



Universidade Federal de Pernambuco  
Centro de Ciências Exatas e da Natureza  
Departamento de Estatística  
Pós-graduação em Estatística

# **Modelos Dinâmicos para Dados Temporais sob Distribuição Simétrica Condicional: Estimação e Diagnóstico**

VINÍCIUS QUINTAS SOUTO MAIOR

Recife  
2016

VINÍCIUS QUINTAS SOUTO MAIOR

**Modelos Dinâmicos para Dados Temporais sob  
Distribuição Simétrica Condicional: Estimação e  
Diagnóstico**

Tese apresentada ao Programa de Pós-  
Graduação em Estatística da Universidade  
Federal de Pernambuco, como requisito  
parcial para a obtenção do título de  
Doutor em Estatística.

Orientador:  
Prof. Dr. Francisco José de Azevêdo Cysneiros

Recife  
2016

Catálogo na fonte  
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S728m Souto Maior, Vinícius Quintas.  
Modelos dinâmicos para dados temporais sob distribuição simétrica condicional: estimação e diagnóstico / Vinícius Quintas Souto Maior. – 2016.  
145 f.: il., fig., tab.

Orientador: Francisco José de Azevêdo Cysneiros.  
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2016.  
Inclui referências e apêndices.

1. Estatística matemática. 2. Probabilidades. 3. Séries temporais. I. Cysneiros, Francisco José de Azevêdo (orientador). II. Título.

519.5 CDD (23. ed.) UFPE- MEI 2016-039

VINÍCIUS QUINTAS SOUTO MAIOR

MODELOS DINÂMICOS PARA DADOS TEMPORAIS SOB DISTRIBUIÇÃO  
SIMÉTRICA CONDICIONAL: ESTIMAÇÃO E DIAGNÓSTICO

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Estatística.

Aprovada em: 26 de fevereiro de 2016.

BANCA EXAMINADORA

---

Prof. Francisco José de Azevedo Cysneiros  
UFPE

---

Prof.<sup>a</sup> Audrey Helen Mariz de Aquino Cysneiros  
UFPE

---

Prof. Francisco Cribari Neto  
UFPE

---

Prof. Gilberto Alvarenga Paula  
USP

---

Prof. Manoel Ferreira do Santos Neto  
UFCG



*Dedico este trabalho a minha mãe Patrícia e ao meu pai  
Abílio (in memorian).*

# Agradecimentos

A conclusão desse trabalho foi fruto de muito esforço e dedicação, porém, sem o apoio e a parceria de algumas pessoas, tudo se tornaria mais árduo e penoso.

Agradeço primeiramente à minha mãe Patrícia, que educou a mim, ao meu irmão Danilo e a minha irmã Flávia com maestria e muita sabedoria. Em especial, agradeço ao meu pai Abílio (*in memoriam*), que foi o melhor pai que poderia ter tido. Soube, ao seu modo, impor limites e dispensar amor e cuidados no grau certo. A ele dedico esse trabalho, pois foi e será minha inspiração para a carreira da docência e, por que não dizer, para os “números”.

Agradeço à minha esposa, Karine, pela dedicação e incentivo para dar continuidade a essa trajetória acadêmica. Agradeço pela compreensão que sempre teve, quando precisávamos adiar alguns planos para “depois do doutorado”.

Agradeço ao professor Francisco Cysneiros, que, além de meu orientador, é, sem dúvida, um grande amigo. Obrigado pela disponibilidade e dedicação, não poderia ter tido melhor parceiro na construção desse estudo.

Agradeço à professora Audrey Cysneiros e ao professor Getúlio Amaral que, trouxeram considerações e reflexões importantes na banca de qualificação e que, certamente, foram bastante pertinentes e enriqueceram este trabalho.

Agradeço ao professor Francisco Cribari pelos ensinamentos passados durante alguns momentos do meu percurso enquanto estudante.

Agradeço aos colegas da UFPE que, direta ou indiretamente contribuíram nessa trajetória. O meu abraço em todos e todas.

Agradeço à UFPE, por ter sido a minha “segunda casa” durante 11 anos.

Agradeço à Valéria por ser sempre solícita e estar em prontidão para resolver questões burocráticas que eu sempre teimava em deixar para a última hora.

Agradeço à CAPES, pelo apoio financeiro, o que me possibilitou dedicação exclusiva ao Doutorado.

*A ciência nunca resolve um problema sem criar pelo menos outros dez.*

—GEORGE BERNARD SHAW

## Resumo

Nossa abordagem é direcionada a variáveis aleatórias simétricas observadas ao longo do tempo. Nesse sentido, avaliamos os procedimentos de estimação e discutimos o uso da metodologia de diagnóstico sob o enfoque de influência local para classe de modelos autorregressivos de médias móveis simétrico, SYMARMA. Modelos sazonais também são abordados neste trabalho. A estimação dos parâmetros do modelo SYMARMA é feita através da maximização do logaritmo da função de verossimilhança condicional utilizando o algoritmo escore de Fisher. Apresentamos um estudo de robustez baseado na função de influência para avaliar a qualidade do procedimento de estimação. Além disso, conduzimos um estudo de simulação para avaliar a consistência e normalidade assintótica do estimador de máxima verossimilhança condicional. Derivamos expressões mais simples para as funções escore e a matriz informação de Fisher. Desenvolvemos medidas de diagnóstico sob o enfoque de influência local baseado nas medidas de curvatura de Cook (1986), inclinação de Billor e Loynes (1993) e curvatura de Lesaffre e Verbeke (1998). Derivamos, através de simulações, marcas de referência (limiares) para determinar se uma observação é influente. Aplicações de dados reais foram abordadas neste trabalho.

**Palavras-chave:** Distribuição simétrica. Influência local. Modelos SYMARMA. Outlier. Séries temporais.

# Abstract

Our approach is applied to symmetric random variables on over time. In this sense, we develop estimation procedures and discuss the use of local influence diagnostic methodology to class of the autoregressive and moving average symmetric models, SYMARMA. Sazonal models also are considered. The Fisher scoring algorithm is used to find the estimations of parameters SYMARMA model maximizing the logarithm of the conditional likelihood function. We present an robustness study based on influence function to assess the quality of the estimation procedure and we conduct simulation studies to evaluate the consistency and asymptotic normality of the conditional maximum likelihood estimator. We derive simpler expressions for the score function and Fisher information matrix. In order to assess local influence we develop diagnostic measures based on Cook's curvature (1986), slope of Billor and Loynes (1993) and curvature of Lesaffre and Verbeke (1998). We evaluate benchmarks by simulation to identify influential observations. Application are used to illustrate of the proposed methodology.

**Keywords:** Local influence. Outlier. SYMARMA models. Symmetric distributions. Time series.

## Lista de Figuras

1.1	Série excesso de retorno da <i>Microsoft</i> .	21
1.2	Série inflacionária no Brasil de acordo com IGP-OG.	21
2.1	Funções escore do modelo SYMARMA-normal para $\varphi$ e $\mu$ .	37
2.2	Funções escore do modelo SYMARMA- $t_4$ para $\varphi$ e $\mu$ .	38
2.3	QQ-plot normal dos EMVC para os parâmetros $\beta_0$ , $\phi$ e $\varphi$ do modelo SYMARMA-normal (a) e do modelo SYMARMA- $t_4$ (b), baseados em 10.000 réplicas.	40
2.4	QQ-plot normal dos EMVC para os parâmetros $\beta_0$ , $\phi$ e $\varphi$ do modelo SYMARMA-logística I, baseados em 10.000 réplicas.	41
2.5	Temperatura média mensal, Dubuque, Iowa.	43
2.6	Níveis de dióxido de carbono mensais em Alert, NWT, Canadá.	43
3.1	Série autorregressiva de ordem 1 simulada como $\phi = 0,3$ , perturbada (linha descontínua) com <i>outliers</i> aditivos.	63
3.2	Diagnóstico de influência local na série 1 ( $\phi = 0,3$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA-normal.	76
3.3	Diagnóstico de influência local na série 1 ( $\phi = 0,3$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA- $t_4$ .	77
3.4	Diagnóstico de influência local na série 2 ( $\phi = 0,5$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA-normal.	78
3.5	Diagnóstico de influência local na série 2 ( $\phi = 0,5$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA- $t_4$ .	79
3.6	Diagnóstico de influência local na série 3 ( $\phi = 0,7$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA-normal.	80
3.7	Diagnóstico de influência local na série 3 ( $\phi = 0,7$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA- $t_4$ .	81
3.8	Diagnóstico de influência local na série 4 ( $\phi = 0,9$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA-normal.	82
3.9	Diagnóstico de influência local na série 4 ( $\phi = 0,9$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA- $t_4$ .	83
3.10	Diagnóstico de influência local na série 1 ( $\phi = 0,3$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA-normal.	84
3.11	Diagnóstico de influência local na série 1 ( $\phi = 0,3$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA- $t_4$ .	85

3.12	Diagnóstico de influência local na série 2 ( $\phi = 0,5$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA-normal.	86
3.13	Diagnóstico de influência local na série 2 ( $\phi = 0,5$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA- $t_4$ .	87
3.14	Diagnóstico de influência local na série 3 ( $\phi = 0,7$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA-normal.	88
3.15	Diagnóstico de influência local na série 3 ( $\phi = 0,7$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA- $t_4$ .	89
3.16	Diagnóstico de influência local na série 4 ( $\phi = 0,9$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA-normal.	90
3.17	Diagnóstico de influência local na série 4 ( $\phi = 0,9$ , $\varphi = 1$ e $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA- $t_4$ .	91
4.1	Série excesso de retorno da <i>Microsoft</i> .	94
4.2	Excesso de retorno da S&P500 versus excesso de retorno da <i>Microsoft</i> .	94
4.3	Função de autocorrelação, autocorrelação parcial e boxplot da série excesso de retorno da <i>Microsoft</i> .	95
4.4	Histograma (a) e boxplot (b) dos resíduos obtidos pelo ajuste do modelo SYMARMA-normal	96
4.5	Gráfico normal de probabilidades com envelope para o ajuste dos modelos SYMARMA-normal (a) e SYMARMA- $t_5$ (b).	96
4.6	Função de autocorrelação e autocorrelação parcial dos resíduos obtidos do ajuste do modelo SYMARMA- $t_5$ - série excesso de retorno da <i>Microsoft</i> .	99
4.7	Série original (linha preta) e série ajustada pelo modelo SYMARMA- $t_5$ (linha azul) - série excesso de retorno da <i>Microsoft</i> .	99
4.8	Diagnóstico de influência local na série excesso de retorno da <i>Microsoft</i> sob o esquema de perturbação nos dados para os modelos SYMARMA-normal (lado esquerdo) e SYMARMA- $t_5$ (lado direito).	107
4.9	Diagnóstico de influência local na série excesso de retorno da <i>Microsoft</i> sob o esquema de perturbação no parâmetro de dispersão para os modelos SYMARMA-normal (lado esquerdo) e SYMARMA- $t_5$ (lado direito).	108
4.10	Série inflacionária no Brasil de acordo com IGP-OG.	110
4.11	Função de autocorrelação e autocorrelação parcial da série IGP-OG.	110
4.12	Função de autocorrelação e autocorrelação parcial dos resíduos obtidos do ajuste do modelo SYMARMA- $t_3$ - série IGP-OG.	113
4.13	Série original (linha preta), série ajustada pelo modelo SYMARMA-normal (linha laranja) e série ajustada pelo modelo SYMARMA- $t_3$ (linha azul) - série IGP-OG.	113
4.14	Diagnóstico de influência local na série inflacionária IGP-OG sob o esquema de perturbação nos dados para os modelos SYMARMA-normal (lado esquerdo) e SYMARMA- $t_3$ (lado direito).	120

- 4.15 Diagnóstico de influência local na série inflacionária IGP-OG sob o esquema de perturbação no parâmetro de dispersão para os modelos SYMARMA-normal (lado esquerdo) e SYMARMA- $t_3$  (lado direito).



## Lista de Tabelas

2.1	Valores de $\xi$ para algumas distribuições simétricas.	26
2.2	Variância e covariância marginal para os modelos SYMARMA(1,0), SYMARMA(0,1) e SYMARMA(1,1).	30
2.3	Expressões para $W_g(u)$ e $W'_g(u)$ para algumas distribuições simétricas.	33
2.4	Valores de $d_g$ e $f_g$ para algumas distribuições simétricas.	34
2.5	Medidas descritivas empíricas para os EMVC dos parâmetros $\beta_0 = 20$ , $\phi = 0,6$ e $\varphi = 1$ obtidos de 10.000 modelos SYMARMA simulados.	39
3.1	Configurações das séries simuladas com $r_t \sim N(0, 1)$ .	63
3.2	Introdução de <i>outliers</i> nas séries simuladas.	64
3.3	Estatísticas globais $BS_0$ e $BC_0$ para o modelo SYMARMA-normal sob o esquema de perturbação nos dados.	67
3.4	Estatísticas individuais $BS_1$ , $BC_1^{Cook}$ e $BC_1^{LV}$ para o modelo SYMARMA-normal sob o esquema de perturbação nos dados.	67
3.5	Estatísticas individuais $BS_2$ , $BC_2^{Cook}$ e $BC_2^{LV}$ para o modelo SYMARMA-normal sob o esquema de perturbação nos dados.	68
3.6	Estatísticas globais $BS_0$ e $BC_0$ para o modelo SYMARMA- $t_4$ sob o esquema de perturbação nos dados.	69
3.7	Estatísticas individuais $BS_1$ , $BC_1^{Cook}$ e $BC_1^{LV}$ para o modelo SYMARMA- $t_4$ sob o esquema de perturbação nos dados.	69
3.8	Estatísticas individuais $BS_2$ , $BC_2^{Cook}$ e $BC_2^{LV}$ para o modelo SYMARMA- $t_4$ sob o esquema de perturbação nos dados.	70
3.9	Estatísticas globais $BS_0$ e $BC_0$ para o modelo SYMARMA-normal sob o esquema de perturbação no parâmetro de dispersão.	71
3.10	Estatísticas individuais $BS_1$ , $BC_1^{Cook}$ e $BC_1^{LV}$ para o modelo SYMARMA-normal sob o esquema de perturbação no parâmetro de dispersão.	71
3.11	Estatísticas individuais $BS_2$ , $BC_2^{Cook}$ e $BC_2^{LV}$ para o modelo SYMARMA-normal sob o esquema de perturbação no parâmetro de dispersão.	72
3.12	Estatísticas globais $BS_0$ e $BC_0$ para o modelo SYMARMA- $t_4$ sob o esquema de perturbação no parâmetro de dispersão.	73
3.13	Estatísticas individuais $BS_1$ , $BC_1^{Cook}$ e $BC_1^{LV}$ para o modelo SYMARMA- $t_4$ sob o esquema de perturbação no parâmetro de dispersão.	73
3.14	Estatísticas individuais $BS_2$ , $BC_2^{Cook}$ e $BC_2^{LV}$ para o modelo SYMARMA- $t_4$ sob o esquema de perturbação no parâmetro de dispersão.	74

3.15	Estatísticas globais $O_s$ e $O_c$ e seus respectivos limiares $BS_0$ e $BC_0$ para o ajuste do modelo SYMARMA-normal sob o esquema de perturbação nos dados. ( $n = 100$ )	75
3.16	Estatísticas globais $O_s$ e $O_c$ e seus respectivos limiares $BS_0$ e $BC_0$ para o ajuste do modelo SYMARMA-normal sob o esquema de perturbação no parâmetro de dispersão. ( $n = 100$ )	75
3.17	Estatísticas globais $O_s$ e $O_c$ e seus respectivos limiares $BS_0$ e $BC_0$ para o ajuste do modelo SYMARMA- $t_4$ sob o esquema de perturbação nos dados. ( $n = 100$ )	75
3.18	Estatísticas globais $O_s$ e $O_c$ e seus respectivos limiares $BS_0$ e $BC_0$ para o ajuste do modelo SYMARMA- $t_4$ sob o esquema de perturbação no parâmetro de dispersão. ( $n = 100$ )	75
4.1	Valores para algumas medidas de comparação obtidas do ajuste de modelos SYMARMA- $t$ para a série excesso de retorno da <i>Microsoft</i> .	97
4.2	Valores para algumas medidas de comparação obtidas do ajuste de modelos SYMARMA-EP para a série excesso de retorno da <i>Microsoft</i> .	97
4.3	Estimativas e erros-padrão assintóticos (em parênteses) das estimativas dos parâmetros dos três modelos ajustados e algumas medias de comparação - série excesso de retorno da <i>Microsoft</i> .	98
4.4	Previsões 12 passos à frente, RMSE e limites do intervalo com 95% de confiança via bootstrap para o ajuste dos modelos SYMARMA-normal e SYMARMA- $t_5$ .	101
4.5	Limiares para o modelo SYMARMA-normal calculados para série excesso de retorno da <i>Microsoft</i> .	103
4.6	Limiares para o modelo SYMARMA- $t_5$ calculados para série excesso de retorno da <i>Microsoft</i> .	104
4.7	Diagnóstico de influência local para a série excesso de retorno da <i>Microsoft</i> considerando o modelo SYMARMA-normal. $r_t$ são os resíduos.	105
4.8	Diagnóstico de influência local para a série excesso de retorno da <i>Microsoft</i> considerando o modelo SYMARMA- $t_5$ . $r_t$ são os resíduos.	106
4.9	Modelos SYMARMA-normal ajustados a série inflacionária no Brasil, para diferentes configurações, e seus respectivos valores do critério BIC e o logaritmo da função de verossimilhança.	111
4.10	Valores para algumas medidas de comparação obtidas do ajuste de modelos SYMARMA- $t$ para a série IGP-OG.	112
4.11	Valores para algumas medidas de comparação obtidas do ajuste de modelos SYMARMA-EP para a série IGP-OG.	112
4.12	Estimativas e erros-padrão assintóticos (em parênteses) das estimativas dos parâmetros dos três modelos ajustados e algumas medias de comparação - série IGP-OG.	114
4.13	Limiares para o modelo SYMARMA-normal calculados para série inflacionária IGP-OG.	116
4.14	Limiares para o modelo SYMARMA- $t_3$ calculados para série inflacionária IGP-OG.	117

4.15	Diagnóstico de influência local para a série inflacionária IGP-OG considerando o modelo SYMARMA-normal. $r_t$ são os resíduos.	118
4.16	Diagnóstico de influência local para a série inflacionária IGP-OG considerando o modelo SYMARMA- $t_3$ . $r_t$ são os resíduos.	119

# Sumário

<b>1</b>	<b>Introdução</b>	<b>17</b>
1.1	Motivação	20
1.2	Apresentação dos capítulos	22
<b>2</b>	<b>Modelo Autorregressivo de Médias Móveis Simétrico</b>	<b>23</b>
2.1	Introdução	23
2.2	Modelo autorregressivo de médias móveis simétrico	25
2.3	Média, variância e covariância marginal	27
2.3.1	Casos particulares	29
2.4	Estimação dos parâmetros	30
2.5	Inferência estatística	34
2.6	Robustez, consistência e normalidade assintótica	36
2.6.1	Função de influência	36
2.6.2	Consistência em erro quadrático médio e <i>QQ-plot</i>	37
2.7	Previsões	39
2.8	Componente sazonal	42
2.8.1	Modelo SYMARMA sazonal	45
<b>3</b>	<b>Influência Local em Modelos Autorregressivos de Médias Móveis Simétricos</b>	<b>48</b>
3.1	Introdução	48
3.2	Medidas de influência local	50
3.2.1	Curvatura de Cook	50
3.2.2	Curvatura de Lesaffre e Verbeke	52
3.2.3	Inclinação de Billor e Loynes	52
3.3	Metodologia de limiares	54
3.3.1	Limiares para as medidas de influência global	55
3.3.2	Limiares para as medidas de influência individual	55
3.4	Esquemas de perturbação	56
3.4.1	Perturbação aditiva nos dados	57
3.4.2	Perturbação no parâmetro de dispersão	59
3.4.3	Perturbação no vetor de parâmetros autorregressivos	59
3.4.4	Perturbação no vetor de parâmetros de médias móveis	60
3.5	Simulações	61

<b>4</b>	<b>Aplicações a Dados Reais</b>	<b>92</b>
4.1	Introdução	92
4.2	Excesso de retorno da <i>Microsoft</i>	93
4.3	Índice Geral de Preços - Oferta Global	109
<b>5</b>	<b>Rotinas <i>elliptical.ts</i> e <i>influence.ts</i></b>	<b>122</b>
<b>6</b>	<b>Conclusões e Considerações Finais</b>	<b>130</b>
	<b>Referências</b>	<b>133</b>
	<b>Apêndice A Prova do Teorema 1</b>	<b>136</b>
	<b>Apêndice B Prova do Teorema 2</b>	<b>137</b>
	<b>Apêndice C Prova do Teorema 3</b>	<b>138</b>
	<b>Apêndice D Matriz de informação condicional esperada de Fisher</b>	<b>140</b>
D.1	Resultados gerais	141
D.2	Elementos da matriz $\mathbf{K}_{\delta\delta}$	141
D.3	Elementos da matriz $\mathbf{K}_{\varphi\varphi}$	142
D.4	Elementos da matriz $\mathbf{K}_{\delta\varphi}$	142
	<b>Apêndice E Matriz de informação condicional observada de Fisher</b>	<b>143</b>
E.1	Resultados gerais	143
E.2	Derivadas de segunda ordem entre elementos do vetor $\delta$	144
E.3	Derivadas de segunda ordem entre elementos do vetor $\delta$ e $\varphi$	144
E.4	Derivadas de segunda ordem do parâmetro $\varphi$	145

## Capítulo 1

### Introdução

A suposição de normalidade para os erros em uma análise de regressão tem sido constantemente utilizada na descrição e investigação de fenômenos considerados aleatórios. No entanto, em muitas situações, a suposição de normalidade dos erros não é satisfeita e, além disso, muitas vezes nos deparamos com acontecimentos inesperados ou incontroláveis que dão origem a observações atípicas que, de alguma forma, são inconsistentes com o restante da série. Tais observações podem comprometer os procedimentos inferenciais de métodos estatísticos mais convencionais, que assumem a distribuição normal para os erros, tal como a classe de modelos autorregressivos de médias móveis (ARMA), podendo fornecer estimativas bastante viesadas para os parâmetros do modelo ou, até mesmo, uma identificação incorreta do mesmo. Após o ajuste do modelo, técnicas de diagnóstico devem ser utilizadas para checar se as suposições feitas para o modelo são verdadeiras. Dentre as técnicas de diagnóstico utilizadas para modelos de séries temporais, a medida de influência local tem um papel importante por não ser baseada na exclusão de observações do conjunto de dados que, a princípio, são dependentes.

