rho_p \eta_{l(i-p)_j}) \\ &\quad \times \{W'_g(\delta_i) \delta_i + W_g(\delta_i)\}, \quad l = 0, \dots, k, \quad \text{e} \quad j = 1, \dots, r_l, \end{aligned}$$

e

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\gamma_{l_j} \rho_{j'}} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{l_j} \partial \rho_{j'}} \\ &= \frac{4}{\phi} \sum_{i=1}^n W'_g(\delta_i) \delta_i (\eta_{li_j} - \rho_1 \eta_{l(i-1)_j} - \dots - \rho_p \eta_{l(i-p)_j}) (\epsilon_{i-j'}) \\ &\quad + \frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i) (\eta_{li_j} - \rho_1 \eta_{l(i-1)_j} - \dots - \rho_p \eta_{l(i-p)_j}) (\epsilon_{i-j'}) \\ &\quad + \frac{2}{\phi} \sum_{i=1}^n W_g(\delta_i) (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{l(i-j')_j}), \\ &\quad l = 0, \dots, k, \quad j = 1, \dots, r_l, \quad j' = 1, \dots, p. \end{aligned}$$

Na forma matricial temos que,

$$\begin{aligned} \ddot{\mathbf{L}}_p^{\gamma_l \phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_l \partial \phi} = \frac{1}{\phi^2} \left[(\mathbf{AN}_l)^\top (2\mathbf{D}_d - \mathbf{D}_v) (\mathbf{A}\boldsymbol{\epsilon}) \right], \\ \ddot{\mathbf{L}}_p^{\gamma_l \rho_{j'}} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_l \partial \rho_{j'}} = \frac{1}{\phi} \left[(\mathbf{AN}_l)^\top (\mathbf{D}_v - 4\mathbf{D}_d) (\mathbf{B}_{j'} \boldsymbol{\epsilon}) + (\mathbf{B}_{j'} \mathbf{N}_l)^\top \mathbf{D}_v (\mathbf{A}\boldsymbol{\epsilon}) \right]. \end{aligned}$$

E, por fim, temos a derivada de $L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ com respeito a (ρ_j, ϕ) e $(\rho_j, \rho_{j'})$. Portanto,

$$\begin{aligned}\ddot{L}_p^{\rho_j \phi} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho_j \partial \phi} = \frac{2}{\phi^2} \sum_{i=1}^n (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\epsilon_{i-j}) \{W_g(\delta_i) + W'_g(\delta_i) \delta_i\} \\ &= \frac{1}{\phi^2} [(\mathbf{B}_j \boldsymbol{\epsilon})^\top (\mathbf{D}_v - 2\mathbf{D}_d)(\mathbf{A} \boldsymbol{\epsilon})], \quad j = 1, \dots, p. \\ \ddot{L}_p^{\rho_j \rho_{j'}} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho_j \partial \rho_{j'}} = -\frac{2}{\phi} \sum_{i=1}^n (-\epsilon_{i-j}) (\epsilon_{i-j'}) \{W_g(\delta_i) + 2W'_g(\delta_i) \delta_i\} \\ &= \frac{1}{\phi} [(\mathbf{B}_{j'} \boldsymbol{\epsilon})^\top (4\mathbf{D}_d - \mathbf{D}_v)(\mathbf{B}_j \boldsymbol{\epsilon})], \quad j \neq j' = 1, \dots, p.\end{aligned}$$

B.3 Matriz informação de Fisher penalizada

Na Estatística, a informação de Fisher (às vezes chamada de informação, [Lehmann e Casella \(2006, p.115\)](#)) é uma maneira de medir a quantidade de informação que uma variável aleatória observável Y carrega sobre um parâmetro desconhecido $\boldsymbol{\theta}$ sobre o qual a probabilidade de Y depende, ou seja, de uma distribuição que modela Y . Formalmente, é a variação do escore ou o valor esperado das informações observadas. O papel da informação de Fisher na teoria assintótica da estimativa de máxima verossimilhança foi enfatizado pelo Estatístico Ronald Fisher (segundo alguns resultados iniciais de Francis Ysidro Edgeworth).

A matriz de informações de Fisher é usada para calcular as matrizes de variância-covariância associadas às estimativas de máxima verossimilhança. Também pode ser usada na formulação de estatísticas de teste, como o teste de Wald.

Seja $h(y_i; \boldsymbol{\theta})$ a função de densidade de probabilidade (ou função de massa de probabilidade) para Y condicional ao valor de $\boldsymbol{\theta}$. Descreve a probabilidade de observarmos um dado resultado de Y , dado um valor conhecido de $\boldsymbol{\theta}$. Se h tiver um pico acentuado em relação às mudanças em $\boldsymbol{\theta}$, é fácil indicar o valor “correto” de $\boldsymbol{\theta}$ a partir dos dados, ou equivalentemente, que os dados Y forneçam muitas informações sobre o parâmetro $\boldsymbol{\theta}$. Se a probabilidade h for plana e espalhada, seriam necessárias muitas amostras de Y para estimar o “verdadeiro” valor real de $\boldsymbol{\theta}$ que seria obtido usando toda a população amostrada. Isso sugere estudar algum tipo de variação em relação a $\boldsymbol{\theta}$.

Formalmente, a derivada parcial em relação a $\boldsymbol{\theta}$ do logaritmo natural da função de verossimilhança é chamada de escore. Sob certas condições de regularidade, se $\boldsymbol{\theta}$ é o parâmetro verdadeiro (ou seja, Y é realmente distribuído como $h(Y; \boldsymbol{\theta})$), o valor esperado (o primeiro momento) da função escore é 0, ou seja, $E(\mathbf{U}_\boldsymbol{\theta}) = \mathbf{0}$. A variância da função escore é definida como a informação de Fisher, e é dada por

$$\mathcal{I}(\boldsymbol{\theta}) = E \left[\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log h(Y; \boldsymbol{\theta}) \right)^2 \right].$$

Note que $0 \leq \mathcal{I}(\boldsymbol{\theta})$. Uma variável aleatória que carrega muitas informações de Fisher implica que o valor absoluto da função escore geralmente é importante. A informação de Fisher não é uma função de uma observação específica, pois a variável aleatória Y é calculada em média.

Se $\log [h(y; \boldsymbol{\theta})]$ for duas vezes diferenciável em relação a $\boldsymbol{\theta}$, e sob certas condições de regularidade, a informação de Fisher também poderá ser escritas como [Lehmann e Casella \(2006, Lemma 5.3, p.116\)](#)

$$\mathcal{I}(\boldsymbol{\theta}) = -E \left[\left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \log h(Y; \boldsymbol{\theta}) \right) \right].$$

Assim, a informação de Fisher pode ser vista como a curvatura da curva de suporte (o gráfico da probabilidade logarítmica). Perto da estimativa de máxima verossimilhança, as informações baixas de Fisher indicam, portanto, que o máximo parece “contudente”, ou seja, o máximo é raso e há muitos valores próximos com uma log-verossimilhança semelhante. Por outro lado, informações altas de Fisher indicam que o máximo é nítido. Duas condições de regularidade serão usadas para

encontrar a matriz informação de Fisher dos parâmetros $\boldsymbol{\theta}$ e são dadas por

$$(i) \quad \mathbb{E}[\mathbf{U}_{\boldsymbol{\theta}}] = \mathbf{0},$$

$$(ii) \quad -\mathbb{E}\left[\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right] = \mathbb{E}\left[\left(\frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta}}\right)^2\right].$$

No primeiro momento iremos calcular as medidas puras da informação, ou seja, a diagonal principal da matriz de informação. Considere algumas medidas já discutidas anteriormente como $d_g = \mathbb{E}[W_g^2(Z^2)Z^2]$ e $f_g = \mathbb{E}[W_g^2(Z^2)Z^4]$ com $Z \sim S(0, 1)$. Sob certas condições de regularidade, se $\boldsymbol{\theta}$ é o verdadeiro parâmetro, isto é, Y é na realidade distribuído como $h(Y; \boldsymbol{\theta})$, pode ser mostrado que o valor esperado (o primeiro momento) da função escore de ϕ é zero, ou seja, $\mathbb{E}[U_\phi(\boldsymbol{\theta})] = 0$. Portanto, conclui-se que $\mathbb{E}[W_g(\delta_i)\delta_i] = -\frac{1}{2}$, para qualquer i .

A informação para o parâmetro ϕ é dado por

$$\begin{aligned} \mathbb{E}\left[\left(\frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi}\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{2\phi} + \frac{1}{\phi}W_g(\delta_i)\delta_i\right)^2\right] = \frac{1}{\phi^2}\left[\frac{1}{4} + \mathbb{E}(W_g(\delta_i)\delta_i) + \mathbb{E}(W_g^2(\delta_i)\delta_i^2)\right] \\ &= \frac{1}{\phi^2}(f_g - 1/4). \end{aligned}$$

Logo, podemos concluir que a informação para ϕ é

$$\mathcal{I}_p^{\phi\phi} = \frac{n}{4\phi^2}(4f_g - 1).$$

O próximo passo é encontrar a matriz de informação dos parâmetros relacionados às r_0 covariáveis γ_0 e os parâmetros γ_l da l -ésima variável não paramétrica, em que $l = 1, \dots, k$. As derivações são similares para os vetores de parâmetros, portanto, estabelecemos o passo a passo apenas para o vetor γ_0 . Dessa forma, temos que,

$$\begin{aligned} \mathbb{E}\left[\left(\frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \gamma_{0_j}}\right)^2\right] &= \mathbb{E}\left[\left(\frac{2}{\phi}W_g(\delta_i)(\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(\eta_{0i_j} - \rho_1\eta_{0(i-1)_j} - \dots - \rho_p\eta_{0(i-p)_j})\right)^2\right] \\ &= \mathbb{E}\left[\frac{4}{\phi}W_g^2(\delta_i)\delta_i(\eta_{0i_j} - \rho_1\eta_{0(i-1)_j} - \dots - \rho_p\eta_{0(i-p)_j})^2\right] \\ &= \frac{4}{\phi}d_g(\eta_{0i_j} - \rho_1\eta_{0(i-1)_j} - \dots - \rho_p\eta_{0(i-p)_j})^2. \end{aligned}$$

Na forma matricial, temos que a informação para γ_0 e γ_l são dadas por, respectivamente,

$$\begin{aligned} \mathcal{I}_p^{\gamma_0\gamma_0} &= \frac{4d_g}{\phi}(\mathbf{AN}_0)^\top(\mathbf{AN}_0), \\ \mathcal{I}_p^{\gamma_l\gamma_l} &= \frac{4d_g}{\phi}(\mathbf{AN}_l)^\top(\mathbf{AN}_l) - \lambda_l\mathbf{M}_l, \quad \forall l = 1, \dots, k. \end{aligned}$$

Neste momento, vamos encontrar a informação para os parâmetros ρ_1, \dots, ρ_p . De acordo com a segunda condição de regularidade, temos que

$$\begin{aligned}
 \mathbb{E} \left[\left(\frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho_j} \right)^2 \right] &= \mathbb{E} \left[\left(\frac{2}{\phi} W_g(\delta_i) (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\epsilon_{i-j}) \right)^2 \right] \\
 &= \mathbb{E} \left[\frac{4}{\phi} W_g^2(\delta_i) \delta_i \epsilon_{i-j}^2 \right] \\
 &= \mathbb{E} \left\{ \mathbb{E} \left[\frac{4}{\phi} W_g^2(\delta_i) \delta_i \epsilon_{i-j}^2 \mid y_{i-1}, \dots, y_{i-p} \right] \right\} \\
 &= \mathbb{E} \left[\frac{4}{\phi} d_g \epsilon_{i-j}^2 \right] \quad \forall j = 1, \dots, p.
 \end{aligned}$$

Para os erros autoregressivos de primeira ordem, ou seja AR(1), temos que $\mathbb{E}(\epsilon_i^2) = \phi \xi \frac{(1-\rho_1^{2i})}{(1-\rho_1^2)}$, caso $\epsilon_0 = 0$, a informação se resume a

$$\begin{aligned}
 I_p^{\rho_1 \rho_1} &= \frac{4}{\phi} d_g \sum_{i=2}^n \mathbb{E}(\epsilon_{i-1}^2) = \frac{4}{\phi} d_g \sum_{i=1}^{n-1} \mathbb{E}(\epsilon_i^2) = 4d_g \xi \sum_{i=1}^{n-1} \frac{1 - \rho_1^{2i}}{(1 - \rho_1^2)} \\
 &= \frac{4d_g \xi}{1 - \rho_1^2} \left[(n-1) - \left(\frac{\rho_1^2(1 - \rho_1^{2(n-1)})}{1 - \rho_1^2} \right) \right].
 \end{aligned}$$

Já para o caso em que não definimos ϵ_0 , temos que $\mathbb{E}(\epsilon_i^2) = \phi \xi \frac{1}{(1-\rho_1^2)}$, logo,

$$\begin{aligned}
 I_p^{\rho_1 \rho_1} &= \frac{4}{\phi} d_g \sum_{i=2}^n \mathbb{E}(\epsilon_{i-1}^2) = \frac{4}{\phi} d_g \sum_{i=1}^{n-1} \mathbb{E}(\epsilon_i^2) = 4d_g \xi \sum_{i=1}^{n-1} \frac{1}{(1 - \rho_1^2)} \\
 &= \frac{4d_g \xi (n-1)}{1 - \rho_1^2}.
 \end{aligned}$$

Para o modelo com erros autoregressivos de ordem p , ou seja, AR(p), sob os momentos de ϵ_{i-j} , podemos escrever

$$I_p^{\rho_j \rho_j} = \frac{4d_g}{\phi} \sum_{i=j+1}^n \mathbb{E}[\epsilon_{i-j}^2] = \frac{4d_g}{\phi} \sum_{i=1}^{n-j} \mathbb{E}[\epsilon_i^2] = \frac{4d_g(n-j)}{\phi} \phi \epsilon,$$

em que $\mathbb{E}(\epsilon_i^2)$ é discutida no Capítulo 4 com maiores detalhes.

O próximo passo é encontrar as esperanças mistas ou cruzadas entre os parâmetros $\boldsymbol{\theta}$ do modelo postulado. Nesse momento aplicaremos a segunda condição de regularidade com as informações já discutidas anteriormente. Sabemos que $\mathcal{I}_p^{\gamma_0 \gamma_0} = \frac{4}{\phi} d_g (\mathbf{AN}_0)^\top (\mathbf{AN}_0)$, também, temos que,

$$\mathcal{I}_p^{\gamma_0 \gamma_0} = -\mathbb{E} \left\{ \frac{1}{\phi} \left[(\mathbf{AN}_0)^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{AN}_0) \right] \right\}.$$

Dessa maneira, é fácil ver que $\mathbb{E}[4\mathbf{D}_d - \mathbf{D}_v] = -4d_g$. Assim, calculamos a informação de Fisher mista de γ_0 e γ_l , dada por

$$\mathcal{I}_p^{\gamma_0 \gamma_l} = -\mathbb{E} \left\{ \frac{1}{\phi} \left[(\mathbf{AN}_0)^\top (4\mathbf{D}_d - \mathbf{D}_v) (\mathbf{AN}_l) \right] \right\} = \frac{1}{\phi} 4d_g \left[(\mathbf{AN}_0)^\top (\mathbf{AN}_l) \right], \quad \forall l = 1, \dots, k.$$

Podemos concluir que γ_0 e γ_l não são ortogonais. Por semelhança aos modelos simétricos tra-

dicionais, já discutido anteriormente, derivamos a seguinte expressão:

$$E \left\{ [W_g(\delta_i) + W'_g(\delta_i)\delta_i] (\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p}) \mid y_{i-1}, \dots, y_{i-p} \right\} = 0.$$

Com esse resultado, podemos encontrar as seguintes esperanças mistas

$$\begin{aligned} \mathcal{I}_p^{\phi\gamma_0} &= -E \left\{ \frac{1}{\phi^2} [(\mathbf{AN}_0)^\top (2\mathbf{D}_d - \mathbf{D}_v)(\mathbf{A}\epsilon)] \right\} = \mathbf{0}, \\ \mathcal{I}_p^{\phi\gamma_l} &= -E \left\{ \frac{1}{\phi^2} [(\mathbf{AN}_l)^\top (2\mathbf{D}_d - \mathbf{D}_v)(\mathbf{A}\epsilon)] \right\} = \mathbf{0}, \quad \forall l = 1, \dots, k, \\ \mathcal{I}_p^{\phi\rho_j} &= -E \left\{ \frac{1}{\phi^2} [(\mathbf{B}_j\epsilon)^\top (\mathbf{D}_v - 2\mathbf{D}_d)(\mathbf{A}\epsilon)] \right\} = \mathbf{0}, \quad \forall j = 1, \dots, p. \end{aligned}$$

Como $E[\mathbf{U}\gamma_0] = \mathbf{0}$, segue que,

$$E \left[\frac{1}{\phi} (\mathbf{AN}_0)^\top \mathbf{D}_v \mathbf{A}\epsilon \right] = 0 \Leftrightarrow E[\mathbf{D}_v \mathbf{A}\epsilon] = \mathbf{0}.$$

Logo, podemos concluir que,

$$\begin{aligned} \mathcal{I}_p^{\rho_j\gamma_0} &= -\frac{1}{\phi} E \left\{ [(\mathbf{AN}_0)^\top (\mathbf{D}_v - 4\mathbf{D}_d)(\mathbf{B}_j\epsilon)] + (\mathbf{B}_j\mathbf{N}_0)^\top \mathbf{D}_v (\mathbf{A}\epsilon) \right\} \\ &= -\frac{1}{\phi} E \left[(\mathbf{AN}_0)^\top (\mathbf{D}_v - 4\mathbf{D}_d)(\mathbf{B}_j\epsilon) \right] \\ &= -\frac{1}{\phi} E \left\{ E \left[(\mathbf{AN}_0)^\top (\mathbf{D}_v - 4\mathbf{D}_d)(\mathbf{B}_j\epsilon) \mid y_{i-1}, \dots, y_{i-p} \right] \right\} \\ &= \frac{1}{\phi} E \left[(\mathbf{AN}_0)^\top 4d_g(\mathbf{B}_j\epsilon) \right] \\ &= \mathbf{0}, \quad \forall j = 1, \dots, p, \end{aligned}$$

já que $E[(\mathbf{B}_j\epsilon)] = \mathbf{0}$. Logo, por analogia, temos que $\mathcal{I}_p^{\rho_j\gamma_l} = 0, \forall l = 1, \dots, k$ e $\forall j = 1, \dots, p$.

Por fim, temos a informação de ρ_j e $\rho_{j'}$ para $j \neq j' : 1, \dots, p$ que é dada por

$$\begin{aligned} -E \left[\frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \rho_j \partial \rho_{j'}} \right] &= E \left[\frac{4}{\phi} W'_g(\delta_i)\delta_i(\epsilon_{i-j})(\epsilon_{i-j'}) + \frac{2}{\phi} W_g(\delta_i)(\epsilon_{i-j})(\epsilon_{i-j'}) \right] \\ &= E \left[\frac{1}{\phi} (4W'_g(\delta_i)\delta_i + 2W_g(\delta_i)) (\epsilon_{i-j})(\epsilon_{i-j'}) \right] \\ &= -\frac{1}{\phi} E \left\{ E \left[(4W'_g(\delta_i)\delta_i + 2W_g(\delta_i)) (\epsilon_{i-j})(\epsilon_{i-j'}) \mid y_{i-1}, \dots, y_{i-p} \right] \right\} \\ &= \frac{4d_g}{\phi} E(\epsilon_{i-j}\epsilon_{i-j'}). \end{aligned}$$

Considere $j' < j$, portanto, podemos escrever a informação de ρ_j e $\rho_{j'}$ da seguinte maneira,

$$I_p^{\rho_j\rho_{j'}} = \frac{4d_g}{\phi} \sum_{i=j+1}^{n+j'} E(\epsilon_{i-j}\epsilon_{i-j'}) = \frac{4d_g}{\phi} \sum_{i=1}^{n-j+j'} E(\epsilon_i\epsilon_{i+j-j'}) = \frac{4d_g(n-j+j')}{\phi} \phi_{j-j'},$$

em que $E(\epsilon\epsilon^\top)$ é discutida no capítulo 4. A expressão para Υ_p torna-se progressivamente mais complicada. Uma expressão geral para \mathbf{V} é dada por Wise (1955).

Apêndice C

Análise de sensibilidade

Nesta Seção apresentaremos os cálculos da matriz $\mathbf{\Delta}$, conforme definida no Capítulo 5 as quais são utilizadas para detectar eventuais variações nas estimativas dos parâmetros sob pequenas perturbações no modelo ou nos dados. Através desses esquemas é possível averiguar a sensibilidade de algumas suposições do modelo, e assimilar melhor o processo de modelagem. A matriz $\mathbf{\Delta}$ é derivada para cinco diferentes tipos de perturbação as quais serão apresentadas em subseções. O primeiro, supomos uma ponderação de casos na verossimilhança do modelo. O segundo, há uma suposta perturbação no parâmetro de dispersão. O terceiro, consideramos que a variável resposta seja acrescida de uma constante. O quarto, supomos que uma das variáveis explicativas (contínua) seja aumentado por uma constante. E, o quinto, supomos que o coeficiente de autocorrelação seja acrescido de uma constante. Cada um desses esquemas é conhecido, respectivamente, como, perturbação ponderação de casos, perturbação ponderada no parâmetro de dispersão, perturbação na variável resposta, perturbação na variável explicativa e perturbação no coeficiente de autocorrelação. Considere a verossimilhança do modelo sem perturbação dada por

$$L_p(\boldsymbol{\theta}, \lambda) = \sum_{i=1}^n L_{p_i}(\boldsymbol{\theta}, \lambda) = \sum_{i=1}^n L_i(\boldsymbol{\theta}) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k. \quad (\text{C.1})$$

C.1 Ponderação de casos

Considere que cada observação possui um peso diferente no processo de estimação, ou seja, atribua uma ponderação arbitrária para a verossimilhança individual de tal forma que

$$\begin{aligned} L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega}) &= \sum_{i=1}^n \omega_i L_i(\boldsymbol{\theta}) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k \\ &= -\frac{n\omega_i}{2} \log(\phi) + \sum_{i=1}^n \omega_i \log(g(\delta_i)) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k, \end{aligned}$$

em que $L_i(\boldsymbol{\theta}) = -\frac{1}{2} \log(\phi) + \log[g(\delta_i)]$, e $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ é o vetor de pesos (ponderações), tal que $0 \leq \omega_i \leq 1$, para $i = 1, \dots, n$. Note que, para $\omega_i = 0$ e $\omega_j = 1$ para $j \neq i$ excluimos o i -ésimo indivíduo da expressão do logaritmo da verossimilhança. Além disso, o vetor de não perturbação é definida por $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$. Para este esquema, temos a seguinte quantidade:

$$\delta_i = \frac{(\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p})^2}{\phi}.$$

As funções escores são dadas por

$$\begin{aligned}
\frac{\partial L_p(\theta, \lambda|\omega)}{\partial \gamma_{0_j}} &= +\omega_i \frac{g'(\delta_i)}{g(\delta_i)} 2\phi^{-1}(\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(-\eta_{0_{ij}} + \rho_1\eta_{0_{(i-1)j}} + \dots + \rho_p\eta_{0_{(i-p)j}}) \\
&= -2 \frac{W_g(\delta_i)}{\phi} \omega_i (\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(\eta_{0_{ij}} - \rho_1\eta_{0_{(i-1)j}} - \dots - \rho_p\eta_{0_{(i-p)j}}), \\
\frac{\partial L_p(\theta, \lambda|\omega)}{\partial \gamma_{1_j}} &= +\omega_i \frac{g'(\delta_i)}{g(\delta_i)} 2\phi^{-1}(\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(-\eta_{1_{ij}} + \rho_1\eta_{1_{(i-1)j}} + \dots + \rho_p\eta_{1_{(i-p)j}}) \\
&\quad - [\lambda_1 \mathbf{M}_1 \gamma_1]_j \\
&= -2 \frac{W_g(\delta_i)}{\phi} \omega_i (\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(\eta_{1_{ij}} - \rho_1\eta_{1_{(i-1)j}} - \dots - \rho_p\eta_{1_{(i-p)j}}) \\
&\quad - [\lambda_1 \mathbf{M}_1 \gamma_1]_j, \\
\frac{\partial L_p(\theta, \lambda|\omega)}{\partial \gamma_{k_j}} &= +\omega_i \frac{g'(\delta_i)}{g(\delta_i)} 2\phi^{-1}(\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(-\eta_{k_{ij}} + \rho_1\eta_{k_{(i-1)j}} + \dots + \rho_p\eta_{k_{(i-p)j}}) \\
&\quad - [\lambda_k \mathbf{M}_k \gamma_k]_j \\
&= -2 \frac{W_g(\delta_i)}{\phi} \omega_i (\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(\eta_{k_{ij}} - \rho_1\eta_{k_{(i-1)j}} - \dots - \rho_p\eta_{k_{(i-p)j}}) \\
&\quad - [\lambda_k \mathbf{M}_k \gamma_k]_j, \\
\frac{\partial L_p(\theta, \lambda|\omega)}{\partial \phi} &= -\frac{\omega_i}{2\phi} - \omega_i \frac{g'(\delta_i)}{g(\delta_i)} \phi^{-2}(\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})^2, \\
\frac{\partial L_p(\theta, \lambda|\omega)}{\partial \rho_1} &= +\omega_i \frac{g'(\delta_i)}{g(\delta_i)} \phi^{-1} 2(\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(-\epsilon_{i-1}), \\
\frac{\partial L_p(\theta, \lambda|\omega)}{\partial \rho_p} &= +\omega_i \frac{g'(\delta_i)}{g(\delta_i)} \phi^{-1} 2(\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(-\epsilon_{i-p}).
\end{aligned}$$

A partir dos termos definidos acima derivamos em relação a ω para obter a matriz de curvatura normal. Assim,

$$\begin{aligned}
\frac{\partial^2 L_p(\theta, \lambda|\omega)}{\partial \omega_i \partial \gamma_{0_j}} &= -2 \frac{W_g(\delta_i)}{\phi} (\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(\eta_{0_{ij}} - \rho_1\eta_{0_{(i-1)j}} - \dots - \rho_p\eta_{0_{(i-p)j}}), \\
\frac{\partial^2 L_p(\theta, \lambda|\omega)}{\partial \omega_i \partial \gamma_{1_j}} &= -2 \frac{W_g(\delta_i)}{\phi} (\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(\eta_{1_{ij}} - \rho_1\eta_{1_{(i-1)j}} - \dots - \rho_p\eta_{1_{(i-p)j}}), \\
\frac{\partial^2 L_p(\theta, \lambda|\omega)}{\partial \omega_i \partial \gamma_{k_j}} &= -2 \frac{W_g(\delta_i)}{\phi} (\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(\eta_{k_{ij}} - \rho_1\eta_{k_{(i-1)j}} - \dots - \rho_p\eta_{k_{(i-p)j}}), \\
\frac{\partial^2 L_p(\theta, \lambda|\omega)}{\partial \omega_i \partial \phi} &= -\frac{1}{2\phi} - 2 \frac{W_g(\delta_i)\delta_i}{2\phi}, \\
\frac{\partial^2 L_p(\theta, \lambda|\omega)}{\partial \omega_i \partial \rho_1} &= +2 \frac{W_g(\delta_i)}{\phi} (\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(-\epsilon_{i-1}), \\
\frac{\partial^2 L_p(\theta, \lambda|\omega)}{\partial \omega_i \partial \rho_p} &= +2 \frac{W_g(\delta_i)}{\phi} (\epsilon_i - \rho_1\epsilon_{i-1} - \dots - \rho_p\epsilon_{i-p})(-\epsilon_{i-p}).
\end{aligned}$$

Esse esquema de perturbação permite identificar aqueles indivíduos que exercem um grande impacto no processo de estimação. Em notação matricial e avaliando os elementos acima em $\hat{\theta}$ e ω_0 , temos que os elementos da matriz $\mathbf{\Delta}$ são escritos como:

$$\begin{aligned}
 \Delta_{0(r_0 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_0)^\top \mathbf{D}(\hat{\mathbf{v}}) \mathbf{D}(\hat{\mathbf{A}}_\epsilon), \\
 \Delta_{1(r_1 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_1)^\top \mathbf{D}(\hat{\mathbf{v}}) \mathbf{D}(\hat{\mathbf{A}}_\epsilon), \\
 &\vdots \\
 \Delta_{k(r_k \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_k)^\top \mathbf{D}(\hat{\mathbf{v}}) \mathbf{D}(\hat{\mathbf{A}}_\epsilon), \\
 \Delta_{k+1(1 \times n)} &= -\frac{\mathbf{1}_n^\top}{2\hat{\phi}} + \frac{\mathbf{1}_n^\top \mathbf{D}(\hat{\mathbf{m}})}{2\hat{\phi}} = \frac{\mathbf{1}_n^\top}{2\hat{\phi}} [\mathbf{D}(\hat{\mathbf{m}}) - \mathbf{I}_n], \\
 \Delta_{k+1+1(1 \times n)} &= -\frac{1}{\hat{\phi}} (\hat{\mathbf{B}}_1 \epsilon)^\top \mathbf{D}(\hat{\mathbf{v}}) \mathbf{D}(\hat{\mathbf{A}}_\epsilon), \\
 &\vdots \\
 \Delta_{k+p+1(1 \times n)} &= -\frac{1}{\hat{\phi}} (\hat{\mathbf{B}}_p \epsilon)^\top \mathbf{D}(\hat{\mathbf{v}}) \mathbf{D}(\hat{\mathbf{A}}_\epsilon),
 \end{aligned}$$

em que $\mathbf{1}_n = (1, \dots, 1)^\top$ e $\mathbf{D}(\hat{\mathbf{A}}_\epsilon)$ é uma matriz diagonal com elementos dados por $\hat{\mathbf{A}}\hat{\epsilon}$.