Alguns trabalhos desenvolvidos ao longo das últimas décadas visam alternativas a suposição de normalidade. O estudo pioneiro para séries não-gaussianas foi desenvolvido por Heyde e Feigin (1975), que definiram a família exponencial condicional de distribuições e usaram um modelo simples que tinha um único termo autorregressivo. Cox (1981) estudou a autocorrelação dos dados, caracterizando duas classes de modelos dependentes do tempo: modelos

condicionados a observações passadas e modelos baseados em processos latentes. Na linha dos modelos condicionados a observações passadas, notamos que Zeger e Qaqish (1988) desenvolveram um trabalho que aborda os modelos de Poisson e gama condicionais autorregressivos, adicionando aos modelos de Heyde e Feigin (1975) termos autorregressivos e incluindo o passado e o presente de covariáveis, Li (1991) desenvolveu testes de qualidade do ajuste para tais modelos e, em seguida, Li (1994) apresenta uma estrutura de médias móveis aos modelos de Zeger e Qaqish (1988). Chen e Liu (1993) desenvolveram um procedimento para detecção de *outliers* na classe de modelos autorregressivos integrados de médias móveis (ARIMA). Extensões da classe de modelos ARMA podem ser encontradas em Shephard (1995) e Benjamin *et al.* (2003) que consideram a distribuição dos dados, condicional ao passado do processo, pertencente à família exponencial. Na mesma direção Rocha e Cribari-Neto (2009) desenvolveram modelos dinâmicos para variáveis aleatórias na família de distribuição beta ( $\beta$ ARMA). Cordeiro e Andrade (2009) incorporam a ideia de variável resposta transformada ao modelo GARMA introduzido por Benjamin *et al.* (2003). Mais recentemente, Maior (2012) desenvolve uma nova classe de modelos para dados temporais pertencentes a classe simétrica de distribuições e Creal *et al.* (2013) propõem a classe de modelos GAS (*Generalized Autoregressive Score*).

Em se tratando da análise de diagnóstico, uma das ferramentas mais úteis para dados temporais é a análise de influência local, proposta inicialmente por Cook (1986). A proposta de Cook (1986) é baseada em uma medida de afastamento da função de verossimilhança. Na literatura, através da Curvatura de Cook, podemos citar os trabalhos de Lesaffre e Verbeke (1998) que utilizaram o enfoque de influência local para avaliar modelos lineares mistos, Liu (2002) que utilizou o método de influência local para modelos de regressão linear elípticos multivariados considerando perturbações de casos ponderados na variável explicativa e na variável resposta e, mais recentemente, Paula *et al.* (2009) que consideraram os erros elípticos autorregressivos de primeira ordem e Paula e Cysneiros (2009) que propuseram a Curvatura de Cook como método de diagnóstico de influência local para modelos CAPM (*Capital Asset Pricing Model*) sob a

suposição de distribuições simétricas.

Billor e Loynes (1993) apontam algumas dificuldades práticas e teóricas que surgem no enfoque de influência local introduzido por Cook (1986), entre elas, a falta de invariância da curvatura sobre reparametrizações do esquema de perturbação. Dessa forma, propõem utilizar a máxima inclinação baseada no afatamento da função de verossimilhança modificada como medida de influência local e aplicam a sua metodologia para avaliar modelos de regressão linear, considerando o esquema de perturbação na matriz de variâncias. Zhang e King (2005) simularam a distribuição da medida de influência local proposta por Billor e Loynes (1993) em modelos heterocedásticos autorregressivos generalizados (GARCH) com erros gaussianos, usando a primeira derivada (inclinação de Billor e Loynes) e a segunda derivada do gráfico de influência modificado. Esse procedimento foi denominado de limiares (*benchmarks*). O trabalho de Zhang e King (2005) foi o pioneiro na aplicação de marcas de referência que determinam, estatisticamente, que observações da série podem ser caracterizadas como influentes, na maioria das vezes essa caracterização é feita por meio de uma simples análise exploratória. Scall e Dunne (1991) apresentaram um estudo para investigar a influência local em modelos de regressão com erros ARMA, utilizando diversos esquemas de perturbação. Finalmente, Zevallos e Hotta (2012) utilizaram a metodologia dos limiares para análise de influência local em modelos GARCH considerando, alternativamente, a curvatura de Cook ao invés da segunda derivada do gráfico de influência modificado, obtendo uma grande vantagem de ordem computacional.

Os objetivos desse trabalho são desenvolver modelos dinâmicos para dados temporais sob distribuição simétrica condicional e propor métodos de diagnóstico sob o enfoque de influência local. Ampliar o leque de opções para distribuição condicional da série temporal, considerando, por exemplo, a distribuição  $t$ -Student para os dados, tem-se mostrado uma boa alternativa no ajuste de séries temporais na presença de observações atípicas. Neste trabalho obtemos expressões mais simplificadas para os estimadores dos parâmetros do modelo autorregressivo de médias móveis simétrico, SYMARMA, e, pela primeira vez, avaliamos a convergência assintó-



tica e a robustez qualitativa dos mesmos. Além disso, apresentamos propriedades relacionadas à média, variância e covariância marginal. Para cada modelo proposto desenvolvemos análises de diagnóstico baseadas nas medidas de influência local desenvolvidas por Cook (1986), Billor e Loynes (1993) e Lesaffre e Verbeke (1998). Dentre os esquemas de perturbação considerados temos a perturbação aditiva na variável resposta e a perturbação no parâmetro de dispersão. Aplicamos a metodologia de limiares, proposta por Zhang e King (2005), para classificação de observações influentes. Para ilustrar a metodologia desenvolvida apresentamos aplicações com dados reais e estudos de simulação.

## 1.1 Motivação

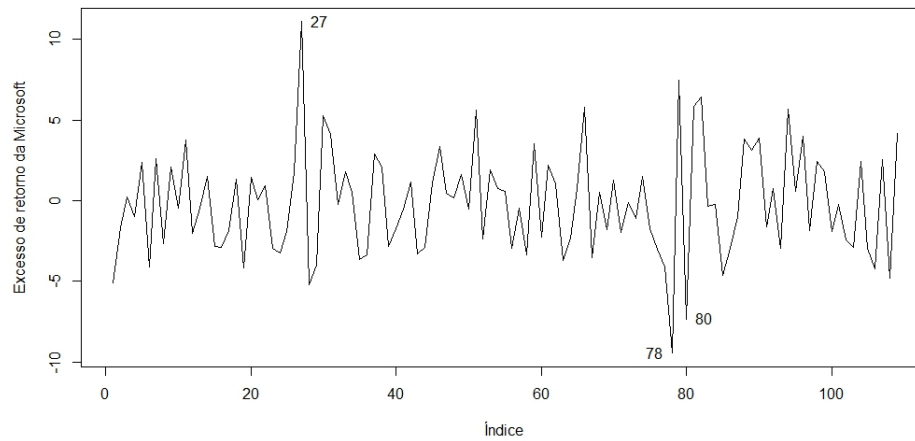
Frequentemente, em séries temporais, são encontradas observações que surgem como atípicas face às restantes em consequência de erros de medição, influência de intervenções exógenas, alterações inesperadas em certas condições de um sistema físico, entre outras. Essas observações atípicas são comumente denominadas de *outliers* (FOX, 1972).

Dentre os autores que discutem a influência de *outliers* na estimação e inferência dos parâmetros em modelos estatísticos podemos citar Martin e Yohai (1985), Ota (1996) e Maior (2012).

Para ilustrar a presença de observações atípicas em séries temporais, apresentamos, nas Figuras 1.1 e 1.2, duas séries com essas características. A primeira série corresponde ao excesso de retorno nos preços de fechamento diário da *Microsoft*, abrangendo o período entre 1 de abril e 5 de setembro de 2002. Notamos que a série possui três pontos atípicos correspondentes as observações  $y_{27} = 11,10\%$ ,  $y_{78} = -9,47\%$  e  $y_{80} = -7,34\%$ .

A segunda série refere-se à inflação no Brasil avaliada anualmente pelo Índice Geral de Preços - Oferta Global (IGP-OG) durante o período de 1970 e 2014. Diversas observações atípicas, correspondentes à época de hiperinflação no Brasil, estão presentes nesta série, entre elas, podemos destacar os anos de 1986 (plano Cruzado), 1987 (plano Bresser), 1989 (plano

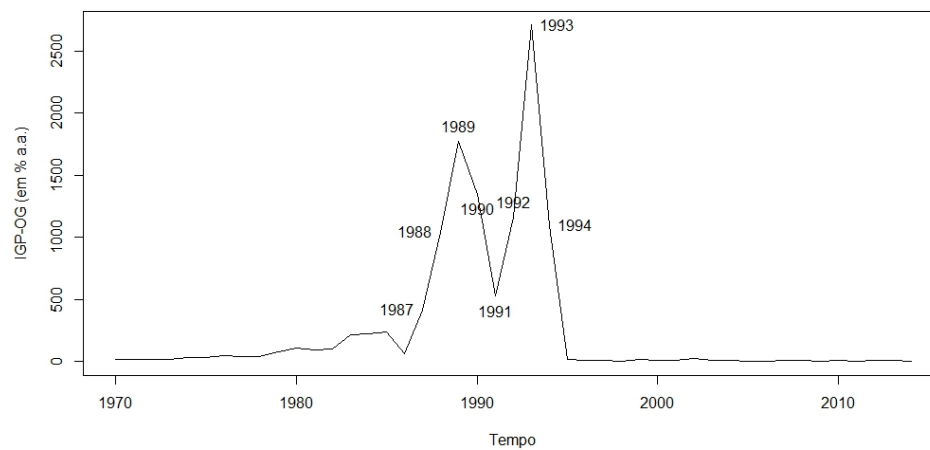
**Figura 1.1** Série excesso de retorno da *Microsoft*.



Fonte: Autoria própria.

Verão), 1990 (plano Collor I), 1991 (plano Collor II) e 1994 (plano Real).

**Figura 1.2** Série inflacionária no Brasil de acordo com IGP-OG.



Fonte: Autoria própria.

Estudaremos essas séries em mais detalhes no decorrer desse trabalho e, como veremos, considerar modelos com distribuições mais flexíveis pode ser bastante útil para modelagem de séries na presença de *outliers*.

## 1.2 Apresentação dos capítulos

No Capítulo 2, definimos o modelo SYMARMA e apresentamos algumas propriedades relacionadas à média, variância e covariância marginal. Tratamos do problema de estimação dos parâmetros fornecendo expressões para o vetor escore e para a matriz informação esperada de Fisher condicional. Com o objetivo de observar o comportamento dos estimadores, realizamos alguns estudos de simulação e apresentamos resultados relacionados a robustez, obtidos através da função de influência. Discutimos como realizar testes de hipóteses para os parâmetros do modelo e previsões. Finalizamos o capítulo com uma explicação sobre modelos SYMARMA sazonais.

No Capítulo 3, apresentamos as medidas de influência local que serão aplicadas à classe de modelos SYMARMA e, em seguida, descrevemos detalhadamente a metodologia de limiares proposta por Zhang e King (2005) para identificar observações influentes baseada nas distribuições estimadas das estatísticas de Inclinação de Billor e Loynes (1993), Curvatura de Cook (1986) e Curvatura de Lesaffre e Verbeke (1998). Consideramos os esquemas de perturbação aditiva nos dados, perturbação no parâmetro de dispersão, perturbação no vetor de parâmetros autorregressivos e perturbação no vetor de parâmetros de médias móveis. Ilustramos a metodologia proposta através de estudos de simulação e tecemos comentários a respeito da aplicação dessa metodologia nos modelos SYMARMA sazonais.

No Capítulo 4, apresentamos a análise de dois conjuntos de dados reais.

No Capítulo 5, evidenciamos detalhes da sintaxe de comandos das rotinas *elliptical.ts* e *influence.ts*, desenvolvidas pelo autor na plataforma R para modelagem e diagnóstico de influência local de dados de séries temporais sob distribuição simétrica condicional.

No Capítulo 6, apresentamos as contribuições, algumas conclusões e futuras linhas de pesquisa.

Por fim, apresentamos alguns detalhes técnicos nos Apêndices.

## Capítulo 2

### Modelo Autorregressivo de Médias Móveis Simétrico

#### 2.1 Introdução

Séries temporais evoluem em torno de uma média que, geralmente, apresenta algum tipo de tendência totalmente arbitrária em função do tempo. Essas tendências, quando determinísticas, podem ser ajustadas por modelos bastantes simples como, por exemplo,

$$Y_t = \mu_t + r_t, \quad (1)$$

em que  $\mu_t$  é uma função determinística que pode ser constante, linear, quadrática, sazonal, entre outras formas, e  $r_t$  é uma variável aleatória independente e identicamente distribuída (i.i.d.) com média zero, para todo  $t$ , ou seja, um ruído branco. Esses modelos determinísticos levam em consideração que a tendência da série é a mesma em qualquer período considerado. Porém, em diversas aplicações, particularmente em economia e negócios, notamos que a tendência varia de forma aleatória ao longo do tempo. Nesses casos, considerar uma função estocástica para média tem sido uma boa alternativa. Tais modelos são chamados de modelos estocásticos.

Os modelos estocásticos, em sua maioria, consideram que a tendência estocástica da média pode ser representada por uma dependência linear entre as observações passadas e/ou uma combinação linear de impactos aleatórios. A classe de modelos autorregressivos de médias mó-

veis, denotada por ARMA, é, sem dúvida, a mais utilizada na modelagem de séries temporais estacionárias com tendência estocástica, sendo a sua estrutura definida como

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j r_{t-j} + r_t,$$

em que  $\{r_t\}$  representa uma série de variáveis aleatórias i.i.d. com média zero e as constantes  $p$  e  $q$  são as ordens dos polinômios autorregressivos ( $\phi$ 's) e de médias móveis ( $\theta$ 's), respectivamente. Geralmente consideramos a distribuição normal de probabilidade para as variáveis aleatórias  $r_t$ 's.

Para séries não estacionárias integradas, uma alternativa é induzir estacionariedade através de transformações nos dados como, por exemplo, o uso da escala logarítmica, o uso da transformação de Box-Cox (BOX; COX, 1964), o uso da série diferenciada, entre outras.

A proposta da classe de modelos autorregressivos de médias móveis simétricos, denotada por SYMARMA, é ampliar o leque de opções para a distribuição das variáveis aleatórias  $r_t$ 's nos modelos (1), que passam a pertencer a classe de distribuições simétricas. Dessa forma, temos que, dado o conjunto de informações passadas  $H_{t-1} = \{y_{t-1}, \dots, y_1, \mu_{t-1}, \dots, \mu_1\}$ , a distribuição condicional de  $Y_t|H_{t-1}$  também pertence a classe simétrica na qual encontramos distribuições com caudas mais pesadas do que as da normal como, por exemplo, as distribuições  $t$ -Student e a logística II. Distribuições com caudas pesadas tendem a acomodar melhor observações atípicas. Diante disso, a classe de modelos SYMARMA pode ser vista como uma alternativa para modelagem de séries temporais na presença de *outliers*.

Neste capítulo introduzimos o modelo SYMARMA juntamente com algumas propriedades relacionadas a média, variância e covariância marginal. Sugerimos um método para estimação dos parâmetros e apresentamos resultados relacionados a robustez, obtidos através da função de influência. Finalizamos o capítulo com alguns estudos de simulação e a extensão dos modelos SYMARMA para dados sazonais.

## 2.2 Modelo autorregressivo de médias móveis simétrico

Suponhamos que os ruídos brancos  $\{r_t\}$  em (1), para  $t = 1, \dots, n$ , são variáveis aleatórias simétricas contínuas e, conseqüentemente, a distribuição condicional de cada  $Y_t$ , dado o conjunto de informações passadas  $H_{t-1} = \{y_{t-1}, \dots, y_1, \mu_{t-1}, \dots, \mu_1\}$ , segue uma distribuição simétrica contínua com função de densidade de probabilidade definida como

$$f(y_t|H_{t-1}, \mu_t, \varphi) = \frac{1}{\sqrt{\varphi}} g\left\{\frac{(y_t - \mu_t)^2}{\varphi}\right\}, \quad y_t \in \mathbb{R}, \mu_t \in \mathbb{R} \text{ e } \varphi > 0, \quad (2)$$

em que  $\mu_t$  é o parâmetro de locação,  $\varphi$  é o parâmetro de dispersão e  $g(\cdot)$  é a função geradora de densidades, com  $g(u) > 0$ , para  $u > 0$ , sendo tal que  $\int_0^\infty u^{-\frac{1}{2}} g(u) du = 1$ . Essa condição é necessária para que  $f(y_t|H_{t-1}, \mu_t, \varphi)$  seja uma função de densidade de probabilidade. Em particular, temos que:  $g(u) = (2\pi)^{-1} \exp\{-u/2\}$  para a distribuição normal;  $g(u) = \frac{\nu^{\nu/2}}{B(1/2, \nu/2)} (\nu + u)^{-\frac{\nu+1}{2}}$  para a distribuição  $t$ -Student com  $\nu$  graus de liberdade, em que  $B(\cdot, \cdot)$  é a função beta; e  $g(u) = c \frac{e^{-u}}{(1+e^{-u})^2}$  para a distribuição logística I, em que  $c \approx 1,484200029$  é a constante normalizadora obtida da relação  $\int_0^\infty u^{-\frac{1}{2}} g(u) du = 1$ .

Utilizamos a notação  $Y_t|H_{t-1} \sim S(\mu_t, \varphi, g)$  para indicar que a variável  $Y_t$ , condicional ao conjunto de observações passadas,  $H_{t-1}$ , segue distribuição simétrica com parâmetro de locação  $\mu_t$ , parâmetro de dispersão  $\varphi$  e função geradora de densidades  $g$ .

Se  $Y_t|H_{t-1} \sim S(\mu_t, \varphi, g)$ , então a função característica de  $Y_t$ , definida por  $\zeta(t) = E(\exp(itY))$ , fica expressa como  $\zeta(t) = \exp(it\mu_t)h(t^2\varphi)$ ,  $t \in \mathbb{R}$ , para alguma função  $h$  que depende da distribuição simétrica considerada, com  $h(u) \in \mathbb{R}$  para todo  $u > 0$ . Quando existem, a média e a variância de  $Y_t|H_{t-1}$ , são expressas, respectivamente, por

$$E(Y_t|H_{t-1}) = \mu_t \quad \text{e} \quad \text{Var}(Y_t|H_{t-1}) = \xi \varphi,$$

em que  $\xi > 0$  é uma constante igual a  $\xi = -2h'(0)$ , com  $h'(0) = \partial h(u)/\partial u|_{u=0}$ . Kelker (1970) observa que se  $u^{-\frac{1}{2}(k+1)}f(u)$  for integrável, então o  $k$ -ésimo momento de  $Y_t|H_{t-1}$  existe. Dis-

tribuições pertencentes à classe simétrica que não possuem os dois primeiros momentos finitos não são consideradas nesse trabalho como, por exemplo, a distribuição Cauchy. Alguns valores de  $\xi$  são apresentados para algumas distribuições simétricas na Tabela 2.1.

**Tabela 2.1** Valores de  $\xi$  para algumas distribuições simétricas.

Distribuição	$\xi$
Normal	1
$t$ -Student	$\frac{v}{v-2}, \quad v > 2$
$t$ -Student generalizada	$\frac{s}{r-2}, \quad s > 0, r > 2$
Logística-I	0,79569
Logística-II	$\frac{\pi^2}{3}$
Logística generalizada	$2\psi'(m)$
Exponencial potência	$2^{(1+k)} \frac{\Gamma\{3(k+1)/2\}}{\Gamma\{(k+1)/2\}}$

Fonte: Cysneiros (2004)

Diferentemente da classe de modelos simétricos de regressão, em que  $\mu_t = \mathbf{x}_t^\top \beta$ , temos na especificação da média dos modelos SYMARMA uma componente dinâmica adicional,  $\tau_t$ , com termos autorregressivos e de médias móveis incluídos aditivamente. Definimos  $\mu_t$  como

$$\mu_t = \mathbf{x}_t^\top \beta + \tau_t, \quad (3)$$

em que  $\mathbf{x}_t^\top = (1, x_{t1}, x_{t2}, \dots, x_{tk})$  é um vetor que contém os valores das  $k$  variáveis explicativas,  $k < n$ , sendo  $n$  o tamanho da amostra,  $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^\top$  é um vetor de parâmetros desconhecidos e  $\tau_t$  é uma componente ARMA expressa por

$$\tau_t = \sum_{i=1}^p \phi_i \left\{ y_{t-i} - \mathbf{x}_{t-i}^\top \beta \right\} + \sum_{j=1}^q \theta_j r_{t-j}, \quad (4)$$

sendo  $\phi = (\phi_1, \phi_2, \dots, \phi_p)^\top$  e  $\theta = (\theta_1, \theta_2, \dots, \theta_q)^\top$  os vetores de parâmetros autorregressivos e de médias móveis, respectivamente. Os termos de médias móveis,  $r_t$ 's, são ruídos brancos que

podem ser, por exemplo, erros de medida na escala original (isto é,  $y_t - \mu_t$ ), ou, resíduos de Pearson, entre outros.

A motivação da definição da componente ARMA,  $\tau_t$ , em (3), consiste inicialmente em considerar  $\omega_t = y_t - \mathbf{x}_t^\top \beta$  com sendo um processo ARMA( $p, q$ ), isto é,

$$\omega_t = \sum_{i=1}^p \phi_i \omega_{t-i} + \sum_{j=1}^q \theta_j r_{t-j} + r_t. \quad (5)$$

Assumindo que  $E(r_t | H_{t-1}) = 0$  e tomando a esperança condicional a  $H_{t-1}$  para ambos os lados de (5), obtemos a expressão (4).

Portanto,

$$\mu_t = \mathbf{x}_t^\top \beta + \sum_{i=1}^p \phi_i \left\{ y_{t-i} - \mathbf{x}_{t-i}^\top \beta \right\} + \sum_{j=1}^q \theta_j r_{t-j}. \quad (6)$$

O modelo SYMARMA é definido pela componente aleatória dada em (2) e pela componente dinâmica dada em (6). Os modelos ARMA e ARMAX são casos particulares do modelo SYMARMA sob distribuição condicional normal, SYMARMA-normal, quando consideramos, respectivamente, a ausência ou a presença de variáveis explicativas. Os parâmetros do modelo podem ser selecionados utilizando, por exemplo, o critério de informação de Akaike, que em inglês é designado pela sigla AIC - *Akaike Information Criterion*, introduzido por Akaike (1973, 1974) ou, alternativamente, o critério de informação bayesiano, que em inglês é designado pela sigla BIC - *Bayesian Information Criterion*, de Schwarz (1978).

## 2.3 Média, variância e covariância marginal

Nessa seção denotamos por  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  o polinômio autorregressivo, por  $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  o polinômio de médias móveis e  $B^k y_t = y_{t-k}$  o operador de defasagens.



**Teorema 1.** *A média marginal de  $Y_t$  nos modelos SYMARMA é expressa por*

$$E(Y_t) = \mathbf{x}_t^\top \boldsymbol{\beta},$$

*desde que  $\Phi(B)$  seja invertível.*

**Teorema 2.** *A variância marginal de  $Y_t$  nos modelos SYMARMA é expressa por*

$$\text{Var}(Y_t) = \xi \varphi \sum_{i=0}^{\infty} \psi_i^2,$$

*em que  $\psi_i$  é obtido da divisão de  $\Psi(B) = \Theta(B)\Phi(B)^{-1} = \psi_0 B^0 + \psi_1 B^1 + \psi_2 B^2 + \dots$  sendo  $\psi_0 = 1$ , assumindo que  $\Phi(B)$  é invertível.*

**Teorema 3.** *A covariância e a correlação marginal entre  $Y_t$  e  $Y_{t-k}$  nos modelos SYMARMA são, respectivamente, expressas por*

$$\text{Cov}(Y_t, Y_{t-k}) = \xi \varphi \sum_{i=0}^{\infty} \psi_i \psi_{i+k} \quad \text{e} \quad \text{Corr}(Y_t, Y_{t-k}) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}.$$

De acordo com o Teorema 1, temos que a esperança marginal de  $Y_t$  nos modelos SYMARMA é estacionária desde que  $\Phi(B)$  seja invertível e  $\mathbf{x}_t^\top \boldsymbol{\beta} = c$  para todo  $t$ , sendo  $c$  uma constante. Enquanto que, pelo Teorema 2, a condição de estacionariedade para variância marginal de  $Y_t$  nos modelos SYMARMA é que  $\Phi(B)$  seja invertível. As provas dos Teoremas 1-3 são apresentadas nos Apêndice A, B e C.