C.2 Perturbação ponderada no parâmetro de dispersão

Esse esquema de perturbação é introduzido por considerar $\phi_\omega = \phi/\omega_i$, para $i = 1, \dots, n$, em que $\omega = (\omega_1, \dots, \omega_n)^\top$, $\omega_i > 0$ e $\omega_0 = \mathbf{1}_n$. Neste caso o logaritmo da função de verossimilhança resulta em

$$L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \boldsymbol{\omega}) = -\frac{n}{2} \log \left(\frac{\phi}{\omega_i} \right) + \sum_{i=1}^n \log(g(\delta_{\omega_i})) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k.$$

Para tal esquema temos a quantidade

$$\delta_{\omega_i} = \omega_i \delta_i = \frac{\omega_i (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p})^2}{\phi}.$$

As funções escore da verossimilhança perturbada em relação aos parâmetros do modelo são

$$\begin{aligned}
 \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \boldsymbol{\omega})}{\partial \gamma_{0j}} &= -2\omega_i \frac{W_g(\omega_i \delta_i)}{\phi} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{0ij} - \rho_1 \eta_{0(i-1)j} - \dots - \rho_p \eta_{0(i-p)j}), \\
 \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \boldsymbol{\omega})}{\partial \gamma_{lj}} &= -2\omega_i \frac{W_g(\omega_i \delta_i)}{\phi} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{lij} - \rho_1 \eta_{l(i-1)j} - \dots - \rho_p \eta_{l(i-p)j}) \\
 &\quad - [\lambda_l \mathbf{M}_l \boldsymbol{\gamma}_l]_j, \quad \forall l = 1, \dots, k, \\
 \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \boldsymbol{\omega})}{\partial \phi} &= -\frac{1}{2\phi} - \omega_i \delta_i \frac{W_g(\omega_i \delta_i)}{\phi} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p})^2, \\
 \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \boldsymbol{\omega})}{\partial \rho_j} &= +2\omega_i \frac{W_g(\omega_i \delta_i)}{\phi} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (-\epsilon_{i-j}), \quad \forall j = 1, \dots, p.
 \end{aligned}$$

Calculando a derivada dessa função para o vetor $\boldsymbol{\omega}$, encontramos

$$\begin{aligned} \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega})}{\partial \omega_i \partial \gamma_{0j}} &= -2 \frac{W_g(\omega_i \delta_i)}{\phi} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{0ij} - \rho_1 \eta_{0(i-1)j} - \dots - \rho_p \eta_{0(i-p)j}) \\ &\quad - 2\omega_i \delta_i \frac{W'_g(\omega_i \delta_i)}{\phi} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{0ij} - \rho_1 \eta_{0(i-1)j} - \dots - \rho_p \eta_{0(i-p)j}), \\ \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega})}{\partial \omega_i \partial \gamma_{lj}} &= -2 \frac{W_g(\omega_i \delta_i)}{\phi} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{lij} - \rho_1 \eta_{l(i-1)j} - \dots - \rho_p \eta_{l(i-p)j}) \\ &\quad - 2\omega_i \delta_i \frac{W'_g(\omega_i \delta_i)}{\phi} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (\eta_{lij} - \rho_1 \eta_{l(i-1)j} - \dots - \rho_p \eta_{l(i-p)j}), \\ \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega})}{\partial \omega_i \partial \phi} &= -\frac{2W_g(\omega_i \delta_i) \delta_i}{2\phi} - \frac{2W'_g(\omega_i \delta_i) \omega_i \delta_i \delta_i}{2\phi}, \\ \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega})}{\partial \omega_i \partial \rho_j} &= +2 \frac{W_g(\omega_i \delta_i)}{\phi} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (-\epsilon_{i-j}) \\ &\quad + 2\omega_i \delta_i \frac{W'_g(\omega_i \delta_i)}{\phi} (\epsilon_i - \rho_1 \epsilon_{i-1} - \dots - \rho_p \epsilon_{i-p}) (-\epsilon_{i-j}). \end{aligned}$$

O interesse do esquema de perturbação o parâmetro de dispersão pode revelar indivíduos que interferem principalmente na modelagem da estrutura de escala, assim como indicar aqueles indivíduos influentes na estimação dos parâmetros. Com esse último cálculo, completamos a matriz $\boldsymbol{\Delta}$ da curvatura conformal. Diferenciando $L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega})$ com relação a $\boldsymbol{\theta}$ e ω_i e avaliando em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, obtemos,

$$\begin{aligned} \boldsymbol{\Delta}_{0(r_0 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_0)^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 2\mathbf{D}(\hat{\mathbf{d}}) \right] \mathbf{D}(\hat{\mathbf{A}}\boldsymbol{\epsilon}), \\ \boldsymbol{\Delta}_{1(r_1 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_1)^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 2\mathbf{D}(\hat{\mathbf{d}}) \right] \mathbf{D}(\hat{\mathbf{A}}\boldsymbol{\epsilon}), \\ &\vdots \\ \boldsymbol{\Delta}_{k(r_k \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_k)^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 2\mathbf{D}(\hat{\mathbf{d}}) \right] \mathbf{D}(\hat{\mathbf{A}}\boldsymbol{\epsilon}), \\ \boldsymbol{\Delta}_{k+1(1 \times n)} &= \frac{1}{2\hat{\phi}} \boldsymbol{\delta}^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 2\mathbf{D}(\hat{\mathbf{d}}) \right], \\ \boldsymbol{\Delta}_{k+1+1(1 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{B}}_1 \boldsymbol{\epsilon})^\top \left[2\mathbf{D}(\hat{\mathbf{d}}) - \mathbf{D}(\hat{\mathbf{v}}) \right] \mathbf{D}(\hat{\mathbf{A}}\boldsymbol{\epsilon}), \\ &\vdots \\ \boldsymbol{\Delta}_{k+p+1(1 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{B}}_p \boldsymbol{\epsilon})^\top \left[2\mathbf{D}(\hat{\mathbf{d}}) - \mathbf{D}(\hat{\mathbf{v}}) \right] \mathbf{D}(\hat{\mathbf{A}}\boldsymbol{\epsilon}), \end{aligned}$$

em que $\mathbf{D}(\hat{\mathbf{A}}\boldsymbol{\epsilon})$ é uma matriz diagonal com elementos dados por $\hat{\mathbf{A}}\hat{\boldsymbol{\epsilon}}$.

C.3 Perturbação na variável resposta

A perturbação na variável resposta é introduzida mediante alteração de y_i por $y_{\omega_i} = y_i + \omega_i$ em que ω_i é um vetor de perturbações para $i = 1, \dots, n$. Portanto, é necessário derivar as funções escore em relação ao vetor $\boldsymbol{\omega}$. Antes, definimos as seguintes quantidades:

$$\begin{aligned}\epsilon_{\omega_i} &= y_i + \omega_i - \eta_{0_i}^\top \gamma_0 - \eta_{1_i}^\top \gamma_1 - \eta_{k_i}^\top \gamma_k, \\ \delta_{\omega_i} &= \frac{(\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}})^2}{\phi}.\end{aligned}$$

Além disso, temos que

$$\frac{\partial \epsilon_{\omega_i}}{\partial \omega_l} = \begin{cases} 1, & \text{se } l = i, \\ \rho_1, & \text{se } l = i - 1, \\ \rho_p, & \text{se } l = i - p, \\ \vdots & \vdots \\ 0, & \text{caso contrário} \end{cases}$$

e

$$\frac{\partial W_g(\delta_{\omega_i})}{\partial \omega_l} = \begin{cases} +\frac{2}{\phi} W'_g(\delta_{\omega_i})(\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}), & \text{se } l = i, \\ -\frac{2\rho_1}{\phi} W'_g(\delta_{\omega_i})(\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}), & \text{se } l = i - 1, \\ \vdots & \vdots \\ -\frac{2\rho_p}{\phi} W'_g(\delta_{\omega_i})(\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}), & \text{se } l = i - p, \\ 0, & \text{caso contrário.} \end{cases}$$

Portanto, a log-verossimilhança penalizada é dada por

$$\begin{aligned}L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega}) &= \sum_{i=1}^n L_i(\boldsymbol{\theta}|\boldsymbol{\omega}) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k \\ &= -\frac{n}{2} \log(\phi) + \sum_{i=1}^n \log(g(\delta_{\omega_i})) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k.\end{aligned}$$

As funções escore em relação aos parâmetros do modelo supondo esse esquema de perturbação são dadas por

$$\begin{aligned}\frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \gamma_{0_j}} &= + \frac{g'(\delta_{\omega_i})}{g(\delta_{\omega_i})} 2\phi^{-1} (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}) (-\eta_{\omega 0_{(i,j)}} + \rho_1 \eta_{\omega 0_{(i-1,j)}} + \dots + \rho_p \eta_{\omega 0_{(i-p,j)}}) \\ &= -2 \frac{W_g(\delta_{\omega_i})}{\phi} (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}) (\eta_{\omega 0_{(i,j)}} - \rho_1 \eta_{\omega 0_{(i-1,j)}} - \dots - \rho_p \eta_{\omega 0_{(i-p,j)}}), \\ \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \gamma_{l_j}} &= -2 \frac{W_g(\delta_{\omega_i})}{\phi} (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}) (\eta_{\omega l_{(i,j)}} - \rho_1 \eta_{\omega l_{(i-1,j)}} - \dots - \rho_p \eta_{\omega l_{(i-p,j)}}) \\ &\quad - [\lambda_l \mathbf{M}_l \boldsymbol{\gamma}_l]_l, \quad \forall l = 1, \dots, k, \\ \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \phi} &= -\frac{1}{2\phi} + \frac{g'(\delta_{\omega_i})}{g(\delta_{\omega_i})} (-\phi^{-2}) (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}})^2 \\ &= -\frac{1}{2\phi} - \frac{W_g(\delta_{\omega_i}) \delta_{\omega_i}}{\phi}, \\ \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \rho_l} &= + \frac{g'(\delta_{\omega_i})}{g(\delta_{\omega_i})} (2\phi^{-1}) (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}) (-\epsilon_{\omega_{(i-l)}}) \\ &= + \frac{2}{\phi} W_g(\delta_{\omega_i}) (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}) (-\epsilon_{\omega_{(i-l)}}), \quad \forall l = 1, \dots, p.\end{aligned}$$

Calculando a derivada dessa função para o vetor $\boldsymbol{\omega}$, encontramos

$$\begin{aligned}
\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega})}{\partial \omega_l \partial \gamma_{l'j}} &= -\frac{4}{\phi} W'_g(\delta_{\omega_l}) \delta_{\omega_l} (\eta_{\omega l' l_j} - \rho_1 \eta_{\omega l' (l-1)_j} - \dots - \rho_p \eta_{\omega l' (l-p)_j}) \\
&+ \frac{4\rho_1}{\phi} W'_g(\delta_{\omega_{(l+1)}}) \delta_{\omega_{(l+1)}} (\eta_{\omega l' (l+1)_j} - \rho_1 \eta_{\omega l' l_j} - \dots - \rho_p \eta_{\omega l' (l+1-p)_j}) \\
&\vdots \\
&+ \frac{4\rho_p}{\phi} W'_g(\delta_{\omega_{(l+p)}}) \delta_{\omega_{(l+p)}} (\eta_{\omega l' (l+p)_j} - \rho_1 \eta_{\omega l' (l+p-1)_j} - \dots - \rho_p \eta_{\omega l' l_j}) \\
&- \frac{2}{\phi} W_g(\delta_{\omega_l}) (\eta_{\omega l' l_j} - \rho_1 \eta_{\omega l' (l-1)_j} - \dots - \rho_p \eta_{\omega l' (l-p)_j}) \\
&+ \frac{2\rho_1}{\phi} W_g(\delta_{\omega_{(l+1)}}) (\eta_{\omega l' (l+1)_j} - \rho_1 \eta_{\omega l' l_j} - \dots - \rho_p \eta_{\omega l' (l+1-p)_j}) \\
&\vdots \\
&+ \frac{2\rho_p}{\phi} W_g(\delta_{\omega_{(l+p)}}) (\eta_{\omega l' (l+p)_j} - \rho_1 \eta_{\omega l' (l+p-1)_j} - \dots - \rho_p \eta_{\omega l' l_j}), \quad \forall l' = 0, 1, \dots, k,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega})}{\partial \omega_l \partial \phi} &= -\frac{2}{\phi^2} W'_g(\delta_{\omega_l}) \delta_{\omega_l} (\epsilon_{\omega_l} - \rho_1 \epsilon_{\omega_{(l-1)}} - \dots - \rho_p \epsilon_{\omega_{(l-p)}}) \\
&+ \frac{2\rho_1}{\phi^2} W'_g(\delta_{\omega_{(l+1)}}) \delta_{\omega_{(l+1)}} (\epsilon_{\omega_{(l+1)}} - \rho_1 \epsilon_{\omega_l} - \dots - \rho_p \epsilon_{\omega_{(l+1-p)}}) \\
&\vdots \\
&+ \frac{2\rho_p}{\phi^2} W'_g(\delta_{\omega_{(l+p)}}) \delta_{\omega_{(l+p)}} (\epsilon_{\omega_{(l+p)}} - \rho_1 \epsilon_{\omega_{(l+p-1)}} - \dots - \rho_p \epsilon_{\omega_l}) \\
&- \frac{2}{\phi^2} W_g(\delta_{\omega_l}) (\epsilon_{\omega_l} - \rho_1 \epsilon_{\omega_{(l-1)}} - \dots - \rho_p \epsilon_{\omega_{(l-p)}}) \\
&+ \frac{2\rho_1}{\phi^2} W_g(\delta_{\omega_{(l+1)}}) (\epsilon_{\omega_{(l+1)}} - \rho_1 \epsilon_{\omega_l} - \dots - \rho_p \epsilon_{\omega_{(l+1-p)}}) \\
&\vdots \\
&+ \frac{2\rho_p}{\phi^2} W_g(\delta_{\omega_{(l+p)}}) (\epsilon_{\omega_{(l+p)}} - \rho_1 \epsilon_{\omega_{(l+p-1)}} - \dots - \rho_p \epsilon_{\omega_l}),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega})}{\partial \omega_l \partial \rho_{l'}} &= +\frac{4}{\phi} W'_g(\delta_{\omega_l}) \delta_{\omega_l} (-\epsilon_{\omega_{(l-l')}}) + \\
&- \frac{4\rho_1}{\phi} W'_g(\delta_{\omega_{(l+1)}}) \delta_{\omega_{(l+1)}} (-\epsilon_{\omega_l}) - \dots - \frac{4\rho_p}{\phi} W'_g(\delta_{\omega_{(l+p)}}) \delta_{\omega_{(l+p)}} (-\epsilon_{\omega_{(l+p-1)}}) \\
&+ \frac{2}{\phi} W_g(\delta_{\omega_l}) (-\epsilon_{\omega_{(l-1)}}) - \frac{2\rho_1}{\phi} W_g(\delta_{\omega_{(l+1)}}) (-\epsilon_{\omega_l}) \dots - \frac{2\rho_p}{\phi} W_g(\delta_{\omega_{(l+p)}}) (-\epsilon_{\omega_{(l+p-1)}}) \\
&- \frac{2}{\phi} W_g(\delta_{\omega_{(l+1)}}) (\epsilon_{\omega_{(l+1)}} - \rho_1 \epsilon_{\omega_l} - \dots - \rho_p \epsilon_{\omega_{(l+1-p)}}), \quad \forall l' = 1, \dots, p.
\end{aligned}$$

Além disso, devemos avaliar as matrizes $\mathbf{D}(d_\omega)$ e $\mathbf{D}(v_\omega)$ em função de $\hat{\boldsymbol{\delta}}$. Denotamos essas matrizes diagonal por $\mathbf{D}(\hat{\mathbf{d}})$ e $\mathbf{D}(\hat{\mathbf{v}})$. Ainda, como \mathbf{A} depende de $\boldsymbol{\rho}$, chamaremos de $\hat{\mathbf{A}}$ essas matrizes utilizando $\hat{\boldsymbol{\rho}}$. Assim, podemos escrever os elementos de $\hat{\boldsymbol{\Delta}}$, avaliados em $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$, como

$$\begin{aligned}
 \Delta_{0(r_0 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}} \mathbf{N}_0)^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 4\mathbf{D}(\hat{\mathbf{d}}) \right] \hat{\mathbf{A}}, \\
 \Delta_{1(r_1 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}} \mathbf{N}_1)^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 4\mathbf{D}(\hat{\mathbf{d}}) \right] \hat{\mathbf{A}}, \\
 &\vdots \\
 \Delta_{k(r_k \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}} \mathbf{N}_k)^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 4\mathbf{D}(\hat{\mathbf{d}}) \right] \hat{\mathbf{A}}, \\
 \Delta_{k+1(1 \times n)} &= \frac{1}{\hat{\phi}^2} (\hat{\mathbf{A}} \epsilon)^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 2\mathbf{D}(\hat{\mathbf{d}}) \right] \hat{\mathbf{A}}, \\
 \Delta_{k+1+1(1 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{B}}_1 \epsilon)^\top \left[4\mathbf{D}(\hat{\mathbf{d}}) - \mathbf{D}(\hat{\mathbf{v}}) \right] \hat{\mathbf{A}} - \frac{1}{\hat{\phi}} (\hat{\mathbf{A}} \epsilon)^\top \mathbf{D}(\hat{\mathbf{v}}) \mathbf{B}_1, \\
 &\vdots \\
 \Delta_{k+p+1(1 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{B}}_p \epsilon)^\top \left[4\mathbf{D}(\hat{\mathbf{d}}) - \mathbf{D}(\hat{\mathbf{v}}) \right] \hat{\mathbf{A}} - \frac{1}{\hat{\phi}} (\hat{\mathbf{A}} \epsilon)^\top \mathbf{D}(\hat{\mathbf{v}}) \mathbf{B}_p.
 \end{aligned}$$

C.4 Perturbação na variável explicativa

Neste caso, o interesse é perturbar a r -ésima variável explicativa contínua $\eta_{\omega_0_i}^{(r)} = \eta_{0_i}^{(r)} + \omega_i$, $\forall i = 1, \dots, n$. Para este caso, consideramos as seguintes quantidades:

$$\begin{aligned}
 \epsilon_{\omega_i} &= y_i - \eta_{0_i}^\top \gamma_0 - \eta_{1_i}^\top \gamma_1 - \eta_{k_i}^\top \gamma_k - \omega_i \gamma_0^{(r)}, \\
 \delta_{\omega_i} &= \frac{(\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}})^2}{\hat{\phi}}.
 \end{aligned}$$

Logo, obtemos as seguintes derivadas básicas:

$$\frac{\partial \epsilon_{\omega_i}}{\partial \omega_l} = \begin{cases} -\gamma_0^{(r)}, & \text{se } l = i, \\ -\rho_1 \gamma_0^{(r)}, & \text{se } l = i - 1, \\ -\rho_p \gamma_0^{(r)}, & \text{se } l = i - p, \\ \vdots & \vdots \\ 0, & \text{caso contrário} \end{cases}$$

e

$$\frac{\partial W_g(\delta_{\omega_i})}{\partial \omega_l} = \begin{cases} -W'_g(\delta_{\omega_i}) \frac{2\gamma_0^{(r)}}{\hat{\phi}} (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}), & \text{se } l = i \\ +W'_g(\delta_{\omega_i}) \frac{2\rho_1 \gamma_0^{(r)}}{\hat{\phi}} (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}), & \text{se } l = i - 1, \\ \vdots & \vdots \\ +W'_g(\delta_{\omega_i}) \frac{2\rho_p \gamma_0^{(r)}}{\hat{\phi}} (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}), & \text{se } l = i - p, \\ 0, & \text{caso contrário.} \end{cases}$$

Dessa forma, a função log-verossimilhança é expressa da seguinte maneira

$$\begin{aligned} L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega}) &= \sum_{i=1}^n L_i(\boldsymbol{\theta}|\boldsymbol{\omega}) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k \\ &= -\frac{n}{2} \log(\phi) + \sum_{i=1}^n \log(g(\delta_{\omega_i})) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k. \end{aligned}$$

As funções escore em relação aos parâmetros do modelo sob o esquema de perturbação adotado são

$$\begin{aligned} \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \gamma_{0j}} &= + \frac{g'(\delta_{\omega_i})}{g(\delta_{\omega_i})} 2\phi^{-1} (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}) (-\eta_{\omega 0_{(i,j)}} + \rho_1 \eta_{\omega 0_{(i-1,j)}} + \dots + \rho_p \eta_{\omega 0_{(i-p,j)}}) \\ &= -2 \frac{W_g(\delta_{\omega_i})}{\phi} (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}) (\eta_{\omega 0_{(i,j)}} - \rho_1 \eta_{\omega 0_{(i-1,j)}} - \dots - \rho_p \eta_{\omega 0_{(i-p,j)}}), \\ \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \gamma_{lj}} &= -2 \frac{W_g(\delta_{\omega_i})}{\phi} (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}) (\eta_{\omega l_{(i,j)}} - \rho_1 \eta_{\omega l_{(i-1,j)}} - \dots - \rho_p \eta_{\omega l_{(i-p,j)}}) \\ &\quad - [\lambda_l \mathbf{M}_l \boldsymbol{\gamma}_l]_j, \quad \forall l = 1, \dots, k, \\ \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \phi} &= -\frac{1}{2\phi} + \frac{g'(\delta_{\omega_i})}{g(\delta_{\omega_i})} (-\phi^{-2}) (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}})^2 \\ &= -\frac{1}{2\phi} - \frac{W_g(\delta_{\omega_i}) \delta_{\omega_i}}{\phi}, \\ \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \rho_l} &= + \frac{g'(\delta_{\omega_i})}{g(\delta_{\omega_i})} (2\phi^{-1}) (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}) (-\epsilon_{\omega_{(i-l)}}) \\ &= + \frac{2}{\phi} W_g(\delta_{\omega_i}) (\epsilon_{\omega_i} - \rho_1 \epsilon_{\omega_{(i-1)}} - \dots - \rho_p \epsilon_{\omega_{(i-p)}}) (-\epsilon_{\omega_{(i-l)}}), \quad \forall l = 1, \dots, p. \end{aligned}$$

Calculando a derivada dessa função em relação ao vetor $\boldsymbol{\omega}$, temos que

$$\begin{aligned} \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \omega_l \partial \gamma_{l'j}} &= + W'_g(\delta_{\omega_l}) \delta_{\omega_l} \frac{4\gamma_0^{(r)}}{\phi} (\eta_{\omega l'_{lj}} - \rho_1 \eta_{\omega l'_{(l-1)j}} - \dots - \rho_p \eta_{\omega l'_{(l-p)j}}) \\ &\quad - W'_g(\delta_{\omega_{(l+1)}}) \delta_{\omega_{(l+1)}} \frac{4\rho_1 \gamma_0^{(r)}}{\phi} (\eta_{\omega l'_{(l+1)j}} - \rho_1 \eta_{\omega l'_{lj}} - \dots - \rho_p \eta_{\omega l'_{(l+1-p)j}}) \\ &\quad \vdots \\ &\quad - W'_g(\delta_{\omega_{(l+p)}}) \delta_{\omega_{(l+p)}} \frac{4\rho_p \gamma_0^{(r)}}{\phi} (\eta_{\omega l'_{(l+p)j}} - \rho_1 \eta_{\omega l'_{(l+p-1)j}} - \dots - \rho_p \eta_{\omega l'_{lj}}) \\ &\quad + W_g(\delta_{\omega_l}) \frac{2\gamma_0^{(r)}}{\phi} (\eta_{\omega l'_{lj}} - \rho_1 \eta_{\omega l'_{(l-1)j}} - \dots - \rho_p \eta_{\omega l'_{(l-p)j}}) \\ &\quad - W_g(\delta_{\omega_{(l+1)}}) \frac{2\rho_1 \gamma_0^{(r)}}{\phi} (\eta_{\omega l'_{(l+1)j}} - \rho_1 \eta_{\omega 0_{lj}} - \dots - \rho_p \eta_{\omega l'_{(l+1-p)j}}) \\ &\quad \vdots \\ &\quad - W_g(\delta_{\omega_{(l+p)}}) \frac{2\rho_p \gamma_0^{(r)}}{\phi} (\eta_{\omega l'_{(l+p)j}} - \rho_1 \eta_{\omega l'_{(l+p-1)j}} - \dots - \rho_p \eta_{\omega l'_{lj}}), \quad \forall l' = 0, 1, \dots, k. \end{aligned}$$

$$\begin{aligned}
 \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega})}{\partial \omega_l \partial \phi} &= + W'_g(\delta_{\omega_l}) \delta_{\omega_l} \frac{2\gamma_0^{(r)}}{\phi^2} (\epsilon_{\omega_l} - \rho_1 \epsilon_{\omega_{(l-1)}} - \dots - \rho_p \epsilon_{\omega_{(l-p)}}) \\
 &\quad - W'_g(\delta_{\omega_{(l+1)}}) \delta_{\omega_{(l+1)}} \frac{2\rho_1 \gamma_0^{(r)}}{\phi^2} (\epsilon_{\omega_{(l+1)}} - \rho_1 \epsilon_{\omega_l} - \dots - \rho_p \epsilon_{\omega_{(l+1-p)}}) \\
 &\quad \vdots \\
 &\quad - W'_g(\delta_{\omega_{(l+p)}}) \delta_{\omega_{(l+p)}} \frac{2\rho_p \gamma_0^{(r)}}{\phi^2} (\epsilon_{\omega_{(l+p)}} - \rho_1 \epsilon_{\omega_{(l+p-1)}} - \dots - \rho_p \epsilon_{\omega_l}) \\
 &\quad + W_g(\delta_{\omega_l}) \frac{2\gamma_0^{(r)}}{\phi^2} (\epsilon_{\omega_l} - \rho_1 \epsilon_{\omega_{(l-1)}} - \dots - \rho_p \epsilon_{\omega_{(l-p)}}) \\
 &\quad - W_g(\delta_{\omega_{(l+1)}}) \frac{2\rho_1 \gamma_0^{(r)}}{\phi^2} (\epsilon_{\omega_{(l+1)}} - \rho_1 \epsilon_{\omega_l} - \dots - \rho_p \epsilon_{\omega_{(l+1-p)}}) \\
 &\quad \vdots \\
 &\quad - W_g(\delta_{\omega_{(l+p)}}) \frac{2\rho_p \gamma_0^{(r)}}{\phi^2} (\epsilon_{\omega_{(l+p)}} - \rho_1 \epsilon_{\omega_{(l+p-1)}} - \dots - \rho_p \epsilon_{\omega_l}), \\
 \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega})}{\partial \omega_l \partial \rho_{l'}} &= - W'_g(\delta_{\omega_l}) \delta_{\omega_l} \frac{4\gamma_0^{(r)}}{\phi} (-\epsilon_{\omega_{(l-l')}}) + \\
 &\quad + W'_g(\delta_{\omega_{(l+1)}}) \delta_{\omega_{(l+1)}} \frac{4\rho_1 \gamma_0^{(r)}}{\phi} (-\epsilon_{\omega_l}) + \dots + W'_g(\delta_{\omega_{(l+p)}}) \delta_{\omega_{(l+p)}} \frac{4\rho_p \gamma_0^{(r)}}{\phi} (-\epsilon_{\omega_{(l+p-1)}}) \\
 &\quad - W_g(\delta_{\omega_l}) \frac{2\gamma_0^{(r)}}{\phi} (-\epsilon_{\omega_{(l-1)}}) + W_g(\delta_{\omega_{(l+1)}}) \frac{2\rho_1 \gamma_0^{(r)}}{\phi} (-\epsilon_{\omega_l}) \dots + W_g(\delta_{\omega_{(l+p)}}) \frac{2\rho_p \gamma_0^{(r)}}{\phi} (-\epsilon_{\omega_{(l+p-1)}}) \\
 &\quad + W_g(\delta_{\omega_{(l+1)}}) \frac{2\gamma_0^{(r)}}{\phi} (\epsilon_{\omega_{(l+1)}} - \rho_1 \epsilon_{\omega_l} - \dots - \rho_p \epsilon_{\omega_{(l+1-p)}}), \quad \forall l' = 1, \dots, p.
 \end{aligned}$$

Esse esquema de perturbação permite, por exemplo, detectar possíveis maus condicionamentos entre algumas colunas da matriz de planeamento \mathbf{X}_i , ou \mathbf{N}_0 como definimos na estrutura desta tese. Avaliando este esquema sob $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$, podemos escrever os elementos de $\boldsymbol{\Delta}$ da seguinte maneira:

$$\begin{aligned}
\Delta_{0(r_0 \times n)} &= \frac{\gamma_0^{(r)}}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_0)^\top (4\mathbf{D}(\hat{\mathbf{d}}) - \mathbf{D}(\hat{\mathbf{v}}))\hat{\mathbf{A}}, \\
\Delta_{1(r_1 \times n)} &= \frac{\gamma_0^{(r)}}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_1)^\top (4\mathbf{D}(\hat{\mathbf{d}}) - \mathbf{D}(\hat{\mathbf{v}}))\hat{\mathbf{A}}, \\
&\vdots \\
\Delta_{k(r_k \times n)} &= \frac{\gamma_0^{(r)}}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_k)^\top (4\mathbf{D}(\hat{\mathbf{d}}) - \mathbf{D}(\hat{\mathbf{v}}))\hat{\mathbf{A}}, \\
\Delta_{k+1(1 \times n)} &= \frac{\gamma_0^{(r)}}{\hat{\phi}^2} (\hat{\mathbf{A}}\epsilon)^\top (2\mathbf{D}(\hat{\mathbf{d}}) - \mathbf{D}(\hat{\mathbf{v}}))\hat{\mathbf{A}}, \\
\Delta_{k+1+1(1 \times n)} &= \frac{\gamma_0^{(r)}}{\hat{\phi}} (\hat{\mathbf{B}}_1\epsilon)^\top (\mathbf{D}(\hat{\mathbf{v}}) - 4\mathbf{D}(\hat{\mathbf{d}}))\hat{\mathbf{A}} + \frac{\gamma_0^{(r)}}{\hat{\phi}} (\hat{\mathbf{A}}\epsilon)^\top \mathbf{D}(\hat{\mathbf{v}})\mathbf{B}_1, \\
&\vdots \\
\Delta_{k+p+1(1 \times n)} &= \frac{\gamma_0^{(r)}}{\hat{\phi}} (\hat{\mathbf{B}}_p\epsilon)^\top (\mathbf{D}(\hat{\mathbf{v}}) - 4\mathbf{D}(\hat{\mathbf{d}}))\hat{\mathbf{A}} + \frac{\gamma_0^{(r)}}{\hat{\phi}} (\hat{\mathbf{A}}\epsilon)^\top \mathbf{D}(\hat{\mathbf{v}})\mathbf{B}_p.
\end{aligned}$$

C.5 Perturbação no coeficiente de autocorrelação

Tendo em vista a importância da autocorrelação na estrutura de covariância dos dados, aqui estudamos o diagnóstico de perturbação do coeficiente de autocorrelação em modelos de regressão com erros simétricos. Para tanto, considere a perturbação no r -ésimo coeficiente de autocorrelação $\rho_{r\omega} = \rho_r + \omega_i$. Além disso, temos que, $\epsilon_i = y_i - \eta_{0_i}^\top \gamma_0 - \eta_{1_i}^\top \gamma_1 - \eta_{k_i}^\top \gamma_k$. Para facilitar a compreensão das fórmulas, vamos supor que $r = 1$, mas os resultados podem ser expandidos para qualquer $r = 1, \dots, p$. Considere

$$\delta_{\omega_i} = \frac{(\epsilon_i - \rho_1 \epsilon_{(i-1)} - \omega_i \epsilon_{(i-1)} - \dots - \rho_p \epsilon_{(i-p)})^2}{\hat{\phi}}.$$

A função log-verossimilhança perturbada pode ser reescrita como

$$\begin{aligned}
L_p(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{\omega}) &= \sum L_i(\boldsymbol{\theta} | \boldsymbol{\omega}) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k \\
&= -\frac{n}{2} \log(\hat{\phi}) + \sum_i \log(g(\delta_{\omega_i})) - \frac{\lambda_1}{2} \boldsymbol{\gamma}_1^\top \mathbf{M}_1 \boldsymbol{\gamma}_1 - \dots - \frac{\lambda_k}{2} \boldsymbol{\gamma}_k^\top \mathbf{M}_k \boldsymbol{\gamma}_k.
\end{aligned}$$

As funções score em relação aos parâmetros do modelo sob o esquema de perturbação adotado

são

$$\begin{aligned}
 \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \gamma_{0j}} &= + \frac{g'(\delta_{\omega_i})}{g(\delta_{\omega_i})} 2\phi^{-1} (\epsilon_i - (\rho_{1'} + \omega_i)\epsilon_{(i-1')} - \dots - \rho_p \epsilon_{(i-p)}) \times \\
 &\quad \times (-\eta_{0(i,j)} + (\rho_{1'} + \omega_i)\eta_{0(i-1',j)} + \dots + \rho_p \eta_{0(i-p,j)}) \\
 &= - 2 \frac{W_g(\delta_{\omega_i})}{\phi} (\epsilon_i - (\rho_{1'} + \omega_i)\epsilon_{(i-1')} - \dots - \rho_p \epsilon_{(i-p)}) \times \\
 &\quad \times (\eta_{0(i,j)} - (\rho_{1'} + \omega_i)\eta_{0(i-1',j)} - \dots - \rho_p \eta_{0(i-p,j)}), \\
 \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \gamma_{lj}} &= - 2 \frac{W_g(\delta_{\omega_i})}{\phi} (\epsilon_i - (\rho_{1'} + \omega_i)\epsilon_{(i-1')} - \dots - \rho_p \epsilon_{(i-p)}) \times \\
 &\quad \times (\eta_{l(i,j)} - (\rho_{1'} + \omega_i)\eta_{l(i-1',j)} - \dots - \rho_p \eta_{l(i-p,j)}) - [\lambda_l \mathbf{M}_l \boldsymbol{\gamma}_l]_j \quad \forall l = 1, \dots, k, \\
 \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \phi} &= - \frac{1}{2\phi} + \frac{g'(\delta_{\omega_i})}{g(\delta_{\omega_i})} (-\phi^{-2}) (\epsilon_i - (\rho_{1'} + \omega_i)\epsilon_{(i-1')} - \dots - \rho_p \epsilon_{(i-p)})^2 \\
 &= - \frac{1}{2\phi} - \frac{W_g(\delta_{\omega_i})\delta_{\omega_i}}{\phi}, \\
 \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \rho_l} &= + \frac{g'(\delta_{\omega_i})}{g(\delta_{\omega_i})} (2\phi^{-1}) (\epsilon_i - (\rho_{1'} + \omega_i)\epsilon_{(i-1')} - \dots - \rho_p \epsilon_{(i-p)}) (-\epsilon_{(i-l)}) \\
 &= + \frac{2}{\phi} W_g(\delta_{\omega_i}) (\epsilon_i - (\rho_{1'} + \omega_i)\epsilon_{(i-1')} - \dots - \rho_p \epsilon_{(i-p)}) (-\epsilon_{(i-l)}), \quad \forall l = 1, \dots, p,
 \end{aligned}$$

em que l' representa que estamos lidando com a perturbação do primeiro coeficiente de autocorrelação. No entanto, esse valor pode ser substituído por qualquer valor de $r = 1, \dots, p$.

Calculando a derivada dessa função em relação ao vetor $\boldsymbol{\omega}$, temos que

$$\begin{aligned}
 \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \omega_l \partial \gamma_{l'j}} &= - \frac{4}{\phi} W'_g(\delta_{\omega_i}) \delta_{\omega_i} (-\epsilon_{(i-1')}) (\eta'_{l'(i,j)} - (\rho_{1'} + \omega_i)\eta'_{l'(i-1',j)} - \dots - \rho_p \eta'_{l'(i-p,j)}) \\
 &\quad - \frac{2}{\phi} W_g(\delta_{\omega_i}) (-\epsilon_{(i-1')}) (\eta'_{l'(i,j)} - (\rho_{1'} + \omega_i)\eta'_{l'(i-1',j)} - \dots - \rho_p \eta'_{l'(i-p,j)}) \\
 &\quad - \frac{2}{\phi} W_g(\delta_{\omega_i}) (\epsilon_i - (\rho_{1'} + \omega_i)\epsilon_{(i-1')} - \dots - \rho_p \epsilon_{(i-p)}) (-\eta'_{l'(i-1')}), \quad \forall l' = 1, \dots, k, \\
 \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \omega_l \partial \phi} &= - \frac{2}{\phi^2} W'_g(\delta_{\omega_i}) \delta_{\omega_i} (\epsilon_i - (\rho_{1'} + \omega_i)\epsilon_{(i-1')} - \dots - \rho_p \epsilon_{(i-p)}) (-\epsilon_{(i-1')}) \\
 &\quad - \frac{2}{\phi^2} W_g(\delta_{\omega_i}) (\epsilon_i - (\rho_{1'} + \omega_i)\epsilon_{(i-1')} - \dots - \rho_p \epsilon_{(i-p)}) (-\epsilon_{(i-1')}), \\
 \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \omega_l \partial \rho_{l'}} &= + \frac{4}{\phi} W'_g(\delta_{\omega_i}) \delta_{\omega_i} (-\epsilon_{(i-l')}) (-\epsilon_{(i-1')}) \\
 &\quad + \frac{2}{\phi} W_g(\delta_{\omega_i}) (-\epsilon_{(i-l')}) (-\epsilon_{(i-1')}), \quad \forall l' = 1, \dots, p.
 \end{aligned}$$

Com esse último cálculo, completamos a matriz $\boldsymbol{\Delta}$ para esse esquema de perturbação, tornando-se possível encontrar a curvatura normal. Lembramos que essas matrizes devem ser avaliadas em $\hat{\boldsymbol{\theta}}$ e no vetor de não perturbação $\boldsymbol{\omega}_0$, que nesse caso, é dado por $(0, \dots, 0)^\top$. Assim, temos que $\mathbf{D}(\mathbf{d}_\omega)$ e $\mathbf{D}(\mathbf{v}_\omega)$ em função de $\hat{\boldsymbol{\delta}}$. Denotamos essas matrizes diagonal por $\mathbf{D}(\hat{\mathbf{d}})$ e $\mathbf{D}(\hat{\mathbf{v}})$. Ainda, como \mathbf{A}_ω é função de $\boldsymbol{\rho}$, chamaremos de $\hat{\mathbf{A}}$ essas matrizes utilizando $\hat{\boldsymbol{\rho}}$ sob $\boldsymbol{\omega}_0$. Assim, podemos escrever os elementos de $\boldsymbol{\Delta}$, avaliados em $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$, como

$$\begin{aligned}
\Delta_{0(r_0 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_0)^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 4\mathbf{D}(\hat{\mathbf{d}}) \right] \mathbf{D}(\mathbf{B}_{1'\hat{\epsilon}}) + \frac{1}{\hat{\phi}} (\mathbf{B}_{1'\mathbf{N}_0})^\top \mathbf{D}(\hat{\mathbf{v}}) \mathbf{D}(\hat{\mathbf{A}}_{\hat{\epsilon}}), \\
\Delta_{1(r_1 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_1)^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 4\mathbf{D}(\hat{\mathbf{d}}) \right] \mathbf{D}(\mathbf{B}_{1'\hat{\epsilon}}) + \frac{1}{\hat{\phi}} (\mathbf{B}_{1'\mathbf{N}_1})^\top \mathbf{D}(\hat{\mathbf{v}}) \mathbf{D}(\hat{\mathbf{A}}_{\hat{\epsilon}}), \\
&\vdots \\
\Delta_{k(r_k \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{A}}\mathbf{N}_k)^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 4\mathbf{D}(\hat{\mathbf{d}}) \right] \mathbf{D}(\mathbf{B}_{1'\hat{\epsilon}}) + \frac{1}{\hat{\phi}} (\mathbf{B}_{1'\mathbf{N}_k})^\top \mathbf{D}(\hat{\mathbf{v}}) \mathbf{D}(\hat{\mathbf{A}}_{\hat{\epsilon}}), \\
\Delta_{k+1(1 \times n)} &= \frac{1}{\hat{\phi}^2} (\hat{\mathbf{A}}_{\hat{\epsilon}})^\top \left[\mathbf{D}(\hat{\mathbf{v}}) - 2\mathbf{D}(\hat{\mathbf{d}}) \right] \mathbf{D}(\mathbf{B}_{1'\hat{\epsilon}}), \\
\Delta_{k+1+1(1 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{B}}_1 \hat{\epsilon})^\top \left[4\mathbf{D}(\hat{\mathbf{d}}) - \mathbf{D}(\hat{\mathbf{v}}) \right] \mathbf{D}(\mathbf{B}_{1'\hat{\epsilon}}), \\
&\vdots \\
\Delta_{k+p+1(1 \times n)} &= \frac{1}{\hat{\phi}} (\hat{\mathbf{B}}_p \hat{\epsilon})^\top \left[4\mathbf{D}(\hat{\mathbf{d}}) - \mathbf{D}(\hat{\mathbf{v}}) \right] \mathbf{D}(\mathbf{B}_{1'\hat{\epsilon}}).
\end{aligned}$$

em que $\mathbf{D}(\hat{\mathbf{A}}_{\hat{\epsilon}})$ é uma matriz diagonal com elementos dados por $\hat{\mathbf{A}}_{\hat{\epsilon}}$ e $\mathbf{D}(\hat{\mathbf{B}}_{1'\hat{\epsilon}})$ é uma matriz diagonal com elementos dados por $\hat{\mathbf{B}}_{1'\hat{\epsilon}}$, dado que $1'$ representa que estamos lidando com a perturbação do primeiro coeficiente de autocorrelação, porém, esse valor pode ser substituído por qualquer valor de $r = 1, 2, \dots, p$.

Apêndice D

Análises Simulações

D.1 Tabelas e gráficos da simulação AR(1)

Tabela D.1: Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro normal e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 e PE_2 , NG e L_2 .

ρ	Erro	n=100					n=300					n=500				
		AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM
-0,75	N	11,8	-0,7391	0,0841	0,0109	0,0049	63,5	-0,7311	0,0132	0,0189	0,0021	82,9	-0,7314	0,4698	0,0186	0,0014
	t_3	2,0	-0,7406	0,0163	0,0094	0,0057	0,0	-0,7310	0,0239	0,0190	0,0024	0,0	-0,7319	0,6730	0,0181	0,0015
	PE_1	65,8	-0,7390	0,0032	0,0110	0,0055	26,8	-0,7313	0,2424	0,0187	0,0023	13,1	-0,7314	0,5050	0,0186	0,0015
	PE_2	1,9	-0,7402	0,0040	0,0098	0,0054	1,0	-0,7309	0,0058	0,0191	0,0023	0,3	-0,7321	0,6157	0,0179	0,0015
	NG	3,6	-0,7106	0,0053	0,0394	0,0086	0,3	-0,7235	0,3623	0,0265	0,0028	0,0	-0,7269	0,3779	0,0231	0,0017
	L_2	14,9	-0,7397	0,1259	0,0103	0,0051	8,4	-0,7310	0,0037	0,0190	0,0022	3,7	-0,7316	0,8367	0,0184	0,0014
-0,25	N	11,5	-0,2789	0,3998	-0,0289	0,0104	63,6	-0,2394	0,5455	0,0106	0,0036	83,1	-0,2307	0,9918	0,0193	0,0025
	t_3	1,1	-0,2815	0,0169	-0,0315	0,0122	0,0	-0,2400	0,3567	0,0100	0,0042	0,0	-0,2317	0,7003	0,0183	0,0028
	PE_1	65,4	-0,2792	0,7183	-0,0292	0,0117	27,4	-0,2397	0,1500	0,0103	0,0041	13,1	-0,2304	0,9211	0,0196	0,0026
	PE_2	2,0	-0,2801	0,0891	-0,0301	0,0116	0,7	-0,2395	0,0983	0,0105	0,0041	0,4	-0,2316	0,8826	0,0184	0,0028
	NG	4,2	-0,2126	0,0065	0,0374	0,0191	0,3	-0,2233	0,4779	0,0267	0,0052	0,0	-0,2219	0,6738	0,0281	0,0032
	L_2	15,8	-0,2797	0,0446	-0,0297	0,0108	8,0	-0,2397	0,3733	0,0103	0,0038	3,4	-0,2311	0,9555	0,0189	0,0026
0,25	N	12,9	0,1606	0,8829	-0,0894	0,0206	63,0	0,2267	0,3070	-0,0233	0,0042	82,8	0,2433	0,5149	-0,0067	0,0022
	t_3	1,1	0,1593	0,9754	-0,0907	0,0235	0,0	0,2251	0,8793	-0,0249	0,0048	0,0	0,2418	0,6550	-0,0082	0,0026
	PE_1	65,4	0,1595	0,8757	-0,0905	0,0219	27,7	0,2265	0,6594	-0,0235	0,0046	13,3	0,2437	0,7721	-0,0063	0,0023
	PE_2	2,0	0,1608	0,9499	-0,0892	0,0225	0,3	0,2261	0,7379	-0,0239	0,0047	0,3	0,2424	0,6632	-0,0076	0,0025
	NG	5,2	0,2548	0,0041	0,0048	0,0288	0,4	0,2448	0,6390	-0,0052	0,0047	0,1	0,2536	0,6486	0,0036	0,0025
	L_2	13,4	0,1606	0,7218	-0,0894	0,0214	8,6	0,2260	0,4752	-0,0240	0,0044	3,5	0,2425	0,9004	-0,0075	0,0023
0,75	N	17,8	0,6566	0,0423	-0,0934	0,0188	65,1	0,7036	0,1678	-0,0464	0,0046	83,8	0,7179	0,4510	-0,0321	0,0023
	t_3	1,0	0,6566	0,0054	-0,0934	0,0206	0,0	0,7023	0,2857	-0,0477	0,0051	0,0	0,7168	0,2782	-0,0332	0,0026
	PE_1	58,2	0,6553	0,0871	-0,0947	0,0195	27,0	0,7030	0,2074	-0,0470	0,0048	13,0	0,7175	0,0594	-0,0325	0,0024
	PE_2	3,2	0,6583	0,0067	-0,0917	0,0199	0,3	0,7029	0,5673	-0,0471	0,0050	0,0	0,7171	0,3408	-0,0329	0,0026
	NG	6,7	0,7383	0,1580	-0,0117	0,0137	0,5	0,7261	0,7191	-0,0239	0,0046	0,1	0,7301	0,1540	-0,0199	0,0023
	L_2	13,1	0,6569	0,0428	-0,0931	0,0194	7,1	0,7031	0,3844	-0,0469	0,0048	3,1	0,7174	0,2614	-0,0326	0,0024
ρ	Erro	n=100					n=300					n=500				
		AIC	$\hat{\phi}$	KS	viés	EQM	AIC	$\hat{\phi}$	KS	viés	EQM	AIC	$\hat{\phi}$	KS	viés	EQM
-0,75	N	11,8	0,9886	0,0080	-0,0114	0,0186	63,5	1,0505	0,8844	0,0505	0,0097	82,9	1,0626	0,5537	0,0626	0,0082
	t_3	2,0	0,5803	0,0083	-0,4197	0,1849	0,0	0,6458	0,7104	-0,3542	0,1289	0,0	0,6597	0,6981	-0,3403	0,1179
	PE_1	65,8	1,6312	0,0015	0,6312	0,4529	26,8	1,7563	0,5389	0,7563	0,5938	13,1	1,7824	0,6435	0,7824	0,6251
	PE_2	1,9	0,3993	0,0032	-0,6007	0,3639	1,0	0,4297	0,7093	-0,5703	0,3264	0,3	0,4359	0,6688	-0,5641	0,3189
	NG	3,6	3,0225	0,0123	2,0225	4,3125	0,3	3,0366	0,4431	2,0366	4,2184	0,0	3,0471	0,3681	2,0471	4,2302
	L_2	14,9	0,3182	0,0016	-0,6818	0,4669	8,4	0,3415	0,7357	-0,6585	0,4345	3,7	0,3462	0,7013	-0,6538	0,4279
-0,25	N	11,5	0,9486	0,0002	-0,0514	0,0188	63,6	1,0089	0,8579	0,0089	0,0066	83,1	1,0251	0,4105	0,0251	0,0046
	t_3	1,1	0,5594	0,0053	-0,4406	0,2016	0,0	0,6211	0,7846	-0,3789	0,1467	0,0	0,6367	0,7115	-0,3633	0,1339
	PE_1	65,4	1,5668	0,0015	0,5668	0,3695	27,4	1,6873	0,4139	0,6873	0,4924	13,1	1,7200	0,7987	0,7200	0,5304
	PE_2	2,0	0,3835	0,0007	-0,6165	0,3829	0,7	0,4129	0,5175	-0,5871	0,3458	0,4	0,4206	0,8377	-0,5794	0,3364
	NG	4,2	2,8714	0,0010	1,8714	3,7242	0,3	2,8969	0,1230	1,8969	3,6606	0,0	2,9265	0,4969	1,9265	3,7478
	L_2	15,8	0,3056	0,0020	-0,6944	0,4840	8,0	0,3280	0,6286	-0,6720	0,4523	3,4	0,3341	0,8168	-0,6659	0,4439
0,25	N	12,9	0,9060	0,0390	-0,0940	0,0261	63,0	0,9788	0,8239	-0,0212	0,0066	82,8	0,9961	0,6843	-0,0039	0,0037
	t_3	1,1	0,5375	0,1687	-0,4625	0,2215	0,0	0,6037	0,8856	-0,3963	0,1599	0,0	0,6194	0,5280	-0,3806	0,1466
	PE_1	65,4	1,4974	0,0179	0,4974	0,2989	27,7	1,6378	0,3719	0,6378	0,4257	13,3	1,6715	0,4617	0,6715	0,4622
	PE_2	2,0	0,3669	0,1594	-0,6331	0,4037	0,3	0,4008	0,6006	-0,5992	0,3601	0,3	0,4089	0,8529	-0,5911	0,3501
	NG	5,2	2,7343	0,0012	1,7343	3,2607	0,4	2,7949	0,3725	1,7949	3,2785	0,1	2,8357	0,5644	1,8357	3,4028
	L_2	13,4	0,2923	0,2242	-0,7077	0,5027	8,6	0,3184	0,6949	-0,6816	0,4653	3,5	0,3247	0,8148	-0,6753	0,4565
0,75	N	17,8	0,9070	0,3612	-0,0930	0,0259	65,1	0,9632	0,5460	-0,0368	0,0072	83,8	0,9801	0,8832	-0,0199	0,0039
	t_3	1,0	0,5402	0,1221	-0,4598	0,2191	0,0	0,5963	0,9934	-0,4037	0,1657	0,0	0,6111	0,5794	-0,3889	0,1529
	PE_1	58,2	1,5037	0,1485	0,5037	0,3054	27,0	1,6134	0,5108	0,6134	0,3944	13,0	1,6456	0,3938	0,6456	0,4276
	PE_2	3,2	0,3673	0,2647	-0,6327	0,4032	0,3	0,3948	0,8651	-0,6052	0,3673	0,0	0,4026	0,6034	-0,5974	0,3575
	NG	6,7	2,7923	0,0000	1,7923	3,5642	0,5	2,7853	0,0000	1,7853	3,2870	0,1	2,7989	0,0137	1,7989	3,2730
	L_2	13,1	0,2926	0,2591	-0,7074	0,5022	7,1	0,3136	0,8730	-0,6864	0,4718	3,1	0,3197	0,6568	-0,6803	0,4632

Tabela D.2: Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro t -Student ($\nu = 3$) e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 e PE_2 , NG e L_2 .

ρ	Erro	n=100					n=300					n=500				
		AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM
-0,75	N	2,3	-0,7534	0,0240	-0,0034	0,0043	0,0	-0,7450	0,0047	0,0050	0,0015	0,0	-0,7433	0,0809	0,0067	0,0009
	t_3	75,9	-0,7486	0,2210	0,0014	0,0034	91,5	-0,7426	0,6717	0,0074	0,0011	97,7	-0,7426	0,4280	0,0074	0,0006
	PE_1	3,8	-0,7617	0,0015	-0,0117	0,0084	0,0	-0,7513	0,0121	-0,0013	0,0040	0,0	-0,7453	0,1740	0,0047	0,0029
	PE_2	2,0	-0,7494	0,1040	0,0006	0,0034	2,7	-0,7435	0,1657	0,0065	0,0011	1,0	-0,7425	0,0907	0,0075	0,0007
	NG	0,0	-0,7361	0,0008	0,0139	0,0109	0,0	-0,7459	0,0059	0,0041	0,0048	0,0	-0,7418	0,2201	0,0082	0,0035
	L_2	16,0	-0,7497	0,0662	0,0003	0,0033	5,8	-0,7431	0,3949	0,0069	0,0011	1,3	-0,7428	0,4370	0,0072	0,0007
	-0,25	N	1,6	-0,3503	0,0743	-0,1003	0,0156	0,0	-0,2787	0,9194	-0,0287	0,0035	0,0	-0,2584	0,9944	-0,0084
t_3		78,0	-0,3103	0,5275	-0,0603	0,0094	93,1	-0,2620	0,7928	-0,0120	0,0022	98,0	-0,2498	0,4679	-0,0002	0,0012
PE_1		3,2	-0,3977	0,3608	-0,1477	0,0330	0,0	-0,3115	0,8522	-0,0615	0,0114	0,0	-0,2764	0,5193	-0,0264	0,0064
PE_2		1,5	-0,3204	0,4200	-0,0704	0,0103	2,4	-0,2666	0,8317	-0,0166	0,0024	1,0	-0,2515	0,5478	-0,0015	0,0012
NG		0,0	-0,3535	0,9213	-0,1035	0,0272	0,0	-0,3050	0,9706	-0,0550	0,0123	0,0	-0,2736	0,3301	-0,0236	0,0075
L_2		15,7	-0,3230	0,2828	-0,0730	0,0104	4,5	-0,2664	0,8776	-0,0164	0,0023	1,0	-0,2523	0,8765	-0,0023	0,0012
0,25		N	2,3	0,2185	0,2610	-0,0315	0,0080	0,0	0,2307	0,4563	-0,0193	0,0034	0,0	0,2402	0,7783	-0,0098
	t_3	79,8	0,2355	0,1779	-0,0145	0,0063	93,7	0,2424	0,7304	-0,0076	0,0021	98,3	0,2475	0,9967	-0,0025	0,0012
	PE_1	2,0	0,1877	0,7871	-0,0623	0,0212	0,0	0,1981	0,9139	-0,0519	0,0121	0,0	0,2202	0,4105	-0,0298	0,0073
	PE_2	1,6	0,2317	0,8905	-0,0183	0,0063	1,7	0,2398	0,4562	-0,0102	0,0023	0,9	0,2464	0,7801	-0,0036	0,0012
	NG	0,3	0,2446	0,1310	-0,0054	0,0265	0,0	0,2025	0,5135	-0,0475	0,0137	0,0	0,2204	0,7933	-0,0296	0,0087
	L_2	14,0	0,2314	0,6933	-0,0186	0,0060	4,6	0,2400	0,3142	-0,0100	0,0022	0,8	0,2454	0,9849	-0,0046	0,0012
	0,75	N	2,0	0,6846	0,0037	-0,0654	0,0128	0,0	0,7095	0,4005	-0,0405	0,0038	0,0	0,7232	0,0719	-0,0268
t_3		79,6	0,7089	0,0006	-0,0411	0,0073	93,9	0,7262	0,0170	-0,0238	0,0019	98,5	0,7345	0,1457	-0,0155	0,0009
PE_1		1,9	0,6497	0,0411	-0,1003	0,0281	0,0	0,6693	0,2545	-0,0807	0,0135	0,0	0,6924	0,2077	-0,0576	0,0077
PE_2		1,8	0,7029	0,0001	-0,0471	0,0082	1,6	0,7225	0,0413	-0,0275	0,0022	0,7	0,7322	0,1153	-0,0178	0,0010
NG		0,2	0,6889	0,0046	-0,0611	0,0248	0,0	0,6695	0,4078	-0,0805	0,0159	0,0	0,6881	0,4222	-0,0619	0,0094
L_2		14,5	0,7015	0,0000	-0,0485	0,0084	4,5	0,7219	0,1477	-0,0281	0,0022	0,8	0,7315	0,1022	-0,0185	0,0011
-0,75		N	2,3	2,5687	0,0000	1,5687	3,6043	0,0	2,7101	0,0000	1,7101	3,2781	0,0	2,7294	0,0076	1,7294
	t_3	75,9	0,9358	0,0707	-0,0642	0,0458	91,5	1,0294	0,4628	0,0294	0,0153	97,7	1,0420	0,1885	0,0420	0,0099
	PE_1	3,8	5,2308	0,0000	4,2308	27,2654	0,0	5,9707	0,0000	4,9707	28,6592	0,0	6,1310	0,0000	5,1310	28,7046
	PE_2	2,0	0,8426	0,0000	-0,1574	0,0808	2,7	0,8881	0,0163	-0,1119	0,0299	1,0	0,8949	0,1237	-0,1051	0,0207
	NG	0,0	9,7813	0,0000	8,7813	118,6714	0,0	10,6044	0,0000	9,6044	106,3785	0,0	10,8524	0,0000	9,8524	105,4287
	L_2	16,0	0,6227	0,0025	-0,3773	0,1657	5,8	0,6563	0,0843	-0,3437	0,1256	1,3	0,6614	0,1184	-0,3386	0,1188
	-0,25	N	1,6	2,5687	0,0000	1,5687	3,7361	0,0	2,6765	0,0000	1,6765	3,1890	0,0	2,6890	0,0121	1,6890
t_3		78,0	0,9098	0,0477	-0,0902	0,0488	93,1	0,9929	0,2344	-0,0071	0,0141	98,0	1,0084	0,3919	0,0084	0,0075
PE_1		3,2	5,3255	0,0000	4,3255	29,8133	0,0	5,9680	0,0000	4,9680	29,0941	0,0	6,0930	0,0001	5,0930	28,3575
PE_2		1,5	0,8299	0,0001	-0,1701	0,0862	2,4	0,8679	0,0025	-0,1321	0,0351	1,0	0,8751	0,1552	-0,1249	0,0250
NG		0,0	9,7166	0,0000	8,7166	118,9890	0,0	10,5505	0,0000	9,5505	106,6125	0,0	10,7661	0,0000	9,7661	103,7499
L_2		15,7	0,6097	0,0236	-0,3903	0,1755	4,5	0,6391	0,0289	-0,3609	0,1378	1,0	0,6451	0,2040	-0,3549	0,1300
0,25		N	2,3	2,5777	0,0000	1,5777	3,8078	0,0	2,6355	0,0000	1,6355	3,0661	0,0	2,6545	0,0006	1,6545
	t_3	79,8	0,8916	0,0095	-0,1084	0,0503	93,7	0,9685	0,2193	-0,0315	0,0149	98,3	0,9851	0,4829	-0,0149	0,0078
	PE_1	2,0	5,4493	0,0000	4,4493	31,6027	0,0	5,9239	0,0000	4,9239	28,8757	0,0	6,0522	0,0000	5,0522	28,0310
	PE_2	1,6	0,8226	0,0000	-0,1774	0,0896	1,7	0,8509	0,0027	-0,1491	0,0403	0,9	0,8601	0,0372	-0,1399	0,0294
	NG	0,3	9,7681	0,0000	8,7681	117,4357	0,0	10,4359	0,0000	9,4359	104,8142	0,0	10,6803	0,0000	9,6803	102,3037
	L_2	14,0	0,6018	0,0003	-0,3982	0,1817	4,6	0,6256	0,0682	-0,3744	0,1478	0,8	0,6331	0,1357	-0,3669	0,1388
	0,75	N	2,0	2,4984	0,0000	1,4984	3,2865	0,0	2,5944	0,0000	1,5944	2,8755	0,0	2,6522	0,0002	1,6522
t_3		79,6	0,8968	0,0122	-0,1032	0,0485	93,9	0,9593	0,2687	-0,0407	0,0147	98,5	0,9746	0,5345	-0,0254	0,0076
PE_1		1,9	5,2152	0,0000	4,2152	26,7981	0,0	5,8632	0,0000	4,8632	27,6914	0,0	6,1379	0,0000	5,1379	29,1244
PE_2		1,8	0,8123	0,0000	-0,1877	0,0864	1,6	0,8398	0,0180	-0,1602	0,0419	0,7	0,8543	0,0519	-0,1457	0,0308
NG		0,2	9,5193	0,0000	8,5193	104,7854	0,0	10,3911	0,0000	9,3911	102,6555	0,0	10,8623	0,0000	9,8623	106,8165
L_2		14,5	0,5976	0,0001	-0,4024	0,1833	4,5	0,6181	0,0858	-0,3819	0,1528	0,8	0,6277	0,0321	-0,3723	0,1426

Tabela D.3: Percentual de acerto pelo método AIC, média das estimativas, valor-P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro exponencial potência ($\nu = -0,3$) e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 e PE_2 , NG e L_2 .