### 2.3.1 Casos particulares

O modelo SYMARMA(1,1), definido como

$$Y_t = \phi Y_{t-1} + \theta r_{t-1} + r_t, \quad (7)$$

pode ser reescrito na forma  $Y_t = \Psi(B)r_t$ , em que  $\Psi(B) = \Theta(B)\Phi(B)^{-1} = 1 + \psi_1 B^1 + \psi_2 B^2 + \dots$

Utilizando operações recursivas a partir (7) é possível mostrar que o modelo SYMARMA(1,1) pode ser expresso por meio de um processo linear geral dado por

$$Y_t = r_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} r_{t-j}.$$

Diante disso, temos que

$$\psi_j = (\phi + \theta)\phi^{j-1} \quad \text{para } j \geq 1. \quad (8)$$

A Tabela 2.2 apresenta as expressões da variância e da covariância marginal de  $Y_t$  para os modelos SYMARMA(1,0), SYMARMA(0,1) e SYMARMA(1,1), obtidas pelos Teoremas 2 e 3 e pela equação (8). Além disso, podemos mostrar que as funções de correlação para esses modelos são expressas por:

Modelo SYMARMA(1,0):

$$\text{Corr}(Y_t, Y_{t-k}) = \phi^k \quad \text{para } k \geq 1.$$

Modelo SYMARMA(0,1):

$$\text{Corr}(Y_t, Y_{t-k}) = \begin{cases} 0 & \text{para } k = 1 \\ -\frac{\theta}{1+\theta^2} & \text{para } k \geq 2 \end{cases}.$$

Modelo SYMARMA(1,1):

$$\text{Corr}(Y_t, Y_{t-k}) = \frac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2} \phi^{k-1} \quad \text{para } k \geq 1.$$

**Tabela 2.2** Variância e covariância marginal para os modelos SYMARMA(1,0), SYMARMA(0,1) e SYMARMA(1,1).

Modelo	Var( $Y_t$ )	Cov( $Y_t, Y_{t-k}$ )	
SYMARMA(1,0)	$\frac{\xi\varphi}{1-\phi^2}$	$\frac{\xi\varphi}{1-\phi^2} \phi^k$	para $k = 1, 2, 3, \dots$
SYMARMA(0,1)	$\xi\varphi(1 + \theta^2)$	$-\xi\varphi\theta$ 0	para $k = 1$ para $k = 2, 3, 4, \dots$
SYMARMA(1,1)	$\xi\varphi \frac{1+2\phi\theta+\theta^2}{(1-\phi^2)}$	$\xi\varphi \frac{(\phi+\theta)(1+\phi\theta)}{(1-\phi^2)} \phi^{k-1}$	para $k = 1, 2, 3, \dots$

## 2.4 Estimação dos parâmetros

Nessa seção tratamos do problema de estimar os parâmetros do modelo SYMARMA com base na série histórica observada  $y_1, y_2, \dots, y_n$ . Assumimos que um modelo já foi especificado, isto é, que já foram especificados os valores para  $p$  e  $q$ . A proposta é utilizar o método de máxima verossimilhança condicional às  $m$  primeiras observações, em que  $m = \max\{p, q\}$ . Para o modelo SYMARMA, a função de verossimilhança condicional é uma função dos parâmetros  $\beta$ 's,  $\phi$ 's,  $\theta$ 's e  $\varphi$ , expressa por

$$L(\delta, \varphi) = \prod_{t=m+1}^n f(y_t | \mathbf{H}_{t-1}, \mu_t, \varphi),$$

em que  $\delta = (\beta^\top, \phi^\top, \theta^\top)^\top$  e  $f(y_t | H_{t-1}, \mu_t, \varphi)$  é definida como em (2). Dessa forma, o logaritmo da função de verossimilhança condicional é

$$\ell(\delta, \varphi) = -\frac{(n-m)}{2} \log(\varphi) + \sum_{t=m+1}^n \log(g(u_t)), \quad (9)$$

em que  $u_t = (y_t - \mu_t)^2 / \varphi$  e  $\mu_t$  é definido como em (6).

Assumimos que o logaritmo da função de verossimilhança,  $\ell$ , é regular (COX; HINKLEY, 1974) em relação a  $\delta$  e  $\varphi$  e que é possível derivar  $\ell$  em relação aos parâmetros desconhecidos e calcular alguns momentos dessas derivadas. Algumas distribuições pertencentes à classe simétrica não satisfazem as condições de regularidade e não têm derivadas ou momentos finitos como, por exemplo, as distribuições de Kotz, Kotz generalizada e exponencial dupla, não sendo consideradas nesse trabalho.

Na construção do logaritmo da função verossimilhança condicional, assumimos que os  $q$  primeiros erros são iguais a zero, desde que, condicional a  $H_m$ , os  $m$  primeiros erros são iguais a zero (ou aproximadamente zero). As expressões para o vetor escore ( $\mathbf{U}$ ) e para a matriz de informação de Fisher condicional ( $\mathbf{K}$ ) são obtidas, respectivamente, pelas derivadas de primeira e segunda ordem do logaritmo da função de verossimilhança condicional apresentada em (9). Sendo a função  $g(\cdot)$  contínua e diferenciável, os vetores escore para  $\delta$  e  $\varphi$  são dados, respectivamente, por

$$\begin{aligned} \mathbf{U}_\delta(\delta, \varphi) &= \varphi^{-1} \mathbf{O}^\top \mathbf{D}(\mathbf{y} - \boldsymbol{\mu}) \quad \text{e} \\ \mathbf{U}_\varphi(\delta, \varphi) &= \frac{1}{2\varphi} [-(n-m) + \varphi^{-1} (\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{D}(\mathbf{y} - \boldsymbol{\mu})], \end{aligned}$$

em que  $\mathbf{y} = (y_{m+1}, \dots, y_n)^\top$ ,  $\boldsymbol{\mu} = (\mu_{m+1}, \dots, \mu_n)^\top$ ,  $\mathbf{D} = \text{diag}(v_{m+1}, \dots, v_n)$  com  $v_t = v(u_t) = -2W_g(u_t)$  e  $W_g(u_t) = \frac{d \log(g(u_t))}{du_t}$ . Temos ainda que  $\mathbf{O} = (\mathbf{C}, \mathbf{A}, \mathbf{B})$  é uma matriz  $(n-m) \times (p+q+k+1)$  com elementos dados por

$$\mathbf{A} = \begin{pmatrix} y_m - \mathbf{x}_m^\top \boldsymbol{\beta} & \cdots & y_{m+1-p} - \mathbf{x}_{m+1-p}^\top \boldsymbol{\beta} \\ y_{m+1} - \mathbf{x}_{m+1}^\top \boldsymbol{\beta} & \cdots & y_{m+2-p} - \mathbf{x}_{m+2-p}^\top \boldsymbol{\beta} \\ \vdots & \ddots & \vdots \\ y_{n-1} - \mathbf{x}_{n-1}^\top \boldsymbol{\beta} & \cdots & y_{n-p} - \mathbf{x}_{n-p}^\top \boldsymbol{\beta} \end{pmatrix}_{(n-m) \times p},$$

$$\mathbf{B} = \begin{pmatrix} y_m - \mu_m & \cdots & y_{m+1-q} - \mu_{m+1-q} \\ y_{m+1} - \mu_{m+1} & \cdots & y_{m+2-q} - \mu_{m+2-q} \\ \vdots & \ddots & \vdots \\ y_{n-1} - \mu_{n-1} & \cdots & y_{n-q} - \mu_{n-q} \end{pmatrix}_{(n-m) \times q} \quad \mathbf{e}$$

$$\mathbf{C} = \begin{pmatrix} 1 - \sum_{i=1}^p \phi_i & \cdots & x_{(m+1)k} - \sum_{i=1}^p \phi_i x_{(m+1-i)k} \\ 1 - \sum_{i=1}^p \phi_i & \cdots & x_{(m+2)k} - \sum_{i=1}^p \phi_i x_{(m+2-i)k} \\ \vdots & \ddots & \vdots \\ 1 - \sum_{i=1}^p \phi_i & \cdots & x_{(n)k} - \sum_{i=1}^p \phi_i x_{(n-i)k} \end{pmatrix}_{(n-m) \times (k+1)}.$$

Expressões para  $W_g(u_t)$  e  $W'_g(u_t) = \frac{dW_g(u_t)}{du_t}$  para algumas distribuições simétricas encontram-se na Tabela 2.3.

Os estimadores de máxima verossimilhança condicional (EMVC) para o vetor de parâmetros  $\boldsymbol{\delta}$  e  $\boldsymbol{\phi}$ , dados, respectivamente, por  $\hat{\boldsymbol{\delta}}$  e  $\hat{\boldsymbol{\phi}}$ , são obtidos igualando-se  $\mathbf{U}$  ao vetor de zeros. Exceto para  $\mathbf{U}_\phi$ , o sistema de equações não possui solução explícita e, dessa forma, devemos utilizar um procedimento iterativo para estimar os parâmetros do modelo (NOCEDAL; WRIGHT, 1999). O método score de Fisher tem sido utilizado por diversos autores para estimar os parâmetros dos modelos simétricos de regressão (GALEA et al, 2003; CYSNEIROS; PAULA, 2005). Nesse trabalho também optamos por utilizar o método score de Fisher para estimar os parâmetros do modelo SYMARMA. O processo iterativo é dado por

$$\boldsymbol{\delta}^{(i+1)} = \boldsymbol{\delta}^{(i)} + (\mathbf{K}^{(i)})^{-1} \mathbf{U}^{(i)},$$

**Tabela 2.3** Expressões para  $W_g(u)$  e  $W'_g(u)$  para algumas distribuições simétricas.

Distribuição	$W_g(u)$	$W'_g(u)$
Normal	$-\frac{1}{2}$	0
$t$ -Student	$-\frac{(v+1)}{2(v+u)}$	$\frac{(v+1)}{2(v+u)^2}$
$t$ -Student generalizada	$-\frac{(r+1)}{2(s+u)}$	$\frac{(r+1)}{2(s+u)^2}$
Logística-I	$-\tanh\left(\frac{u}{2}\right)$	$-\operatorname{sech}\left(\frac{u}{2}\right)/2$
Logística-II	$-\frac{\exp(-\sqrt{u})-1}{(-2\sqrt{u})[1+\exp(-\sqrt{u})]}$	$\frac{2\exp(-\sqrt{u})\sqrt{u}+\exp(-2\sqrt{u})-1}{-4u^{3/2}[1+\exp(-\sqrt{u})]^2}$
Logística generalizada	$-\frac{\alpha m[\exp(-\alpha\sqrt{u})-1]}{(-2\sqrt{u})[1+\exp(-\alpha\sqrt{u})]}$	$-\frac{\alpha m}{4} \frac{2\alpha\exp(-\alpha\sqrt{u})\sqrt{u}+\exp(-2\alpha\sqrt{u})-1}{u^{3/2}[1+\exp(-\alpha\sqrt{u})]^2}$
Exponencial potência	$-\frac{1}{2(1+k)u^{k/(k+1)}}$	$\frac{k}{2(1+k)^2u^{(2k+1)/(k+1)}}$

Fonte: Cysneiros (2004)

em que  $\mathbf{K}$  é a matriz de informação condicional esperada de Fisher para  $(\delta, \varphi)$ .

A matriz  $\mathbf{K}$  é bloco-diagonal,

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_{\delta\delta} & 0 \\ 0 & \mathbf{K}_{\varphi\varphi} \end{pmatrix},$$

ou seja, o vetor de parâmetros  $\delta$  e  $\varphi$  são globalmente ortogonais. Conforme detalhado no Apêndice D, as matrizes de informação condicional esperada de Fisher  $\mathbf{K}_{\delta\delta}$  e  $\mathbf{K}_{\varphi\varphi}$  são expressas, respectivamente, por

$$\mathbf{K}_{\delta\delta} = 4d_g\varphi^{-1}\mathbf{O}^\top\mathbf{O} \quad \text{e} \quad \mathbf{K}_{\varphi\varphi} = \frac{n-m}{4\varphi^2}(4f_g-1),$$

em que  $d_g = E[W_g^2(U^2)U^2]$  e  $f_g = E[W_g^2(U^2)U^4]$ , sendo  $U \sim S(0, 1, g)$ . A Tabela 2.4 apresenta os valores de  $d_g$  e  $f_g$  para algumas distribuições simétricas.

As estimativas de máxima verossimilhança condicional para  $\delta$  e  $\varphi$  podem ser obtidas

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

Distribuição	$d_g$	$f_g$
Normal	$\frac{1}{4}$	$\frac{3}{4}$
$t$ -Student	$\frac{(v+1)}{4(v+3)}$	$\frac{3(v+1)}{4(v+3)}$
$t$ -Student generalizada	$\frac{r(r+1)}{4s(r+3)}$	$\frac{3(r+1)}{4(r+3)}$
Logística-I	0,36931	1,00345
Logística-II	$\frac{1}{12}$	0,60749
Logística generalizada	$\frac{\alpha^2 m^2}{4(2m+1)}$	$\frac{2m(2+m^2 \psi'(m))}{4(2m+1)}$
Exponencial potência	$\frac{\Gamma\{(3-k)/2\}}{4(2^{k-1})(1+k)^2 \Gamma\{(k+1)/2\}}$	$\frac{(k+3)}{4(k+1)}$

Fonte: Cysneiros (2004)

resolvendo-se o seguinte sistema de equações:

$$\begin{aligned}\delta^{(i+1)} &= \left( \mathbf{O}^{\top(i)} \mathbf{O}^{(i)} \right)^{-1} \mathbf{O}^{\top(i)} \mathbf{z}_{\delta}^{(i)} \\ \varphi^{(i+1)} &= \frac{1}{n-m} \left( \mathbf{y} - \boldsymbol{\mu}^{(i+1)} \right)^{\top} \mathbf{D}^{(i+1)} \left( \mathbf{y} - \boldsymbol{\mu}^{(i+1)} \right),\end{aligned}$$

em que  $\mathbf{z}_{\delta}$  é um vetor  $(n-m) \times 1$  dado por

$$\mathbf{z}_{\delta} = \mathbf{O}\delta + (4d_g)^{-1} \mathbf{D}(\mathbf{y} - \boldsymbol{\mu}).$$

Os valores iniciais para o vetor de parâmetros  $\delta^{(0)}$  e  $\varphi^{(0)}$  podem ser considerados como sendo as estimativas obtidas através do ajuste do modelo ARMA.

## 2.5 Inferência estatística

Sob as condições de regularidade usuais e para  $n$  suficientemente grande, temos que o EMVC converge em distribuição para uma distribuição normal, isto é,

$$\begin{pmatrix} \hat{\delta} \\ \hat{\varphi} \end{pmatrix} \xrightarrow{D} N_{(k+p+q+2)} \left( \begin{pmatrix} \delta \\ \varphi \end{pmatrix}, \mathbf{K}^{-1} \right), \quad (10)$$

em que  $\hat{\delta}$  e  $\hat{\varphi}$  são os estimadores de máxima verossimilhança condicional de  $\delta$  e  $\varphi$ , respectivamente, e  $N_r$  denota a distribuição normal  $r$ -dimensional. Portanto, a matriz  $\mathbf{K}_{\delta\delta}^{-1}$ , avaliada em  $(\hat{\delta}, \hat{\varphi})$ , é um estimador consistente para a variância assintótica de  $\hat{\delta}$  e a matriz  $\mathbf{K}_{\varphi\varphi}^{-1}$ , avaliada em  $(\hat{\delta}, \hat{\varphi})$ , é um estimador consistente para variância assintótica de  $\hat{\varphi}$ . Dada a normalidade assintótica do EMVC em (10), podemos facilmente construir intervalos de confiança assintóticos para  $\delta$  e  $\varphi$ .

Hipóteses envolvendo o vetor de parâmetros  $\gamma = (\beta^\top, \varphi, \phi^\top, \theta^\top)^\top$  podem ser expressas na forma geral

$$H_0 : \mathbf{Z}\gamma = \mathbf{0} \quad \text{e} \quad H_1 : \mathbf{Z}\gamma \neq \mathbf{0}, \quad (11)$$

em que  $\mathbf{Z}$  é uma matriz  $r \times (k + p + q + 2)$  de posto completo  $r$  ( $r < k + p + q + 2$ ). A hipótese nula pode contemplar situações bastante simples, como, por exemplo, testar  $H_0: \beta = 0$  contra  $H_1: \beta \neq 0$ . A estatística utilizada para testar as hipóteses  $H_0$  e  $H_1$  definidas em (11) é baseada na razão de verossimilhanças condicional dada por

$$\lambda = 2 \{ \ell(\hat{\gamma}) - \ell(\tilde{\gamma}) \},$$

em que  $\ell(\cdot)$  é o logaritmo da função de verossimilhança condicional,  $\tilde{\gamma}$  é o EMVC restrito de  $\gamma$  (sob a hipótese  $H_0$  em (11)) e  $\hat{\gamma}$  é o EMVC irrestrito de  $\gamma$  (sob a hipótese  $H_1$  em (11)).

Sob certas condições de regularidade e sob  $H_0$  temos que  $\lambda \xrightarrow{D} \chi_r^2$ , em que  $\xrightarrow{D}$  denota convergência em distribuição e  $\chi_r^2$  denota a distribuição qui-quadrado com  $r$  graus de liberdade.



## 2.6 Robustez, consistência e normalidade assintótica

Nessa seção avaliamos por meio da função de influência a robustez qualitativa do procedimento de estimação dos parâmetros. Resultados de estudos de simulação para avaliar a consistência e a normalidade assintótica do EMVC também são apresentados.

### 2.6.1 Função de influência

Um procedimento de estimação é considerado qualitativamente robusto quando a função de influência (FI), definida por

$$FI(y, \delta, \varphi, f) = -[E[\mathbf{U}'(\delta, \varphi)]]^{-1} \mathbf{U}(\delta, \varphi), \quad (12)$$

é limitada, em que  $\mathbf{U}$  é a função escore,  $\mathbf{U}'$  é a sua derivada e  $f$  é a função densidade de probabilidade associada ao modelo SYMARMA.

Consideramos a função de influência definida em (12) para os modelos SYMARMA-normal e SYMARMA- $t$ . As funções escores para  $\varphi$  e  $\mu_t$  são dadas, respectivamente, por

$$\mathbf{U}_\varphi = -\frac{1}{2\varphi} - \frac{1}{\varphi} W_g(u_t) u_t \quad \text{e} \quad \mathbf{U}_{\mu_t} = -\frac{2}{\varphi} W_g(u_t) (y_t - \mu).$$

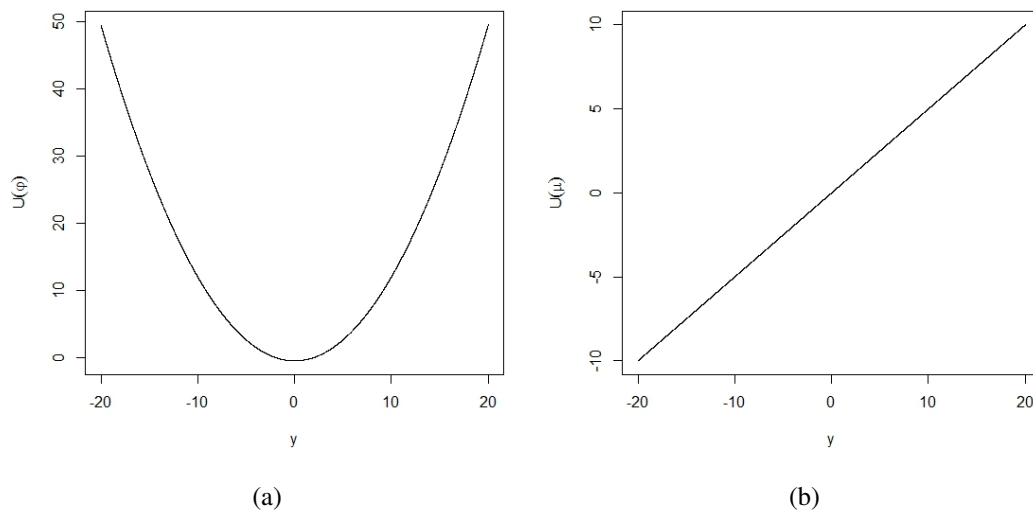
Notamos que, para o modelo SYMARMA-normal, as funções de influência correspondentes são proporcionais às suas funções escore. Então, desde que  $\mathbf{U}_\varphi$  e  $\mathbf{U}_\mu$  são ilimitadas (ver Figura 2.1), as correspondentes FI's também são ilimitadas. Dessa forma, há indícios de que o procedimento para estimação dos parâmetros baseado no modelo SYMARMA-normal não é robusto podendo fornecer estimativas bastantes viesadas na presença de observações atípicas.

No caso do modelo SYMARMA- $t$ , assumindo conhecido os graus de liberdade ( $\nu$ ), as funções de influência para  $\varphi$  e  $\mu$  também são proporcionais às suas funções escore, mas, nesse caso,  $\mathbf{U}_\varphi$  e  $\mathbf{U}_\mu$  são limitadas e, conseqüentemente, as FI's também são limitadas (ver Figura 2.2

para  $\nu = 4$ ). Portanto, o procedimento para estimação dos parâmetros  $\phi$  e  $\mu$  baseado no modelo SYMARMA- $t$  é qualitativamente robusto.

Lucas (1997) desenvolveu um importante estudo sobre aspectos de robustez no modelo  $t$ -Student, demonstrando que a robustez no ajuste de *outliers* é preservada apenas se os graus de liberdade,  $\nu$ , são mantidos fixo. Caso contrário, se os graus de liberdade também são estimados por máxima verossimilhança, as funções de influência para  $\phi$  e  $\nu$  não são limitadas. Diante disso, manteremos fixos todos os parâmetros extras, como, por exemplo,  $\nu$  da distribuição  $t$ -Student.

**Figura 2.1** Funções escore do modelo SYMARMA-normal para  $\phi$  e  $\mu$ .



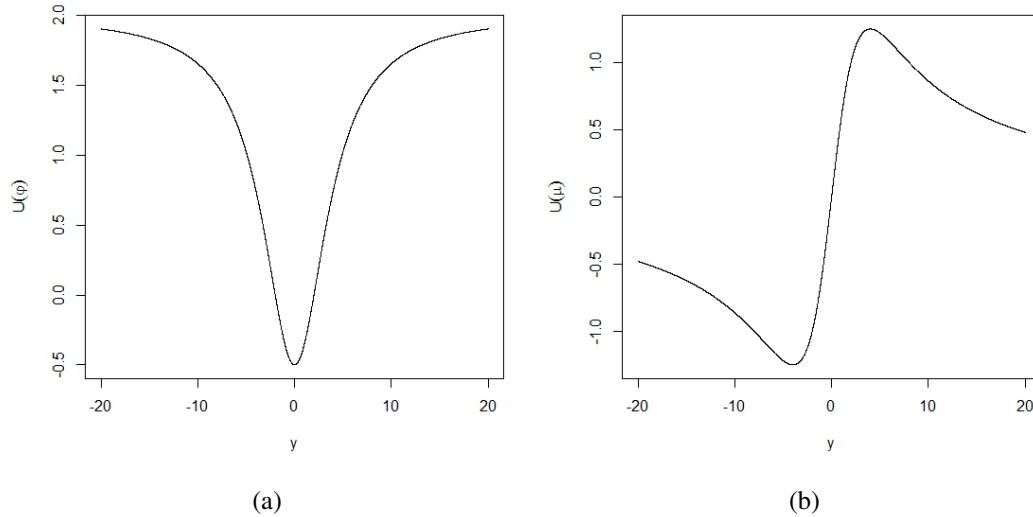
Fonte: Autoria própria.