ρ	Erro	n=100					n=300					n=500				
		AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM
-0,75	N	4,6	-0,7221	0,0167	0,0279	0,0059	13,0	-0,7207	0,0085	0,0293	0,0028	9,1	-0,7207	0,2632	0,0293	0,0019
	t_3	0,3	-0,7248	0,0033	0,0252	0,0073	0,0	-0,7210	0,0033	0,0290	0,0032	0,0	-0,7206	0,1777	0,0294	0,0023
	PE_1	86,5	-0,7207	0,0361	0,0293	0,0061	85,3	-0,7210	0,0135	0,0290	0,0027	90,4	-0,7210	0,4906	0,0290	0,0018
	PE_2	1,1	-0,7243	0,0049	0,0257	0,0070	0,0	-0,7215	0,0017	0,0285	0,0032	0,0	-0,7208	0,3744	0,0292	0,0023
	NG	2,4	-0,6819	0,1539	0,0681	0,0121	1,7	-0,7106	0,0153	0,0394	0,0036	0,5	-0,7146	0,1214	0,0354	0,0023
	L_2	5,1	-0,7234	0,0169	0,0266	0,0064	0,0	-0,7207	0,0017	0,0293	0,0030	0,0	-0,7206	0,1578	0,0294	0,0021
-0,25	N	3,8	-0,3099	0,7975	-0,0599	0,0096	8,7	-0,2341	0,4006	0,0159	0,0034	5,3	-0,2190	0,5785	0,0310	0,0030
	t_3	0,0	-0,3215	0,6189	-0,0715	0,0138	0,0	-0,2380	0,2689	0,0120	0,0045	0,0	-0,2215	0,7741	0,0285	0,0036
	PE_1	86,8	-0,3032	0,8174	-0,0532	0,0090	88,2	-0,2319	0,4627	0,0181	0,0033	91,7	-0,2179	0,8054	0,0321	0,0029
	PE_2	1,0	-0,3192	0,7182	-0,0692	0,0128	0,0	-0,2379	0,5503	0,0121	0,0043	0,0	-0,2216	0,3636	0,0284	0,0036
	NG	5,3	-0,2232	0,0582	0,0268	0,0130	3,1	-0,2094	0,1618	0,0406	0,0053	3,0	-0,2046	0,7839	0,0454	0,0041
	L_2	3,1	-0,3155	0,9498	-0,0655	0,0112	0,0	-0,2360	0,5236	0,0140	0,0038	0,0	-0,2202	0,3486	0,0298	0,0033
0,25	N	3,9	0,2347	0,7343	-0,0153	0,0097	5,1	0,2545	0,9707	0,0045	0,0034	3,3	0,2645	0,9924	0,0145	0,0023
	t_3	0,0	0,2301	0,9772	-0,0199	0,0134	0,0	0,2499	0,9073	-0,0001	0,0047	0,0	0,2614	0,7475	0,0114	0,0030
	PE_1	84,0	0,2360	0,9468	-0,0140	0,0096	88,8	0,2569	0,9726	0,0069	0,0032	91,8	0,2659	0,9934	0,0159	0,0021
	PE_2	0,9	0,2322	0,9650	-0,0178	0,0127	0,0	0,2505	0,9540	0,0005	0,0045	0,0	0,2619	0,6597	0,0119	0,0030
	NG	9,1	0,3705	0,0000	0,1205	0,0446	6,1	0,2830	0,8618	0,0030	0,0050	4,9	0,2802	0,9874	0,0302	0,0030
	L_2	2,1	0,2329	0,8816	-0,0171	0,0111	0,0	0,2522	0,9916	0,0022	0,0040	0,0	0,2630	0,7620	0,0130	0,0026
0,75	N	5,1	0,6738	0,0369	-0,0762	0,0154	4,1	0,7069	0,0567	-0,0431	0,0041	2,1	0,7237	0,2747	-0,0263	0,0019
	t_3	0,2	0,6704	0,1095	-0,0796	0,0189	0,0	0,7003	0,1870	-0,0497	0,0057	0,0	0,7187	0,1265	-0,0313	0,0027
	PE_1	82,1	0,6746	0,0358	-0,0754	0,0146	80,2	0,7107	0,1215	-0,0393	0,0036	84,3	0,7267	0,4775	-0,0233	0,0016
	PE_2	1,1	0,6723	0,1389	-0,0777	0,0180	0,0	0,7013	0,2740	-0,0487	0,0055	0,0	0,7191	0,3901	-0,0309	0,0026
	NG	9,2	0,7739	0,1965	0,0239	0,0155	15,7	0,7400	0,1600	-0,0100	0,0038	13,6	0,7456	0,0020	-0,0044	0,0020
	L_2	2,3	0,6725	0,0705	-0,0775	0,0168	0,0	0,7035	0,1435	-0,0465	0,0049	0,0	0,7211	0,2007	-0,0289	0,0023
-0,75	N	4,6	0,6563	0,0017	-0,3437	0,1242	13,0	0,6816	0,1728	-0,3184	0,1036	9,1	0,6810	0,4551	-0,3190	0,1031
	t_3	0,3	0,4035	0,0285	-0,5965	0,3595	0,0	0,4464	0,1467	-0,5536	0,3078	0,0	0,4524	0,2638	-0,5476	0,3007
	PE_1	86,5	1,0606	0,0061	0,0606	0,0194	85,3	1,1054	0,1610	0,1054	0,0168	90,4	1,1037	0,5563	0,1037	0,0141
	PE_2	1,1	0,2704	0,0073	-0,7296	0,5335	0,0	0,2869	0,0481	-0,7131	0,5089	0,0	0,2884	0,6708	-0,7116	0,5067
	NG	2,4	2,0212	0,0600	1,0212	1,1234	1,7	1,9226	0,0935	0,9226	0,8715	0,5	1,8933	0,5990	0,8933	0,8090
	L_2	5,1	0,2165	0,0101	-0,7835	0,6146	0,0	0,2292	0,0753	-0,7708	0,5945	0,0	0,2302	0,6746	-0,7698	0,5928
-0,25	N	3,8	0,6153	0,0562	-0,3847	0,1540	8,7	0,6447	0,5432	-0,3553	0,1283	5,3	0,6501	0,8447	-0,3499	0,1236
	t_3	0,0	0,3847	0,0065	-0,6153	0,3822	0,0	0,4261	0,7268	-0,5739	0,3307	0,0	0,4353	0,8119	-0,5647	0,3196
	PE_1	86,8	0,9903	0,0243	-0,0097	0,0152	88,2	1,0424	0,4328	0,0424	0,0071	91,7	1,0510	0,7939	0,0510	0,0057
	PE_2	1,0	0,2550	0,0186	-0,7450	0,5562	0,0	0,2724	0,3608	-0,7276	0,5298	0,0	0,2761	0,5958	-0,7239	0,5242
	NG	5,3	1,8578	0,0539	0,8578	0,8166	3,1	1,7979	0,1839	0,7979	0,6547	3,0	1,7912	0,6921	0,7912	0,6356
	L_2	3,1	0,2045	0,0077	-0,7955	0,6336	0,0	0,2177	0,3367	-0,7823	0,6122	0,0	0,2206	0,5403	-0,7794	0,6077
0,25	N	3,9	0,5747	0,3410	-0,4253	0,1869	5,1	0,6153	0,5490	-0,3847	0,1499	3,3	0,6241	0,9064	-0,3759	0,1425
	t_3	0,0	0,3638	0,3916	-0,6362	0,4084	0,0	0,4103	0,8295	-0,5897	0,3490	0,0	0,4212	0,7757	-0,5788	0,3358
	PE_1	84,0	0,9233	0,3413	-0,0767	0,0209	88,8	0,9922	0,5391	-0,0078	0,0049	91,8	1,0063	0,5729	0,0063	0,0029
	PE_2	0,9	0,2392	0,4919	-0,7608	0,5800	0,0	0,2609	0,5045	-0,7391	0,5466	0,0	0,2659	0,6667	-0,7341	0,5391
	NG	9,1	1,7537	0,0000	0,7537	0,7028	6,1	1,6968	0,5118	0,6968	0,5006	4,9	1,7046	0,4761	0,7046	0,5049
	L_2	2,1	0,1919	0,5691	-0,8081	0,6538	0,0	0,2087	0,3180	-0,7913	0,6265	0,0	0,2125	0,7205	-0,7875	0,6203
0,75	N	5,1	0,5729	0,5816	-0,4271	0,1880	4,1	0,6008	0,6563	-0,3992	0,1611	2,1	0,6087	0,4471	-0,3913	0,1542
	t_3	0,2	0,3669	0,1383	-0,6331	0,4044	0,0	0,4042	0,2615	-0,5958	0,3561	0,0	0,4139	0,3402	-0,5861	0,3443
	PE_1	82,1	0,9202	0,4968	-0,0798	0,0203	80,2	0,9669	0,6249	-0,0331	0,0054	84,3	0,9796	0,4386	-0,0204	0,0030
	PE_2	1,1	0,2394	0,2589	-0,7606	0,5797	0,0	0,2556	0,5222	-0,7444	0,5544	0,0	0,2601	0,5602	-0,7399	0,5477
	NG	9,2	1,7268	0,0000	0,7268	0,6465	15,7	1,6710	0,0000	0,6710	0,4819	13,6	1,6676	0,0002	0,6676	0,4583
	L_2	2,3	0,1920	0,5070	-0,8080	0,6536	0,0	0,2046	0,4385	-0,7954	0,6330	0,0	0,2080	0,6067	-0,7920	0,6274

Tabela D.4: Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro exponencial potência ($\nu = 0,5$) e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 e PE_2 , NG e L_2 .

ρ	Erro	n=100					n=300					n=500				
		AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM
-0,75	N	10,2	-0,7543	0,0071	-0,0043	0,0041	8,0	-0,7461	0,0414	0,0039	0,0016	2,8	-0,7427	0,1913	0,0073	0,0009
	t_3	20,3	-0,7521	0,0368	-0,0021	0,0040	4,8	-0,7451	0,1187	0,0049	0,0014	0,9	-0,7426	0,3942	0,0074	0,0009
	PE_1	24,6	-0,7579	0,0794	-0,0079	0,0056	0,1	-0,7479	0,0195	0,0021	0,0024	0,0	-0,7430	0,0196	0,0070	0,0014
	PE_2	6,4	-0,7526	0,0175	-0,0026	0,0038	33,4	-0,7455	0,0770	0,0045	0,0014	42,1	-0,7426	0,3825	0,0074	0,0009
	NG	1,0	-0,7405	0,2597	0,0095	0,0066	0,0	-0,7443	0,0229	0,0057	0,0027	0,0	-0,7408	0,0701	0,0092	0,0016
	L_2	37,5	-0,7527	0,0022	-0,0027	0,0038	53,7	-0,7453	0,0686	0,0047	0,0014	54,2	-0,7426	0,1916	0,0074	0,0009
	-0,25	N	10,5	-0,3545	0,1186	-0,1045	0,0161	7,6	-0,2758	0,7398	-0,0258	0,0036	2,4	-0,2565	0,9078	-0,0065
t_3		21,9	-0,3377	0,3987	-0,0877	0,0136	4,8	-0,2690	0,6652	-0,0190	0,0031	1,1	-0,2534	0,9394	-0,0034	0,0017
PE_1		21,8	-0,3697	0,4705	-0,1197	0,0222	0,1	-0,2836	0,7176	-0,0336	0,0055	0,0	-0,2604	0,8947	-0,0104	0,0029
PE_2		5,6	-0,3418	0,2248	-0,0918	0,0138	35,5	-0,2700	0,8098	-0,0200	0,0031	47,5	-0,2536	0,7631	-0,0036	0,0017
NG		0,7	-0,3345	0,9416	-0,0845	0,0176	0,0	-0,2762	0,4719	-0,0262	0,0056	0,0	-0,2561	0,8543	-0,0061	0,0032
L_2		39,5	-0,3443	0,1103	-0,0943	0,0139	52,0	-0,2715	0,8887	-0,0215	0,0031	49,0	-0,2545	0,9025	-0,0045	0,0017
0,25		N	12,9	0,2229	0,6705	-0,0271	0,0086	6,7	0,2309	0,8261	-0,0191	0,0035	2,4	0,2415	0,5220	-0,0085
	t_3	21,8	0,2278	0,9691	-0,0222	0,0087	5,1	0,2353	0,7506	-0,0147	0,0031	1,1	0,2442	0,4582	-0,0058	0,0018
	PE_1	19,9	0,2160	0,7948	-0,0340	0,0125	0,0	0,2241	0,2459	-0,0259	0,0055	0,0	0,2370	0,6886	-0,0130	0,0032
	PE_2	5,0	0,2261	0,9478	-0,0239	0,0087	34,6	0,2348	0,5953	-0,0152	0,0030	49,2	0,2439	0,3267	-0,0061	0,0018
	NG	1,6	0,2756	0,0073	0,0256	0,0208	0,0	0,2332	0,5695	-0,0168	0,0056	0,0	0,2416	0,7148	-0,0084	0,0034
	L_2	38,8	0,2264	0,7924	-0,0236	0,0080	53,6	0,2339	0,9187	-0,0161	0,0031	47,3	0,2433	0,2286	-0,0067	0,0017
	0,75	N	12,1	0,6870	0,0003	-0,0630	0,0122	7,7	0,7078	0,1442	-0,0422	0,0039	1,7	0,7232	0,1027	-0,0268
t_3		21,5	0,6988	0,0046	-0,0512	0,0102	6,0	0,7144	0,4139	-0,0356	0,0031	1,1	0,7270	0,2751	-0,0230	0,0015
PE_1		16,4	0,6760	0,0014	-0,0740	0,0169	0,0	0,6987	0,0920	-0,0513	0,0059	0,0	0,7173	0,1844	-0,0327	0,0030
PE_2		6,2	0,6969	0,0027	-0,0531	0,0103	27,3	0,7137	0,2766	-0,0363	0,0031	47,4	0,7267	0,1042	-0,0233	0,0015
NG		1,4	0,7163	0,0053	-0,0337	0,0145	0,0	0,7075	0,5518	-0,0425	0,0058	0,0	0,7212	0,6700	-0,0288	0,0030
L_2		42,4	0,6940	0,0013	-0,0560	0,0106	59,0	0,7120	0,1493	-0,0380	0,0033	49,8	0,7256	0,1213	-0,0244	0,0016
-0,75		N	10,2	2,4325	0,0269	1,4325	2,2486	8,0	2,6214	0,1397	1,6214	2,6936	2,8	2,6442	0,2752	1,6442
	t_3	20,3	1,2645	0,0427	0,2645	0,1312	4,8	1,3938	0,4532	0,3938	0,1753	0,9	1,4178	0,5219	0,4178	0,1880
	PE_1	24,6	4,2095	0,0177	3,2095	10,9800	0,1	4,6740	0,2076	3,6740	13,7511	0,0	4,7440	0,1769	3,7440	14,1773
	PE_2	6,4	0,9352	0,0762	-0,0648	0,0322	33,4	1,0079	0,0694	0,0079	0,0091	42,1	1,0185	0,6278	0,0185	0,0061
	NG	1,0	7,5478	0,0339	6,5478	45,0687	0,0	8,0360	0,0304	7,0360	50,2761	0,0	8,1232	0,2004	7,1232	51,2256
	L_2	37,5	0,7346	0,0812	-0,2654	0,0879	53,7	0,7891	0,0638	-0,2109	0,0501	54,2	0,7970	0,6622	-0,2030	0,0448
	-0,25	N	10,5	2,3785	0,0445	1,3785	2,0970	7,6	2,5664	0,2665	1,5664	2,5223	2,4	2,5984	0,3137	1,5984
t_3		21,9	1,2297	0,0349	0,2297	0,1132	4,8	1,3588	0,2519	0,3588	0,1510	1,1	1,3846	0,6724	0,3846	0,1603
PE_1		21,8	4,1360	0,0504	3,1360	10,5358	0,1	4,5869	0,1573	3,5869	13,1270	0,0	4,6751	0,1498	3,6751	13,6629
PE_2		5,6	0,9116	0,0991	-0,0884	0,0354	35,5	0,9850	0,2600	-0,0150	0,0100	47,5	0,9981	0,6392	-0,0019	0,0055
NG		0,7	7,3321	0,0877	6,3321	42,2960	0,0	7,8596	0,0788	6,8596	47,8388	0,0	7,9925	0,1388	6,9925	49,3701
L_2		39,5	0,7153	0,0664	-0,2847	0,0983	52,0	0,7707	0,3775	-0,2293	0,0587	49,0	0,7806	0,6884	-0,2194	0,0515
0,25		N	12,9	2,3501	0,0096	1,3501	2,0221	6,7	2,5377	0,0454	1,5377	2,4337	2,4	2,5751	0,3401	1,5751
	t_3	21,8	1,2184	0,0240	0,2184	0,1067	5,1	1,3385	0,0943	0,3385	0,1362	1,1	1,3662	0,7478	0,3662	0,1468
	PE_1	19,9	4,1026	0,0056	3,1026	10,3502	0,0	4,5504	0,0281	3,5504	12,8701	0,0	4,6446	0,4176	3,6446	13,4428
	PE_2	5,0	0,9004	0,0434	-0,0996	0,0374	34,6	0,9720	0,1967	-0,0280	0,0105	49,2	0,9871	0,6793	-0,0129	0,0058
	NG	1,6	7,2598	0,0016	6,2598	41,5198	0,0	7,7813	0,0236	6,7813	46,7797	0,0	7,9283	0,3441	6,9283	48,4847
	L_2	38,8	0,7062	0,0436	-0,2938	0,1034	53,6	0,7602	0,1635	-0,2398	0,0635	47,3	0,7718	0,6708	-0,2282	0,0556
	0,75	N	12,1	2,3739	0,0356	1,3739	2,0853	7,7	2,5385	0,2054	1,5385	2,4367	1,7	2,5651	0,6215	1,5651
t_3		21,5	1,2238	0,0042	0,2238	0,1078	6,0	1,3376	0,1914	0,3376	0,1351	1,1	1,3586	0,2431	0,3586	0,1404
PE_1		16,4	4,1631	0,0131	3,1631	10,7266	0,0	4,5707	0,1870	3,5707	13,0270	0,0	4,6402	0,2152	3,6402	13,4076
PE_2		6,2	0,9072	0,0536	-0,0928	0,0358	27,3	0,9711	0,1832	-0,0289	0,0105	47,4	0,9819	0,6951	-0,0181	0,0056
NG		1,4	7,4860	0,0000	6,4860	46,0110	0,0	7,8345	0,0849	6,8345	47,5915	0,0	7,9243	0,1084	6,9243	48,4332
L_2		42,4	0,7108	0,0241	-0,2892	0,1004	59,0	0,7592	0,2190	-0,2408	0,0640	49,8	0,7677	0,6948	-0,2323	0,0572

Tabela D.5: Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro normal generalizada e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 e PE_2 , NG e L_2 .

ρ	Erro	n=100					n=300					n=500				
		AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM
-0,75	N	4,6	-0,6962	0,1300	0,0538	0,0092	11,8	-0,7001	0,0295	0,0499	0,0045	6,9	-0,7055	0,2405	0,0445	0,0032
	t_3	0,4	-0,6970	0,0228	0,0530	0,0109	0,0	-0,6992	0,0657	0,0508	0,0051	0,0	-0,7041	0,1206	0,0459	0,0037
	PE_1	86,8	-0,6971	0,1424	0,0529	0,0092	87,7	-0,7007	0,0284	0,0493	0,0043	92,5	-0,7066	0,1723	0,0434	0,0031
	PE_2	1,2	-0,6962	0,0147	0,0538	0,0104	0,0	-0,6992	0,0307	0,0508	0,0050	0,0	-0,7038	0,0747	0,0462	0,0037
	NG	1,9	-0,6443	0,0430	0,1057	0,0210	0,4	-0,6860	0,0211	0,0640	0,0063	0,6	-0,6981	0,1533	0,0519	0,0040
	L_2	5,1	-0,6962	0,0715	0,0538	0,0099	0,1	-0,6995	0,0231	0,0505	0,0048	0,0	-0,7046	0,1057	0,0454	0,0035
-0,25	N	3,8	-0,2740	0,6733	-0,0240	0,0084	6,5	-0,1977	0,9990	0,0523	0,0060	4,9	-0,1921	0,9717	0,0579	0,0054
	t_3	0,1	-0,2841	0,7162	-0,0341	0,0120	0,0	-0,2007	0,8972	0,0493	0,0068	0,0	-0,1927	0,6667	0,0573	0,0059
	PE_1	88,6	-0,2687	0,7547	-0,0187	0,0082	91,1	-0,1960	0,8452	0,0540	0,0060	93,4	-0,1920	0,8878	0,0580	0,0053
	PE_2	1,1	-0,2800	0,6428	-0,0300	0,0110	0,0	-0,1997	0,7389	0,0503	0,0068	0,0	-0,1922	0,6200	0,0578	0,0059
	NG	2,6	-0,1582	0,0200	0,0918	0,0257	2,2	-0,1660	0,6183	0,0840	0,0109	1,7	-0,1754	0,6851	0,0746	0,0078
	L_2	3,8	-0,2786	0,7999	-0,0286	0,0098	0,2	-0,1991	0,8923	0,0509	0,0063	0,0	-0,1923	0,9258	0,0577	0,0056
0,25	N	3,9	0,2617	0,9874	0,0117	0,0085	3,8	0,2799	0,8846	0,0299	0,0042	2,3	0,2850	0,9116	0,0350	0,0033
	t_3	0,1	0,2576	0,9477	0,0076	0,0119	0,0	0,2761	0,3030	0,0261	0,0052	0,0	0,2824	0,9837	0,0324	0,0038
	PE_1	88,7	0,2619	0,9401	0,0119	0,0086	89,5	0,2818	0,8872	0,0318	0,0041	91,9	0,2863	0,7540	0,0363	0,0032
	PE_2	0,7	0,2610	0,9255	0,0110	0,0111	0,0	0,2773	0,8115	0,0273	0,0052	0,0	0,2833	0,9472	0,0333	0,0038
	NG	4,6	0,4088	0,0013	0,1588	0,0519	6,6	0,3140	0,1731	0,0640	0,0083	5,8	0,3045	0,5903	0,0545	0,0052
	L_2	2,0	0,2603	0,9949	0,0103	0,0097	0,1	0,2781	0,6780	0,0281	0,0046	0,0	0,2838	0,9896	0,0338	0,0035
0,75	N	5,9	0,6721	0,0129	-0,0779	0,0146	1,4	0,7131	0,1099	-0,0369	0,0035	0,7	0,7237	0,3523	-0,0263	0,0019
	t_3	0,2	0,6707	0,1008	-0,0793	0,0183	0,0	0,7063	0,0798	-0,0437	0,0049	0,0	0,7184	0,0457	-0,0316	0,0027
	PE_1	80,7	0,6717	0,0368	-0,0783	0,0142	84,6	0,7169	0,4991	-0,0331	0,0031	83,9	0,7268	0,4905	-0,0232	0,0016
	PE_2	0,7	0,6737	0,0492	-0,0763	0,0168	0,0	0,7074	0,0150	-0,0426	0,0047	0,0	0,7193	0,1437	-0,0307	0,0026
	NG	9,0	0,7725	0,2579	0,0225	0,0127	13,9	0,7608	0,0008	0,0108	0,0046	15,4	0,7511	0,0000	0,0011	0,0025
	L_2	3,5	0,6716	0,0478	-0,0784	0,0161	0,1	0,7097	0,0742	-0,0403	0,0042	0,0	0,7210	0,0786	-0,0290	0,0023
ρ	Erro	n=100					n=300					n=500				
		AIC	$\hat{\phi}$	KS	viés	EQM	AIC	$\hat{\phi}$	KS	viés	EQM	AIC	$\hat{\phi}$	KS	viés	EQM
-0,75	N	4,6	0,4232	0,4846	-0,5768	0,3353	11,8	0,4380	0,2653	-0,5620	0,3168	6,9	0,4324	0,3710	-0,5676	0,3227
	t_3	0,4	0,2579	0,0837	-0,7421	0,5522	0,0	0,2860	0,2812	-0,7140	0,5104	0,0	0,2873	0,9217	-0,7127	0,5082
	PE_1	86,8	0,6851	0,7111	-0,3149	0,1060	87,7	0,7101	0,7183	-0,2899	0,0866	92,5	0,6996	0,2775	-0,3004	0,0917
	PE_2	1,2	0,1739	0,1168	-0,8261	0,6829	0,0	0,1843	0,1916	-0,8157	0,6655	0,0	0,1833	0,6433	-0,8167	0,6671
	NG	1,9	1,3396	0,0280	0,3396	0,1518	0,4	1,2500	0,4192	0,2500	0,0719	0,6	1,2068	0,1358	0,2068	0,0479
	L_2	5,1	0,1391	0,1044	-0,8609	0,7415	0,1	0,1471	0,1469	-0,8529	0,7275	0,0	0,1462	0,7422	-0,8538	0,7290
-0,25	N	3,8	0,3891	0,0723	-0,6109	0,3752	6,5	0,4068	0,6723	-0,5932	0,3526	4,9	0,4060	0,7137	-0,5940	0,3533
	t_3	0,1	0,2415	0,0678	-0,7585	0,5766	0,0	0,2690	0,3729	-0,7310	0,5348	0,0	0,2722	0,8538	-0,7278	0,5300
	PE_1	88,6	0,6269	0,0634	-0,3731	0,1441	91,1	0,6568	0,8536	-0,3432	0,1197	93,4	0,6552	0,9048	-0,3448	0,1201
	PE_2	1,1	0,1609	0,0798	-0,8391	0,7044	0,0	0,1720	0,8011	-0,8280	0,6857	0,0	0,1726	0,9085	-0,8274	0,6847
	NG	2,6	1,2160	0,0064	0,2160	0,0793	2,2	1,1425	0,1585	0,1425	0,0275	1,7	1,1206	0,5850	0,1206	0,0183
	L_2	3,8	0,1289	0,0700	-0,8711	0,7590	0,2	0,1375	0,7606	-0,8625	0,7440	0,0	0,1379	0,8483	-0,8621	0,7433
0,25	N	3,9	0,3687	0,1822	-0,6313	0,4004	3,8	0,3796	0,8562	-0,6204	0,3855	2,3	0,3838	0,9479	-0,6162	0,3801
	t_3	0,1	0,2327	0,0219	-0,7673	0,5899	0,0	0,2546	0,3276	-0,7454	0,5561	0,0	0,2602	0,6528	-0,7398	0,5476
	PE_1	88,7	0,5918	0,3809	-0,4082	0,1711	89,5	0,6101	0,9405	-0,3899	0,1536	91,9	0,6171	0,7617	-0,3829	0,1476
	PE_2	0,7	0,1534	0,0328	-0,8466	0,7171	0,0	0,1615	0,4029	-0,8385	0,7033	0,0	0,1639	0,9096	-0,8361	0,6992
	NG	4,6	1,1260	0,0000	0,1260	0,0530	6,6	1,0486	0,5087	0,0486	0,0080	5,8	1,0481	0,7937	0,0481	0,0054
	L_2	2,0	0,1231	0,0220	-0,8769	0,7693	0,1	0,1291	0,2088	-0,8709	0,7585	0,0	0,1310	0,9243	-0,8690	0,7551
0,75	N	5,9	0,3675	0,1340	-0,6325	0,4018	1,4	0,3742	0,3407	-0,6258	0,3922	0,7	0,3771	0,0129	-0,6229	0,3883
	t_3	0,2	0,2343	0,4121	-0,7657	0,5875	0,0	0,2539	0,3014	-0,7461	0,5570	0,0	0,2584	0,4231	-0,7416	0,5502
	PE_1	80,7	0,5906	0,1245	-0,4094	0,1716	84,6	0,5992	0,1347	-0,4008	0,1619	83,9	0,6043	0,0334	-0,3957	0,1573
	PE_2	0,7	0,1533	0,2200	-0,8467	0,7172	0,0	0,1599	0,4061	-0,8401	0,7059	0,0	0,1617	0,3064	-0,8383	0,7028
	NG	9,0	1,1237	0,0000	0,1237	0,0757	13,9	1,0695	0,0000	0,0695	0,0285	15,4	1,0374	0,0000	0,0374	0,0077
	L_2	3,5	0,1230	0,1927	-0,8770	0,7694	0,1	0,1280	0,4441	-0,8720	0,7605	0,0	0,1294	0,3807	-0,8706	0,7580

Tabela D.6: Percentual de acerto pelo método AIC, média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\rho}$ e $\hat{\phi}$ do estudo de simulação no qual os dados são gerados a partir do modelo (7.1) sob erro L_2 e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 e PE_2 , NG e L_2 .