### 2.6.2 Consistência em erro quadrático médio e QQ-plot

Para avaliar a consistência e a normalidade assintótica dos EMVC para os parâmetros do modelo SYMARMA, apresentamos um estudo de simulação em modelos SYMARMA com intercepto e um parâmetro autorregressivo, dado por

$$y_t = \beta_0 + \phi(y_{t-1} - \beta_0) + r_t, \quad t = 2, \dots, n, \quad (13)$$

**Figura 2.2** Funções escore do modelo SYMARMA- $t_4$  para  $\varphi$  e  $\mu$ .



Fonte: Autoria própria.

em que  $\beta_0 = 20$ ,  $\phi = 0,6$  e  $\varphi = 1$ . Foram geradas 10.000 réplicas de Monte Carlo do modelo (13) considerando as distribuições normal,  $t$ -Student e logística I para  $y_t | H_{t-1}$ . Para cada cenário e combinação de tamanhos amostrais  $n = 50, 100$  e  $200$ , avaliamos o viés e o erro quadrático médio (EQM) e apresentamos os gráficos quantil-quantil (QQ) para a distribuição empírica de cada parâmetro do modelo.

De acordo com os resultados apresentados na Tabela 2.5, os vieses são próximos a zero e os EQM decrescem com o aumento do tamanho amostral  $n$ . Esse fato evidencia a consistência dos EMVC obtidos pelo método escore de Fisher. Observamos também que os EQM para  $\hat{\phi}$  no modelo SYMARMA-logística I é maior do que nos modelos SYMARMA- $t_4$  e SYMARMA-normal.

As Figuras 2.3 e 2.4 mostram os gráficos QQ para a distribuição empírica de cada estimador. Para uma melhor visualização, os quantis empíricos são representados por diferentes tipos de linhas que correspondem aos diferentes tamanhos amostrais. A linha reta de  $45^\circ$  indica onde os pontos de uma distribuição normal padrão seriam representados no gráfico QQ. Notamos que as distribuições dos estimadores  $\hat{\beta}_0$ ,  $\hat{\phi}$  e  $\hat{\varphi}$  se aproximam da distribuição normal à medida

**Tabela 2.5** Medidas descritivas empíricas para os EMVC dos parâmetros  $\beta_0 = 20$ ,  $\phi = 0,6$  e  $\varphi = 1$  obtidos de 10.000 modelos SYMARMA simulados.

Parâmetros	Medidas descritivas empíricas					
	Viés			EQM		
	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
Normal						
$\beta_0$	0,0062	0,0044	0,0044	0,1273	0,0631	0,0319
$\phi$	0,0574	0,0278	0,0138	0,0184	0,0075	0,0035
$\varphi$	0,0429	0,0214	0,0116	0,0403	0,0202	0,0100
Student $t_4$						
$\beta_0$	0,0041	0,0024	0,0001	0,1882	0,0918	0,0451
$\phi$	0,0441	0,0211	0,0097	0,0139	0,0057	0,0025
$\varphi$	0,0277	0,0145	0,0046	0,0725	0,0352	0,0170
Logística I						
$\beta_0$	0,0060	0,0045	0,0026	0,0619	0,0307	0,0152
$\phi$	0,0644	0,0330	0,0169	0,0255	0,0114	0,0055
$\varphi$	0,6070	0,5964	0,5920	0,3805	0,3617	0,3534

que o tamanho da série aumenta. Contudo, no modelo SYMARMA-logística I a aproximação requer tamanhos amostrais maiores.

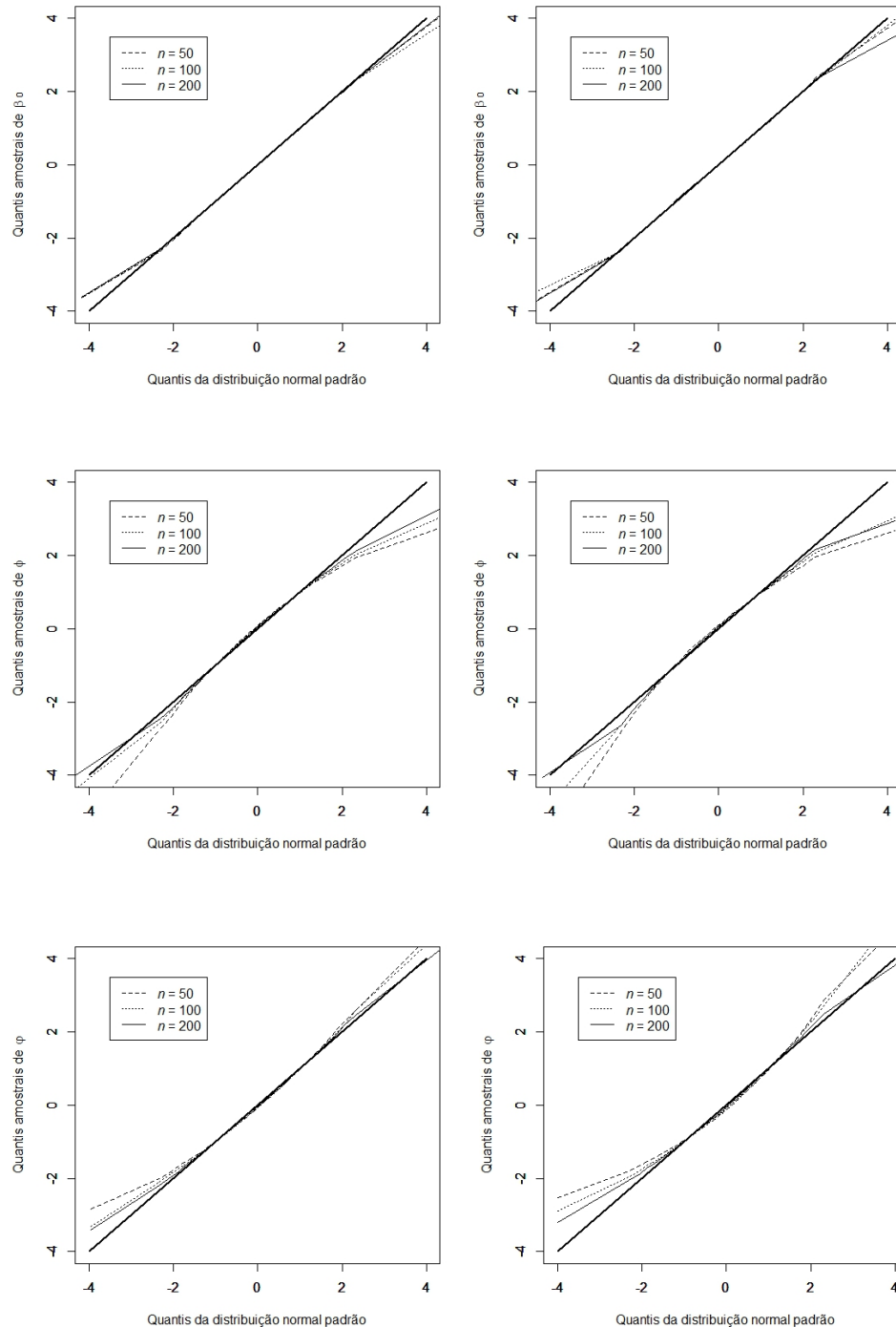
## 2.7 Previsões

Para realização de previsões, devemos utilizar os EMVC de  $\delta$  ( $\hat{\delta}$ ) e  $\varphi$  ( $\hat{\varphi}$ ) para estimar  $\mu_t$ , para  $t = m + 1, \dots, n$ . Essas estimativas serão denotadas por  $\hat{\mu}_t$ . Temos que

$$\hat{\mu}_t = \mathbf{x}_t^\top \hat{\beta} + \sum_{i=1}^p \hat{\phi}_i \left\{ y_{t-i} - \mathbf{x}_{t-i}^\top \hat{\beta} \right\} + \sum_{j=1}^q \hat{\theta}_j \hat{r}_{t-j}.$$

Utilizando  $\hat{\mu}_t$ , podemos obtemos as estimativas de  $r_t$ ,  $\hat{r}_t$ , para  $t = m + 1, \dots, n$ . Por exemplo, se  $r_t = y_t - \mu_t$ , isto é, erros mensurados na escada original, teremos  $\hat{r}_t = y_t - \hat{\mu}_t$ . Para  $N > n$ , a previsão do erro  $r_N$  é igual a zero. Assim, para prever o valor médio do processo no tempo  $N > n$ , devemos utilizar:

**Figura 2.3** QQ-plot normal dos EMVC para os parâmetros  $\beta_0$ ,  $\phi$  e  $\varphi$  do modelo SYMARMA-normal (a) e do modelo SYMARMA- $t_4$  (b), baseados em 10.000 réplicas.

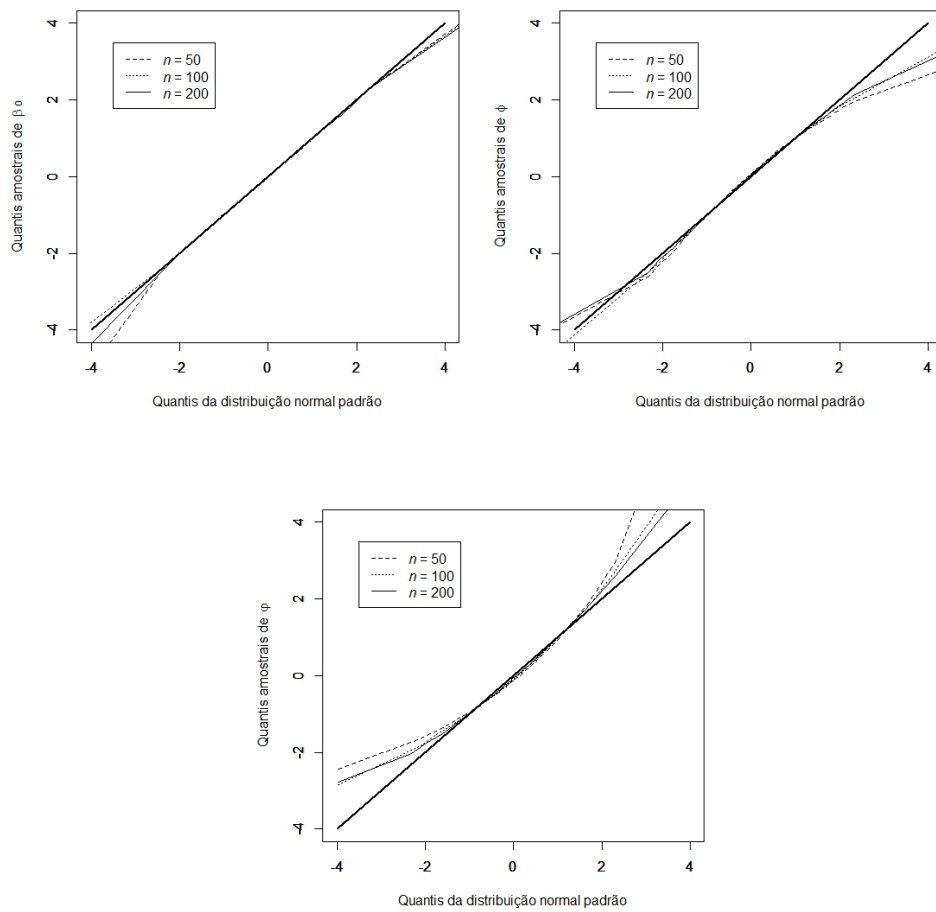


(a) Modelo SYMARMA-normal

(b) Modelo SYMARMA- $t_4$

Fonte: Autoria própria.

**Figura 2.4** QQ-plot normal dos EMVC para os parâmetros  $\beta_0$ ,  $\phi$  e  $\varphi$  do modelo SYMARMA-logística I, baseados em 10.000 réplicas.



(a) Modelo SYMARMA-logística I

Fonte: Autoria própria.

- (a) os EMVC  $\hat{\delta}$  e  $\hat{\phi}$ ;
- (b) as estimativas de  $\mu_t$ ,  $\hat{\mu}_t$ , para  $t = m + 1, \dots, n$ ;
- (c) as estimativas de  $r_t$ ,  $\hat{r}_t$ , para  $t = m + 1, \dots, n$  (lembrando que  $E(r_t) = 0$  para  $N > n$ ).

Essas quantidades são suficientes para obter  $\hat{\mu}_{t+1}$  e, procedendo de forma análoga, podemos obter  $\hat{\mu}_{t+2}$ , e assim por diante. Por fim, devemos substituir  $y_t$  por  $\hat{\mu}_t$  se  $n < t < N$ . Dessa forma, a estimativa para resposta média no tempo  $n + 1$  é dada por

$$\hat{y}_{n+1} = \mathbf{x}_{n+1}^\top \hat{\beta} + \sum_{i=1}^p \hat{\phi}_i \left\{ y_{n+1-i} - \mathbf{x}_{n+1-i}^\top \hat{\beta} \right\} + \sum_{j=1}^q \hat{\theta}_j \hat{r}_{n+1-j}.$$

No tempo  $n + 2$ , obtemos que

$$\hat{y}_{n+2} = \mathbf{x}_{n+2}^\top \hat{\beta} + \sum_{i=1}^p \hat{\phi}_i \left\{ y_{n+2-i} - \mathbf{x}_{n+2-i}^\top \hat{\beta} \right\} + \sum_{j=1}^q \hat{\theta}_j \hat{r}_{n+2-j},$$

e assim por diante.

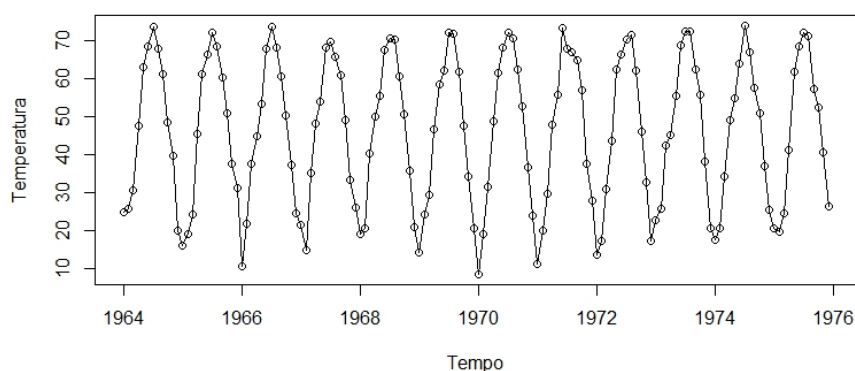
O método *bootstrap* (EFRON; TIBSHIRANI, 1993) pode ser usado para obter a distribuição empírica  $\hat{F}$  dos valores futuros ( $y_{n+\ell}$ ). Podemos construir um intervalo de previsão com nível de aproximadamente  $(1 - \alpha)$ , definindo os percentis  $\alpha/2$  e  $1 - \alpha/2$  de  $\hat{y}_{n+\ell}$ . O intervalo é  $(L_{inf}, L_{sup}) = (\hat{F}(\alpha/2), \hat{F}(1 - \alpha/2))$ . Geramos B séries temporais *bootstrap* (por exemplo,  $B = 1.000$ )  $Y_1^f, \dots, Y_2^f, \dots, Y_n^f$  por simulação do modelo SYMARMA( $p, q$ ) ajustado. Para cada série *bootstrap*  $\hat{y}_{n+\ell}$  é calculado. Das B estimativas *bootstrap*  $\hat{y}_{n+\ell}^f$ , podemos calcular a distribuição empírica  $\hat{F}$ .

## 2.8 Componente sazonal

A sazonalidade em uma série corresponde às oscilações de subida e de queda que sempre ocorrem em um determinado período do ano, do mês, da semana ou do dia. A sazonalidade é

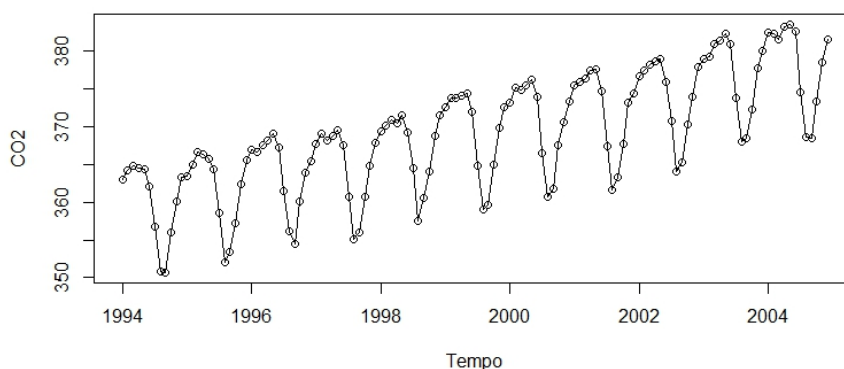
considerada determinística quando pressupomos um padrão sazonal regular e estável no tempo, desta forma podemos prever o comportamento sazonal perfeitamente a partir de dados anteriores. Quando a componente sazonal da série varia com o tempo a sazonalidade é considerada estocástica. As Figuras 2.5 e 2.6 apresentam, respectivamente, exemplos de séries com sazonalidade determinística e com sazonalidade estocástica.

**Figura 2.5** Temperatura média mensal, Dubuque, Iowa.



Fonte: Autoria própria.

**Figura 2.6** Níveis de dióxido de carbono mensais em Alert, NWT, Canadá.



Fonte: Autoria própria.

Modelos de regressão produzem bons ajustes para séries com sazonalidade determinística.



Podemos assumir, por exemplo, que a série observada é representada por

$$Y_t = \mu_t + r_t,$$

em que  $E(r_t) = 0$ , para todo  $t$ , e  $\mu_t$  é uma função com  $s$  parâmetros constantes,  $\beta_1, \beta_2, \dots, \beta_s$ , que fornecem o valor esperado para cada período. Por exemplo,  $s = 12$  para sazonalidade mensal e  $s = 2$  para sazonalidade semestral. Desta forma,  $\mu_t$  pode ser escrito como

$$\mu_t = \begin{cases} \beta_1, & \text{para } t = 1, s+1, 2s+1, \dots; \\ \beta_2, & \text{para } t = 2, s+2, 2s+2, \dots; \\ \vdots & \\ \beta_s, & \text{para } t = s, 2s, 3s, \dots \end{cases}$$

Em alguns casos, séries com sazonalidade determinística podem ser modeladas através de curvas cossenoides que incorporam a mudança esperada de forma suave de um período para o outro, preservando a sazonalidade. Consideramos que

$$\mu_t = \beta \cos(2\pi ft + \Phi), \quad (14)$$

em que  $\beta (> 0)$  é a amplitude da curva,  $f$  a frequência da curva e  $\Phi$  a fase da curva. Como  $t$  varia, a curva oscila entre um máximo de  $\beta$  e um mínimo de  $-\beta$ . Uma reparametrização mais conveniente para (14) é dada por

$$\beta \cos(2\pi ft + \Phi) = \beta_1 \cos(2\pi ft) + \beta_2 \sin(2\pi ft),$$

em que

$$\beta = \sqrt{\beta_1^2 + \beta_2^2}, \quad \Phi = \text{atan}(-\beta_2/\beta_1)$$

e, reciprocamente,

$$\beta_1 = \beta \cos(\Phi) \quad \text{e} \quad \beta_2 = \beta \sin(\Phi).$$

No contexto de séries temporais com sazonalidade estocástica encontramos a classe de modelos ARMA sazonal (SARMA), para séries estacionárias, e ARIMA sazonal (SARIMA), para séries não estacionárias. O modelo  $\text{ARMA}(p,q) \times (P,Q)$  com período sazonal  $s$  é expresso por

$$\phi(x)\Phi(x)Y_t = \theta(x)\Theta(x)r_t, \quad (15)$$

em que  $\phi(\cdot)$ ,  $\Phi(\cdot)$ ,  $\theta(\cdot)$  e  $\Theta(\cdot)$  são polinômios característicos, expressos por

$$\left. \begin{aligned} \phi(x) &= 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p \\ \Phi(x) &= 1 - \Phi_1 x^s - \Phi_2 x^{2s} - \dots - \Phi_P x^{Ps} \end{aligned} \right\} \text{AR}$$

e

$$\left. \begin{aligned} \theta(x) &= 1 + \theta_1 x + \theta_2 x^2 + \dots + \theta_q x^q \\ \Theta(x) &= 1 + \Theta_1 x^s + \Theta_2 x^{2s} + \dots + \Theta_Q x^{Qs} \end{aligned} \right\} \text{MA},$$

as constantes  $p$  e  $q$  são as ordens dos parâmetros autorregressivos e de médias móveis, respectivamente, e as constantes  $P$  e  $Q$  são as ordens dos parâmetros sazonais autorregressivos e de médias móveis, respectivamente.

### 2.8.1 Modelo SYMARMA sazonal

Nosso objetivo agora é adicionar aos modelos SYMARMA componentes sazonais estocásticas. Esse fato ampliará o leque de opções para a distribuição condicional de séries sazonais que, a partir deste momento, é tomada como pertencente a classe de distribuições simétricas.

Considere a distribuição de  $Y_t$ , dado o conjunto de observações passadas,  $H_{t-1}$ , pertencente a classe de distribuições simétricas com função de densidade condicional dada por (2). A

estrutura da média  $\mu_t$  nos modelos SYMARMA sazonais é definida como

$$\mu_t = \mathbf{x}_t^\top \beta + \tau_t, \quad (16)$$

em que  $\mathbf{x}_t^\top = (1, x_{t1}, x_{t2}, \dots, x_{tk})$  é o vetor que contém os valores das  $k$  variáveis explicativas,  $k < n$ , sendo  $n$  o tamanho da amostra,  $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^\top$  é um vetor de parâmetros desconhecidos e  $\tau_t$  é uma componente SARMA com período sazonal  $s$ , expressa por

$$\begin{aligned} \tau_t = & \sum_{i=1}^p \phi_i \left\{ y_{t-i} - \mathbf{x}_{t-i}^\top \beta \right\} + \sum_{j=1}^P \Phi_j \left\{ y_{t-js} - \mathbf{x}_{t-js}^\top \beta \right\} - \sum_{i=1}^p \phi_i \sum_{j=1}^P \Phi_j \left\{ y_{t-i-js} - \mathbf{x}_{t-i-js}^\top \beta \right\} + \\ & \sum_{l=1}^q \theta_l r_{t-l} + \sum_{m=1}^Q \Theta_m r_{t-ms} + \sum_{l=1}^q \theta_l \sum_{m=1}^Q \Theta_m r_{t-l-ms}, \end{aligned}$$

em que os termos  $r_t$ 's são ruídos brancos que podem ser, por exemplo, erros de medida na escala original (isto é,  $y_t - \mu_t$ ), ou, resíduos de Pearson, entre outros.

O modelo SYMARMA sazonal é definido pela componente aleatória dada em (2) e pela componente dinâmica dada em (16). Denotaremos o modelo SYMARMA sazonal por SYMARMA( $p, q$ ) $\times$ ( $P, Q$ ). Podemos verificar que, da forma como definido, os modelos SYMARMA sazonais podem ser apresentados como modelos SYMARMA gerais. Por exemplo, a média  $\mu_t$  no modelo SYMARMA(1,0) $\times$ (1,1) com periodicidade  $s = 12$  pode ser reescrita como

$$\mu_t = \mathbf{x}_t^\top \beta + \phi_1 \left\{ y_{t-1} - \mathbf{x}_{t-1}^\top \beta \right\} + \theta_1 r_{t-1} + \Theta_1 r_{t-12} + \theta_1 \Theta_1 r_{t-13}.$$

Dessa forma, o modelo SYMARMA(1,0) $\times$ (1,1) com periodicidade  $s = 12$  pode ser apresentado como um modelo SYMARMA(1,13) com  $\theta_2 = \theta_3 = \dots = \theta_{11} = 0$ . No geral, o modelo SYMARMA( $p, q$ ) $\times$ ( $P, Q$ ) com periodicidade  $s$  pode ser apresentado como um modelo SYMARMA( $p + Ps, q + Qs$ ).

Tendo especificado um modelo SYMARMA sazonal para uma série temporal, o interesse passa a ser estimar os parâmetros do modelo. Como pôde ser observado anteriormente, os modelos SYMARMA sazonais são casos especiais dos modelos SYMARMA. Como tal, o

mesmo procedimento adotado para estimar os parâmetros do modelo SYMARMA pode ser adotado para estimar os parâmetros dos modelos SYMARMA sazonais.

## **Capítulo 3**

# **Influência Local em Modelos Autorregressivos de Médias Móveis Simétricos**

### **3.1 Introdução**

Quando consideramos um modelo de regressão para uma aplicação, normalmente não temos certeza se o mesmo é apropriado. Algumas, ou muitas, das características do modelo, como, por exemplo, a distribuição adotada para o erro, podem não ser apropriadas para o conjunto de dados em estudo e, portanto, conclusões baseadas nesses modelos podem estar erradas. Diante disso, é importante examinar o quanto o modelo que está sendo considerado é adequado para os dados antes de fazer inferências. Essa avaliação do modelo é conhecida como análise de resíduos e diagnóstico, que consiste na verificação de possíveis afastamentos relevantes das suposições feitas para o mesmo bem como, na verificação da existência de observações com alguma interferência desproporcional nos resultados do ajuste.

Dentre as técnicas de diagnóstico utilizadas para modelos de séries temporais, a medida de influência local tem um papel importante por não ser baseada na exclusão de observações do conjunto de dados que, a princípio, são dependentes. A dependência das observações coletadas ao longo do tempo faz com que técnicas como a distância de Cook (COOK; WEISBERG, 1982) e a medida DFFITS (BESLEY et al, 1980) não sejam convenientes para avaliar pontos

influentes em séries temporais.

A metodologia de diagnóstico de influência local foi proposta inicialmente por Cook (1986), que sugeriu avaliar a influência de pequenas perturbações no modelo e/ou nos dados através do afastamento da função de verossimilhança e, conseqüentemente, na análise da curvatura do gráfico de influência. Contudo, a curvatura não é invariante a reparametrizações do modelo e, por essa razão, Billor e Loynes (1993) sugerem uma medida alternativa baseada no uso da inclinação do gráfico de influência modificado. Uma outra medida de influência local, também baseada na curvatura normal, pode ser encontrada em Lesaffre e Verbeke (1998), que utilizaram o enfoque de influência local para avaliar modelos lineares mistos.

Em todo caso, o maior desafio na análise de diagnóstico através de influência local é encontrar a distribuição das medidas de diagnóstico de inclinação e de curvatura, para que seja possível caracterizar uma observação como influente. Na maioria das vezes, essa caracterização é feita por meio de uma simples análise exploratória. O primeiro trabalho a discutir a aplicação de marcas de referência para determinar se, estatisticamente, as observações são influentes, foi proposto por Zhang e King (2005), que simularam a distribuição das medidas de influência nos modelos GARCH(1,1) com erros gaussianos, usando a primeira derivada (inclinação de Billor e Loynes) e a segunda derivada do gráfico de influência modificado. Esse procedimento foi denominado de limiares (*benchmarks*). Mais recentemente, Zevallos e Hotta (2012) utilizaram a metodologia dos limiares para análise de influência local em modelos GARCH considerando, alternativamente, a curvatura de Cook ao invés da segunda derivada do gráfico de influência modificado, obtendo uma grande vantagem de ordem computacional.

Esse capítulo tem por objetivo desenvolver uma metodologia para análise de diagnóstico em modelos autorregressivos de médias móveis simétricos (SYMARMA) sob o enfoque de influência local. Discutimos a aplicação da metodologia de marcas de referência (limiares), proposta por Zhang e King (2005), na análise de influência local dos modelos SYMARMA, através das medidas de influência desenvolvidas por Cook (1986), Billor e Loynes (1993) e Lesaffre e Verbeke (1998). É a primeira vez que a metodologia de limiares é aplicada à medida

de influência local proposta por Lesaffre e Verbeke (1998). Consideramos quatro esquemas de perturbação: perturbação aditiva na variável resposta, perturbação no parâmetro de dispersão, perturbação no vetor de parâmetros autorregressivos e perturbação no vetor de parâmetros de médias móveis. Para ilustrar a metodologia proposta apresentamos alguns resultados de simulação.

## 3.2 Medidas de influência local

As três medidas de influência local estudadas neste capítulo são: curvatura de Cook (COOK, 1986), curvatura de Lesaffre e Verbeke (LESAFFRE; VERBEKE, 1998) e inclinação de Billor e Loynes (BILLOR; LOYNES, 1993). Aplicamos a metodologia de limiares, baseada em níveis de referência, para avaliar a significância relacionada a influência local.

### 3.2.1 Curvatura de Cook

Na literatura, o primeiro trabalho sobre influência local foi introduzido por Cook (1986). A proposta é avaliar a influência de pequenas perturbações nos dados ou no modelo utilizando o afastamento da função de verossimilhança, definido por

$$LD(\omega) = -2\{L(\hat{\delta}, \hat{\phi}) - L(\hat{\delta}_\omega, \hat{\phi}_\omega)\},$$

em que  $L(\hat{\delta}, \hat{\phi})$  é a função de verossimilhança do modelo *postulado*,  $L(\hat{\delta}_\omega, \hat{\phi}_\omega)$  é a função de verossimilhança do modelo *perturbado* pelo vetor  $\omega = (\omega_1, \dots, \omega_n)^\top \in \mathbb{R}^n$  e,  $(\hat{\delta}, \hat{\phi})$  e  $(\hat{\delta}_\omega, \hat{\phi}_\omega)$  são as estimativas de máxima verossimilhança de  $L(\delta, \phi)$  e  $L(\delta, \phi|\omega)$ , respectivamente. Sendo  $\mathbf{y} = (y_1, \dots, y_n)^\top$  o vetor de observações, podemos considerar, por exemplo, que a perturbação seja inserida na variável resposta  $y_t$  na forma  $\tilde{y}_t = y_t + \omega_t$ , onde  $\tilde{y}_t$  é a variável resposta perturbada por  $\omega_t$ .

Em geral, o vetor  $\omega$  é introduzido no modelo via  $\omega = \omega_0 + a\ell$ , em que  $\ell, \omega_0 \in \mathbb{R}^n$  e

$a$  medem a magnitude da perturbação na direção  $\ell$ . Assumimos que existe um  $\omega_0$  tal que  $L(\delta, \varphi | \omega_0) = L(\delta, \varphi)$  para todo  $\delta$  e  $\varphi$ .

A idéia de influência local é estudar o comportamento da função  $LD(\omega)$  numa vizinhança de  $\omega_0$ . Para tanto, consideramos a superfície geométrica  $(n+1)$ -dimensional formada pelos valores do vetor

$$\lambda(\omega) = \begin{pmatrix} \omega \\ LD(\omega) \end{pmatrix}.$$

Essa superfície é denominada gráfico de influência e fornece informações importantes sobre a sensibilidade de  $LD$  sob pequenas perturbações. A proposta é avaliar a maior mudança local na curvatura de  $LD$ . Dessa forma, tomamos a segunda derivada de  $LD$  com relação a  $a$ , isto é,  $\partial^2 LD(\omega) / \partial a \partial a = C_\ell(\delta, \varphi) = 2(\ell^\top \ddot{\mathbf{F}} \ell)$  em que  $\ddot{\mathbf{F}}$  é uma matriz  $(n \times n)$  definida como

$$\ddot{\mathbf{F}} = \frac{\partial^2 L(\hat{\delta}_\omega, \hat{\varphi}_\omega)}{\partial \omega \partial \omega^\top},$$

que pode ser expressa da forma

$$\ddot{\mathbf{F}} = \Delta^\top \ddot{\mathbf{L}}^{-1} \Delta, \quad (17)$$

em que  $-\ddot{\mathbf{L}}$  é a matriz de informação condicional observada para o modelo *postulado* ( $\omega = \omega_0$ ),  $\ddot{\mathbf{L}}^{-1}$  é a matriz inversa de  $\ddot{\mathbf{L}}$  e  $\Delta$  é uma matriz com elementos

$$\Delta_{ij} = \frac{\partial^2 L(\delta, \varphi | \omega)}{\partial \delta_i \partial \omega_j},$$

avaliados em  $\hat{\delta}$ ,  $\hat{\varphi}$  e  $\omega_0$ , com  $i = 1, \dots, p+q+k$  e  $j = 1, \dots, n$ . Consideramos então o maior autovalor de  $2\ddot{\mathbf{F}}$ , denotado por  $O_c$ , como medida de influência global e o autovetor associado ao maior autovalor de  $2\ddot{\mathbf{F}}$ ,  $\mathbf{C} = (c_1, \dots, c_n)^\top$ , como medida de influência individual, onde a  $i$ -ésima componente do vetor  $\mathbf{C}$  indica a influência da  $i$ -ésima observação. Adicionalmente, considerando o modelo de regressão linear, Cook propõe relacionar a  $i$ -ésima componente do



vetor  $\mathbf{C}$  com a  $i$ -ésima componente do vetor de resíduos.

### 3.2.2 Curvatura de Lesaffre e Verbeke

Alternativamente, Lesaffre e Verbeke (1998) sugerem avaliar a curvatura normal na direção da  $i$ -ésima observação, que consiste na avaliação de  $C_\ell(\boldsymbol{\delta}, \boldsymbol{\varphi})$  no vetor  $n$ -dimensional,  $\ell_i$ , formado por zeros com um na  $i$ -ésima posição. Essa curvatura é denominada por

$$C_{\ell_i} = 2|\ddot{\mathbf{F}}_{ii}|.$$

Lesaffre e Verbeke sugerem que as observações tais que  $C_{\ell_i} > 2\overline{C_\ell}$  tenham uma atenção especial.

### 3.2.3 Inclinação de Billor e Loynes

Conforme apontado por Billor e Loynes (1993), dificuldades práticas e teóricas surgem no enfoque de influência local introduzido por Cook (1986), entre elas: i) a escolha de uma referência para a medida de influência global  $O_c$ , ii) a fórmula explícita para  $O_c$  é difícil de ser encontrada analiticamente e iii) a falta de invariância da curvatura sobre reparametrizações do esquema de perturbação.

Para contornar problemas como a falta de invariância da curvatura de Cook sobre reparametrizações do esquema de perturbação, Billor e Loynes (1993) consideram uma modificação no afastamento da função de verossimilhança, definida por

$$\text{MLD}(\boldsymbol{\omega}) = -2\{\text{L}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\varphi}}) - \text{L}(\hat{\boldsymbol{\delta}}|\boldsymbol{\omega}, \hat{\boldsymbol{\varphi}}|\boldsymbol{\omega})\},$$

em que  $\text{L}(\hat{\boldsymbol{\delta}}|\boldsymbol{\omega}, \hat{\boldsymbol{\varphi}}|\boldsymbol{\omega})$  é a função de verossimilhança do modelo *perturbado* avaliada sobre  $(\hat{\boldsymbol{\delta}}|\boldsymbol{\omega}, \hat{\boldsymbol{\varphi}})$ , que é o estimador de máxima verossimilhança de  $(\boldsymbol{\delta}|\boldsymbol{\omega}, \boldsymbol{\varphi})$  quando perturbações são introduzidas no modelo *postulado*. Assim podemos estudar a superfície geométrica  $(n+1)$ -dimensional formada pelos valores do vetor

$$\lambda(\omega) = \begin{pmatrix} \omega \\ \text{MLD}(\omega) \end{pmatrix},$$

para avaliar a influência de pequenas perturbações no modelo. A proposta é avaliar a maior mudança local na inclinação de MLD. Dessa forma, tomamos a primeira derivada de MLD com relação a  $a$ , isto é,  $\partial \text{MLD}(\omega)/\partial a = 2(\ell^\top \mathbf{S})$ , em que

$$\mathbf{S} = \frac{\partial L(\hat{\delta}\omega, \hat{\phi}\omega|\omega)}{\partial \omega}. \quad (18)$$

Billor e Loynes propõem utilizar a norma do vetor  $\mathbf{S} = (s_1, \dots, s_n)^\top$ ,  $\|\mathbf{S}\| = \sqrt{\sum_{i=1}^n s_i^2}$ , como medida de influência global e duas vezes as componentes do vetor  $\mathbf{S}$ ,  $2\mathbf{S}$ , como medida de influência individual, onde a  $2s_i$  indica a influência da  $i$ -ésima observação.

As estatísticas para as medidas de influência local definidas nas Seções 3.2.1-3.2.3 podem ser resumidas como:

Diagnóstico de inclinação ( <i>Billor e Loynes</i> )	{	<b>Critério global:</b>
		dado por $O_s = \ \mathbf{S}\ $ . Dessa forma, a série temporal em estudo tem pelo menos um ponto influente se $O_s$ for grande.
		<b>Critério individual:</b>
		a observação $y_t$ é considerada como influente se $d_t = 2s_t$ for grande, para $t = 1, \dots, n$ . O vetor $\mathbf{d} = (d_1, \dots, d_n)^\top$ é chamado vetor de inclinação.

Diagnóstico de curvatura (Cook) {

**Critério global:**  
 dado por  $O_c$ , o maior autovalor de  $2\ddot{\mathbf{F}}$ . Dessa forma, a série temporal em estudo tem pelo menos um ponto influente se  $O_c$  for grande.

**Critério individual:**  
 a observação  $y_t$  é considerada como influente se a  $i$ -ésima componente do autovetor associado ao maior autovalor de  $2\ddot{\mathbf{F}}$ ,  $\mathbf{C} = (c_1, \dots, c_n)^\top$ , for grande. O vetor  $\mathbf{C}$  é chamado vetor de curvatura.

Diagnóstico de curvatura (Lesaffre e Verbeke) {

**Critério global:**  
 dado por  $O_c$ , o maior autovalor de  $2\ddot{\mathbf{F}}$ . Dessa forma, a série temporal em estudo tem pelo menos um ponto influente se  $O_c$  for grande.

**Critério individual:**  
 a observação  $y_t$  é considerada como influente se a  $i$ -ésima componente do vetor  $C_{\ell_i} = 2|\ddot{\mathbf{F}}_{ii}|$  for grande.

### 3.3 Metodologia de limiares

Diferentes critérios de diagnóstico foram propostos por Cook (1986), Lesaffre e Verbeke (1998) e Billor e Loynes (1993) para avaliar a influência local dos dados. Em todos os casos, as técnicas de diagnóstico propostas apresentam dois tipos de problemas. O primeiro trata das aproximações bastante imprecisas apresentadas por Billor e Loynes (1993) para  $O_s$  e por Cook (1986) para  $O_c$ . A segunda é a determinação de que uma observação específica é influente através da simples análise exploratória de gráficos.

Uma proposta para contornar esses problemas foi apresentada por Zhang e King (2005) que, baseados em simulações de Monte Carlo para o modelo GARCH sob diferentes esquemas de perturbação, construíram níveis de referência para os vetores de diagnóstico da inclinação e da curvatura. Esta metodologia ficou conhecida como limiares (*benchmarks*).

Considerando o modelo *postulado*, ou seja, sem perturbação, as simulações de Monte Carlo se tornam uma importante ferramenta, pois, permitem estimar as distribuições das medidas de influência local. Assim, através de parâmetros predeterminados, para cada replicação é gerada uma amostra de tamanho  $n$ , e conseguinte são computadas as estatísticas globais  $O_s$  e  $O_c$  e os vetores  $\mathbf{d}$  e  $\mathbf{C}$ . A partir dessas replicações, podemos estimar a distribuição das estatísticas  $O_s$  e  $O_c$ . No nosso trabalho, são realizadas 2.000 replicações para cada simulação. É a primeira vez que limiares são utilizados para simular a distribuição da estatística do diagnóstico de curvatura sugerido por Lesaffre e Verbeke, o que trará níveis de referências mais confiáveis.

Para obtenção das marcas de referência por meio de limiares, primeiramente simulamos 2.000 vezes o modelo *postulado*, isto é, séries temporais de tamanho  $n$  sem perturbação. Para a  $k$ -ésima série temporal simulada, calculamos a estatística  $\mathbf{S}$  de acordo com a Equação (18), a estatística  $O_s$ , o vetor de inclinação  $\mathbf{d}_k = (d_{k1}, \dots, d_{kn})$ , a matriz  $\ddot{\mathbf{F}}$  de acordo com a Equação (17), a estatística  $O_c$  e os vetores de curvatura  $\mathbf{C}_k = (c_{k1}, \dots, c_{kn})$  e  $\mathbf{C}_{\ell_k} = (c_{\ell_{k1}}, \dots, c_{\ell_{kn}})$ , para  $k = 1, \dots, 2.000$ . Consideramos  $BS_j$  como sendo os limiares para o diagnóstico de inclinação,  $BC_j^{Cook}$  como sendo os limiares para o diagnóstico de curvatura de Cook e  $BC_j^{LV}$  como sendo os limiares para o diagnóstico de curvatura de Lesaffre e Verbeke, em que  $j = 0, 1, 2$ . Agora, a partir dos resultados obtidos das simulações de Monte Carlo, as marcas de referência são calculadas da seguinte maneira.

### 3.3.1 Limiares para as medidas de influência global

$$BS_0 = \text{Percentil } 95\% \text{ de } \{O_{s1}, \dots, O_{s2.000}\}$$

$$BC_0^{Cook} = BC_0^{LV} = \text{Percentil } 95\% \text{ de } \{O_{c1}, \dots, O_{c2.000}\}$$

### 3.3.2 Limiares para as medidas de influência individual

Seja  $\gamma_k = \max\{|d_{k1}|, \dots, |d_{kn}|\}$  para  $k = 1, \dots, 2.000$ .

$$BS_1 = \text{Percentil } 95\% \text{ de } \{\gamma_1, \dots, \gamma_{2.000}\}.$$

$BS_2 = \text{Percentil } 5\% \text{ de } \{\gamma_{k_1}, \dots, \gamma_{k_m}\}$ , em que  $\{k_1, \dots, k_m\} \in \{1, \dots, 2.000\}$  corresponde a série

gerada que satisfaz  $O_{sk_j} > BS_0$ ,  $j = 1, \dots, m$ .

Seja  $\xi_k = \max\{|c_{k(1)}|, \dots, |c_{k(n)}|\}$  para  $k = 1, \dots, 2.000$ .

$BC_1^{Cook} = \text{Percentil } 95\% \text{ de } \{\xi_1, \dots, \xi_{2.000}\}$ .

$BC_2^{Cook} = \text{Percentil } 5\% \text{ de } \{\xi_{k_1}, \dots, \xi_{k_m}\}$ , em que  $\{k_1, \dots, k_m\} \in \{1, \dots, 2.000\}$  corresponde a série gerada que satisfaz  $O_{ck_j} > BC_0^{Cook}$ ,  $j = 1, \dots, m$ .

Seja  $\xi_k = \max\{|c_{\ell_k(1)}|, \dots, |c_{\ell_k(n)}|\}$  para  $k = 1, \dots, 2.000$ .

$BC_1^{LV} = \text{Percentil } 95\% \text{ de } \{\xi_1, \dots, \xi_{2.000}\}$ .

$BC_2^{LV} = \text{Percentil } 5\% \text{ de } \{\xi_{k_1}, \dots, \xi_{k_m}\}$ , em que  $\{k_1, \dots, k_m\} \in \{1, \dots, 2.000\}$  corresponde a série gerada que satisfaz  $O_{ck_j} > BC_0^{LV}$ ,  $j = 1, \dots, m$ .

Sendo  $y_1, \dots, y_n$  a série temporal observada, primeiro calculamos o vetor de inclinação  $\mathbf{d} = (d_1, \dots, d_n)$  e a estatística  $O_s$ . Se  $O_s$  for maior que  $BS_0$ , consideramos a existência de pelo menos uma observação influente no conjunto de dados. Para identificar que observações específicas são influentes, comparamos os valores  $(d_1, \dots, d_n)$  com o limiar  $BS_1$ . Por exemplo, se  $|d_j| > BS_1$ , então  $y_j$  é individualmente influente. No entanto, esse valor de referência é calculado levando em conta séries temporais simuladas que não são consideradas globalmente influentes usando os critérios gerais  $O_s$ . Dessa forma, devemos comparar  $|d_j|$  com o limiar  $BS_2$  se a proposta for usar as séries temporais globalmente influentes em termos de inclinação. Note que é possível definir limiares  $BS_j$ ,  $BC_j^{Cook}$ ,  $BC_j^{LV}$ ,  $j = 0, 1$  utilizando, por exemplo, o percentil 90% ou 99%, ao invés do percentil 95%. A aplicação para o diagnóstico de curvatura é similar ao apresentado para o diagnóstico de inclinação.

### 3.4 Esquemas de perturbação

Conforme resumido por Billor e Loynes (1993), vários esquemas de perturbação podem ser introduzidos através de  $\omega$ , e esses são divididos em dois grupos:

- **Perturbação no modelo:** Este tipo de perturbação visa a modificação das suposições propostas para o modelo. Por exemplo, uma suposição de homoscedasticidade (variância constante) nos erros normalmente distribuídos pode ser substituída por uma suposição heteroscedástica, ou seja,  $\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$  é substituída por  $\varepsilon \sim N(\mathbf{0}, \sigma^2 \text{diag}^{-1}(\omega_i))$ , onde  $\omega_i$  são as perturbações especificadas e  $\text{diag}^{-1}(\omega_i)$  é a inversa da matriz diagonal de componentes  $\omega_i$ , para  $i = 1, \dots, n$ .
- **Perturbação nos dados:** Perturbar a variável resposta ou as variáveis explicativas são exemplos de perturbação nos dados. As duas razões para considerar a perturbação nos dados são os possíveis erros de medida e a existência de observações aberrantes (*outliers*), em uma proporção relativamente pequena das observações.

Para o estudo de influência local em modelos SYMARMA, consideramos quatro esquemas de perturbação: perturbação aditiva na variável resposta, perturbação no parâmetro de dispersão, perturbação no vetor de parâmetros autorregressivos e perturbação no vetor de parâmetros de médias móveis. Para cada caso, derivamos expressões analíticas para o vetor  $\mathbf{S}$  de acordo com a Equação (18) e  $\ddot{\mathbf{F}}$  de acordo com a Equação (17), que são a base para encontrar os vetores de inclinação e de curvatura bem como as medidas globais. A matriz de informação condicional observada  $-\ddot{\mathbf{L}} = \partial^2 \mathbf{L}(\delta) / \partial \delta \partial \delta^\top$ , que não depende do esquema de perturbação adotado, é apresentada no Apêndice E.

Toda metodologia desenvolvida nessa seção pode ser estendida aos modelos SYMARMA sazonais. Conforme pôde ser observado no Seção 2.8.1 do Capítulo 2, os modelos SYMARMA sazonais são casos especiais dos modelos SYMARMA.