ρ	Erro	n=100					n=300					n=500				
		AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM	AIC	$\hat{\rho}$	KS	viés	EQM
-0,75	N	11,9	-0,7557	0,0038	-0,0057	0,0041	16,8	-0,7478	0,0373	0,0022	0,0015	9,3	-0,7446	0,1661	0,0054	0,0009
	t_3	15,5	-0,7540	0,0194	-0,0040	0,0042	2,3	-0,7469	0,0666	0,0031	0,0014	0,2	-0,7444	0,2833	0,0056	0,0009
	PE_1	28,8	-0,7590	0,0532	-0,0090	0,0055	0,2	-0,7496	0,0332	0,0004	0,0022	0,0	-0,7451	0,0238	0,0049	0,0013
	PE_2	2,6	-0,7545	0,0047	-0,0045	0,0041	12,6	-0,7474	0,1054	0,0026	0,0014	11,4	-0,7444	0,3286	0,0056	0,0009
	NG	1,6	-0,7433	0,1243	0,0067	0,0064	0,0	-0,7461	0,0516	0,0039	0,0025	0,0	-0,7430	0,0915	0,0070	0,0015
	L_2	39,6	-0,7542	0,0039	-0,0042	0,0039	68,1	-0,7470	0,0796	0,0030	0,0014	79,1	-0,7445	0,3925	0,0055	0,0009
-0,25	N	12,7	-0,3558	0,0796	-0,1058	0,0163	17,3	-0,2788	0,9923	-0,0288	0,0037	8,7	-0,2593	0,9823	-0,0093	0,0019
	t_3	15,4	-0,3431	0,4088	-0,0931	0,0148	2,6	-0,2737	0,9748	-0,0237	0,0035	0,6	-0,2567	0,7668	-0,0067	0,0018
	PE_1	27,1	-0,3684	0,2780	-0,1184	0,0218	0,1	-0,2853	0,8529	-0,0353	0,0054	0,0	-0,2631	0,9028	-0,0131	0,0028
	PE_2	2,1	-0,3471	0,1940	-0,0971	0,0150	11,6	-0,2748	0,9428	-0,0248	0,0035	11,8	-0,2574	0,6606	-0,0074	0,0019
	NG	1,4	-0,3355	0,9977	-0,0855	0,0175	0,0	-0,2783	0,6967	-0,0283	0,0055	0,0	-0,2591	0,9660	-0,0091	0,0029
	L_2	41,3	-0,3476	0,0404	-0,0976	0,0146	68,4	-0,2753	0,8812	-0,0253	0,0034	78,9	-0,2575	0,5766	-0,0075	0,0017
0,25	N	14,5	0,2237	0,3905	-0,0263	0,0083	16,0	0,2295	0,9555	-0,0205	0,0036	8,1	0,2395	0,4954	-0,0105	0,0020
	t_3	15,3	0,2267	0,9986	-0,0233	0,0089	1,9	0,2326	0,6276	-0,0174	0,0035	0,8	0,2414	0,4340	-0,0086	0,0019
	PE_1	25,6	0,2179	0,4056	-0,0321	0,0118	0,1	0,2238	0,3844	-0,0262	0,0053	0,0	0,2354	0,8369	-0,0146	0,0030
	PE_2	2,6	0,2255	0,8092	-0,0245	0,0088	9,6	0,2319	0,6478	-0,0181	0,0034	10,6	0,2409	0,4518	-0,0091	0,0020
	NG	2,4	0,2732	0,0041	0,0232	0,0198	0,0	0,2329	0,7044	-0,0171	0,0055	0,0	0,2399	0,8404	-0,0101	0,0032
	L_2	39,6	0,2263	0,6228	-0,0237	0,0079	72,4	0,2319	0,8367	-0,0181	0,0033	80,5	0,2409	0,2990	-0,0091	0,0019
0,75	N	15,3	0,6880	0,0045	-0,0620	0,0121	15,9	0,7082	0,0962	-0,0418	0,0039	7,7	0,7239	0,0439	-0,0261	0,0019
	t_3	15,3	0,6968	0,0067	-0,0532	0,0108	2,6	0,7129	0,2231	-0,0371	0,0034	0,3	0,7263	0,2354	-0,0237	0,0017
	PE_1	20,6	0,6787	0,0017	-0,0713	0,0162	0,0	0,7006	0,0986	-0,0494	0,0055	0,0	0,7192	0,1503	-0,0308	0,0027
	PE_2	2,7	0,6949	0,0055	-0,0551	0,0109	7,0	0,7118	0,2577	-0,0382	0,0035	7,8	0,7256	0,1051	-0,0244	0,0017
	NG	2,9	0,7140	0,0088	-0,0360	0,0143	0,0	0,7079	0,4367	-0,0421	0,0055	0,0	0,7231	0,5540	-0,0269	0,0028
	L_2	43,2	0,6935	0,0058	-0,0565	0,0108	74,5	0,7115	0,1603	-0,0385	0,0034	84,2	0,7257	0,0794	-0,0243	0,0017
ρ	Erro	n=100					n=300					n=500				
		AIC	$\hat{\phi}$	KS	viés	EQM	AIC	$\hat{\phi}$	KS	viés	EQM	AIC	$\hat{\phi}$	KS	viés	EQM
-0,75	N	11,9	3,0403	0,0184	2,0403	4,4670	16,8	3,2669	0,0668	2,2669	5,2390	9,3	3,2960	0,5548	2,2960	5,3323
	t_3	15,5	1,6213	0,0823	0,6213	0,4795	2,3	1,7928	0,4056	0,7928	0,6596	0,2	1,8270	0,6192	0,8270	0,7039
	PE_1	28,8	5,2419	0,0108	4,2419	19,0898	0,2	5,7856	0,0823	4,7856	23,3078	0,0	5,8668	0,0567	4,8668	23,9322
	PE_2	2,6	1,1792	0,1442	0,1792	0,0749	12,6	1,2723	0,0492	0,2723	0,0880	11,4	1,2873	0,3703	0,2873	0,0912
	NG	1,6	9,3594	0,0189	8,3594	73,4833	0,0	9,9429	0,0092	8,9429	81,2242	0,0	10,0389	0,0962	9,0389	82,4543
	L_2	39,6	0,9268	0,1019	-0,0732	0,0317	68,1	0,9965	0,0776	-0,0035	0,0085	79,1	1,0077	0,6914	0,0077	0,0054
-0,25	N	12,7	2,9861	0,0237	1,9861	4,2544	17,3	3,2108	0,0988	2,2108	4,9956	8,7	3,2459	0,2278	2,2459	5,1056
	t_3	15,4	1,5900	0,0253	0,5900	0,4418	2,6	1,7604	0,1791	0,7604	0,6124	0,6	1,7963	0,9210	0,7963	0,6538
	PE_1	27,1	5,1697	0,0244	4,1697	18,5327	0,1	5,6961	0,0492	4,6961	22,4835	0,0	5,7846	0,0739	4,7846	23,1410
	PE_2	2,1	1,1563	0,0321	0,1563	0,0674	11,6	1,2498	0,3393	0,2498	0,0776	11,8	1,2667	0,5629	0,2667	0,0798
	NG	1,4	9,1405	0,0147	8,1405	69,9027	0,0	9,7552	0,0541	8,7552	77,9571	0,0	9,8819	0,0600	8,8819	79,6451
	L_2	41,3	0,9080	0,0311	-0,0920	0,0349	68,4	0,9784	0,2575	-0,0216	0,0098	78,9	0,9913	0,5217	-0,0087	0,0054
0,25	N	14,5	2,9718	0,0116	1,9718	4,2009	16,0	3,1835	0,0341	2,1835	4,8751	8,1	3,2295	0,2100	2,2295	5,0334
	t_3	15,3	1,5905	0,0339	0,5905	0,4404	1,9	1,7436	0,0751	0,7436	0,5864	0,8	1,7805	0,6393	0,7805	0,6286
	PE_1	25,6	5,1613	0,0015	4,1613	18,4991	0,1	5,6615	0,0194	4,6615	22,1559	0,0	5,7735	0,1427	4,7735	23,0453
	PE_2	2,6	1,1519	0,0572	0,1519	0,0657	9,6	1,2381	0,1669	0,2381	0,0718	10,6	1,2579	0,3548	0,2579	0,0753
	NG	2,4	9,1126	0,0013	8,1126	69,6411	0,0	9,6775	0,0425	8,6775	76,5810	0,0	9,8502	0,1789	8,8502	79,1144
	L_2	39,6	0,9041	0,0644	-0,0959	0,0353	72,4	0,9689	0,1228	-0,0311	0,0102	80,5	0,9838	0,5102	-0,0162	0,0056
0,75	N	15,3	2,9875	0,0378	1,9875	4,2582	15,9	3,1889	0,0810	2,1889	4,9018	7,7	3,2169	0,5810	2,2169	4,9745
	t_3	15,3	1,5941	0,0112	0,5941	0,4422	2,6	1,7501	0,1337	0,7501	0,5957	0,3	1,7782	0,6765	0,7782	0,6242
	PE_1	20,6	5,2106	0,0019	4,2106	18,9011	0,0	5,6877	0,0413	4,6877	22,4237	0,0	5,7561	0,2298	4,7561	22,8669
	PE_2	2,7	1,1559	0,1124	0,1559	0,0662	7,0	1,2407	0,1233	0,2407	0,0731	7,8	1,2540	0,7194	0,2540	0,0728
	NG	2,9	9,2931	0,0000	8,2931	73,6465	0,0	9,7244	0,0473	8,7244	77,4982	0,0	9,8196	0,2250	8,8196	78,5387
	L_2	43,2	0,9064	0,0672	-0,0936	0,0340	74,5	0,9703	0,1267	-0,0297	0,0101	84,2	0,9808	0,7386	-0,0192	0,0055

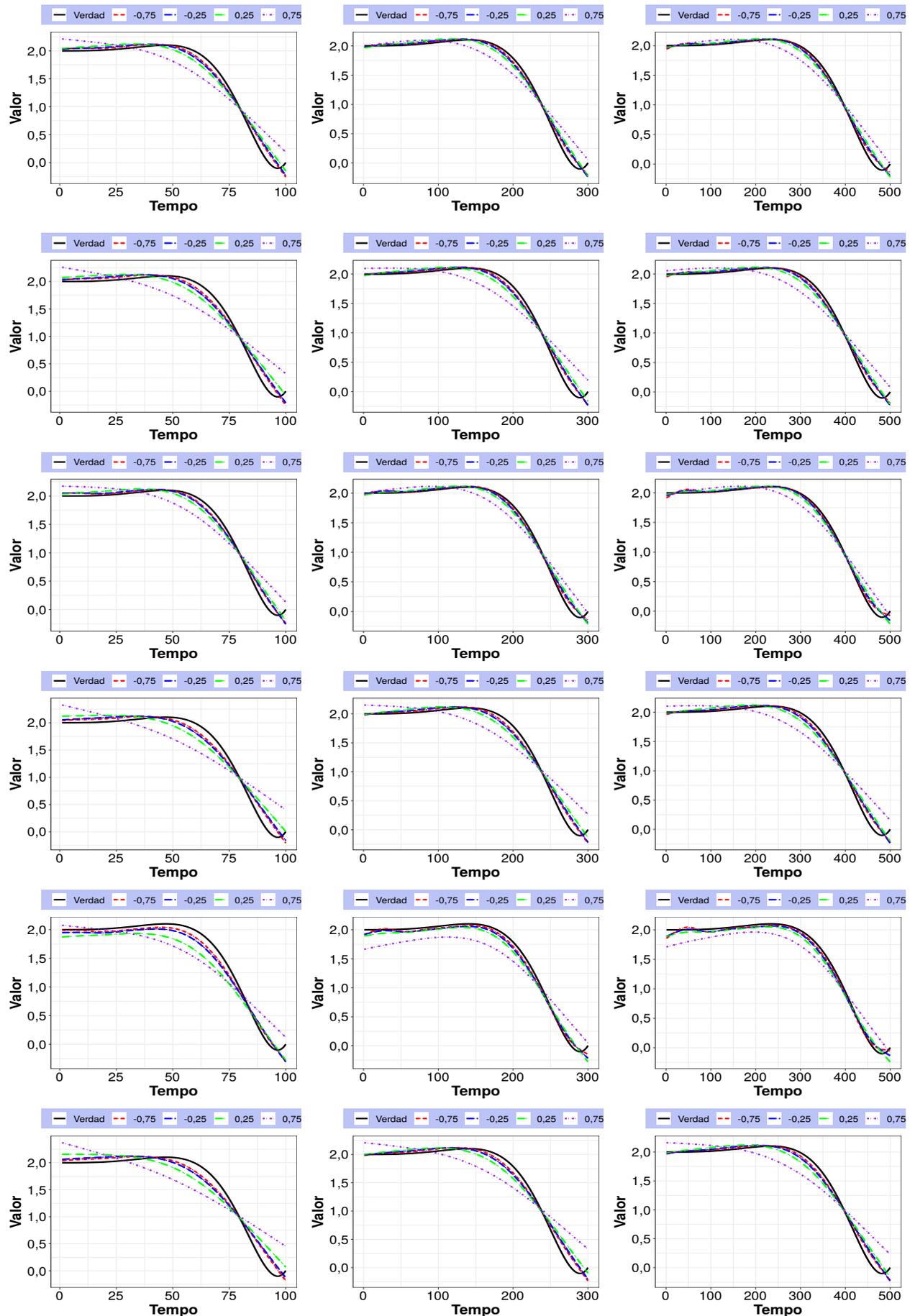


Figura D.1: Gráficos da média das estimativas para a função $f_1(t)$ sob $\phi = 1$ com coeficientes de autocorrelação $\rho = -0,75, -0,25, 0,25$ e $0,75$ a partir do estudo de simulação a qual os dados foram gerados do modelo (7.1) sob os erros com distribuição normal (primeiro), t_3 (segundo), PE_1 (terceiro), PE_2 (quarto), NG (quinto) e L_2 (sexto) e ajustado sob o mesmo erro do modelo. Tamanho amostrais de $n = 100, 300$ e 500 , na esquerda, meio e direita, respectivamente.

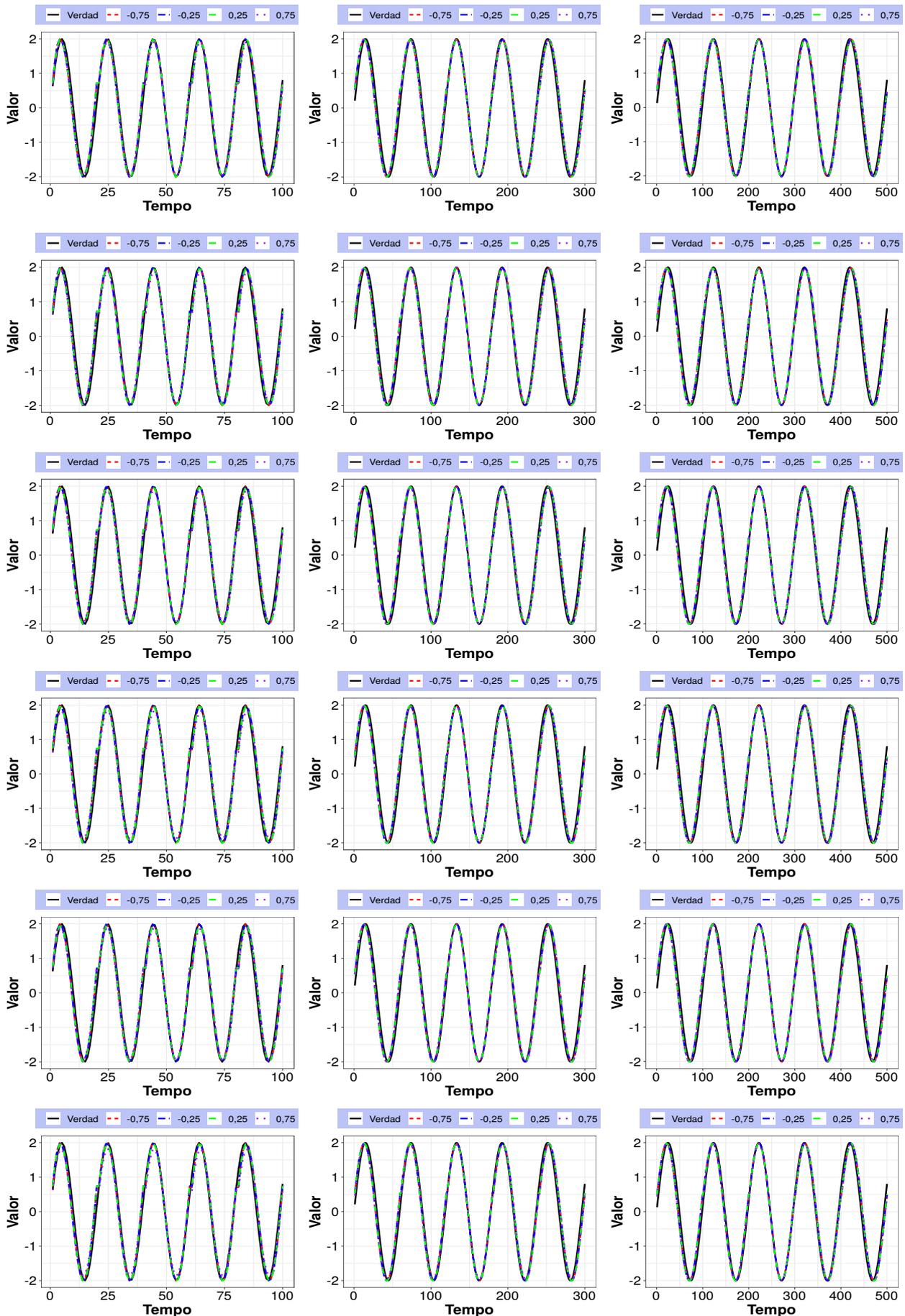


Figura D.2: Gráficos da média das estimativas para a função $f_2(s)$ sob $\phi = 1$ com coeficientes de autocorrelação $\rho = -0,75, -0,25, 0,25$ e $0,75$ a partir do estudo de simulação a qual os dados foram gerados do modelo (7.1) sob os erros com distribuição normal (primeiro), t_3 (segundo), PE_1 (terceiro), PE_2 (quarto), NG (quinto) e L_2 (sexto) e ajustado sob o mesmo erro do modelo. Tamanho amostrais de $n = 100, 300$ e 500 , na esquerda, meio e direita, respectivamente.

D.2 Tabelas e gráficos da simulação AR(2)

Tabela D.7: Percentual de acerto pelo método AIC no estudo de simulação no qual os dados são gerados sob a distribuição dos erros, normal, t_3 , PE_1 e PE_2 , NG e L_2 .

ρ	Erro	N			t_3			PE_1			PE_2			Ng			L_2		
		100	300	500	100	300	500	100	300	500	100	300	500	100	300	500	100	300	500
1	N	3,6	52,3	76,1	2,6	0,0	0,0	81,8	40,0	20,4	0,0	0,1	0,0	1,8	0,0	0,0	10,2	7,6	3,5
	t_3	2,3	0,0	0,0	77,5	93,4	97,6	6,0	0,0	0,0	0,0	0,9	0,5	0,0	0,0	0,0	14,2	5,7	1,9
	PE_1	0,2	1,5	1,3	0,0	0,0	0,0	96,6	97,3	97,7	0,0	0,0	0,0	2,2	1,2	1,0	1,0	0,0	0,0
	PE_2	9,5	7,4	1,8	22,1	8,6	2,7	26,9	0,0	0,0	0,0	14,1	28,8	0,5	0,0	0,0	41,0	69,9	66,7
	NG	0,2	0,1	0,4	0,2	0,0	0,0	98,0	98,7	98,5	0,0	0,0	0,0	1,1	1,2	1,1	0,5	0,0	0,0
	L_2	11,3	16,9	7,4	16,3	3,0	0,5	30,8	0,6	0,0	0,1	3,1	4,5	1,1	0,0	0,0	40,4	76,4	87,6
3	N	10,4	55,5	78,4	1,9	0,0	0,0	69,6	36,7	18,5	0,1	0,2	0,0	4,7	0,1	0,0	13,3	7,5	3,1
	t_3	2,2	0,1	0,0	77,8	92,2	96,9	3,7	0,0	0,0	0,0	1,2	0,6	0,2	0,0	0,0	16,1	6,5	2,5
	PE_1	1,7	1,8	1,2	0,0	0,0	0,0	93,3	95,7	96,0	0,0	0,0	0,0	4,0	2,5	2,8	1,0	0,0	0,0
	PE_2	15,1	6,8	1,6	21,2	7,5	1,9	15,9	0,0	0,0	0,8	15,1	32,0	0,7	0,0	0,0	46,3	70,6	64,5
	NG	0,4	0,2	0,5	0,3	0,0	0,0	95,5	97,8	96,3	0,0	0,0	0,0	2,4	2,0	3,2	1,4	0,0	0,0
	L_2	16,5	16,4	7,7	15,1	2,6	0,3	21,3	0,3	0,0	0,5	3,5	4,9	1,4	0,0	0,0	45,2	77,2	87,1

Tabela D.8: Média das estimativas, valor-*P* do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro normal e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 , PE_2 , NG e L_2 .

		n=100				n=300				n=500			
ϕ	Erro	$\hat{\beta}_0$	KS	viés	EQM	$\hat{\beta}_0$	KS	viés	EQM	$\hat{\beta}_0$	KS	viés	EQM
1	N	2,0611	0,2538	0,0611	0,0173	2,0907	0,2296	0,0907	0,0121	2,0946	0,1003	0,0946	0,0113
	t_3	2,0630	0,3408	0,0630	0,0213	2,0916	0,3065	0,0916	0,0133	2,0942	0,7643	0,0942	0,0120
	PE_1	2,0592	0,3163	0,0592	0,0188	2,0905	0,2315	0,0905	0,0128	2,0948	0,2353	0,0948	0,0118
	PE_2	2,0618	0,3237	0,0618	0,0202	2,0908	0,4127	0,0908	0,0130	2,0943	0,5586	0,0943	0,0119
	NG_3	2,1461	0,0000	0,1461	0,3374	2,1233	0,0000	0,1233	0,1168	2,1038	0,9990	0,1038	0,0412
	L_2	2,0623	0,1449	0,0623	0,0185	2,0912	0,2485	0,0912	0,0125	2,0943	0,2232	0,0943	0,0115
3	N	2,0571	0,2205	0,0571	0,0436	2,0916	0,2143	0,0916	0,0200	2,0951	0,0812	0,0951	0,0161
	t_3	2,0565	0,3011	0,0565	0,0548	2,0933	0,2932	0,0933	0,0239	2,0944	0,7368	0,0944	0,0183
	PE_1	2,0561	0,5883	0,0561	0,0495	2,0911	0,1854	0,0911	0,0222	2,0955	0,1875	0,0955	0,0176
	PE_2	2,0561	0,0994	0,0561	0,0524	2,0916	0,1763	0,0916	0,0231	2,0947	0,6460	0,0947	0,0182
	NG_3	2,1392	0,0000	0,1392	0,6645	2,1415	0,0000	0,1415	0,2594	2,1067	0,4853	0,1067	0,0861
	L_2	2,0572	0,1545	0,0572	0,0469	2,0926	0,2298	0,0926	0,0213	2,0946	0,2673	0,0946	0,0168
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\beta}_1$	KS	viés	EQM	$\hat{\beta}_1$	KS	viés	EQM	$\hat{\beta}_1$	KS	viés	EQM
1	N	6,9983	0,7931	-0,0017	0,0225	6,9990	0,9973	-0,0010	0,0064	7,0001	0,7079	0,0001	0,0040
	t_3	6,9955	0,9447	-0,0045	0,0280	6,9992	0,5748	-0,0008	0,0073	6,9992	0,8737	-0,0008	0,0046
	PE_1	7,0009	0,6008	0,0009	0,0246	6,9991	0,9976	-0,0009	0,0074	7,0003	0,4103	0,0003	0,0046
	PE_2	6,9952	0,9978	-0,0048	0,0264	6,9991	0,8504	-0,0009	0,0071	6,9988	0,9791	-0,0012	0,0045
	NG_3	7,0038	0,8596	0,0038	0,0819	7,0446	0,6617	0,0446	0,0360	7,0388	0,5188	0,0388	0,0269
	L_2	6,9967	0,9498	-0,0033	0,0241	6,9990	0,8984	-0,0010	0,0066	6,9997	0,8118	-0,0003	0,0042
3	N	7,0018	0,7676	0,0018	0,0663	6,9994	0,9991	-0,0006	0,0192	7,0002	0,6091	0,0002	0,0120
	t_3	6,9970	0,7206	-0,0030	0,0818	6,9999	0,7924	-0,0001	0,0222	6,9988	0,9272	-0,0012	0,0138
	PE_1	7,0059	0,6562	0,0059	0,0725	6,9997	0,9875	-0,0003	0,0221	7,0005	0,5759	0,0005	0,0137
	PE_2	6,9967	0,9797	-0,0033	0,0774	7,0002	0,9348	0,0002	0,0214	6,9984	0,8830	-0,0016	0,0137
	NG_3	7,0152	0,6091	0,0152	0,1969	7,0850	0,3785	0,0850	0,0911	7,0816	0,3993	0,0816	0,0648
	L_2	6,9992	0,7689	-0,0008	0,0711	6,9995	0,9348	-0,0005	0,0200	6,9997	0,8458	-0,0003	0,0125
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\phi}$	KS	viés	EQM	$\hat{\phi}$	KS	viés	EQM	$\hat{\phi}$	KS	viés	EQM
1	N	0,8393	0,0578	-0,1607	0,0439	0,9408	0,5884	-0,0592	0,0097	0,9636	0,9956	-0,0364	0,0053
	t_3	0,4857	0,2236	-0,5143	0,2732	0,5774	0,7885	-0,4226	0,1818	0,5962	0,7401	-0,4038	0,1650
	PE_1	1,3808	0,1073	0,3808	0,1957	1,5692	0,1456	0,5692	0,3421	1,6143	0,9824	0,6143	0,3893
	PE_2	0,3375	0,0599	-0,6625	0,4421	0,3846	0,6758	-0,6154	0,3798	0,3949	0,6585	-0,6051	0,3669
	NG_3	2,7178	0,0005	1,7178	3,2823	2,7647	0,2608	1,7647	3,1838	2,7957	0,5604	1,7957	3,2646
	L_2	0,2691	0,1138	-0,7309	0,5363	0,3058	0,4747	-0,6942	0,4826	0,3137	0,9472	-0,6863	0,4715
3	N	2,5971	0,0873	-0,4029	0,3329	2,8596	0,4719	-0,1404	0,0746	2,9108	0,9570	-0,0892	0,0438
	t_3	1,5330	0,2125	-1,4670	2,2334	1,7646	0,3837	-1,2354	1,5543	1,8065	0,8278	-1,1935	1,4419
	PE_1	4,2860	0,1561	1,2860	2,1377	4,7744	0,1675	1,7744	3,3095	4,8799	0,6946	1,8799	3,6407
	PE_2	1,0497	0,0984	-1,9503	3,8333	1,1712	0,4094	-1,8288	3,3543	1,1939	0,9679	-1,8061	3,2682
	NG_3	7,9744	0,0110	4,9744	26,9817	8,3170	0,3147	5,3170	28,8524	8,3901	0,7643	5,3901	29,3921
	L_2	0,8367	0,1061	-2,1633	4,6992	0,9310	0,4483	-2,0690	4,2871	0,9483	0,9106	-2,0517	4,2134
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\rho}_1$	KS	viés	EQM	$\hat{\rho}_1$	KS	viés	EQM	$\hat{\rho}_1$	KS	viés	EQM
1	N	0,6147	0,7545	-0,0413	0,0119	0,6235	0,8628	-0,0325	0,0043	0,6342	0,6000	-0,0218	0,0023
	t_3	0,6123	0,9419	-0,0437	0,0146	0,6216	0,9852	-0,0344	0,0051	0,6331	0,9959	-0,0229	0,0027
	PE_1	0,6128	0,8683	-0,0432	0,0129	0,6233	0,7726	-0,0327	0,0046	0,6341	0,8515	-0,0219	0,0025
	PE_2	0,6158	0,6752	-0,0402	0,0134	0,6221	0,9583	-0,0339	0,0050	0,6332	0,7850	-0,0228	0,0026
	NG_3	0,6748	0,0491	0,0188	0,0172	0,6383	0,4355	-0,0177	0,0049	0,6397	0,4017	-0,0163	0,0026
	L_2	0,6142	0,7868	-0,0418	0,0128	0,6227	0,7966	-0,0333	0,0046	0,6338	0,8336	-0,0222	0,0025
3	N	0,6360	0,8170	-0,0200	0,0092	0,6327	0,9688	-0,0233	0,0037	0,6391	0,6798	-0,0169	0,0021
	t_3	0,6331	0,4879	-0,0229	0,0117	0,6312	0,9976	-0,0248	0,0045	0,6379	0,9898	-0,0181	0,0025
	PE_1	0,6341	0,9842	-0,0219	0,0101	0,6325	0,6383	-0,0235	0,0040	0,6392	0,6995	-0,0168	0,0023
	PE_2	0,6360	0,8272	-0,0200	0,0107	0,6314	0,8974	-0,0246	0,0044	0,6380	0,8044	-0,0180	0,0024
	NG_3	0,6719	0,0940	0,0159	0,0133	0,6418	0,9965	-0,0142	0,0044	0,6436	0,3065	-0,0124	0,0024
	L_2	0,6353	0,4342	-0,0207	0,0100	0,6321	0,9796	-0,0239	0,0040	0,6386	0,8707	-0,0174	0,0023
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\rho}_2$	KS	viés	EQM	$\hat{\rho}_2$	KS	viés	EQM	$\hat{\rho}_2$	KS	viés	EQM
1	N	-0,4212	0,0007	-0,1122	0,0219	-0,3561	0,8856	-0,0471	0,0048	-0,3375	0,7320	-0,0285	0,0023
	t_3	-0,4199	0,0126	-0,1109	0,0237	-0,3561	0,8659	-0,0471	0,0054	-0,3387	0,3578	-0,0297	0,0026
	PE_1	-0,4212	0,0052	-0,1122	0,0232	-0,3573	0,8581	-0,0483	0,0054	-0,3376	0,8534	-0,0286	0,0026
	PE_2	-0,4206	0,0120	-0,1116	0,0232	-0,3556	0,9174	-0,0466	0,0052	-0,3382	0,4939	-0,0292	0,0026
	NG_3	-0,3430	0,0001	-0,0340	0,0196	-0,3317	0,1824	-0,0227	0,0050	-0,3218	0,7696	-0,0128	0,0023
	L_2	-0,4207	0,0032	-0,1117	0,0223	-0,3558	0,7103	-0,0468	0,0050	-0,3380	0,5616	-0,0290	0,0024
3	N	-0,4175	0,0004	-0,1085	0,0207	-0,3511	0,9250	-0,0421	0,0044	-0,3335	0,7570	-0,0245	0,0021
	t_3	-0,4171	0,0128	-0,1081	0,0224	-0,3514	0,7803	-0,0424	0,0050	-0,3346	0,5059	-0,0256	0,0024
	PE_1	-0,4177	0,0042	-0,1087	0,0218	-0,3522	0,6998	-0,0432	0,0049	-0,3337	0,7496	-0,0247	0,0024
	PE_2	-0,4179	0,0160	-0,1089	0,0222	-0,3510	0,7577	-0,0420	0,0049	-0,3342	0,4086	-0,0252	0,0023
	NG_3	-0,3671	0,0033	-0,0581	0,0177	-0,3330	0,1777	-0,0240	0,0047	-0,3221	0,6699	-0,0131	0,0022
	L_2	-0,4173	0,0010	-0,1083	0,0211	-0,3510	0,8725	-0,0420	0,0046	-0,3339	0,4785	-0,0249	0,0021

Tabela D.9: Média das estimativas, valor-P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro t-Student ($\nu = 3$) e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 , PE_2 , NG e L_2 .

ϕ	Erro	n=100				n=300				n=500			
		$\hat{\beta}_0$	KS	viés	EQM	$\hat{\beta}_0$	KS	viés	EQM	$\hat{\beta}_0$	KS	viés	EQM
1	N	2,0641	0,3759	0,0641	0,0378	2,0889	0,0921	0,0889	0,0173	2,0942	0,0323	0,0942	0,0146
	t_3	2,0645	0,9527	0,0645	0,0288	2,0927	0,6995	0,0927	0,0167	2,0914	0,9518	0,0914	0,0132
	PE_1	2,0647	0,1661	0,0647	0,0983	2,0854	0,9265	0,0854	0,0363	2,1001	0,3016	0,1001	0,0296
	PE_2	2,0621	0,7995	0,0621	0,0284	2,0916	0,4818	0,0916	0,0160	2,0923	0,8869	0,0923	0,0132
	NG_3	2,1149	0,0000	0,1149	0,4672	2,1033	0,2367	0,1033	0,1133	2,1164	0,8554	0,1164	0,0721
	L_2	2,0639	0,5351	0,0639	0,0270	2,0912	0,2875	0,0912	0,0157	2,0920	0,5161	0,0920	0,0128
3	N	2,0688	0,3160	0,0688	0,1025	2,0900	0,0923	0,0900	0,0362	2,0945	0,0220	0,0945	0,0259
	t_3	2,0650	0,7797	0,0650	0,0783	2,0938	0,7468	0,0938	0,0333	2,0901	0,6339	0,0901	0,0226
	PE_1	2,0711	0,0038	0,0711	0,3737	2,0856	0,8757	0,0856	0,0909	2,1038	0,4375	0,1038	0,0707
	PE_2	2,0628	0,7730	0,0628	0,0779	2,0929	0,5025	0,0929	0,0317	2,0918	0,7053	0,0918	0,0223
	NG_3	2,0438	0,0000	0,0438	0,9075	2,0880	0,6662	0,0880	0,2049	2,1140	0,7599	0,1140	0,1457
	L_2	2,0649	0,4103	0,0649	0,0733	2,0922	0,5011	0,0922	0,0311	2,0910	0,5709	0,0910	0,0213
ϕ	Erro	n=100				n=300				n=500			
		$\hat{\beta}_1$	KS	viés	EQM	$\hat{\beta}_1$	KS	viés	EQM	$\hat{\beta}_1$	KS	viés	EQM
1	N	6,9923	0,6418	-0,0077	0,0611	6,9940	0,9508	-0,0060	0,0166	6,9961	0,8878	-0,0039	0,0108
	t_3	6,9923	0,9922	-0,0077	0,0402	6,9971	0,5148	-0,0029	0,0099	6,9963	0,7119	-0,0037	0,0062
	PE_1	6,9950	0,0425	-0,0050	0,1426	6,9907	0,8307	-0,0093	0,0516	6,9978	0,8979	-0,0022	0,0331
	PE_2	6,9942	0,9898	-0,0058	0,0426	6,9961	0,9990	-0,0039	0,0105	6,9959	0,2281	-0,0041	0,0068
	NG_3	7,0124	0,1359	0,0124	0,2181	7,0191	0,8044	0,0191	0,0917	7,0075	0,7012	0,0075	0,0548
	L_2	6,9914	0,9036	-0,0086	0,0416	6,9960	0,7729	-0,0040	0,0105	6,9958	0,3508	-0,0042	0,0068
3	N	6,9908	0,6110	-0,0092	0,1754	6,9950	0,9503	-0,0050	0,0481	6,9922	0,7629	-0,0078	0,0327
	t_3	6,9923	0,9486	-0,0077	0,1161	6,9966	0,5840	-0,0034	0,0288	6,9926	0,8217	-0,0074	0,0186
	PE_1	6,9962	0,0175	-0,0038	0,4268	6,9994	0,9456	-0,0006	0,1452	6,9962	0,8452	-0,0038	0,0994
	PE_2	6,9958	0,7359	-0,0042	0,1223	6,9962	0,9726	-0,0038	0,0307	6,9916	0,5353	-0,0084	0,0205
	NG_3	7,0179	0,4050	0,0179	0,5046	7,0334	0,3726	0,0334	0,2249	7,0135	0,5897	0,0135	0,1517
	L_2	6,9899	0,8113	-0,0101	0,1201	6,9953	0,5924	-0,0047	0,0309	6,9917	0,7443	-0,0083	0,0206
ϕ	Erro	n=100				n=300				n=500			
		$\hat{\phi}$	KS	viés	EQM	$\hat{\phi}$	KS	viés	EQM	$\hat{\phi}$	KS	viés	EQM
1	N	2,2883	0,0000	1,2883	2,4144	2,3503	0,0223	1,3503	1,9801	2,3238	0,0645	1,3238	1,8335
	t_3	0,8305	0,0018	-0,1695	0,0678	0,9272	0,4229	-0,0728	0,0167	0,9393	0,4477	-0,0607	0,0108
	PE_1	4,5720	0,0000	3,5720	18,2132	4,8745	0,0015	3,8745	16,1982	4,8180	0,0047	3,8180	15,1408
	PE_2	0,7560	0,0000	-0,2440	0,1038	0,7935	0,1408	-0,2065	0,0535	0,7925	0,3936	-0,2075	0,0492
	NG_3	8,5791	0,0000	7,5791	77,6368	8,6229	0,0014	7,6229	62,0998	8,4523	0,0052	7,4523	57,3546
	L_2	0,5604	0,0015	-0,4396	0,2133	0,5918	0,2847	-0,4082	0,1719	0,5928	0,7022	-0,4072	0,1689
3	N	6,9023	0,0000	3,9023	20,8180	6,9971	0,0503	3,9971	17,2651	6,9613	0,1516	3,9613	16,3637
	t_3	2,6074	0,0014	-0,3926	0,5183	2,8181	0,4287	-0,1819	0,1406	2,8400	0,3414	-0,1600	0,0876
	PE_1	13,7945	0,0000	10,7945	155,2943	14,4260	0,0135	11,4260	139,5424	14,4432	0,0127	11,4432	135,5802
	PE_2	2,3133	0,0001	-0,6867	0,8405	2,3849	0,1750	-0,6151	0,4748	2,3822	0,6610	-0,6178	0,4344
	NG_3	25,0136	0,0000	22,0136	628,7866	25,3028	0,0295	22,3028	526,2669	25,2365	0,0112	22,2365	509,3710
	L_2	1,7221	0,0022	-1,2779	1,8086	1,7841	0,2834	-1,2159	1,5263	1,7840	0,5471	-1,2160	1,5054
ϕ	Erro	n=100				n=300				n=500			
		$\hat{\rho}_1$	KS	viés	EQM	$\hat{\rho}_1$	KS	viés	EQM	$\hat{\rho}_1$	KS	viés	EQM
1	N	0,6269	0,1588	-0,0291	0,0099	0,6310	0,7187	-0,0250	0,0037	0,6363	0,8974	-0,0197	0,0021
	t_3	0,6372	0,0622	-0,0188	0,0067	0,6394	0,7199	-0,0166	0,0023	0,6425	0,1519	-0,0135	0,0014
	PE_1	0,6055	0,8859	-0,0505	0,0197	0,6138	0,7943	-0,0422	0,0096	0,6244	0,9980	-0,0316	0,0058
	PE_2	0,6356	0,4753	-0,0204	0,0072	0,6373	0,9047	-0,0187	0,0025	0,6412	0,1381	-0,0148	0,0015
	NG_3	0,6432	0,9888	-0,0128	0,0215	0,6208	0,7442	-0,0352	0,0104	0,6277	0,9905	-0,0283	0,0065
	L_2	0,6347	0,2168	-0,0213	0,0071	0,6372	0,8008	-0,0188	0,0025	0,6406	0,6582	-0,0154	0,0015
3	N	0,6458	0,2595	-0,0102	0,0083	0,6369	0,8240	-0,0191	0,0035	0,6403	0,8332	-0,0157	0,0019
	t_3	0,6482	0,1191	-0,0078	0,0058	0,6432	0,8497	-0,0128	0,0022	0,6449	0,2217	-0,0111	0,0013
	PE_1	0,6332	0,7661	-0,0228	0,0172	0,6236	0,7989	-0,0324	0,0088	0,6310	0,9870	-0,0250	0,0054
	PE_2	0,6489	0,3356	-0,0071	0,0062	0,6414	0,9784	-0,0146	0,0024	0,6439	0,2529	-0,0121	0,0014
	NG_3	0,6501	0,8656	-0,0059	0,0201	0,6256	0,7801	-0,0304	0,0097	0,6319	0,9931	-0,0241	0,0061
	L_2	0,6483	0,1870	-0,0077	0,0061	0,6415	0,9684	-0,0145	0,0024	0,6435	0,7916	-0,0125	0,0013
ϕ	Erro	n=100				n=300				n=500			
		$\hat{\rho}_2$	KS	viés	EQM	$\hat{\rho}_2$	KS	viés	EQM	$\hat{\rho}_2$	KS	viés	EQM
1	N	-0,4182	0,0023	-0,1092	0,0208	-0,3552	0,9214	-0,0462	0,0046	-0,3337	0,7351	-0,0247	0,0021
	t_3	-0,3819	0,4729	-0,0729	0,0124	-0,3396	0,1149	-0,0306	0,0027	-0,3254	0,5569	-0,0164	0,0014
	PE_1	-0,4552	0,0027	-0,1462	0,0376	-0,3799	0,8559	-0,0709	0,0114	-0,3483	0,9884	-0,0393	0,0058
	PE_2	-0,3917	0,2226	-0,0827	0,0142	-0,3429	0,1507	-0,0339	0,0030	-0,3272	0,6567	-0,0182	0,0015
	NG_3	-0,3983	0,0023	-0,0893	0,0313	-0,3645	0,2014	-0,0555	0,0113	-0,3397	0,7426	-0,0307	0,0061
	L_2	-0,3945	0,0892	-0,0855	0,0143	-0,3445	0,4062	-0,0355	0,0031	-0,3279	0,3822	-0,0189	0,0015
3	N	-0,4124	0,0007	-0,1034	0,0193	-0,3497	0,9495	-0,0407	0,0043	-0,3305	0,5096	-0,0215	0,0019
	t_3	-0,3795	0,5955	-0,0705	0,0117	-0,3370	0,4281	-0,0280	0,0026	-0,3236	0,2764	-0,0146	0,0013
	PE_1	-0,4488	0,0139	-0,1398	0,0361	-0,3699	0,9927	-0,0609	0,0105	-0,3429	0,7797	-0,0339	0,0053
	PE_2	-0,3884	0,0838	-0,0794	0,0131	-0,3398	0,3432	-0,0308	0,0028	-0,3252	0,5279	-0,0162	0,0014
	NG_3	-0,4173	0,0138	-0,1083	0,0335	-0,3617	0,8529	-0,0527	0,0109	-0,3382	0,4629	-0,0292	0,0058
	L_2	-0,3907	0,0383	-0,0817	0,0133	-0,3410	0,5244	-0,0320	0,0029	-0,3257	0,4588	-0,0167	0,0013

Tabela D.10: Média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro exponencial potência ($\nu = -0,3$) - PE_1 e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 , PE_2 , NG e L_2 .

		n=100				n=300				n=500			
ϕ	Erro	$\hat{\beta}_0$	KS	viés	EQM	$\hat{\beta}_0$	KS	viés	EQM	$\hat{\beta}_0$	KS	viés	EQM
1	N	2,0624	0,0354	0,0624	0,0120	2,0887	0,1744	0,0887	0,0104	2,0921	0,0811	0,0921	0,0100
	t_3	2,0626	0,6111	0,0626	0,0150	2,0891	0,3345	0,0891	0,0116	2,0915	0,3340	0,0915	0,0105
	PE_1	2,0633	0,1827	0,0633	0,0120	2,0887	0,1999	0,0887	0,0103	2,0924	0,0949	0,0924	0,0101
	PE_2	2,0623	0,4524	0,0623	0,0143	2,0891	0,2864	0,0891	0,0115	2,0914	0,5118	0,0914	0,0104
	NG_3	2,1318	0,0000	0,1318	0,2452	2,1066	0,0011	0,1066	0,0741	2,1149	0,2572	0,1149	0,0392
	L_2	2,0622	0,0496	0,0622	0,0131	2,0889	0,2120	0,0889	0,0109	2,0917	0,2858	0,0917	0,0101
3	N	2,0634	0,0202	0,0634	0,0280	2,0886	0,1866	0,0886	0,0153	2,0908	0,0901	0,0908	0,0127
	t_3	2,0621	0,2380	0,0621	0,0373	2,0892	0,5380	0,0892	0,0191	2,0897	0,4481	0,0897	0,0144
	PE_1	2,0653	0,1414	0,0653	0,0281	2,0886	0,1968	0,0886	0,0151	2,0916	0,0720	0,0916	0,0129
	PE_2	2,0611	0,3039	0,0611	0,0354	2,0887	0,2605	0,0887	0,0187	2,0894	0,4706	0,0894	0,0141
	NG_3	2,1508	0,0000	0,1508	0,5853	2,1147	0,0004	0,1147	0,1325	2,0961	0,5144	0,0961	0,0674
	L_2	2,0625	0,0288	0,0625	0,0315	2,0888	0,2010	0,0888	0,0168	2,0902	0,2401	0,0902	0,0133
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\beta}_1$	KS	viés	EQM	$\hat{\beta}_1$	KS	viés	EQM	$\hat{\beta}_1$	KS	viés	EQM
1	N	6,9886	0,9244	-0,0114	0,0149	6,9958	0,7012	-0,0042	0,0042	6,9992	0,9909	-0,0008	0,0027
	t_3	6,9903	0,8651	-0,0097	0,0188	6,9951	0,8480	-0,0049	0,0058	6,9984	0,9867	-0,0016	0,0037
	PE_1	6,9866	0,7358	-0,0134	0,0149	6,9961	0,6852	-0,0039	0,0038	6,9997	0,8870	-0,0003	0,0025
	PE_2	6,9911	0,7607	-0,0089	0,0175	6,9951	0,7440	-0,0049	0,0058	6,9985	0,9399	-0,0015	0,0036
	NG_3	6,9662	0,6513	-0,0338	0,0670	7,0433	0,8162	0,0433	0,0249	7,0094	0,5444	0,0094	0,0193
	L_2	6,9897	0,9396	-0,0103	0,0163	6,9955	0,6086	-0,0045	0,0049	6,9988	0,9821	-0,0012	0,0031
3	N	6,9831	0,9708	-0,0169	0,0442	6,9925	0,6953	-0,0075	0,0124	6,9988	0,9924	-0,0012	0,0081
	t_3	6,9851	0,7379	-0,0149	0,0565	6,9915	0,9723	-0,0085	0,0174	6,9975	0,9746	-0,0025	0,0110
	PE_1	6,9799	0,5615	-0,0201	0,0438	6,9927	0,8580	-0,0073	0,0113	6,9996	0,9298	-0,0004	0,0075
	PE_2	6,9887	0,9888	-0,0113	0,0533	6,9918	0,8595	-0,0082	0,0173	6,9977	0,7496	-0,0023	0,0109
	NG_3	6,9721	0,3899	-0,0279	0,1628	7,0752	0,6868	0,0752	0,0620	7,0487	0,9597	0,0487	0,0495
	L_2	6,9845	0,9896	-0,0155	0,0490	6,9921	0,5764	-0,0079	0,0146	6,9982	0,9735	-0,0018	0,0093
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\phi}$	KS	viés	EQM	$\hat{\phi}$	KS	viés	EQM	$\hat{\phi}$	KS	viés	EQM
1	N	0,5113	0,3433	-0,4887	0,2440	0,5782	0,6858	-0,4218	0,1796	0,5938	0,9523	-0,4062	0,1659
	t_3	0,3157	0,1116	-0,6843	0,4712	0,3843	0,7238	-0,6157	0,3802	0,4001	0,8183	-0,5999	0,3606
	PE_1	0,8168	0,2846	-0,1832	0,0465	0,9290	0,7319	-0,0710	0,0093	0,9548	0,9310	-0,0452	0,0044
	PE_2	0,2112	0,1011	-0,7888	0,6232	0,2449	0,3357	-0,7551	0,5705	0,2529	0,9141	-0,7471	0,5584
	NG_3	1,6769	0,0001	0,6769	0,5870	1,6616	0,0358	0,6616	0,4608	1,6654	0,1130	0,6654	0,4525
	L_2	0,1698	0,1693	-0,8302	0,6898	0,1962	0,5325	-0,8038	0,6464	0,2023	0,9522	-0,7977	0,6364
3	N	1,5909	0,3235	-1,4091	2,0364	1,7566	0,6057	-1,2434	1,5617	1,7954	0,8775	-1,2046	1,4601
	t_3	1,0068	0,2996	-1,9932	4,0025	1,1753	0,7644	-1,8247	3,3401	1,2144	0,5536	-1,7856	3,1944
	PE_1	2,5458	0,5206	-0,4542	0,3334	2,8223	0,9205	-0,1777	0,0696	2,8869	0,7413	-0,1131	0,0347
	PE_2	0,6623	0,0444	-2,3377	5,4748	0,7459	0,5347	-2,2541	5,0842	0,7657	0,7144	-2,2343	4,9941
	NG_3	4,9521	0,0003	1,9521	4,9893	4,9606	0,4004	1,9606	4,0058	4,9842	0,9006	1,9842	4,0176
	L_2	0,5320	0,1570	-2,4680	6,0977	0,5972	0,6205	-2,4028	5,7755	0,6125	0,8913	-2,3875	5,7013
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\rho}_1$	KS	viés	EQM	$\hat{\rho}_1$	KS	viés	EQM	$\hat{\rho}_1$	KS	viés	EQM
1	N	0,6113	0,7458	-0,0447	0,0120	0,6179	0,6234	-0,0381	0,0048	0,6311	0,9865	-0,0249	0,0026
	t_3	0,6063	0,7915	-0,0497	0,0169	0,6117	0,9869	-0,0443	0,0066	0,6276	0,9185	-0,0284	0,0035
	PE_1	0,6123	0,5832	-0,0437	0,0112	0,6211	0,9614	-0,0349	0,0044	0,6330	0,9935	-0,0230	0,0024
	PE_2	0,6101	0,4781	-0,0459	0,0156	0,6127	0,9266	-0,0433	0,0063	0,6280	0,7745	-0,0280	0,0034
	NG_3	0,6831	0,5638	0,0271	0,0168	0,6415	0,1588	-0,0145	0,0050	0,6435	0,5063	-0,0125	0,0025
	L_2	0,6093	0,6650	-0,0467	0,0139	0,6149	0,8074	-0,0411	0,0056	0,6293	0,8935	-0,0267	0,0030
3	N	0,6286	0,1453	-0,0274	0,0102	0,6279	0,6327	-0,0281	0,0041	0,6366	0,9662	-0,0194	0,0023
	t_3	0,6242	0,2143	-0,0318	0,0145	0,6227	0,9953	-0,0333	0,0056	0,6336	0,8542	-0,0224	0,0031
	PE_1	0,6287	0,4428	-0,0273	0,0096	0,6304	0,9924	-0,0256	0,0037	0,6381	0,9768	-0,0179	0,0022
	PE_2	0,6272	0,1141	-0,0288	0,0136	0,6234	0,9788	-0,0326	0,0054	0,6339	0,9134	-0,0221	0,0031
	NG_3	0,6789	0,9634	0,0229	0,0139	0,6436	0,4547	-0,0124	0,0040	0,6450	0,8413	-0,0110	0,0022
	L_2	0,6270	0,3311	-0,0290	0,0120	0,6253	0,8929	-0,0307	0,0047	0,6351	0,9413	-0,0209	0,0027
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\rho}_2$	KS	viés	EQM	$\hat{\rho}_2$	KS	viés	EQM	$\hat{\rho}_2$	KS	viés	EQM
1	N	-0,4174	0,0032	-0,1084	0,0220	-0,3561	0,3055	-0,0471	0,0049	-0,3390	0,5786	-0,0300	0,0027
	t_3	-0,4253	0,0447	-0,1163	0,0280	-0,3612	0,8763	-0,0522	0,0067	-0,3432	0,9273	-0,0342	0,0036
	PE_1	-0,4113	0,0033	-0,1023	0,0202	-0,3529	0,1818	-0,0439	0,0044	-0,3365	0,7497	-0,0275	0,0024
	PE_2	-0,4235	0,0871	-0,1145	0,0266	-0,3600	0,4600	-0,0510	0,0064	-0,3422	0,8966	-0,0332	0,0035
	NG_3	-0,3177	0,0035	-0,0087	0,0165	-0,3185	0,0001	-0,0095	0,0041	-0,3158	0,7269	-0,0068	0,0020
	L_2	-0,4220	0,0176	-0,1130	0,0246	-0,3588	0,6539	-0,0498	0,0057	-0,3412	0,8634	-0,0322	0,0031
3	N	-0,4171	0,0156	-0,1081	0,0215	-0,3518	0,4469	-0,0428	0,0045	-0,3351	0,7798	-0,0261	0,0024
	t_3	-0,4263	0,0744	-0,1173	0,0274	-0,3567	0,8754	-0,0477	0,0063	-0,3390	0,9699	-0,0300	0,0034
	PE_1	-0,4110	0,0011	-0,1020	0,0197	-0,3489	0,1493	-0,0399	0,0040	-0,3330	0,9142	-0,0240	0,0022
	PE_2	-0,4248	0,0138	-0,1158	0,0264	-0,3559	0,6935	-0,0469	0,0060	-0,3380	0,8804	-0,0290	0,0033
	NG_3	-0,3416	0,0002	-0,0326	0,0162	-0,3228	0,0009	-0,0138	0,0039	-0,3172	0,8059	-0,0082	0,0020
	L_2	-0,4220	0,0352	-0,1130	0,0241	-0,3544	0,5767	-0,0454	0,0053	-0,3371	0,9040	-0,0281	0,0029

Tabela D.11: Média das estimativas, valor-P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro exponencial potência ($\nu = 0,5$) - PE_2 e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 , PE_2 , NG e L_2 .

		n=100				n=300				n=500			
ϕ	Erro	$\hat{\beta}_0$	KS	viés	EQM	$\hat{\beta}_0$	KS	viés	EQM	$\hat{\beta}_0$	KS	viés	EQM
1	N	2,0606	0,0649	0,0606	0,0371	2,0856	0,0403	0,0856	0,0180	2,0896	0,0267	0,0896	0,0145
	t_3	2,0616	0,2806	0,0616	0,0381	2,0848	0,8064	0,0848	0,0188	2,0891	0,5099	0,0891	0,0142
	PE_1	2,0601	0,3692	0,0601	0,0535	2,0873	0,4082	0,0873	0,0251	2,0902	0,1578	0,0902	0,0199
	PE_2	2,0603	0,1126	0,0603	0,0358	2,0856	0,2299	0,0856	0,0179	2,0891	0,4636	0,0891	0,0138
	NG_3	2,1109	0,0000	0,1109	0,7923	2,1126	0,0113	0,1126	0,1274	2,1108	0,3309	0,1108	0,0742
	L_2	2,0611	0,0654	0,0611	0,0348	2,0848	0,4611	0,0848	0,0175	2,0892	0,2074	0,0892	0,0138
3	N	2,0565	0,0664	0,0565	0,1033	2,0833	0,0367	0,0833	0,0391	2,0868	0,0261	0,0868	0,0269
	t_3	2,0575	0,2559	0,0575	0,1049	2,0823	0,8263	0,0823	0,0413	2,0864	0,3645	0,0864	0,0265
	PE_1	2,0553	0,6654	0,0553	0,1614	2,0863	0,3780	0,0863	0,0619	2,0875	0,1871	0,0875	0,0435
	PE_2	2,0540	0,1648	0,0540	0,0987	2,0832	0,2887	0,0832	0,0386	2,0864	0,5330	0,0864	0,0251
	NG_3	2,1129	0,0000	0,1129	0,9907	2,1084	0,2082	0,1084	0,2028	2,1051	0,4180	0,1051	0,1483
	L_2	2,0571	0,1050	0,0571	0,0953	2,0821	0,4817	0,0821	0,0375	2,0864	0,1834	0,0864	0,0249
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\beta}_1$	KS	viés	EQM	$\hat{\beta}_1$	KS	viés	EQM	$\hat{\beta}_1$	KS	viés	EQM
1	N	6,9824	0,9897	-0,0176	0,0663	6,9896	0,9624	-0,0104	0,0172	6,9979	0,8847	-0,0021	0,0114
	t_3	6,9907	0,9272	-0,0093	0,0585	6,9907	0,7175	-0,0093	0,0157	6,9969	0,9180	-0,0031	0,0095
	PE_1	6,9767	0,8446	-0,0233	0,0939	6,9878	0,9990	-0,0122	0,0258	6,9993	0,7959	-0,0007	0,0182
	PE_2	6,9894	0,9996	-0,0106	0,0577	6,9904	0,4453	-0,0096	0,0157	6,9966	0,8151	-0,0034	0,0095
	NG_3	6,9747	0,9932	-0,0253	0,1993	7,0407	0,9994	0,0407	0,0731	7,0153	0,2822	0,0153	0,0485
	L_2	6,9866	0,9783	-0,0134	0,0581	6,9905	0,9056	-0,0095	0,0154	6,9973	0,9753	-0,0027	0,0097
3	N	6,9771	0,9640	-0,0229	0,1966	6,9859	0,9871	-0,0141	0,0513	6,9966	0,8646	-0,0034	0,0340
	t_3	6,9889	0,9763	-0,0111	0,1732	6,9874	0,8110	-0,0126	0,0469	6,9946	0,9689	-0,0054	0,0282
	PE_1	6,9673	0,7861	-0,0327	0,2840	6,9827	1,0000	-0,0173	0,0781	6,9992	0,8476	-0,0008	0,0549
	PE_2	6,9906	0,9952	-0,0094	0,1724	6,9877	0,5120	-0,0123	0,0470	6,9941	0,9176	-0,0059	0,0284
	NG_3	6,9972	0,9941	-0,0028	0,4260	7,0461	0,9700	0,0461	0,1695	7,0414	0,1051	0,0414	0,1189
	L_2	6,9831	0,8902	-0,0169	0,1722	6,9873	0,8361	-0,0127	0,0460	6,9955	0,9429	-0,0045	0,0290
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\phi}$	KS	viés	EQM	$\hat{\phi}$	KS	viés	EQM	$\hat{\phi}$	KS	viés	EQM
1	N	2,2515	0,1504	1,2515	1,7602	2,4606	0,5156	1,4606	2,1959	2,5235	0,6963	1,5235	2,3593
	t_3	1,1451	0,0007	0,1451	0,0770	1,2896	0,2808	0,2896	0,1042	1,3307	0,6742	0,3307	0,1211
	PE_1	3,9119	0,0798	2,9119	9,1699	4,3995	0,4274	3,3995	11,7942	4,5418	0,6193	3,5418	12,6972
	PE_2	0,8589	0,0324	-0,1411	0,0465	0,9408	0,4170	-0,0592	0,0124	0,9655	0,9695	-0,0345	0,0064
	NG_3	7,2611	0,0454	6,2611	41,8248	7,6780	0,1328	6,6780	45,3682	7,8410	0,3816	6,8410	47,2764
	L_2	0,6736	0,0319	-0,3264	0,1231	0,7361	0,4958	-0,2639	0,0752	0,7551	0,9592	-0,2449	0,0632
3	N	6,9795	0,1039	3,9795	17,6508	7,4569	0,4193	4,4569	20,4277	7,6178	0,7387	4,6178	21,6696
	t_3	3,5929	0,0038	0,5929	0,8807	3,9190	0,2090	0,9190	1,0292	4,0245	0,7107	1,0245	1,1566
	PE_1	12,2141	0,0158	9,2141	91,4826	13,3774	0,5770	10,3774	109,8604	13,7423	0,6262	10,7423	116,7907
	PE_2	2,6678	0,0132	-0,3322	0,3611	2,8526	0,6838	-0,1474	0,1022	2,9160	0,9956	-0,0840	0,0547
	NG_3	21,7069	0,0362	18,7069	372,0203	23,1026	0,1951	20,1026	410,8432	23,6141	0,4547	20,6141	429,2570
	L_2	2,0907	0,0144	-0,9093	0,9823	2,2313	0,5589	-0,7687	0,6412	2,2803	0,9358	-0,7197	0,5474
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\rho}_1$	KS	viés	EQM	$\hat{\rho}_1$	KS	viés	EQM	$\hat{\rho}_1$	KS	viés	EQM
1	N	0,6297	0,5375	-0,0263	0,0101	0,6324	0,7882	-0,0236	0,0041	0,6389	0,9371	-0,0171	0,0023
	t_3	0,6313	0,5189	-0,0247	0,0101	0,6354	0,8093	-0,0206	0,0034	0,6413	0,9101	-0,0147	0,0020
	PE_1	0,6242	0,5577	-0,0318	0,0139	0,6274	0,6575	-0,0286	0,0062	0,6350	0,9967	-0,0210	0,0037
	PE_2	0,6319	0,2903	-0,0241	0,0099	0,6354	0,8992	-0,0206	0,0034	0,6411	0,9587	-0,0149	0,0020
	NG_3	0,6630	0,8644	0,0070	0,0167	0,6359	0,8153	-0,0201	0,0065	0,6392	0,9868	-0,0168	0,0040
	L_2	0,6315	0,5594	-0,0245	0,0095	0,6345	0,6527	-0,0215	0,0035	0,6405	0,8433	-0,0155	0,0020
3	N	0,6484	0,5256	-0,0076	0,0089	0,6399	0,8886	-0,0161	0,0037	0,6436	0,9626	-0,0124	0,0022
	t_3	0,6468	0,5295	-0,0092	0,0089	0,6419	0,6327	-0,0141	0,0031	0,6453	0,9329	-0,0107	0,0018
	PE_1	0,6457	0,6027	-0,0103	0,0124	0,6359	0,8131	-0,0201	0,0057	0,6405	0,9816	-0,0155	0,0035
	PE_2	0,6475	0,2277	-0,0085	0,0087	0,6420	0,9253	-0,0140	0,0031	0,6452	0,9796	-0,0108	0,0018
	NG_3	0,6642	0,9919	0,0082	0,0141	0,6397	0,8207	-0,0163	0,0060	0,6429	0,9940	-0,0131	0,0038
	L_2	0,6483	0,3408	-0,0077	0,0084	0,6413	0,4576	-0,0147	0,0032	0,6448	0,9280	-0,0112	0,0019
		n=100				n=300				n=500			
ϕ	Erro	$\hat{\rho}_2$	KS	viés	EQM	$\hat{\rho}_2$	KS	viés	EQM	$\hat{\rho}_2$	KS	viés	EQM
1	N	-0,4164	0,0002	-0,1074	0,0207	-0,3516	0,2038	-0,0426	0,0046	-0,3330	0,9837	-0,0240	0,0024
	t_3	-0,3991	0,2285	-0,0901	0,0172	-0,3460	0,4561	-0,0370	0,0040	-0,3300	0,9194	-0,0210	0,0021
	PE_1	-0,4302	0,0000	-0,1212	0,0272	-0,3585	0,1740	-0,0495	0,0066	-0,3373	0,8243	-0,0283	0,0036
	PE_2	-0,4027	0,0218	-0,0937	0,0176	-0,3465	0,5987	-0,0375	0,0040	-0,3300	0,8050	-0,0210	0,0021
	NG_3	-0,3783	0,0002	-0,0693	0,0227	-0,3415	0,0379	-0,0325	0,0065	-0,3278	0,8645	-0,0188	0,0036
	L_2	-0,4065	0,0345	-0,0975	0,0179	-0,3482	0,6308	-0,0392	0,0041	-0,3311	0,9804	-0,0221	0,0021
3	N	-0,4092	0,0001	-0,1002	0,0189	-0,3462	0,2205	-0,0372	0,0042	-0,3291	0,9864	-0,0201	0,0022
	t_3	-0,3929	0,1471	-0,0839	0,0157	-0,3416	0,4437	-0,0326	0,0037	-0,3267	0,9468	-0,0177	0,0019
	PE_1	-0,4235	0,0001	-0,1145	0,0254	-0,3522	0,1242	-0,0432	0,0062	-0,3327	0,9790	-0,0237	0,0034
	PE_2	-0,3958	0,0347	-0,0868	0,0160	-0,3419	0,7339	-0,0329	0,0037	-0,3267	0,8976	-0,0177	0,0019
	NG_3	-0,3960	0,0001	-0,0870	0,0230	-0,3418	0,1704	-0,0328	0,0060	-0,3265	0,9149	-0,0175	0,0035
	L_2	-0,3997	0,0058	-0,0907	0,0164	-0,3434	0,6444	-0,0344	0,0037	-0,3276	0,9632	-0,0186	0,0019