### 3.4.1 Perturbação aditiva nos dados

Assumimos que uma pequena perturbação  $\omega_i$  é adicionada para cada observação da série. As novas observações ( $\tilde{y}_i$ ) e os termos do logaritmo da função de verossimilhança ficam

definidos por

$$\begin{aligned}\tilde{y}_t &= y_t + \omega_t, \\ \ell_t(\delta, \varphi | \omega) &= -\log(\varphi)/2 + \log\{g(\tilde{u}_t)\},\end{aligned}$$

para  $t = m+1, \dots, n$ , em que  $\tilde{u}_t = (\tilde{y}_t - \tilde{\mu}_t)^2/\varphi$ , sendo  $\tilde{\mu}_t$  definido como

$$\tilde{\mu}_t = \mathbf{x}_t^\top \beta + \sum_{i=1}^p \phi_i \left\{ \tilde{y}_{t-i} - \mathbf{x}_{t-i}^\top \beta \right\} + \sum_{j=1}^q \theta_j \{ \tilde{y}_{t-j} - \tilde{\mu}_{t-j} \}.$$

O ponto de não perturbação é  $\omega_0 = (0, \dots, 0)^\top$ . No diagnóstico de inclinação temos que

$$\mathbf{S} | \omega_0 = \frac{\partial \ell(\delta, \varphi | \omega)}{\partial \omega} = -\varphi^{-1} \mathbf{D}(\mathbf{y} - \mu),$$

em que  $\mathbf{D} = \text{diag}\{v_{m+1}, \dots, v_n\}$ ,  $v_t = -2W_g(u_t)$  sendo  $W_g(u_t) = \frac{d \log(g(u_t))}{du_t}$  e  $u_t = \frac{(y_t - \mu_t)^2}{\varphi}$ ,  $\mathbf{y} = (y_{m+1}, \dots, y_n)^\top$  e  $\mu = (\mu_{m+1}, \dots, \mu_n)^\top$ .

No diagnóstico de curvatura, para esse esquema de perturbação, temos que

$$\Delta | \omega_0 = \begin{bmatrix} \frac{\partial^2 \ell(\delta, \varphi | \omega)}{\partial \delta \partial \omega} \\ \frac{\partial^2 \ell(\delta, \varphi | \omega)}{\partial \varphi \partial \omega} \end{bmatrix} = \begin{bmatrix} \varphi^{-1} \mathbf{O}^\top \mathbf{D}_a \\ -2\varphi^{-2} \mathbf{b}^\top \end{bmatrix},$$

em que  $a_t = -2(W_g(u_t) + 2W'_g(u_t)u_t)$  com  $W'_g(u_t) = \frac{dW_g(u_t)}{du_t}$ ,  $\mathbf{D}_a = \text{diag}\{a_{m+1}, \dots, a_n\}$ ,  $\mathbf{b} = (b_{m+1}, \dots, b_n)^\top$  em que  $b_t = (W_g(u_t) + W'_g(u_t)u_t)(y_t - \mu_t)$  e  $\mathbf{O}$  é uma matriz  $(n-m) \times (k+p+q+1)$  definida na Seção 2.4.

### 3.4.2 Perturbação no parâmetro de dispersão

Para esse esquema de perturbação temos que o parâmetro de dispersão e os termos do logaritmo da função de verossimilhança ficam expressos por

$$\begin{aligned}\tilde{\varphi}_t &= \varphi \omega_t, \\ \ell_t(\delta, \varphi | \omega) &= -\log(\tilde{\varphi}_t)/2 + \log\{g(\tilde{u}_t)\},\end{aligned}$$

para  $t = m+1, \dots, n$ , em que  $\tilde{u}_t = (y_t - \mu_t)^2 / \tilde{\varphi}_t$  sendo  $\mu_t$  definido como na Equação (6). O ponto de não perturbação é  $\omega_0 = (1, \dots, 1)^\top$ .

Para o diagnóstico de inclinação temos que

$$\mathbf{S} | \omega_0 = \frac{\partial \ell(\delta, \varphi | \omega)}{\partial \omega} = -\frac{1}{2} + \frac{1}{2} \mathbf{D} \mathbf{u},$$

em que  $\mathbf{D}$  e  $u_t$  são definidos como na Seção 3.4.1 e  $\mathbf{u} = (u_{m+1}, \dots, u_n)^\top$ .

Para o diagnóstico de curvatura,

$$\Delta | \omega_0 = \begin{bmatrix} \frac{\partial^2 \ell(\delta, \varphi | \omega)}{\partial \delta \partial \omega} \\ \frac{\partial^2 \ell(\delta, \varphi | \omega)}{\partial \varphi \partial \omega} \end{bmatrix} = \begin{bmatrix} 2\varphi^{-1} \mathbf{O}^\top \mathbf{D}_b \\ \varphi^{-1} \mathbf{D}_c \mathbf{u}^\top \end{bmatrix},$$

em que  $\mathbf{D}_b = \text{diag}\{b_{m+1}, \dots, b_n\}$ ,  $\mathbf{D}_c = \text{diag}\{c_{m+1}, \dots, c_n\}$  sendo  $c_t = W_g(u_t) + W'_g(u_t)u_t$ . A matriz  $\mathbf{O}$  é definida como na Seção 2.4.

### 3.4.3 Perturbação no vetor de parâmetros autorregressivos

Para esse esquema de perturbação temos que os parâmetros autorregressivos e os termos do logaritmo da função de verossimilhança ficam expressos por

$$\begin{aligned}\tilde{\phi}_{it} &= \phi_i + \omega_t, \\ \ell_t(\delta, \varphi | \omega) &= -\log(\varphi_t)/2 + \log\{g(\tilde{u}_t)\},\end{aligned}$$



para  $t = m + 1, \dots, n$ , em que  $\tilde{u}_t = (y_t - \tilde{\mu}_t)^2 / \varphi$ , sendo  $\tilde{\mu}_t$  definido como

$$\tilde{\mu}_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \tilde{\phi}_{it} \left\{ y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta} \right\} + \sum_{j=1}^q \theta_j \{ y_{t-j} - \mu_{t-j} \}.$$

O ponto de não perturbação é  $\omega_0 = (0, \dots, 0)^\top$ . No diagnóstico de inclinação temos que

$$\mathbf{S}|\omega_0 = \frac{\partial \ell(\delta, \varphi|\omega)}{\partial \omega} = \varphi^{-1} \mathbf{D} \mathbf{D}_q (\mathbf{y} - \boldsymbol{\mu}),$$

em que  $\mathbf{D}$ ,  $\mathbf{y}$  e  $\boldsymbol{\mu}$  são tais como definidos como na Seção 3.4.1 e  $\mathbf{D}_q = \text{diag}\{q_{m+1}, \dots, q_n\}$  com  $q_t = \sum_{i=1}^p (y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta})$ , isto é,  $\mathbf{q} = \mathbf{A} \mathbf{1}$  em que  $\mathbf{1} = (1, \dots, 1)^\top$ .

No diagnóstico de curvatura,

$$\Delta|\omega_0 = \begin{bmatrix} \frac{\partial^2 \ell(\delta, \varphi|\omega)}{\partial \delta \partial \omega} \\ \frac{\partial^2 \ell(\delta, \varphi|\omega)}{\partial \varphi \partial \omega} \end{bmatrix} = \begin{bmatrix} -\varphi^{-1} \mathbf{O}^\top \mathbf{D}_a \mathbf{D}_q \\ 2\varphi^{-2} \mathbf{b}^\top \mathbf{D}_q \end{bmatrix},$$

em que a matriz  $\mathbf{O}$  é definida como na Seção 2.4 e a matriz  $\mathbf{D}_a$  e o vetor  $\mathbf{b}$  são definidos como na Seção 3.4.1.

### 3.4.4 Perturbação no vetor de parâmetros de médias móveis

Nesse caso os parâmetros de médias móveis e os termos do logaritmo da função de verossimilhança ficam expressos por

$$\begin{aligned} \tilde{\theta}_{jt} &= \theta_j + \omega_t, \\ \ell_t(\delta, \varphi|\omega) &= -\log(\varphi_t)/2 + \log\{g(\tilde{u}_t)\}, \end{aligned}$$

para  $t = m + 1, \dots, n$ , em que  $\tilde{u}_t = (y_t - \tilde{\mu}_t)^2 / \varphi$ , sendo  $\tilde{\mu}_t$  definido como

$$\tilde{\mu}_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i \left\{ y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta} \right\} + \sum_{j=1}^q \tilde{\theta}_{jt} \{ y_{t-j} - \mu_{t-j} \}.$$

O ponto de não perturbação é  $\omega_0 = (0, \dots, 0)^\top$ . No diagnóstico de inclinação temos que

$$\mathbf{S}|\omega_0 = \frac{\partial \ell(\delta, \varphi|\omega)}{\partial \omega} = \varphi^{-1} \mathbf{D} \mathbf{D}_r (\mathbf{y} - \mu),$$

em que  $\mathbf{D}$ ,  $\mathbf{y}$  e  $\mu$  são tais como definidos como na Seção 3.4.1 e  $\mathbf{D}_r = \text{diag}\{r_{m+1}, \dots, r_n\}$  com  $r_t = \sum_{j=1}^q (y_{t-j} - \mu_{t-j})$ , isto é,  $\mathbf{r} = \mathbf{B}\mathbf{1}$  em que  $\mathbf{1} = (1, \dots, 1)^\top$ .

No diagnóstico de curvatura,

$$\Delta|\omega_0 = \begin{bmatrix} \frac{\partial^2 \ell(\delta, \varphi|\omega)}{\partial \delta \partial \omega} \\ \frac{\partial^2 \ell(\delta, \varphi|\omega)}{\partial \varphi \partial \omega} \end{bmatrix} = \begin{bmatrix} -\varphi^{-1} \mathbf{O}^\top \mathbf{D}_a \mathbf{D}_r \\ 2\varphi^{-2} \mathbf{b}^\top \mathbf{D}_r \end{bmatrix},$$

em que a matriz  $\mathbf{O}$  é definida na Seção 2.4 e a matriz  $\mathbf{D}_a$  e o vetor  $\mathbf{b}$  são definidos como na Seção 3.4.1.

Valores de  $W_g(u_t)$  e  $W'_g(u_t)$  para algumas distribuições simétricas são apresentados na Tabela 2.3 do Capítulo 2. Em particular, para distribuição normal temos que  $W_g(u_t) = -0,5$  e  $W'_g(u_t) = 0$  e, para distribuição  $t$ -Student com  $\nu$  graus de liberdade,  $W_g(u_t) = -\frac{(v+1)}{2(v+u_t)}$  e  $W'_g(u_t) = \frac{(v+1)}{2(v+u_t)^2}$ .

### 3.5 Simulações

As simulações apresentadas nessa seção têm o objetivo de estudar o comportamento das marcas de referência, obtidas por meio de limiares, para séries temporais autorregressivas de primeira ordem na presença de observações atípicas. O estudo foi realizado considerando o ajuste do modelo SYMARMA sob distribuições normal e  $t$ -Student com  $\nu = 4$ . Por possuir caudas mais pesadas, esperamos que os pontos atípicos não exerçam grande influência no ajuste do modelo sob distribuição  $t$ -Student.

O modelo SYMARMA autorregressivo de primeira ordem sem variáveis explicativas é dado por

$$y_t = \phi y_{t-1} + r_t, \quad t = 2, \dots, n,$$

em que os  $r_t$ 's são ruídos branco, isto é, as variáveis aleatórias  $r_t$  são independente e identicamente distribuídas com  $E(r_t) = 0$  e  $\text{Var}(r_t) = \sigma^2$ , consequentemente  $\text{Cov}(r_t, r_s) = 0$  para  $t \neq s$ . No modelo SYMARMA consideramos que  $y_t | H_{t-1} \sim S(0, \sigma^2, g)$ , em que  $H_{t-1} = \{y_{t-1}, \dots, y_1, \mu_{t-1}, \dots, \mu_1\}$ . De acordo com os resultados apresentados na Seção 2.3.1, podemos mostrar que a função de autocorrelação ( $\rho$ ) e autocorrelação parcial ( $\lambda$ ) são expressas, respectivamente, por

$$\rho_k = \begin{cases} 1 & , k = 0 \\ \phi^k & , k \neq 0 \end{cases} \quad \text{e} \quad \lambda_k = \begin{cases} \phi & , k = 1 \\ 0 & , k > 1 \end{cases},$$

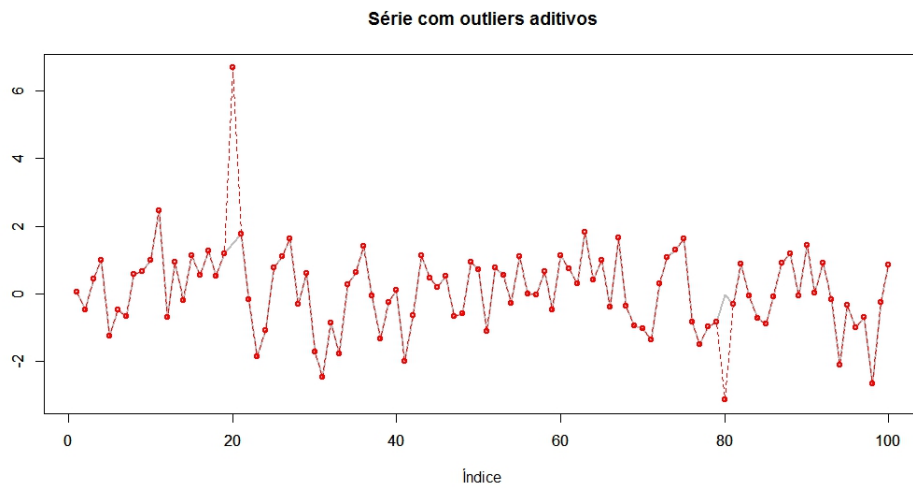
em que  $k$  é o número de defasagens. Dessa forma,  $\phi$  pode ser visto como parâmetro de correlação.

Foram geradas quatro séries temporais com diferentes combinações para  $\phi$  ( $\phi = 0,3$ ,  $\phi = 0,5$ ,  $\phi = 0,7$  e  $\phi = 0,9$ ),  $\varphi = 1$  e tamanho amostral  $n = 100$ . As séries foram escolhidas de tal maneira que as estimativas de  $\phi$  estivessem próximas aos verdadeiros valores,  $|\hat{\phi} - \phi| < 0,001$ . Uma vez simuladas as séries, *outliers* do tipo aditivo foram introduzidos nas observações de ordem  $y_{20}$  e  $y_{80}$  com magnitudes de cinco e três desvios padrões ( $\sigma_y$ ), respectivamente, conforme descrito na Tabela 3.1. *Outliers* aditivos caracterizam-se por afetar a série em apenas uma observação. Para ilustrar o efeito de *outliers* aditivos na série, apresentamos na Figura 3.1 uma série autorregressiva de ordem 1 com  $\phi = 0,3$  sujeita a esse tipo de perturbação. Percebemos que os valores da série mantêm-se inalterados exceto o das observações  $t = 20$  e  $t = 80$  que tiveram os seus valores reais acrescidos de cinco e três desvios padrões, respectivamente, tornando-se assim observações atípicas. A Tabela 3.2 apresenta os valores das séries perturbadas e auxilia na análise das observações influentes.

Os limiares foram calculados utilizando uma rotina desenvolvida pelo autor na linguagem computacional R; [www.R-project.org](http://www.R-project.org). Supondo conhecidos os parâmetros que indexam o modelo, baseado em 2.000 réplicas de Monte Carlo, foram calculadas as marcas de referência  $BS_i$ ,  $BC_i^{Cook}$  e  $BC_i^{LV}$  para  $i = 0, 1$  e  $2$ . Nas Tabelas 3.3 - 3.8 apresentamos os resultados para o esquema de perturbação nos dados e nas Tabelas 3.9 - 3.14 temos os resultados para o esquema de perturbação no parâmetro de dispersão. Nessa etapa foram consideradas amostras de tamanhos  $n = 100$  e  $200$ .

**Tabela 3.1** Configurações das séries simuladas com  $r_t \sim N(0, 1)$ .

	Tamanho amostral	Coefficiente de correlação	Posição dos <i>outliers</i>	Magnitude dos <i>outliers</i> em valor absoluto
Série 1	$n = 100$	$\phi = 0,3$	$y_{20}$ e $y_{80}$	$5\sigma_y$ e $3\sigma_y$
Série 2	$n = 100$	$\phi = 0,5$	$y_{20}$ e $y_{80}$	$5\sigma_y$ e $3\sigma_y$
Série 3	$n = 100$	$\phi = 0,7$	$y_{20}$ e $y_{80}$	$5\sigma_y$ e $3\sigma_y$
Série 4	$n = 100$	$\phi = 0,9$	$y_{20}$ e $y_{80}$	$5\sigma_y$ e $3\sigma_y$

**Figura 3.1** Série autorregressiva de ordem 1 simulada como  $\phi = 0,3$ , perturbada (linha descontínua) com *outliers* aditivos.

Fonte: Autoria própria.

Nas Tabelas 3.3 e 3.6, os resultados obtidos sob o esquema de perturbação nos dados mostram que, para ambas distribuições consideradas (normal ou  $t_4$ ), quando a correlação ( $\phi$ ) e o percentil são mantidos fixos, os limiares para o diagnóstico de inclinação, baseados no critério global  $BS_0$ , aumentam quando o tamanho amostral aumenta. No entanto, para o diagnóstico de curvatura, quando mantemos constantes o percentil e  $\phi$ , os limiares  $BC_0$  decrescem quando o tamanho amostral aumenta. Isto é, as marcas de referência baseadas no critério global variam com o tamanho amostral. Por exemplo, considerando o percentil 95% e  $\phi = 0,5$  a marca de referência  $BS_0$  passa de 11,3460 para 15,4247 quando o tamanho amostral aumenta de  $n = 100$  para  $n = 200$ . Para tamanhos amostrais e percentis fixos, a variação de  $\phi$  não altera o compor-

**Tabela 3.2** Introdução de *outliers* nas séries simuladas.

	$y_{18}$	$y_{19}$	$y_{20}$	$y_{21}$	$y_{22}$	$\omega_y = 5\sigma_y$	$\tilde{y}_{20} = y_{20} + \omega_y$
Série 1	0,5167	1,2029	1,5022	1,7611	-0,1634	5,1780	6,6802
Série 2	-1,1541	-0,6133	-0,1234	2,8437	1,9235	6,8271	-6,9505
Série 3	-0,3967	-1,3079	-1,6943	-0,2116	-0,6658	6,6796	-8,3739
Série 4	-0,3810	0,3035	-1,0938	-1,1296	-0,7429	12,0751	-13,1689

---

	$y_{78}$	$y_{79}$	$y_{80}$	$y_{81}$	$y_{82}$	$\omega_y = 3\sigma_y$	$\tilde{y}_{80} = y_{80} + \omega_y$
Série 1	-0,9767	-0,8202	-0,0236	-0,3030	0,8970	3,1068	-3,1304
Série 2	1,2492	3,2600	2,2446	-0,5822	1,0099	4,0963	6,3409
Série 3	1,7929	1,1222	1,7959	-0,4767	-1,0530	4,0078	5,8037
Série 4	1,3107	2,9523	1,3095	0,2941	-1,2203	7,2451	8,5546

tamento dos limiares  $BS_0$  e  $BC_0$ . Portanto, existem indícios de que os limiares para o critério global não são sensíveis a variação de  $\phi$ . Por fim, de uma maneira geral, os limiares sob distribuição  $t_4$  são menores que os limiares sob distribuição normal. Resultados similares podem ser observados para o esquema de perturbação no parâmetro de dispersão (ver Tabelas 3.9 e 3.12).

Considerando agora os limiares  $BS_1$ ,  $BC_1^{Cook}$  e  $BC_1^{LV}$ , baseados no critério individual, os resultados apresentados nas Tabelas 3.4, 3.7, 3.10 e 3.13 mostram que, fixando o parâmetro  $\phi$  e o percentil, os limiares  $BC_1^{Cook}$  e  $BC_1^{LV}$  decrescem quando o tamanho amostral aumenta. Para as marcas de referência  $BS_1$  temos um comportamento praticamente constante. Considerando agora os valores de  $n$  e o percentil fixos, os limiares sob distribuição normal permanecem uniformes quando a correlação  $\phi$  aumenta, ou seja, há indícios de que o critério  $BS_1$ ,  $BC_1^{Cook}$  e  $BC_1^{LV}$  sejam robustos quanto a variação de  $\phi$ . No entanto, para distribuição  $t_4$ , o limiar  $BC_1^{LV}$  tem comportamento monótono decrescente. Por exemplo, para o esquema de perturbação nos dados (ver Tabela 3.7), considerando o percentil 95% e o tamanho amostral  $n = 200$ , variando  $\phi$  de 0,3 a 0,9 o valor do limiar  $BC_1^{LV}$  passa de 0,7495 para 0,2726. Portanto, existem indícios de que os limiares para o diagnóstico de curvatura sob o enfoque de Lesaffre e Verbeke, baseados no critério individual, são mais sensíveis a variação de  $\phi$ .

Nas Tabelas 3.5 e 3.11, para ambos esquemas de perturbação sob distribuição normal, se a correlação aumenta, mantendo fixos o percentil e tamanho amostral, os limiares baseados nos

critérios individuais  $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$  têm suas magnitudes estáveis; sob distribuição  $t_4$  (ver Tabelas 3.8 e 3.14) o limiar  $BC_2^{LV}$  tem comportamento monótono decrescente.

As Tabelas 3.15 e 3.17 apresentam as estatísticas  $O_s$  e  $O_c$  para o esquema de perturbação nos dados, calculadas para as séries de 1 a 4 considerando o ajuste dos modelos SYMARMA-normal e SYMARMA- $t_4$ , respectivamente. Considerando os critérios  $BS_0$  e  $BC_0$ , avaliados através do percentil 95% (ver Tabelas 3.3 e 3.6), para as quatro séries geradas, não existem indícios de influência global. Esse fato não descarta a hipótese de alguma observações exercer influência de forma individual. Os resultados para o esquema de perturbação no parâmetro de dispersão são apresentados nas Tabelas 3.16 e 3.18 para o ajuste dos modelos SYMARMA-normal e SYMARMA- $t_4$ , respectivamente. Nesse caso, para o modelo sob distribuição normal as estatísticas  $O_s$  e  $O_c$  são maiores que as marcas de referência  $BS_0$  e  $BC_0$ , respectivamente, dando indícios de influência global. Por exemplo, para série 1  $O_s = 14,0948$  e  $O_c = 9,5888$  são maiores que as marcas de referência  $BS_0 = 8,3364$  e  $BC_0 = 3,9331$  apresentadas na Tabela 3.9. Quando consideramos o ajuste sob distribuição  $t_4$  não encontramos indícios de influência global nas séries de 1 a 4.

Os diagnósticos de influência individual para as séries de 1 a 4, avaliados pelos limiares  $BS_i$ ,  $BC_i^{Cook}$  e  $BC_i^{LV}$  para  $i = 1$  e  $2$ , são ilustrados nas Figuras 3.2-3.9 sob o esquema de perturbação nos dados e nas Figuras 3.10-3.17 sob o esquema de perturbação no parâmetro de dispersão. Em cada uma das figuras são apresentados quatro gráficos sendo o primeiro (na parte superior) correspondente à série com *outliers*, o segundo correspondente aos valores obtidos na direção da inclinação de Billor e Loynes, o terceiro correspondente aos valores obtidos na direção da curvatura de Cook e, por fim, o quarto (na parte inferior) correspondente aos valores obtidos na direção da curvatura de Lesaffre e Verbeke.