Tabela D.12: Média das estimativas, valor- P do teste de kolmogorov-Sminorff, viés e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro normal generalizada e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 , PE_2 , NG e L_2 .

ϕ	Erro	n=100				n=300				n=500			
		$\hat{\beta}_0$	KS	viés	EQM	$\hat{\beta}_0$	KS	viés	EQM	$\hat{\beta}_0$	KS	viés	EQM
1	N	2,0636	0,3259	0,0636	0,0089	2,0887	0,0886	0,0887	0,0093	2,0936	0,0264	0,0936	0,0097
	t_3	2,0653	0,9308	0,0653	0,0110	2,0883	0,2422	0,0883	0,0099	2,0931	0,5079	0,0931	0,0100
	PE_1	2,0635	0,3617	0,0635	0,0087	2,0893	0,1781	0,0893	0,0094	2,0939	0,2366	0,0939	0,0097
	PE_2	2,0640	0,9935	0,0640	0,0104	2,0882	0,1842	0,0882	0,0098	2,0934	0,5927	0,0934	0,0101
	NG_3	2,1279	0,0000	0,1279	0,1872	2,1174	0,0003	0,1174	0,0551	2,1057	0,5668	0,1057	0,0292
	L_2	2,0643	0,7115	0,0643	0,0097	2,0884	0,0814	0,0884	0,0095	2,0933	0,1112	0,0933	0,0098
3	N	2,0629	0,3298	0,0629	0,0178	2,0871	0,1311	0,0871	0,0121	2,0941	0,0375	0,0941	0,0115
	t_3	2,0625	0,7644	0,0625	0,0229	2,0853	0,2638	0,0853	0,0138	2,0925	0,4437	0,0925	0,0126
	PE_1	2,0642	0,7200	0,0642	0,0179	2,0888	0,2266	0,0888	0,0122	2,0950	0,2434	0,0950	0,0116
	PE_2	2,0612	0,7457	0,0612	0,0215	2,0856	0,2587	0,0856	0,0138	2,0930	0,2288	0,0930	0,0126
	NG_3	2,1314	0,0000	0,1314	0,5649	2,1175	0,0043	0,1175	0,0875	2,1112	0,4242	0,1112	0,0506
	L_2	2,0625	0,7609	0,0625	0,0196	2,0861	0,0713	0,0861	0,0127	2,0932	0,1096	0,0932	0,0119
		n=100				n=300				n=500			
	ϕ	Erro	$\hat{\beta}_1$	KS	viés	EQM	$\hat{\beta}_1$	KS	viés	EQM	$\hat{\beta}_1$	KS	viés
1	N	6,9919	0,9291	-0,0081	0,0085	6,9965	0,7923	-0,0035	0,0025	7,0010	0,8989	0,0010	0,0016
	t_3	6,9927	0,8080	-0,0073	0,0121	6,9947	0,3923	-0,0053	0,0035	7,0009	0,9464	0,0009	0,0022
	PE_1	6,9915	0,7856	-0,0085	0,0080	6,9979	0,6311	-0,0021	0,0023	7,0011	0,3705	0,0011	0,0014
	PE_2	6,9921	0,9714	-0,0079	0,0113	6,9948	0,3577	-0,0052	0,0033	7,0012	0,7629	0,0012	0,0022
	NG_3	6,9706	0,6134	-0,0294	0,0423	7,0372	0,1219	0,0372	0,0166	7,0135	0,8457	0,0135	0,0122
	L_2	6,9923	0,9577	-0,0077	0,0099	6,9955	0,5069	-0,0045	0,0029	7,0009	0,9576	0,0009	0,0019
3	N	6,9968	0,6806	-0,0032	0,0242	6,9952	0,3987	-0,0048	0,0077	7,0017	0,7441	0,0017	0,0045
	t_3	6,9982	0,9058	-0,0018	0,0343	6,9933	0,3686	-0,0067	0,0107	7,0010	0,9717	0,0010	0,0064
	PE_1	6,9957	0,6386	-0,0043	0,0224	6,9969	0,8374	-0,0031	0,0071	7,0022	0,4812	0,0022	0,0041
	PE_2	6,9978	0,9516	-0,0022	0,0324	6,9934	0,2807	-0,0066	0,0102	7,0019	0,7845	0,0019	0,0063
	NG_3	6,9731	0,3558	-0,0269	0,0999	7,0578	0,8127	0,0578	0,0393	7,0444	0,9178	0,0444	0,0316
	L_2	6,9977	0,4790	-0,0023	0,0282	6,9941	0,3815	-0,0059	0,0090	7,0013	0,7877	0,0013	0,0053
		n=100				n=300				n=500			
	ϕ	Erro	$\hat{\phi}$	KS	viés	EQM	$\hat{\phi}$	KS	viés	EQM	$\hat{\phi}$	KS	viés
1	N	0,3123	0,4807	-0,6877	0,4744	0,3501	0,8923	-0,6499	0,4231	0,3598	0,8796	-0,6402	0,4102
	t_3	0,1921	0,1981	-0,8079	0,6536	0,2339	0,8578	-0,7661	0,5873	0,2440	0,7226	-0,7560	0,5717
	PE_1	0,4968	0,5155	-0,5032	0,2568	0,5601	0,8572	-0,4399	0,1951	0,5759	0,5934	-0,4241	0,1807
	PE_2	0,1290	0,5013	-0,8710	0,7590	0,1487	0,9560	-0,8513	0,7249	0,1537	0,6339	-0,8463	0,7163
	NG_3	1,0479	0,0001	0,0479	0,0412	1,0099	0,2344	0,0099	0,0084	1,0096	0,3102	0,0096	0,0036
	L_2	0,1039	0,4440	-0,8961	0,8033	0,1192	0,9985	-0,8808	0,7760	0,1231	0,5268	-0,8769	0,7690
3	N	0,9284	0,0582	-2,0716	4,3096	1,0521	0,5414	-1,9479	3,8001	1,0784	0,9478	-1,9216	3,6958
	t_3	0,5869	0,5230	-2,4131	5,8340	0,7074	0,6928	-2,2926	5,2597	0,7343	0,7488	-2,2657	5,1356
	PE_1	1,4793	0,0731	-1,5207	2,3561	1,6835	0,8797	-1,3165	1,7468	1,7266	0,7887	-1,2734	1,6298
	PE_2	0,3868	0,2373	-2,6132	6,8324	0,4479	0,6151	-2,5521	6,5144	0,4613	0,9038	-2,5387	6,4455
	NG_3	2,9657	0,0000	-0,0343	0,4623	2,9783	0,0185	-0,0217	0,0658	2,9910	0,5632	-0,0090	0,0312
	L_2	0,3110	0,3137	-2,6890	7,2329	0,3588	0,5326	-2,6412	6,9765	0,3694	0,8919	-2,6306	6,9206
		n=100				n=300				n=500			
	ϕ	Erro	$\hat{\rho}_1$	KS	viés	EQM	$\hat{\rho}_1$	KS	viés	EQM	$\hat{\rho}_1$	KS	viés
1	N	0,6058	0,6625	-0,0502	0,0124	0,6177	0,5325	-0,0383	0,0048	0,6309	0,7636	-0,0251	0,0026
	t_3	0,6012	0,3997	-0,0548	0,0168	0,6121	0,5935	-0,0439	0,0067	0,6263	0,8881	-0,0297	0,0035
	PE_1	0,6071	0,7710	-0,0489	0,0121	0,6206	0,7866	-0,0354	0,0043	0,6333	0,8646	-0,0227	0,0023
	PE_2	0,6053	0,7949	-0,0507	0,0153	0,6130	0,5038	-0,0430	0,0065	0,6273	0,9378	-0,0287	0,0034
	NG_3	0,6926	0,3533	0,0366	0,0196	0,6432	0,3089	-0,0128	0,0047	0,6457	0,9516	-0,0103	0,0025
	L_2	0,6040	0,6417	-0,0520	0,0141	0,6150	0,3358	-0,0410	0,0056	0,6286	0,9350	-0,0274	0,0030
3	N	0,6114	0,7619	-0,0446	0,0118	0,6253	0,5891	-0,0307	0,0042	0,6355	0,9267	-0,0205	0,0024
	t_3	0,6054	0,8164	-0,0506	0,0159	0,6208	0,7136	-0,0352	0,0058	0,6320	0,7494	-0,0240	0,0032
	PE_1	0,6121	0,7609	-0,0439	0,0117	0,6274	0,8397	-0,0286	0,0037	0,6373	0,6989	-0,0187	0,0021
	PE_2	0,6085	0,8615	-0,0475	0,0146	0,6212	0,6689	-0,0348	0,0057	0,6330	0,9138	-0,0230	0,0031
	NG_3	0,6748	0,5892	0,0188	0,0159	0,6424	0,4308	-0,0136	0,0041	0,6449	0,9907	-0,0111	0,0023
	L_2	0,6092	0,4625	-0,0468	0,0133	0,6231	0,3410	-0,0329	0,0049	0,6338	0,9118	-0,0222	0,0027
		n=100				n=300				n=500			
	ϕ	Erro	$\hat{\rho}_2$	KS	viés	EQM	$\hat{\rho}_2$	KS	viés	EQM	$\hat{\rho}_2$	KS	viés
1	N	-0,4188	0,0048	-0,1098	0,0216	-0,3571	0,1101	-0,0481	0,0047	-0,3394	0,6682	-0,0304	0,0026
	t_3	-0,4275	0,0092	-0,1185	0,0273	-0,3633	0,0180	-0,0543	0,0063	-0,3436	0,4038	-0,0346	0,0035
	PE_1	-0,4119	0,0015	-0,1029	0,0198	-0,3531	0,7768	-0,0441	0,0043	-0,3368	0,9882	-0,0278	0,0023
	PE_2	-0,4258	0,0251	-0,1168	0,0260	-0,3618	0,0817	-0,0528	0,0061	-0,3429	0,5387	-0,0339	0,0034
	NG_3	-0,3029	0,0151	0,0061	0,0165	-0,3146	0,0638	-0,0056	0,0040	-0,3127	0,9462	-0,0037	0,0020
	L_2	-0,4239	0,0128	-0,1149	0,0241	-0,3604	0,1032	-0,0514	0,0054	-0,3416	0,4295	-0,0326	0,0030
3	N	-0,4219	0,0449	-0,1129	0,0219	-0,3566	0,2940	-0,0476	0,0048	-0,3366	0,8620	-0,0276	0,0025
	t_3	-0,4321	0,1246	-0,1231	0,0278	-0,3632	0,0103	-0,0542	0,0065	-0,3406	0,5602	-0,0316	0,0033
	PE_1	-0,4148	0,0348	-0,1058	0,0197	-0,3526	0,5484	-0,0436	0,0043	-0,3343	0,9889	-0,0253	0,0023
	PE_2	-0,4294	0,0958	-0,1204	0,0264	-0,3622	0,1243	-0,0532	0,0063	-0,3402	0,7262	-0,0312	0,0032
	NG_3	-0,3315	0,0014	-0,0225	0,0156	-0,3236	0,0279	-0,0146	0,0038	-0,3169	0,6155	-0,0079	0,0020
	L_2	-0,4274	0,0640	-0,1184	0,0246	-0,3601	0,0561	-0,0511	0,0055	-0,3386	0,7704	-0,0296	0,0028

Tabela D.13: Média das estimativas, valor-P do teste de kolmogorov-Sminorff, vies e erro quadrático médio (EQM) de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\phi}$, $\hat{\rho}_1$ e $\hat{\rho}_2$ do estudo de simulação no qual os dados são gerados a partir do modelo (8.5) sob erro L_2 e ajustados sobre a distribuição dos erros normal, t_3 , PE_1 , PE_2 , NG e L_2 .

ϕ	Erro	n=100				n=300				n=500			
		$\hat{\beta}_0$	KS	vies	EQM	$\hat{\beta}_0$	KS	vies	EQM	$\hat{\beta}_0$	KS	vies	EQM
1	N	2,0575	0,0725	0,0575	0,0453	2,0860	0,0475	0,0860	0,0208	2,0902	0,0287	0,0902	0,0160
	t_3	2,0580	0,3045	0,0580	0,0494	2,0860	0,6467	0,0860	0,0229	2,0894	0,4179	0,0894	0,0165
	PE_1	2,0581	0,4848	0,0581	0,0651	2,0873	0,5017	0,0873	0,0290	2,0907	0,1865	0,0907	0,0222
	PE_2	2,0558	0,1617	0,0558	0,0469	2,0863	0,1659	0,0863	0,0221	2,0891	0,2882	0,0891	0,0160
	NG_3	2,1207	0,0000	0,1207	0,6436	2,1123	0,0100	0,1123	0,1551	2,1048	0,4651	0,1048	0,0893
	L_2	2,0577	0,0917	0,0577	0,0438	2,0857	0,3765	0,0857	0,0207	2,0897	0,1920	0,0897	0,0155
3	N	2,0520	0,0746	0,0520	0,1289	2,0836	0,0463	0,0836	0,0472	2,0875	0,0271	0,0875	0,0314
	t_3	2,0538	0,3523	0,0538	0,1421	2,0839	0,7148	0,0839	0,0537	2,0861	0,3675	0,0861	0,0329
	PE_1	2,0509	0,6917	0,0509	0,1971	2,0854	0,4609	0,0854	0,0735	2,0890	0,2040	0,0890	0,0509
	PE_2	2,0492	0,3970	0,0492	0,1360	2,0841	0,3056	0,0841	0,0512	2,0857	0,2177	0,0857	0,0317
	NG_3	2,1016	0,0000	0,1016	0,8707	2,1111	0,1451	0,1111	0,2463	2,1014	0,2603	0,1014	0,1787
	L_2	2,0531	0,1469	0,0531	0,1244	2,0833	0,2995	0,0833	0,0470	2,0865	0,1834	0,0865	0,0299
ϕ	Erro	n=100				n=300				n=500			
		$\hat{\beta}_1$	KS	vies	EQM	$\hat{\beta}_1$	KS	vies	EQM	$\hat{\beta}_1$	KS	vies	EQM
1	N	6,9874	0,9342	-0,0126	0,0828	6,9916	0,6871	-0,0084	0,0209	6,9979	0,9342	-0,0021	0,0144
	t_3	6,9959	0,8686	-0,0041	0,0784	6,9936	0,9664	-0,0064	0,0207	6,9979	0,9506	-0,0021	0,0134
	PE_1	6,9807	0,8147	-0,0193	0,1140	6,9886	0,9997	-0,0114	0,0309	6,9985	0,8145	-0,0015	0,0220
	PE_2	6,9957	0,9872	-0,0043	0,0781	6,9936	0,8122	-0,0064	0,0213	6,9974	0,8245	-0,0026	0,0137
	NG_3	6,9776	0,7693	-0,0224	0,2299	7,0375	0,8966	0,0375	0,0909	7,0242	0,0529	0,0242	0,0595
	L_2	6,9919	0,9968	-0,0081	0,0752	6,9930	0,9392	-0,0070	0,0194	6,9979	0,9836	-0,0021	0,0130
3	N	6,9832	0,9778	-0,0168	0,2457	6,9880	0,7329	-0,0120	0,0624	6,9972	0,8684	-0,0028	0,0428
	t_3	6,9979	0,9029	-0,0021	0,2322	6,9908	0,7640	-0,0092	0,0620	6,9971	0,9067	-0,0029	0,0395
	PE_1	6,9710	0,8646	-0,0290	0,3453	6,9827	0,9965	-0,0173	0,0932	6,9979	0,7922	-0,0021	0,0660
	PE_2	6,9997	0,9981	-0,0003	0,2342	6,9918	0,8974	-0,0082	0,0636	6,9965	0,9751	-0,0035	0,0403
	NG_3	7,0012	0,9048	0,0012	0,5392	7,0454	0,9669	0,0454	0,2043	7,0443	0,0657	0,0443	0,1456
	L_2	6,9910	0,9632	-0,0090	0,2229	6,9902	0,8593	-0,0098	0,0581	6,9972	0,9865	-0,0028	0,0385
ϕ	Erro	n=100				n=300				n=500			
		$\hat{\phi}$	KS	vies	EQM	$\hat{\phi}$	KS	vies	EQM	$\hat{\phi}$	KS	vies	EQM
1	N	2,8596	0,0707	1,8596	3,7690	3,1063	0,3457	2,1063	4,5333	3,1746	0,3172	2,1746	4,7891
	t_3	1,5054	0,0026	0,5054	0,3443	1,6893	0,3360	0,6893	0,5061	1,7436	0,7830	0,7436	0,5712
	PE_1	4,9480	0,0413	3,9480	16,7467	5,5155	0,2024	4,5155	20,7821	5,6619	0,2704	4,6619	21,9838
	PE_2	1,1038	0,0272	0,1038	0,0531	1,2054	0,2974	0,2054	0,0558	1,2354	0,8150	0,2354	0,0637
	NG_3	9,1161	0,0510	8,1161	70,0835	9,6206	0,0291	8,6206	75,5932	9,7713	0,3004	8,7713	77,7167
	L_2	0,8663	0,0333	-0,1337	0,0438	0,9433	0,5560	-0,0567	0,0116	0,9664	0,9188	-0,0336	0,0062
3	N	8,8498	0,0515	5,8498	37,1111	9,4042	0,2614	6,4042	41,8858	9,5840	0,3977	6,5840	43,8976
	t_3	4,7165	0,0068	1,7165	3,7812	5,1343	0,1639	2,1343	4,8360	5,2755	0,7736	2,2755	5,3443
	PE_1	15,4128	0,0343	12,4128	165,1123	16,7453	0,3246	13,7453	192,5028	17,1272	0,3695	14,1272	201,8641
	PE_2	3,4250	0,0117	0,4250	0,5770	3,6531	0,4620	0,6531	0,5492	3,7312	0,8760	0,7312	0,6096
	NG_3	27,2834	0,0976	24,2834	625,3872	28,9153	0,1084	25,9153	682,8874	29,4166	0,2298	26,4166	704,9116
	L_2	2,6858	0,0124	-0,3142	0,3405	2,8583	0,6052	-0,1417	0,0955	2,9190	0,9788	-0,0810	0,0523
ϕ	Erro	n=100				n=300				n=500			
		$\hat{\rho}_1$	KS	vies	EQM	$\hat{\rho}_1$	KS	vies	EQM	$\hat{\rho}_1$	KS	vies	EQM
1	N	0,6353	0,5303	-0,0207	0,0098	0,6329	0,7628	-0,0231	0,0038	0,6403	0,9802	-0,0157	0,0023
	t_3	0,6345	0,8675	-0,0215	0,0103	0,6351	0,8492	-0,0209	0,0036	0,6418	0,8795	-0,0142	0,0021
	PE_1	0,6316	0,5581	-0,0244	0,0131	0,6282	0,8266	-0,0278	0,0057	0,6372	0,9962	-0,0188	0,0036
	PE_2	0,6355	0,3760	-0,0205	0,0101	0,6348	0,8743	-0,0212	0,0036	0,6415	0,9946	-0,0145	0,0022
	NG_3	0,6659	0,7376	0,0099	0,0160	0,6362	0,8712	-0,0198	0,0060	0,6413	0,9873	-0,0147	0,0038
	L_2	0,6358	0,3001	-0,0202	0,0094	0,6346	0,4396	-0,0214	0,0035	0,6414	0,9662	-0,0146	0,0021
3	N	0,6533	0,6179	-0,0027	0,0088	0,6402	0,8174	-0,0158	0,0035	0,6446	0,9921	-0,0114	0,0022
	t_3	0,6509	0,5662	-0,0051	0,0092	0,6415	0,5967	-0,0145	0,0032	0,6457	0,8050	-0,0103	0,0020
	PE_1	0,6515	0,7812	-0,0045	0,0119	0,6367	0,8529	-0,0193	0,0052	0,6421	0,9724	-0,0139	0,0034
	PE_2	0,6517	0,4368	-0,0043	0,0090	0,6414	0,7239	-0,0146	0,0033	0,6455	0,9667	-0,0105	0,0020
	NG_3	0,6676	0,9255	0,0116	0,0134	0,6402	0,9571	-0,0158	0,0054	0,6445	0,9177	-0,0115	0,0037
	L_2	0,6527	0,3496	-0,0033	0,0084	0,6413	0,4256	-0,0147	0,0031	0,6455	0,9895	-0,0105	0,0019
ϕ	Erro	n=100				n=300				n=500			
		$\hat{\rho}_2$	KS	vies	EQM	$\hat{\rho}_2$	KS	vies	EQM	$\hat{\rho}_2$	KS	vies	EQM
1	N	-0,4162	0,0003	-0,1072	0,0205	-0,3496	0,1810	-0,0406	0,0043	-0,3333	0,9836	-0,0243	0,0023
	t_3	-0,4026	0,1098	-0,0936	0,0183	-0,3457	0,4401	-0,0367	0,0041	-0,3316	0,6925	-0,0226	0,0022
	PE_1	-0,4280	0,0001	-0,1190	0,0260	-0,3550	0,0381	-0,0460	0,0061	-0,3364	0,9350	-0,0274	0,0034
	PE_2	-0,4061	0,0147	-0,0971	0,0188	-0,3465	0,5288	-0,0375	0,0041	-0,3319	0,6608	-0,0229	0,0023
	NG_3	-0,3811	0,0009	-0,0721	0,0219	-0,3381	0,0209	-0,0291	0,0061	-0,3267	0,9547	-0,0177	0,0033
	L_2	-0,4081	0,0265	-0,0991	0,0185	-0,3470	0,5399	-0,0380	0,0040	-0,3321	0,8051	-0,0231	0,0022
3	N	-0,4084	0,0003	-0,0994	0,0186	-0,3444	0,1663	-0,0354	0,0040	-0,3295	0,9278	-0,0205	0,0022
	t_3	-0,3956	0,1162	-0,0866	0,0167	-0,3413	0,5160	-0,0323	0,0038	-0,3282	0,7512	-0,0192	0,0021
	PE_1	-0,4207	0,0002	-0,1117	0,0241	-0,3491	0,0489	-0,0401	0,0057	-0,3321	0,9040	-0,0231	0,0031
	PE_2	-0,3986	0,0299	-0,0896	0,0170	-0,3420	0,5025	-0,0330	0,0039	-0,3284	0,8901	-0,0194	0,0021
	NG_3	-0,3954	0,0001	-0,0864	0,0220	-0,3387	0,0740	-0,0297	0,0056	-0,3257	0,9304	-0,0167	0,0032
	L_2	-0,4007	0,0053	-0,0917	0,0168	-0,3423	0,5522	-0,0333	0,0037	-0,3286	0,8503	-0,0196	0,0020

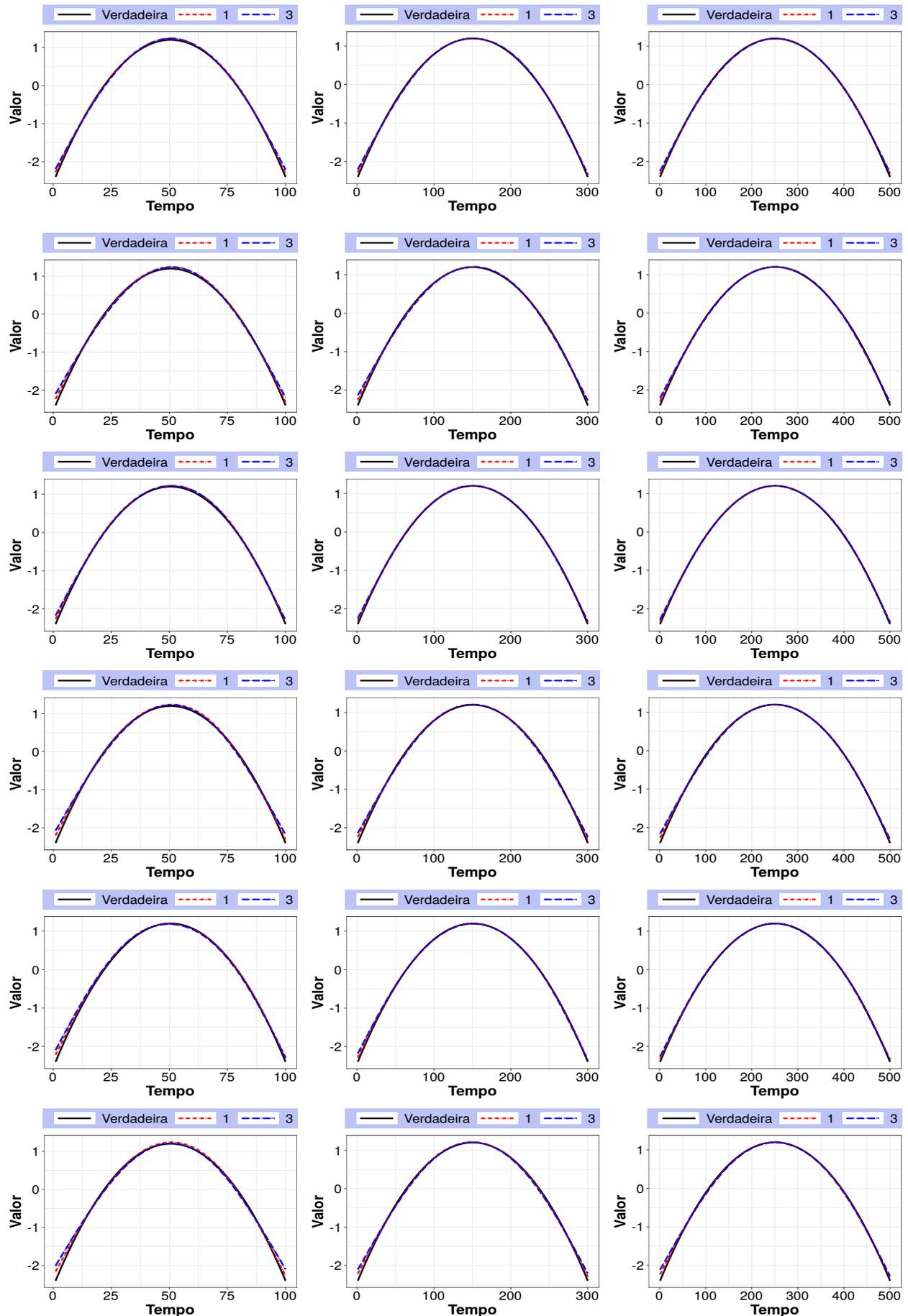


Figura D.3: Gráficos da média das estimativas para a função $f_1(t)$ sob $\phi = 1, 3$ com coeficientes de autocorrelação $\rho_1 = 0,656$ e $\rho_2 = -0,309$ a partir do estudo de simulação a qual os dados foram gerados do modelo (8.5) sob os erros com distribuição normal (primeiro), t_3 (segundo), PE_1 (terceiro), PE_2 (quarto), NG (quinto) e L_2 (sexto) e ajustado sob o mesmo erro do modelo. Tamanhos amostrais de $n = 100, 300$ e 500 , na esquerda, meio e direita, respectivamente.