Discutimos inicialmente os resultados obtidos a partir da perturbação nos dados. Para a série 1, ilustrada nas Figuras 3.2 e 3.3 considerando o ajuste dos modelos SYMARMA-normal e SYMARMA- $t_4$ , respectivamente, notamos que o *outlier* gerado na observação de ordem 20,  $\tilde{y}_{20} = 6,6802$ , é identificado como influente pelos critérios  $BS_2$  e  $BC_2^{Cook}$  apenas no modelo SYMARMA-normal. O mesmo resultado ocorre nas demais séries (ver Figuras 3.4-3.9). Considerando o critério  $BC_2^{LV}$  o *outlier*  $\tilde{y}_{20}$  foi identificado como influente por ambos modelos

apenas na série 1. Baseado nas Figuras 3.2-3.9 podemos notar que no modelo SYMARMA-normal, as observações adjacentes ao *outlier* apresentam valores para os vetores de diagnóstico de inclinação (**d**) e de curvatura (**C**) que tendem a caracterizá-las como influentes, quando na verdade não foram geradas como tais. Enquanto que no ajuste do modelo SYMARMA- $t_4$  esse fato não foi observado. Notamos que a observação  $\tilde{y}_{80}$ , gerada como *outlier*, também foi identificada como influente pelo critério  $BC_2^{Cook}$  no ajuste do modelo SYMARMA-normal.

Para o esquema de perturbação no parâmetro de dispersão (ver Figuras 3.10-3.17), resultados similares aos apresentados para a perturbação nos dados podem ser obtidos. Conforme podemos notar, quando consideramos o ajuste sob distribuição  $t_4$  não encontramos indícios de influência dos *outliers* gerados nas séries de 1 a 4.

**Tabela 3.3** Estatísticas globais  $BS_0$  e  $BC_0$  para o modelo SYMARMA-normal sob o esquema de perturbação nos dados.

Tamanho amostral	Medida	Percentis (%)	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_0$	90	10,9922	10,9910	11,0292	11,0292
		95	11,3119	11,3108	11,3460	11,3628
		99	11,9575	11,9461	11,8294	11,9535
	$BC_0$	90	4,8820	4,8809	4,9148	4,9149
		95	5,1700	5,1691	5,2013	5,2167
		99	5,7771	5,7660	5,6539	5,7732
$n = 200$	$BS_0$	90	15,1428	15,1190	15,1316	15,1432
		95	15,4649	15,4129	15,4247	15,4216
		99	15,9448	15,9631	15,9411	15,8911
	$BC_0$	90	4,6091	4,5947	4,6023	4,6094
		95	4,8073	4,7750	4,7823	4,7804
		99	5,1103	5,1220	5,1079	5,0759

**Tabela 3.4** Estatísticas individuais  $BS_1$ ,  $BC_1^{Cook}$  e  $BC_1^{LV}$  para o modelo SYMARMA-normal sob o esquema de perturbação nos dados.

Tamanho amostral	Medida	Percentis (%)	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_1$	90	6,6985	6,6755	6,6347	6,6552
		95	7,1035	7,0751	7,0698	7,0832
		99	7,9590	7,9537	7,8778	7,8975
	$BC_1^{Cook}$	90	0,3241	0,3238	0,3254	0,3227
		95	0,3430	0,3410	0,3431	0,3414
		99	0,3767	0,3810	0,3816	0,3719
	$BC_1^{LV}$	90	0,4763	0,4759	0,4727	0,4686
		95	0,5332	0,5269	0,5384	0,5350
		99	0,6717	0,6783	0,6761	0,6677
$n = 200$	$BS_1$	90	6,9974	7,0598	7,0333	7,0423
		95	7,3930	7,3968	7,3949	7,4555
		99	8,1134	8,1800	8,0694	8,0928
	$BC_1^{Cook}$	90	0,2435	0,2438	0,2451	0,2459
		95	0,2560	0,2562	0,2562	0,2581
		99	0,2806	0,2800	0,2807	0,2793
	$BC_1^{LV}$	90	0,2584	0,2642	0,2639	0,2633
		95	0,2863	0,2875	0,2899	0,2945
		99	0,3494	0,3491	0,3428	0,3397



**Tabela 3.5** Estatísticas individuais  $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$  para o modelo SYMARMA-normal sob o esquema de perturbação nos dados.

Tamanho amostral	Medida	Estatísticas	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_2$	Mínimo	4,6971	4,7517	4,6668	5,0767
		Quartil (5%)	5,1841	5,2422	5,0741	5,2200
		Máximo	8,5345	10,2087	10,0126	8,7458
		Média	6,4618	6,5726	6,3427	6,4850
		Desvio Padrão	0,9003	0,9138	0,9290	0,8891
	$BC_2^{Cook}$	Mínimo	0,2076	0,2093	0,2050	0,2178
		Quartil (5%)	0,2240	0,2307	0,2195	0,2246
		Máximo	0,3670	0,4464	0,4232	0,3768
		Média	0,2760	0,2812	0,2721	0,2762
		Desvio Padrão	0,0369	0,0389	0,0396	0,0364
	$BC_2^{LV}$	Mínimo	0,2635	0,2457	0,2961	0,2862
		Quartil (5%)	0,3213	0,3136	0,3042	0,3293
		Máximo	0,8569	1,0587	1,0380	0,8370
		Média	0,4673	0,4781	0,4460	0,4706
		Desvio Padrão	0,1279	0,1332	0,1238	0,1219
$n = 200$	$BS_2$	Mínimo	5,2206	5,1500	5,2076	5,3464
		Quartil (5%)	5,6882	5,6897	5,5964	5,6226
		Máximo	9,0150	8,1973	8,9544	9,0096
		Média	6,6356	6,6351	6,6673	6,5779
		Desvio Padrão	0,7074	0,6692	0,8535	0,7324
	$BC_2^{Cook}$	Mínimo	0,1666	0,1668	0,1664	0,1707
		Quartil (5%)	0,1817	0,1823	0,1745	0,1799
		Máximo	0,2892	0,2627	0,2849	0,2877
		Média	0,2107	0,2106	0,2117	0,2094
		Desvio Padrão	0,0226	0,0205	0,0267	0,0230
	$BC_2^{LV}$	Mínimo	0,1582	0,1343	0,1557	0,1578
		Quartil (5%)	0,1760	0,1722	0,1697	0,1658
		Máximo	0,4421	0,3491	0,4035	0,4097
		Média	0,2410	0,2408	0,2421	0,2370
		Desvio Padrão	0,0500	0,0477	0,0587	0,0498

**Tabela 3.6** Estatísticas globais  $BS_0$  e  $BC_0$  para o modelo SYMARMA- $t_4$  sob o esquema de perturbação nos dados.

Tamanho amostral	Medida	Percentis (%)	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_0$	90	9,6469	9,5862	9,5885	9,6158
		95	9,9725	9,9704	9,9410	9,9462
		99	10,6684	10,5475	10,5319	10,6044
	$BC_0$	90	3,1879	3,1806	3,1631	3,1992
		95	3,4431	3,4104	3,3897	3,3884
		99	3,9204	3,8474	3,8076	3,8643
$n = 200$	$BS_0$	90	13,0835	13,0703	13,0630	13,0336
		95	13,3938	13,3919	13,3617	13,3973
		99	14,0754	13,9449	13,9826	13,9035
	$BC_0$	90	2,9176	2,9154	2,9161	2,8983
		95	3,0596	3,0521	3,0476	3,0413
		99	3,3563	3,2719	3,3361	3,2533

**Tabela 3.7** Estatísticas individuais  $BS_1$ ,  $BC_1^{Cook}$  e  $BC_1^{LV}$  para o modelo SYMARMA- $t_4$  sob o esquema de perturbação nos dados.

Tamanho amostral	Medida	Percentis (%)	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_1$	90	2,8430	2,8516	2,8405	2,8530
		95	2,9746	2,9553	2,9423	2,9512
		99	3,1502	3,1280	3,1205	3,1234
	$BC_1^{Cook}$	90	0,1929	0,1889	0,1810	0,1690
		95	0,2267	0,2176	0,2028	0,1812
		99	0,3235	0,3188	0,3064	0,2153
	$BC_1^{LV}$	90	0,7994	0,7093	0,5697	0,3781
		95	1,0589	0,9454	0,7202	0,4407
		99	1,6641	1,3423	1,0883	0,5653
$n = 200$	$BS_1$	90	2,7383	2,7364	2,7372	2,7289
		95	2,8039	2,7957	2,7960	2,7917
		99	2,9303	2,9187	2,9264	2,8968
	$BC_1^{Cook}$	90	0,1278	0,1232	0,1179	0,1143
		95	0,1409	0,1357	0,1273	0,1199
		99	0,1982	0,1871	0,1548	0,1388
	$BC_1^{LV}$	90	0,5808	0,4971	0,3970	0,2312
		95	0,7495	0,6486	0,4982	0,2726
		99	1,2881	1,0064	0,7715	0,4022

**Tabela 3.8** Estatísticas individuais  $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$  para o modelo SYMARMA- $t_4$  sob o esquema de perturbação nos dados.

Tamanho amostral	Medida	Estatísticas	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_2$	Mínimo	2,8732	2,8450	2,8310	2,8605
		Quartil (5%)	2,9123	2,9223	2,8971	2,9024
		Máximo	3,3043	3,5604	3,5350	3,3013
		Média	3,0600	3,0525	3,0472	3,0465
		Desvio Padrão	0,0963	0,1129	0,1162	0,0971
	$BC_2^{Cook}$	Mínimo	0,1290	0,1270	0,1263	0,1302
		Quartil (5%)	0,1336	0,1355	0,1338	0,1353
		Máximo	0,3418	0,2437	0,6134	0,1903
		Média	0,1559	0,1524	0,1651	0,1499
		Desvio Padrão	0,0309	0,0208	0,0597	0,0125
	$BC_2^{LV}$	Mínimo	0,2494	0,2403	0,2005	0,1535
		Quartil (5%)	0,3239	0,2685	0,2697	0,2098
		Máximo	2,4110	2,3235	1,9717	0,6831
		Média	0,7027	0,6131	0,4970	0,3551
		Desvio Padrão	0,4475	0,3838	0,2723	0,1121
$n = 200$	$BS_2$	Mínimo	2,6861	2,7293	2,7173	2,7035
		Quartil (5%)	2,7722	2,7579	2,7657	2,7582
		Máximo	3,1759	3,3020	3,1348	3,3442
		Média	2,8736	2,8689	2,8681	2,8567
		Desvio Padrão	0,0897	0,0932	0,0841	0,1032
	$BC_2^{Cook}$	Mínimo	0,0931	0,0956	0,0951	0,0942
		Quartil (5%)	0,0957	0,0977	0,0962	0,0961
		Máximo	0,2527	0,1991	0,1659	0,1430
		Média	0,1081	0,1089	0,1050	0,1035
		Desvio Padrão	0,0185	0,0162	0,0106	0,0072
	$BC_2^{LV}$	Mínimo	0,1140	0,1172	0,1096	0,0936
		Quartil (5%)	0,1851	0,1614	0,1428	0,1122
		Máximo	2,2744	2,1545	1,1741	0,5267
		Média	0,4329	0,4062	0,3004	0,1977
		Desvio Padrão	0,2801	0,2947	0,1616	0,0866

**Tabela 3.9** Estatísticas globais  $BS_0$  e  $BC_0$  para o modelo SYMARMA-normal sob o esquema de perturbação no parâmetro de dispersão.

Tamanho amostral	Medida	Percentis (%)	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_0$	90	7,9144	7,9012	7,9720	7,8618
		95	8,3364	8,3285	8,3528	8,2347
		99	9,2219	9,1204	9,1410	9,0863
	$BC_0$	90	3,6705	3,6468	3,6991	3,6327
		95	3,9331	3,9742	3,9779	3,9032
		99	4,6249	4,5941	4,6411	4,5363
$n = 200$	$BS_0$	90	10,9668	10,9445	10,9464	10,9537
		95	11,2842	11,3638	11,3173	11,3081
		99	12,0953	12,0338	12,0998	12,1682
	$BC_0$	90	3,4803	3,4833	3,4779	3,4913
		95	3,6229	3,6492	3,6563	3,6446
		99	4,0380	4,0941	4,0335	4,0671

**Tabela 3.10** Estatísticas individuais  $BS_1$ ,  $BC_1^{Cook}$  e  $BC_1^{LV}$  para o modelo SYMARMA-normal sob o esquema de perturbação no parâmetro de dispersão.

Tamanho amostral	Medida	Percentis (%)	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_1$	90	9,4022	9,3776	9,4844	9,3091
		95	10,6505	10,5133	10,6511	10,5380
		99	13,0501	13,3724	13,4190	12,6927
	$BC_1^{Cook}$	90	0,6065	0,6013	0,6067	0,6040
		95	0,6546	0,6457	0,6626	0,6573
		99	0,7324	0,7415	0,7492	0,7402
	$BC_1^{LV}$	90	1,3522	1,3584	1,3829	1,3519
		95	1,6897	1,6034	1,7119	1,6484
		99	2,3631	2,4439	2,5742	2,3684
$n = 200$	$BS_1$	90	10,7993	10,8320	10,9561	11,0309
		95	12,0416	12,0650	12,0619	12,2592
		99	14,6723	14,5972	14,6762	14,5303
	$BC_1^{Cook}$	90	0,4866	0,4902	0,4876	0,4957
		95	0,5302	0,5273	0,5291	0,5372
		99	0,6001	0,6119	0,6141	0,6083
	$BC_1^{LV}$	90	0,8494	0,8697	0,8766	0,9035
		95	1,0167	1,0297	1,0353	1,0663
		99	1,4518	1,4556	1,4839	1,3948

**Tabela 3.11** Estatísticas individuais  $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$  para o modelo SYMARMA-normal sob o esquema de perturbação no parâmetro de dispersão.

Tamanho amostral	Medida	Estatísticas	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_2$	Mínimo	6,9378	7,6498	6,9198	7,2294
		Quartil (5%)	8,4962	8,7358	8,3266	8,1135
		Máximo	17,2156	18,7294	19,4608	16,3588
		Média	11,6099	11,5968	11,6695	11,3375
		Desvio Padrão	1,9704	2,1599	2,3497	1,9671
	$BC_2^{Cook}$	Mínimo	0,4700	0,4456	0,4430	0,4371
		Quartil (5%)	0,5106	0,5157	0,5199	0,4892
		Máximo	0,8221	0,8425	0,8569	0,8248
		Média	0,6467	0,6420	0,6614	0,6495
		Desvio Padrão	0,0846	0,0923	0,0977	0,0943
	$BC_2^{LV}$	Mínimo	0,8947	1,0170	0,9327	0,9299
		Quartil (5%)	1,1451	1,1961	1,2291	1,0417
		Máximo	3,9542	4,7055	4,8919	4,4143
		Média	1,9452	1,9521	2,0813	1,9531
		Desvio Padrão	0,6068	0,6518	0,7098	0,6772
$n = 200$	$BS_2$	Mínimo	7,6851	7,8361	7,7808	7,9401
		Quartil (5%)	8,4430	9,1621	8,8751	9,3422
		Máximo	18,9896	19,6263	18,4937	21,5050
		Média	12,4249	12,5647	12,4716	12,4584
		Desvio Padrão	2,5092	2,3805	2,2745	2,3137
	$BC_2^{Cook}$	Mínimo	0,3179	0,3250	0,3466	0,3248
		Quartil (5%)	0,3710	0,3958	0,3888	0,3878
		Máximo	0,7313	0,7210	0,7118	0,7016
		Média	0,5114	0,5103	0,5187	0,5139
		Desvio Padrão	0,0892	0,0845	0,0841	0,0869
	$BC_2^{LV}$	Mínimo	0,4709	0,4183	0,5656	0,4022
		Quartil (5%)	0,5814	0,6813	0,6471	0,6368
		Máximo	2,5337	2,6514	2,1431	2,6191
		Média	1,1213	1,1229	1,1430	1,1466
		Desvio Padrão	0,4107	0,3973	0,3687	0,4102

**Tabela 3.12** Estatísticas globais  $BS_0$  e  $BC_0$  para o modelo SYMARMA- $t_4$  sob o esquema de perturbação no parâmetro de dispersão.

Tamanho amostral	Medida	Percentis (%)	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_0$	90	5,6829	5,6650	5,6740	5,6756
		95	5,7897	5,7780	5,7768	5,7756
		99	6,0030	5,9828	5,9497	5,9964
	$BC_0$	90	1,3029	1,2837	1,2405	1,1897
		95	1,4266	1,3910	1,3466	1,2598
		99	1,7090	1,5766	1,5513	1,4074
$n = 200$	$BS_0$	90	7,9014	7,8953	7,8920	7,9031
		95	8,0245	8,0195	8,0042	8,0039
		99	8,1939	8,1956	8,1892	8,1714
	$BC_0$	90	1,2180	1,2054	1,1634	1,1403
		95	1,3119	1,2804	1,2363	1,1894
		99	1,4822	1,4481	1,3563	1,2844

**Tabela 3.13** Estatísticas individuais  $BS_1$ ,  $BC_1^{Cook}$  e  $BC_1^{LV}$  para o modelo SYMARMA- $t_4$  sob o esquema de perturbação no parâmetro de dispersão.

Tamanho amostral	Medida	Percentis (%)	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_1$	90	3,7370	3,7344	3,7400	3,7315
		95	3,8039	3,7981	3,8007	3,7999
		99	3,9149	3,9274	3,9333	3,9165
	$BC_1^{Cook}$	90	0,5200	0,4971	0,4327	0,3654
		95	0,5826	0,5489	0,4807	0,3981
		99	0,6982	0,6421	0,5843	0,4423
	$BC_1^{LV}$	90	0,3475	0,3162	0,2473	0,1674
		95	0,4539	0,3992	0,3037	0,1942
		99	0,6517	0,5748	0,4326	0,2543
$n = 200$	$BS_1$	90	3,8008	3,8037	3,7988	3,8133
		95	3,8637	3,8649	3,8603	3,8634
		99	3,9429	3,9397	3,9457	3,9431
	$BC_1^{Cook}$	90	0,4449	0,4162	0,3650	0,2942
		95	0,5081	0,4722	0,4073	0,3194
		99	0,6261	0,5900	0,5233	0,3811
	$BC_1^{LV}$	90	0,2475	0,2125	0,1629	0,1047
		95	0,3203	0,2718	0,1998	0,1222
		99	0,5487	0,4797	0,3420	0,1710

**Tabela 3.14** Estatísticas individuais  $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$  para o modelo SYMARMA- $t_4$  sob o esquema de perturbação no parâmetro de dispersão.

Tamanho amostral	Medida	Estatísticas	$\phi$			
			0,3	0,5	0,7	0,9
$n = 100$	$BS_2$	Mínimo	2,9514	3,0435	3,0422	3,0605
		Quartil (5%)	3,3025	3,2850	3,2233	3,1814
		Máximo	3,9697	3,9767	3,9843	3,9846
		Média	3,6042	3,5789	3,5669	3,5555
		Desvio Padrão	0,1971	0,1946	0,2120	0,2303
	$BC_2^{Cook}$	Mínimo	0,2924	0,2914	0,2513	0,2455
		Quartil (5%)	0,3426	0,3184	0,2884	0,2714
		Máximo	0,8894	0,8948	0,7599	0,4909
		Média	0,5178	0,4683	0,4100	0,3431
		Desvio Padrão	0,1140	0,1129	0,0987	0,0540
	$BC_2^{LV}$	Mínimo	0,1538	0,1433	0,0938	0,0903
		Quartil (5%)	0,1903	0,1593	0,1341	0,1145
		Máximo	2,0596	1,3563	1,2403	0,4143
		Média	0,4739	0,3702	0,2878	0,1806
		Desvio Padrão	0,2642	0,2078	0,1772	0,0605
$n = 200$	$BS_2$	Mínimo	3,0182	3,3041	3,2559	3,2032
		Quartil (5%)	3,3608	3,3700	3,3739	3,3545
		Máximo	3,9815	3,9816	3,9589	3,9832
		Média	3,6593	3,6603	3,6486	3,6583
		Desvio Padrão	0,1737	0,1651	0,1550	0,1712
	$BC_2^{Cook}$	Mínimo	0,2338	0,2196	0,2022	0,1869
		Quartil (5%)	0,2832	0,2490	0,2366	0,2165
		Máximo	0,7748	0,8498	0,7313	0,4728
		Média	0,4599	0,4263	0,3636	0,2793
		Desvio Padrão	0,1358	0,1293	0,1174	0,0527
	$BC_2^{LV}$	Mínimo	0,0859	0,0715	0,0559	0,0511
		Quartil (5%)	0,1099	0,0935	0,0793	0,0641
		Máximo	2,4161	1,0280	1,6984	1,2966
		Média	0,3588	0,2873	0,2133	0,1098
		Desvio Padrão	0,2949	0,1831	0,2065	0,0459

**Tabela 3.15** Estatísticas globais  $O_s$  e  $O_c$  e seus respectivos limiares  $BS_0$  e  $BC_0$  para o ajuste do modelo SYMARMA-normal sob o esquema de perturbação nos dados. ( $n = 100$ )

Medida	$\phi$	0,3	0,5	0,7	0,9
$O_s$		8,2845	6,4144	6,8179	4,6936
$O_c$		2,7730	1,6624	1,8781	0,8901
$BS_0$		11,3119	11,3108	11,3460	11,3628
$BC_0$		5,1700	5,1691	5,2013	5,2167

Limiares  $BS_0$  e  $BC_0$  obtidos do percentil 95%.

**Tabela 3.16** Estatísticas globais  $O_s$  e  $O_c$  e seus respectivos limiares  $BS_0$  e  $BC_0$  para o ajuste do modelo SYMARMA-normal sob o esquema de perturbação no parâmetro de dispersão. ( $n = 100$ )

Medida	$\phi$	0,3	0,5	0,7	0,9
$O_s$		<b>14,0948</b>	<b>12,2482</b>	<b>15,8804</b>	<b>21,6032</b>
$O_c$		<b>9,5888</b>	<b>7,4990</b>	<b>11,2089</b>	<b>20,9583</b>
$BS_0$		8,3364	8,3285	8,3528	8,2347
$BC_0$		3,9331	3,9742	3,9779	3,9032

Limiares  $BS_0$  e  $BC_0$  obtidos do percentil 95%.

**Tabela 3.17** Estatísticas globais  $O_s$  e  $O_c$  e seus respectivos limiares  $BS_0$  e  $BC_0$  para o ajuste do modelo SYMARMA- $t_4$  sob o esquema de perturbação nos dados. ( $n = 100$ )

Medida	$\phi$	0,3	0,5	0,7	0,9
$O_s$		9,7934	7,7477	9,2608	8,5628
$O_c$		3,2140	2,1174	3,2287	2,8541
$BS_0$		9,9725	9,9704	9,9410	9,9462
$BC_0$		3,4431	3,4104	3,3897	3,3884

Limiares  $BS_0$  e  $BC_0$  obtidos do percentil 95%.