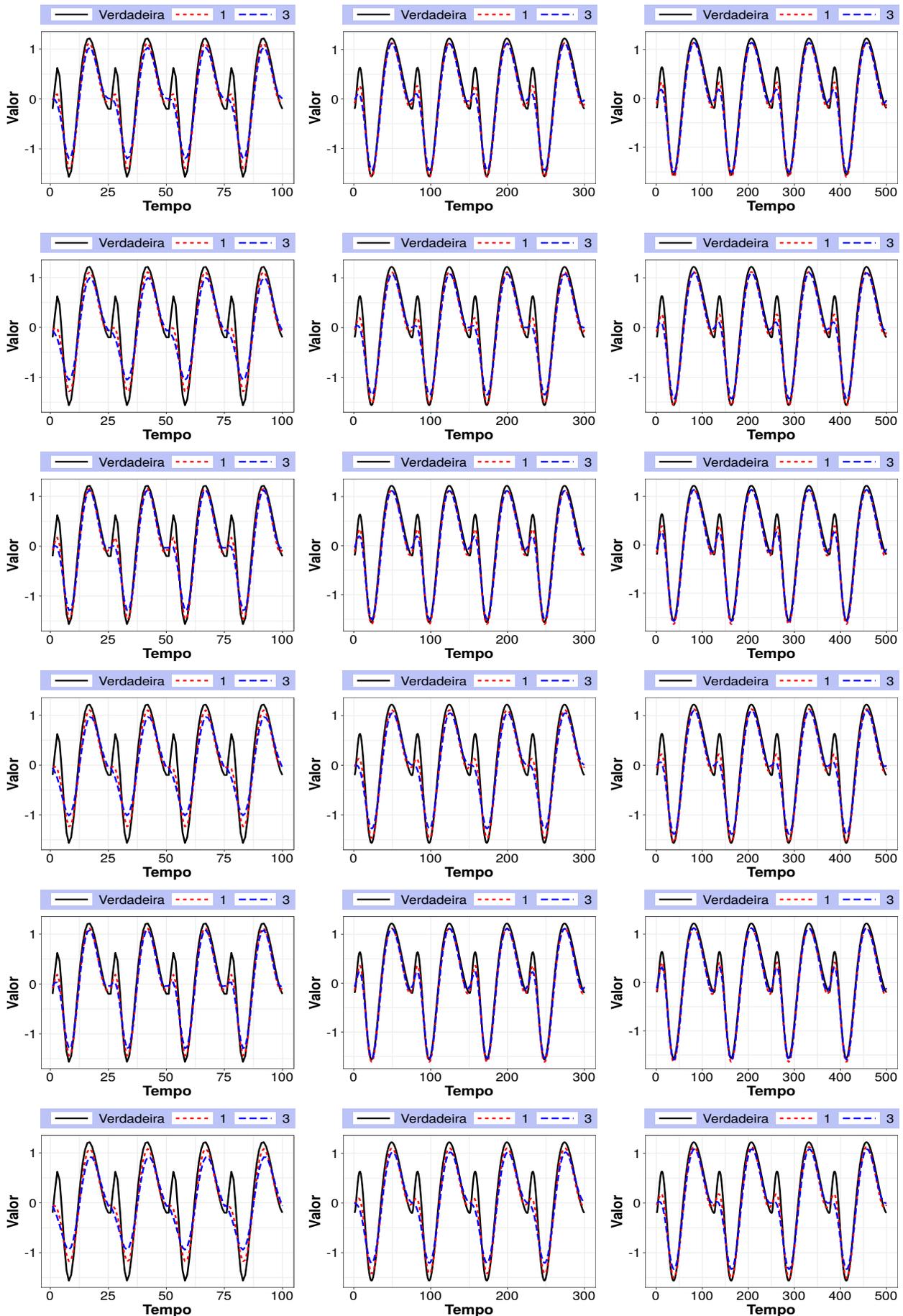


Figura D.4: Gráficos da média das estimativas para a função $f_2(s)$ sob $\phi = 1, 3$ com coeficientes de autocorrelação $\rho_1 = 0,656$ e $\rho_2 = -0,309$ a partir do estudo de simulação a qual os dados foram gerados do modelo (8.5) sob os erros com distribuição normal (primeiro), t_3 (segundo), PE_1 (terceiro), PE_2 (quarto), NG (quinto) e L_2 (sexto) e ajustado sob o mesmo erro do modelo. Tamanho amostrais de $n = 100, 300$ e 500 , na esquerda, meio e direita, respectivamente.

Referências Bibliográficas

- Akaike(1973)** H Akaike. Information theory as an extension of the maximum likelihood principle. á in: Petrov, bn and csaki, f. Em *Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 276-281*. Citado na pág. 65
- Amorim(2019)** William Nilson de Amorim. *Ciência de Dados, Poluição do Ar e Saúde*. Tese de Doutorado, Universidade de São Paulo. Citado na pág. 97
- Andrews e Pregibon(1978)** David F Andrews e Daryl Pregibon. Finding the outliers that matter. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(1):85–93. Citado na pág. 59
- Arellano-Valle(1994)** Reinaldo Boris Arellano-Valle. Distribuições Elípticas: Propriedades, Inferência e Aplicações a Modelos de Regressão. Dissertação de Mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo. Citado na pág. 15, 20
- Barros e Paula(2019)** Michelli Barros e Gilberto A Paula. Discussion of “Birnbbaum-Saunders distributions: A review of models, analysis and applications”. *Applied Stochastic Models in Business and Industry*, 35(1):96–99. Citado na pág. 60
- Belsley et al.(1980)** David A Belsley, Edwin Kuh e Roy E Welsch. Regression Diagnostics: Identifying. Citado na pág. 59
- Borssoi et al.(2020)** Joelmir A Borssoi, Gilberto A Paula e Manuel Galea. Elliptical linear mixed models with a covariate subject to measurement error. *Statistical Papers*, 61(1):31–69. Citado na pág. 62
- Breiman e Friedman(1985)** Leo Breiman e Jerome H Friedman. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80(391):580–598. Citado na pág. 22
- Brook et al.(2010)** Robert D Brook, Sanjay Rajagopalan, C Arden Pope III, Jeffrey R Brook, Aruni Bhatnagar, Ana V Diez-Roux, Fernando Holguin, Yuling Hong, Russell V Luepker, Murray A Mittleman et al. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the american heart association. *Circulation*, 121(21):2331–2378. Citado na pág. 96
- Brunekreef e Holgate(2002)** Bert Brunekreef e Stephen T Holgate. Air pollution and health. *The Lancet*, 360(9341):1233–1242. Citado na pág. 96
- Buja et al.(1989)** Andreas Buja, Trevor Hastie e Robert Tibshirani. Linear smoothers and additive models. *The Annals of Statistics*, 17(2):453–510. Citado na pág. 22
- Byrd et al.(1995)** Richard H Byrd, Peihuang Lu, Jorge Nocedal e Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5): 1190–1208. Citado na pág. 35, 54

- Cahill et al.(2015)** Niamh Cahill, Stefan Rahmstorf e Andrew C Parnell. Change points of global temperature. *Environmental Research Letters*, 10(8):084002. Citado na pág. 3
- Cambanis et al.(1981)** Stamatis Cambanis, Steel Huang e Gordon Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385. Citado na pág. 15
- Cao et al.(2010)** Chun-Zheng Cao, Jin-Guan Lin e Li-Xing Zhu. Heteroscedasticity and/or autocorrelation diagnostics in nonlinear models with AR(1) and symmetrical errors. *Statistical Papers*, 51(4):813–836. Citado na pág. 2, 16, 32, 62
- Cheema et al.(2011)** Sohail Babar Cheema, Ghulam Rasul, Gohar Ali e Dildar Hussain Kazmi. A comparison of minimum temperature trends with model projections. *Pakistan Journal of Meteorology*, 8(15):39–52. Citado na pág. 3
- Chen et al.(2020)** Kai Chen, Alexandra Schneider, Josef Cyrus, Kathrin Wolf, Christa Meisinger, Margit Heier, Wolfgang von Scheidt, Bernhard Kuch, Mike Pitz, Annette Peters et al. Hourly exposure to ultrafine particle metrics and the onset of myocardial infarction in augsburg, germany. *Environmental Health Perspectives*, 128(1):017003. Citado na pág. 96
- Chmielewski(1981)** MA Chmielewski. Elliptically symmetric distributions: a review and bibliography. *International Statistical Review/Revue Internationale de Statistique*, 49(1):67–74. Citado na pág. 15
- Chung et al.(2011)** Eun-Sung Chung, Kyungshin Park e Kil Seong Lee. The relative impacts of climate change and urbanization on the hydrological response of a korean urban watershed. *Hydrological Processes*, 25(4):544–560. Citado na pág. 3
- Cleveland et al.(1990)** Robert B Cleveland, William S Cleveland, Jean E McRae e Irma Terpenning. Stl: a seasonal-trend decomposition. *Journal of Official Statistics*, 6(1):3–73. Citado na pág. 28
- Cook(1986)** R Dennis Cook. Assessment of local influence (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(2):133–169. Citado na pág. 61
- Cook e Weisberg(1982)** R Dennis Cook e Sanford Weisberg. *Residuals and Influence in Regression*. New York: Chapman and Hall. Citado na pág. 61
- Cox e Hinkley(1974)** David R Cox e David V Hinkley. *Theoretical Statistics*. New York: Chapman&Hall/CRC. Citado na pág. 19
- Cysneiros(2004)** F. J. A. Cysneiros. *Métodos Restritos e Validação de Modelos Simétricos de Regressão*. Tese de Doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo. Citado na pág. 15, 16
- Cysneiros et al.(2010)** FJA Cysneiros, GM Cordeiro e AHMA Cysneiros. Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation*, 80(4):451–461. Citado na pág. 2
- Cysneiros(2011)** Francisco José A. Cysneiros. *Regression Models with Symmetrical Errors*, páginas 1211–1212. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-642-04898-2. doi: 10.1007/978-3-642-04898-2_486. URL https://doi.org/10.1007/978-3-642-04898-2_486. Citado na pág. 16
- Cysneiros e Paula(2005)** Francisco José A Cysneiros e Gilberto A Paula. Restricted methods in symmetrical linear regression models. *Computational Statistics & Data Analysis*, 49(3):689–708. Citado na pág. 2, 15, 16, 17, 18, 19, 20, 29, 32, 39

- Cysneiros et al.(2007)** Francisco José A Cysneiros, Gilberto A Paula e Manuel Galea. Heteroscedastic symmetrical linear models. *Statistics & Probability Letters*, 77(11):1084–1090. Citado na pág. 2, 16
- Davidon(1991)** William C Davidon. Variable metric method for minimization. *SIAM Journal on Optimization*, 1(1):1–17. Citado na pág. 35, 54
- De Boor et al.(1978)** Carl De Boor, Carl De Boor, Etats-Unis Mathématicien, Carl De Boor e Carl De Boor. *A Practical Guide to Splines*, volume 27. Springer-Verlag New York. Citado na pág. 11
- De Hoog e Hutchinson(1986)** FR De Hoog e MF Hutchinson. An efficient method for calculating smoothing splines using orthogonal transformations. *Numerische Mathematik*, 50(3):311–319. Citado na pág. 12
- Dominguez-Molina et al.(2003)** J Armando Dominguez-Molina, Graciela González-Farías, Ramón M Rodríguez-Dagnino e ITESM Campus Monterrey. A practical procedure to estimate the shape parameter in the generalized gaussian distribution. *Technique Report I-01-18_eng. pdf, available through http://www.cimat.mx/reportes/enlinea/I-01-18_eng.pdf*, 1. Citado na pág. 20
- Dunn e Smyth(1996)** Peter K Dunn e Gordon K Smyth. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244. Citado na pág. 60
- Eilers e Marx(1996)** Paul HC Eilers e Brian D Marx. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2):89–102. Citado na pág. 14, 36, 57
- Eilers e Marx(2010)** Paul HC Eilers e Brian D Marx. Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(6):637–653. Citado na pág. 14
- Eubank(1999)** Randall L Eubank. *Nonparametric Regression and Spline Smoothing*. CRC Press. Citado na pág. 20
- Fang e Anderson(1990)** Kai-Tai Fang e Theodore Wilbur Anderson. *Statistical Inference in Elliptically Contoured and Related Distributions*. Allerton Press New York. Citado na pág. 15
- Fang et al.(1990)** Kai-Tai Fang, Samuel Kotz e Kai Wang Ng. *Symmetric Multivariate and Related Distributions*. URL <http://hdl.handle.net/10722/130548>. Citado na pág. 15, 16, 29
- Ferrari e Uribe-Opazo(2001)** Silvia LP Ferrari e Miguel A Uribe-Opazo. Corrected likelihood ratio tests in class of symmetric linear regression models. *Brazilian Journal of Probability and Statistics*, 15(1):49–67. Citado na pág. 15
- Fox(2015)** John Fox. *Applied Regression Analysis and Generalized Linear Models*. Sage Publications. Citado na pág. 25, 50
- Friedman e Stuetzle(1981)** Jerome H Friedman e Werner Stuetzle. Projection pursuit regression. *Journal of the American statistical Association*, 76(376):817–823. Citado na pág. 22
- Galea et al.(2003)** Manuel Galea, Gilberto A Paula e Miguel Uribe-Opazo. On influence diagnostic in univariate elliptical linear regression models. *Statistical Papers*, 44(1):23–45. Citado na pág. 2, 62
- Galea et al.(2005)** Manuel Galea, Gilberto A Paula e Francisco José A Cysneiros. On diagnostics in symmetrical nonlinear models. *Statistics & Probability Letters*, 73(4):459–467. Citado na pág. 2, 62
- George G. Judge(1985)** R. Carter Hill Helmut Lutkepohl Tsoung-Chao Lee George G. Judge, William E. Griffiths. *The Theory and Practice of Econometrics*. New York: John Wiley. Citado na pág. 24

- Golub e Van Loan(2012)** Gene H Golub e Charles F Van Loan. *Matrix Computations*, volume 3. Johns Hopkins University Press. Citado na pág. 22
- Gourieroux e Monfort(1995)** Christian Gourieroux e Alain Monfort. *Statistics and Econometric Models*, volume 1. Cambridge University Press. Citado na pág. 19
- Green e Silverman(1994)** Peter J Green e Bernard W Silverman. *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. Chapman and Hall/CRC. Citado na pág. 11, 21, 29, 31, 35, 36, 53, 56, 65, 66
- Greene(2012)** William H Greene. *Econometric Analysis, 7ed.* Prentice Hall. Citado na pág. 2, 28
- Gupta e Varga(2012)** Arjun K Gupta e Tamas Varga. *Elliptically Contoured Models in Statistics*, volume 240. Springer Science & Business Media. Citado na pág. 15
- Hansen et al.(1981)** James Hansen, Donald Johnson, Andrew Lacis, Sergej Lebedeff, Pius Lee, David Rind e Gary Russell. Climate impact of increasing atmospheric carbon dioxide. *Science*, 213(4511):957–966. Citado na pág. 81
- Hansen et al.(2006)** James Hansen, Makiko Sato, Reto Ruedy, Ken Lo, David W Lea e Martin Medina-Elizade. Global temperature change. *Proceedings of the National Academy of Sciences*, 103(39):14288–14293. Citado na pág. 81
- Harvey(1990)** Andrew C Harvey. *The Econometric Analysis of Time Series*. Mit Press. Citado na pág. 24
- Hastie e Tibshirani(1990)** TJ Hastie e RJ Tibshirani. *Generalized Additive Models*. London: Chapman & Hall. Citado na pág. 22, 31, 34, 36, 56
- Hastie e Tibshirani(1986)** Trevor Hastie e Robert Tibshirani. Generalized additive models. *Statistical Science*, 1(3):297–310. Citado na pág. 22
- Hastie e Tibshirani(1987)** Trevor Hastie e Robert Tibshirani. Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398):371–386. Citado na pág. 22
- Henderson e Souto(2018)** Daniel J Henderson e Anne-Charlotte Souto. An introduction to nonparametric regression for labor economists. *Journal of Labor Research*, 39(4):355–382. Citado na pág. 14
- 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. 2
- Ibacache-Pulgar et al.(2012)** Germán Ibacache-Pulgar, Gilberto A Paula e Manuel Galea. Influence diagnostics for elliptical semiparametric mixed models. *Statistical Modelling*, 12(2):165–193. Citado na pág. 62
- Ibacache-Pulgar et al.(2013)** Germán Ibacache-Pulgar, Gilberto A Paula e Francisco José A Cysneiros. Semiparametric additive models under symmetric distributions. *Test*, 22(1):103–121. Citado na pág. 2, 21, 22, 29, 93
- Ihaka e Gentleman(1996)** Ross Ihaka e Robert Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314. Citado na pág. 67
- Judge e Griffiths(1985)** George Judge e William E Griffiths. *The Theory and Practice of Econometrics*, 1985. Citado na pág. 50
- Judge et al.(1982)** GG Judge, RC Hill, WE Griffiths, H. H., Lutkepohl e T. Lee. *Introduction to the Theory and Practice of Econometrics*, 1982. Citado na pág. 47

- Kelker(1970)** Douglas Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, 32(4):419–430. Citado na pág. 16
- Kesavachandran et al.(2013)** C Kesavachandran, BS Pangtey, V Bihari, M Fareed, MK Pathak, AK Srivastava e N Mathur. Particulate matter concentration in ambient air and its effects on lung functions among residents in the national capital region, india. *Environmental Monitoring and Assessment*, 185(2):1265–1272. Citado na pág. 97
- Kissock(1999)** JK Kissock. Ud epa average daily temperature archive, 1999. URL <http://academic.udayton.edu/kissock/http/Weather/default.htm>. Citado na pág. 86
- Kutner et al.(2005)** Michael H Kutner, Christopher J Nachtsheim, John Neter, William Li et al. *Applied Linear Statistical Models*, volume 5. McGraw-Hill Irwin New York. Citado na pág. 25
- Lancaster e Salkauskas(1986)** Peter Lancaster e Kestutis Salkauskas. *Curve and Surface Fitting. An Introduction*. London: Academic Press, 1986. Citado na pág. 12
- Lange et al.(1989)** Kenneth L Lange, Roderick JA Little e Jeremy MG Taylor. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408):881–896. Citado na pág. 15
- Lee e Xu(2004)** Sik-Yum Lee e Liang Xu. Influence analyses of nonlinear mixed-effects models. *Computational Statistics & Data Analysis*, 45(2):321–341. Citado na pág. 62
- Lehmann e Casella(2006)** Erich L Lehmann e George Casella. *Theory of Point Estimation*. Springer Science & Business Media. Citado na pág. 129
- Levi et al.(2002)** Fabio Levi, Francesca Lucchini, Eva Negri e Carlo La Vecchia. Trends in mortality from cardiovascular and cerebrovascular diseases in europe and other areas of the world. *Heart*, 88(2):119–124. Citado na pág. 108
- Little(1988)** Roderick JA Little. Robust estimation of the mean and covariance matrix from data with missing values. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 37(1): 23–38. Citado na pág. 15
- Liu(2004)** Shuangzhe Liu. On diagnostics in conditionally heteroskedastic time series models under elliptical distributions. *Journal of Applied Probability*, 41(A):393–405. Citado na pág. 2
- Machiwal e Jha(2009)** Deepesh Machiwal e Madan K Jha. Time series analysis of hydrologic data for water resources planning and management: a review. *Journal of Hydrology and Hydro-mechanics*, 54(3):237–257. Citado na pág. 3
- McCullagh e Nelder(1989)** P. McCullagh e J. A. Nelder. *Generalized linear models (Second edition)*. London: Chapman & Hall. Citado na pág. 9
- Mensah et al.(2017)** George A Mensah, Gina S Wei, Paul D Sorlie, Lawrence J Fine, Yves Rosenberg, Peter G Kaufmann, Michael E Mussolino, Lucy L Hsu, Ebyan Addou, Michael M Engelgau et al. Decline in cardiovascular mortality: possible causes and implications. *Circulation research*, 120(2):366–380. Citado na pág. 108
- Mikkonen et al.(2015)** S Mikkonen, M Laine, HM Mäkelä, H Gregow, H Tuomenvirta, M Lah-tinen e A Laaksonen. Trends in the average temperature in finland, 1847–2013. *Stochastic Environmental Research and Risk Assessment*, 29(6):1521–1529. Citado na pág. 3
- Mittelhammer et al.(2000)** Ron C Mittelhammer, George G Judge e Douglas J Miller. *Econometric Foundations*. Cambridge University Press. Citado na pág. 35, 54

- Montgomery et al.(2012)** Douglas C Montgomery, Elizabeth A Peck e G Geoffrey Vining. *Introduction to Linear Regression Analysis*, volume 821. John Wiley & Sons. Citado na pág. 15
- Montoya et al.(2014)** Eduardo L Montoya, Nehemias Ulloa e Victoria Miller. A simulation study comparing knot selection methods with equally spaced knots in a penalized regression spline. *International Journal of Statistics and Probability*, 3(3):96. Citado na pág. 14
- Nadarajah(2005)** Saralees Nadarajah. A generalized normal distribution. *Journal of Applied Statistics*, 32(7):685–694. Citado na pág. 16, 20
- Nastos et al.(2010)** Panagiotis T Nastos, Athanasios G Paliatsos, Michael B Anthracopoulos, Eleftheria S Roma e Kostas N Priftis. Outdoor particulate matter and childhood asthma admissions in athens, greece: a time-series study. *Environmental Health*, 9(1):45. Citado na pág. 97
- Osorio et al.(2007)** Felipe Osorio, Gilberto A Paula e Manuel Galea. Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics & Data Analysis*, 51(9):4354–4368. Citado na pág. 62
- Parker e Rice(1985)** RL Parker e JA Rice. Discussion of “some aspects of the spline smoothing approach to nonparametric curve fitting” by bw silverman. *Journal of the Royal Statistical Society, Series B*, 47:40–42. Citado na pág. 12
- Paula et al.(2009)** Gilberto A Paula, Marcio Medeiros e Filidor E Vilca-Labra. Influence diagnostics for linear models with first-order autoregressive elliptical errors. *Statistics & probability letters*, 79(3):339–346. Citado na pág. 2
- Poon e Poon(1999)** W-Y Poon e Yat Sun Poon. Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1): 51–61. Citado na pág. 61
- Pope III e Dockery(2006)** C Arden Pope III e Douglas W Dockery. Health effects of fine particulate air pollution: lines that connect. *Journal of the Air & Waste Management Association*, 56(6):709–742. Citado na pág. 96
- Rahmstorf et al.(2017)** Stefan Rahmstorf, Grant Foster e Niamh Cahill. Global temperature evolution: recent trends and some pitfalls. *Environmental Research Letters*, 12(5):054001. Citado na pág. 3
- Ralston e Rabinowitz(2001)** Anthony Ralston e Philip Rabinowitz. *A First Course in Numerical Analysis*. Courier Corporation. Citado na pág. 22
- Rao et al.(1997)** K Srinivasa Rao, CVRS Vijaya Kmmnar e J Lakshmi Narayana. On a new symmetrical distribution. *Journal of the Indian Society of Agricultural Statistics*, 50(1):95–102. Citado na pág. 15
- Reinsch(1967)** Christian H Reinsch. Smoothing by spline functions. *Numerische mathematik*, 10 (3):177–183. Citado na pág. 12
- Relvas e Paula(2016)** Carlos Eduardo M Relvas e Gilberto A Paula. Partially linear models with first-order autoregressive symmetric errors. *Statistical Papers*, 57(3):795–825. Citado na pág. 2, 3, 11, 16, 29, 30, 39, 62, 122
- Relvas(2013)** Carlos Eduardo Martins Relvas. Modelos Parcialmente Lineares com Erros simétricos Autoregressivos de Primeira Ordem. Dissertação de Mestrado, Universidade de São Paulo. Citado na pág. 25, 40, 41
- Ruppert et al.(2003)** David Ruppert, Matt P Wand e Raymond J Carroll. *Semiparametric regression*. Number 12. Cambridge University Press. Citado na pág. 14

- Schoenberg(1964)** I. J. Schoenberg. Spline functions and the problem of graduation. *Proceedings of the National Academy of Sciences*, 52(4):947–950. ISSN 0027-8424. doi: 10.1073/pnas.52.4.947. URL <https://www.pnas.org/content/52/4/947>. Citado na pág. 11
- Schwarz(1978)** Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464. Citado na pág. 65
- Seber e Lee(2012)** George AF Seber e Alan J Lee. *Linear Regression Analysis*, volume 329. John Wiley & Sons. Citado na pág. 15
- Shumway et al.(1988)** RH Shumway, AS Azari e Y Pawitan. Modeling mortality fluctuations in los angeles as functions of pollution and weather effects. *Environmental Research*, 45(2):224–241. Citado na pág. 108
- Shumway e Stoffer(2016)** Robert H Shumway e David S Stoffer. *Time Series Analysis and Its Applications*. Springer, 4 edição. Citado na pág. 108
- Soltani et al.(2007)** S Soltani, R Modarres e SS Eslamian. The use of time series modeling for the determination of rainfall climates of iran. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 27(6):819–829. Citado na pág. 3
- Stasinopoulos et al.(2015)** DM Stasinopoulos, RA Rigby, V Voudouris, G Heller e F De Bastiani. Flexible regression and smoothing. the gamlss packages in r. *GAMLSS for Statistical Modelling*. *GAMLSS for Statistical Modeling*. Citado na pág. 65
- Sun(2017)** Huihui Sun. Perturbation diagnostics of autocorrelation coefficients in non linear mixed-effects models with ar (1) errors based on m-estimation. *Communications in Statistics-Theory and Methods*, 46(16):8269–8277. Citado na pág. 64
- Taylor(1992)** Jeremy MG Taylor. Properties of modelling the error distribution with an extra shape parameter. *Computational Statistics & Data Analysis*, 13(1):33–46. Citado na pág. 15
- Thomas e Wallis(1971)** J James Thomas e Kenneth F Wallis. Seasonal variation in regression analysis. *Journal of the Royal Statistical Society: Series A (General)*, 134(1):57–72. Citado na pág. 52
- Tukey(1977)** John W Tukey. *Exploratory Data Analysis*, volume 2. Reading, Mass. Citado na pág. 9
- Uribe-Opazo et al.(2008)** Miguel A Uribe-Opazo, Silvia LP Ferrari e Gauss M Cordeiro. Improved score tests in symmetric linear regression models. *Communications in Statistics-Theory and Methods*, 37(2):261–276. Citado na pág. 15
- Vanegas e Paula(2016)** Luis H Vanegas e Gilberto A Paula. An extension of log-symmetric regression models: R codes and applications. *Journal of Statistical Computation and Simulation*, 86(9):1709–1735. Citado na pág. 36
- Vanegas e Cysneiros(2010)** Luis Hernando Vanegas e Francisco José A Cysneiros. Assessment of diagnostic procedures in symmetrical nonlinear regression models. *Computational Statistics & Data Analysis*, 54(4):1002–1016. Citado na pág. 2
- Varanasi e Aazhang(1989)** Mahesh K Varanasi e Behnaam Aazhang. Parametric generalized gaussian density estimation. *The Journal of the Acoustical Society of America*, 86(4):1404–1415. Citado na pág. 20
- Wahba(1980)** Grace Wahba. Spline bases, regularization, and generalized cross-validation for solving approximation problems with large quantities of noisy data. *Approximation Theory III*, 2. Citado na pág. 12

- Wahba(1983)** Grace Wahba. Bayesian confidence intervals for the cross-validated smoothing spline. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(1):133–150. Citado na pág. [55](#)
- Weisberg(2005)** Sanford Weisberg. *Applied Linear Regression*, volume 528. John Wiley & Sons. Citado na pág. [15](#)
- Wise(1955)** J Wise. The autocorrelation function and the spectral density function. *Biometrika*, 42(1/2):151–159. Citado na pág. [52](#), [132](#)
- Wood(2017)** Simon N Wood. *Generalized Additive Models: an Introduction with R, Second Edition*. Chapman and Hall/CRC. Citado na pág. [3](#), [10](#), [11](#), [12](#), [13](#), [23](#), [28](#), [29](#), [31](#), [34](#), [35](#), [36](#), [40](#), [43](#), [55](#), [56](#), [65](#), [69](#), [86](#), [87](#)
- Yamaguchi(1990)** K Yamaguchi. Generalized EM algorithm for model with contaminated error term. Em *Proceedings of the Seven Japan and Korea Joint Conference of Statistics*, páginas 107–114. Citado na pág. [15](#)