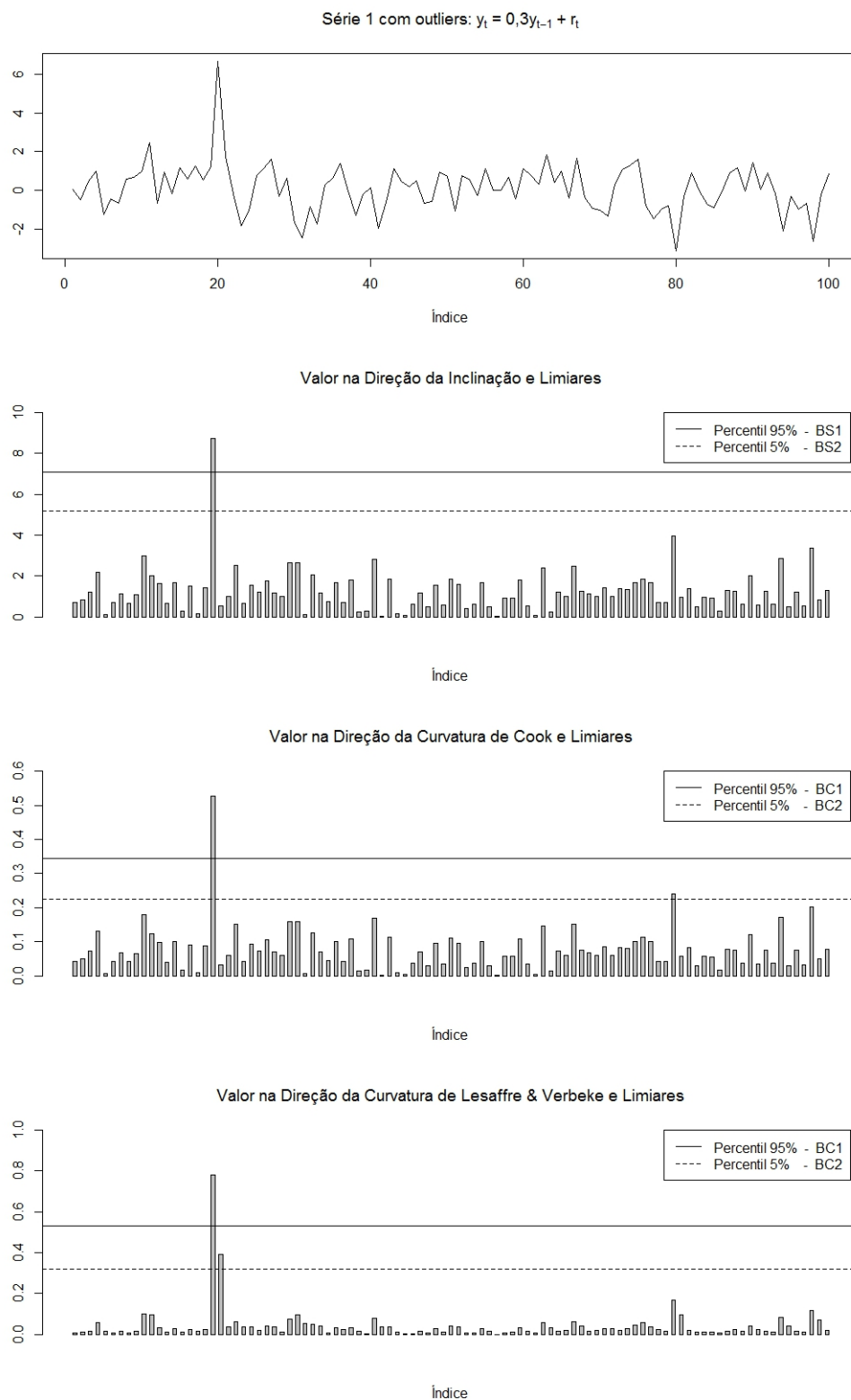
**Tabela 3.18** Estatísticas globais  $O_s$  e  $O_c$  e seus respectivos limiares  $BS_0$  e  $BC_0$  para o ajuste do modelo SYMARMA- $t_4$  sob o esquema de perturbação no parâmetro de dispersão. ( $n = 100$ )

Medida	$\phi$	0,3	0,5	0,7	0,9
$O_s$		4,6160	5,0710	5,4948	5,5152
$O_c$		0,9029	1,4951	0,9975	0,9066
$BS_0$		5,7897	5,7780	5,7768	5,7756
$BC_0$		1,4266	1,3910	1,3466	1,2598

Limiares  $BS_0$  e  $BC_0$  obtidos do percentil 95%.

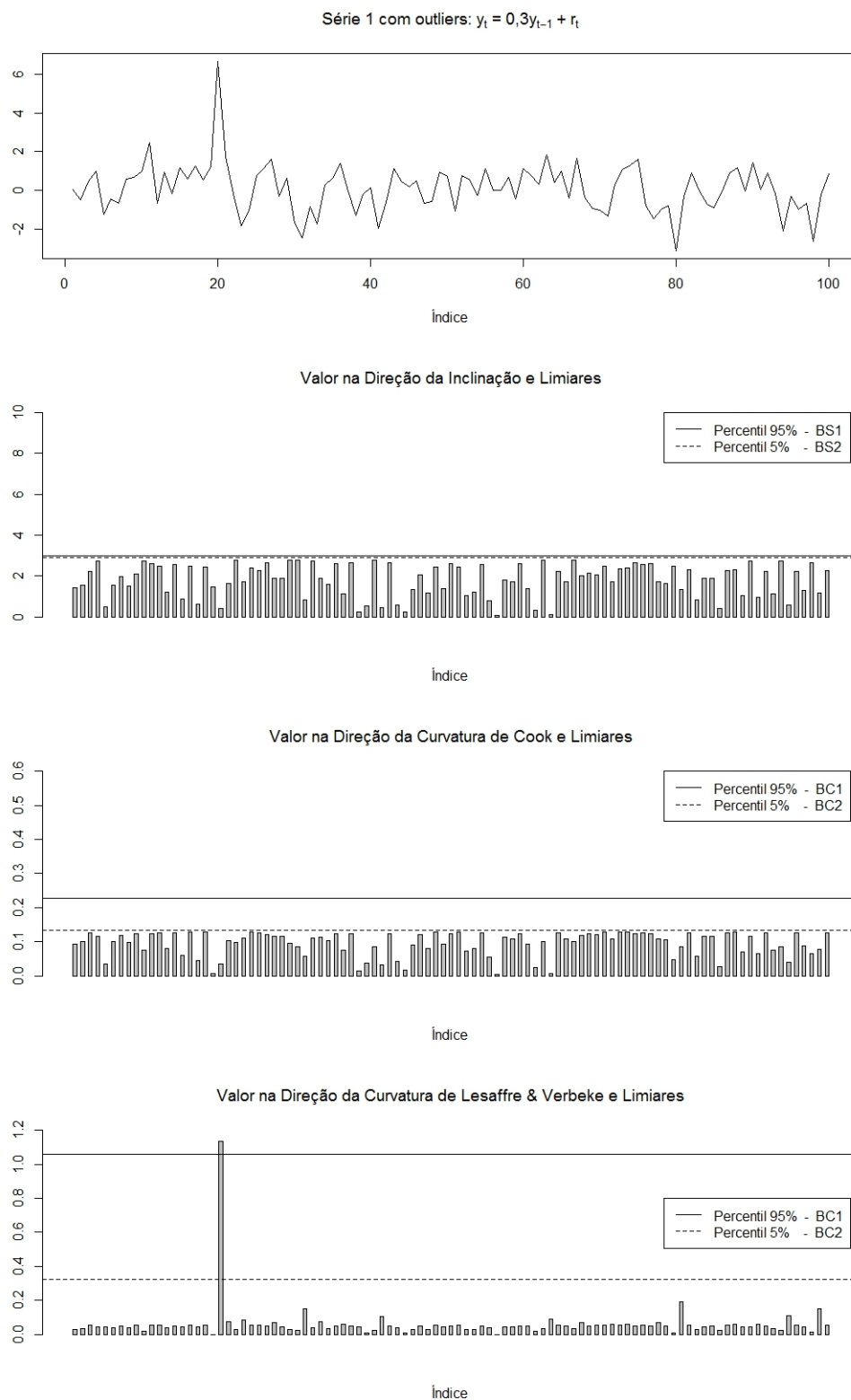


**Figura 3.2** Diagnóstico de influência local na série 1 ( $\phi = 0,3$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA-normal.



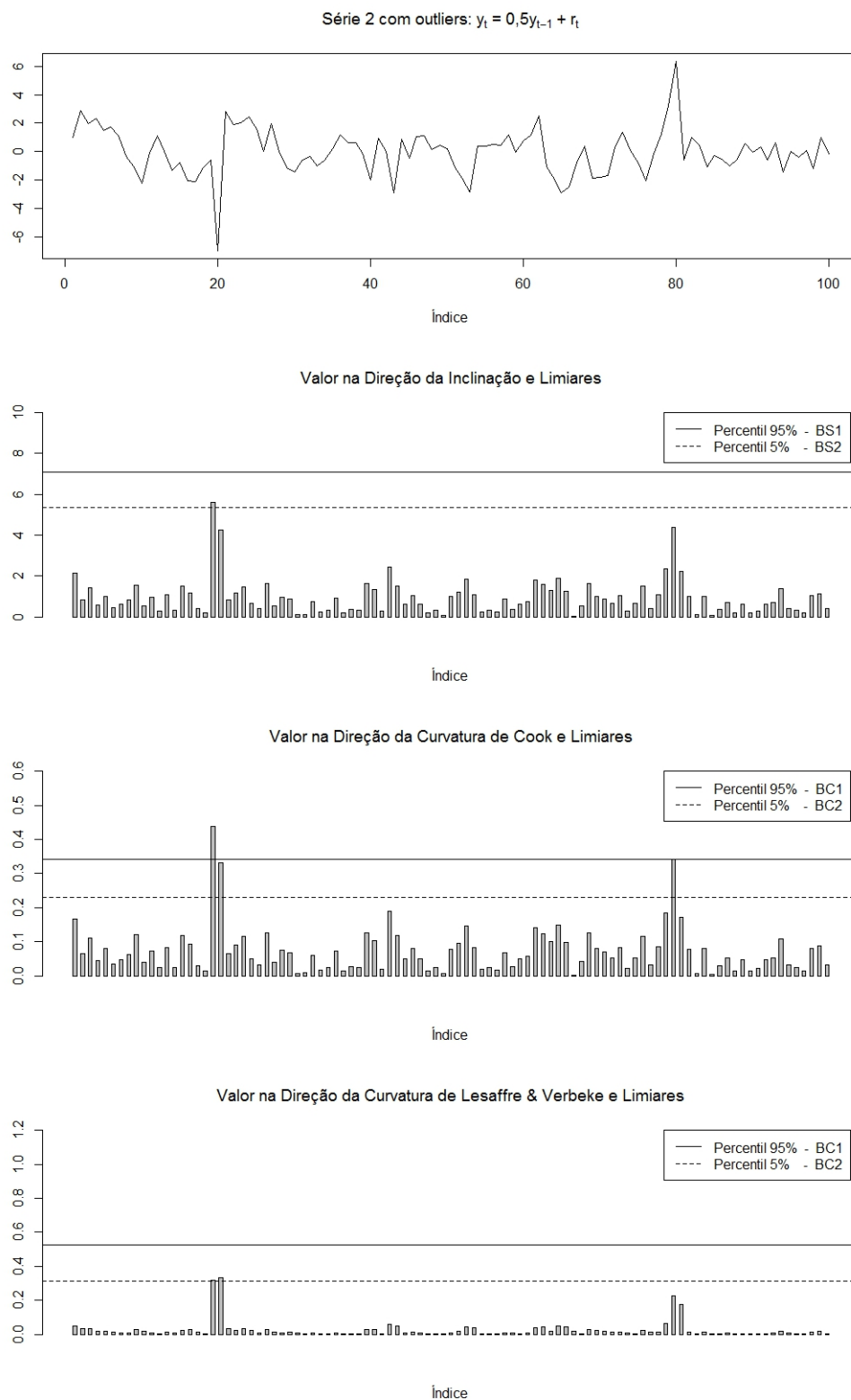
Fonte: Autoria própria.

**Figura 3.3** Diagnóstico de influência local na série 1 ( $\phi = 0,3$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA- $t_4$ .



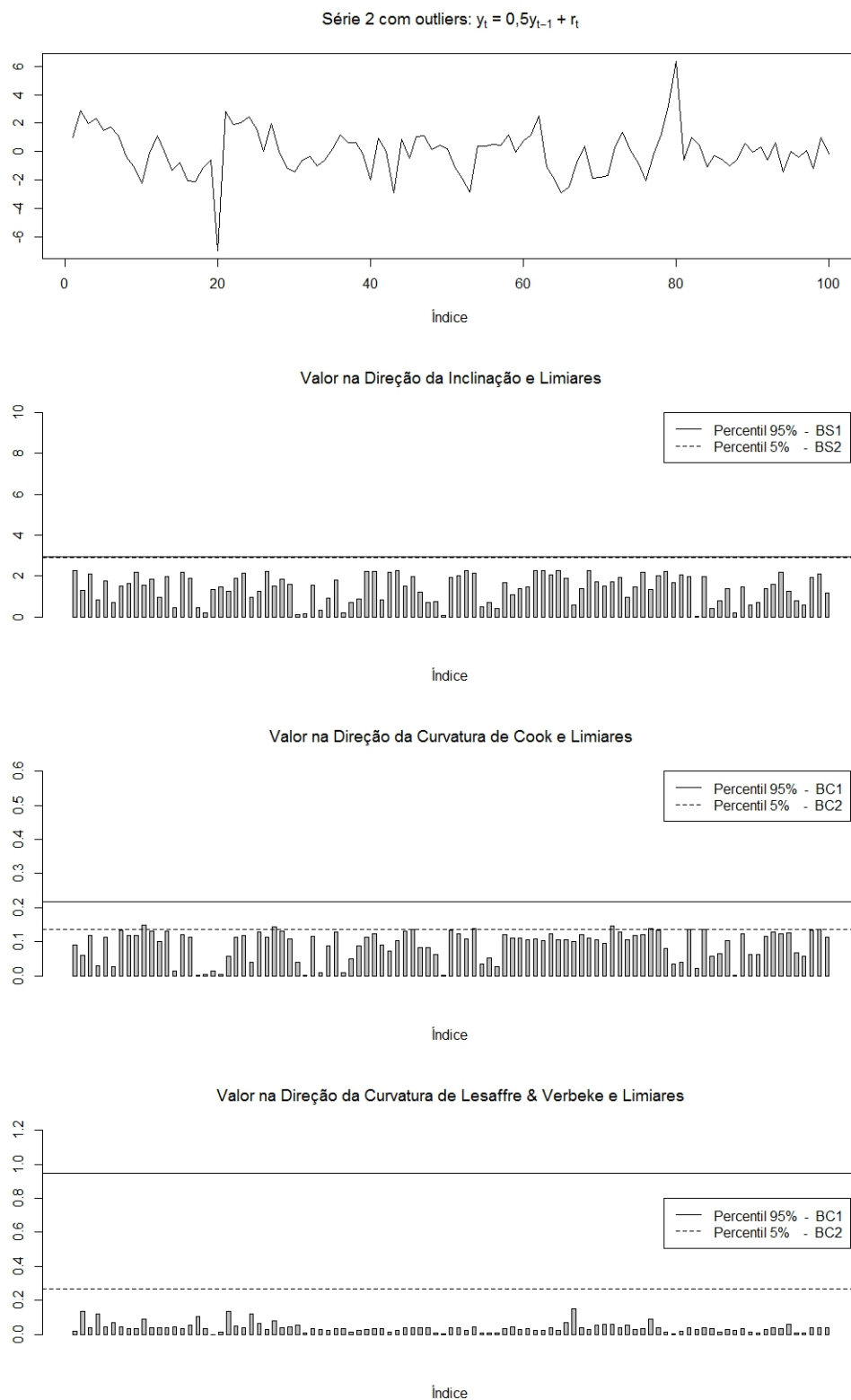
Fonte: Autoria própria.

**Figura 3.4** Diagnóstico de influência local na série 2 ( $\phi = 0,5$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA-normal.



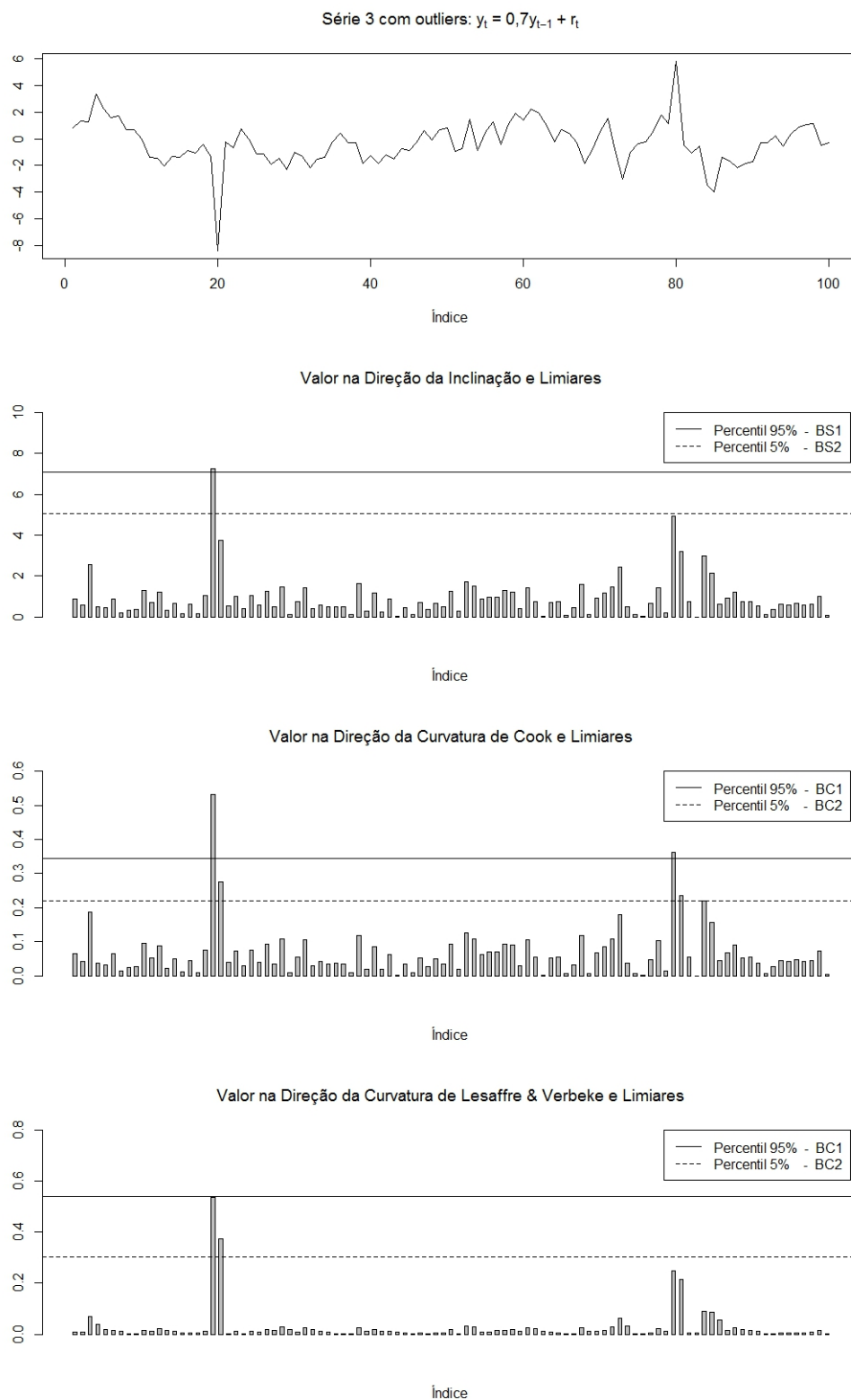
Fonte: Autoria própria.

**Figura 3.5** Diagnóstico de influência local na série 2 ( $\phi = 0,5$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA- $t_4$ .



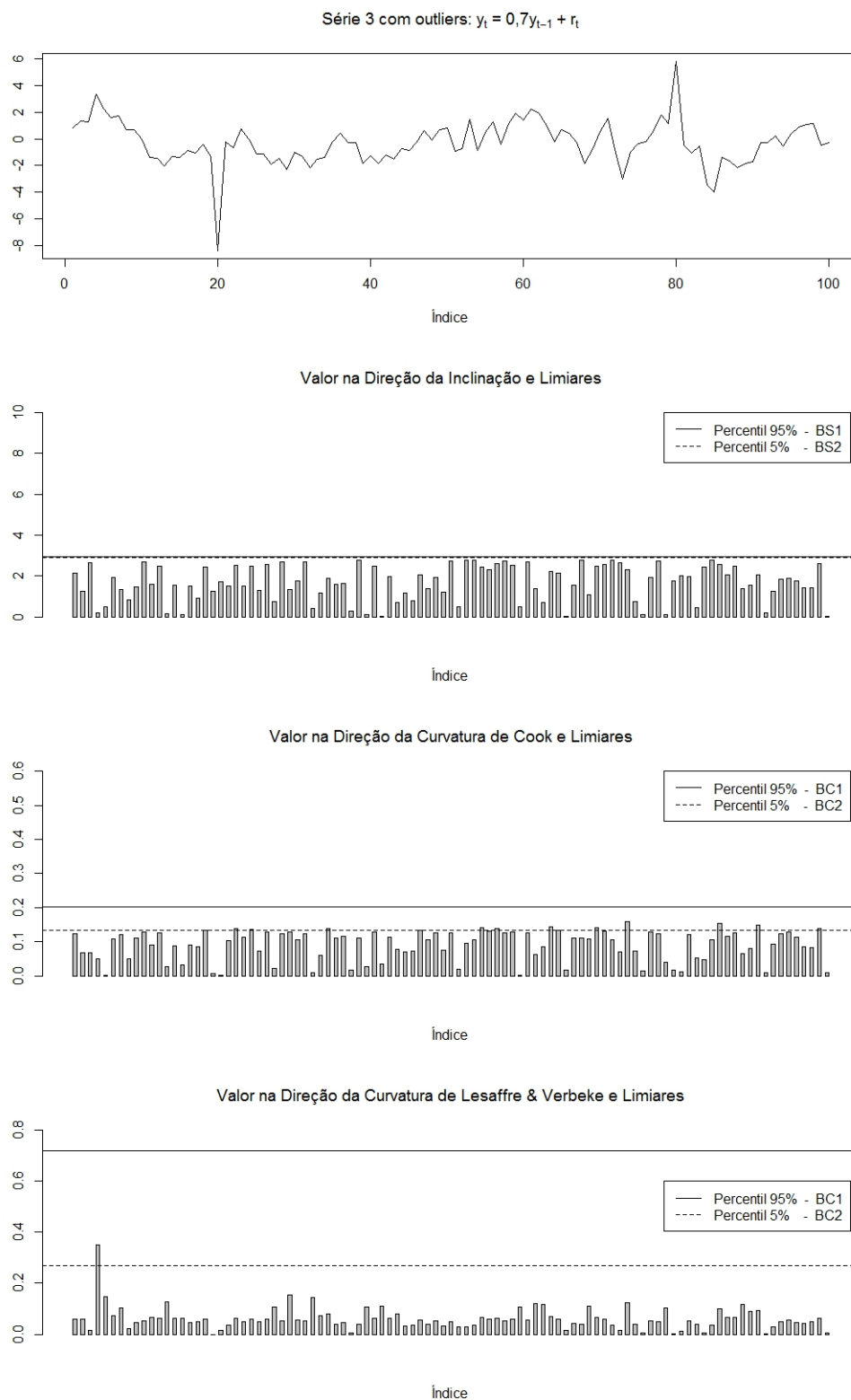
Fonte: Autoria própria.

**Figura 3.6** Diagnóstico de influência local na série 3 ( $\phi = 0,7$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA-normal.



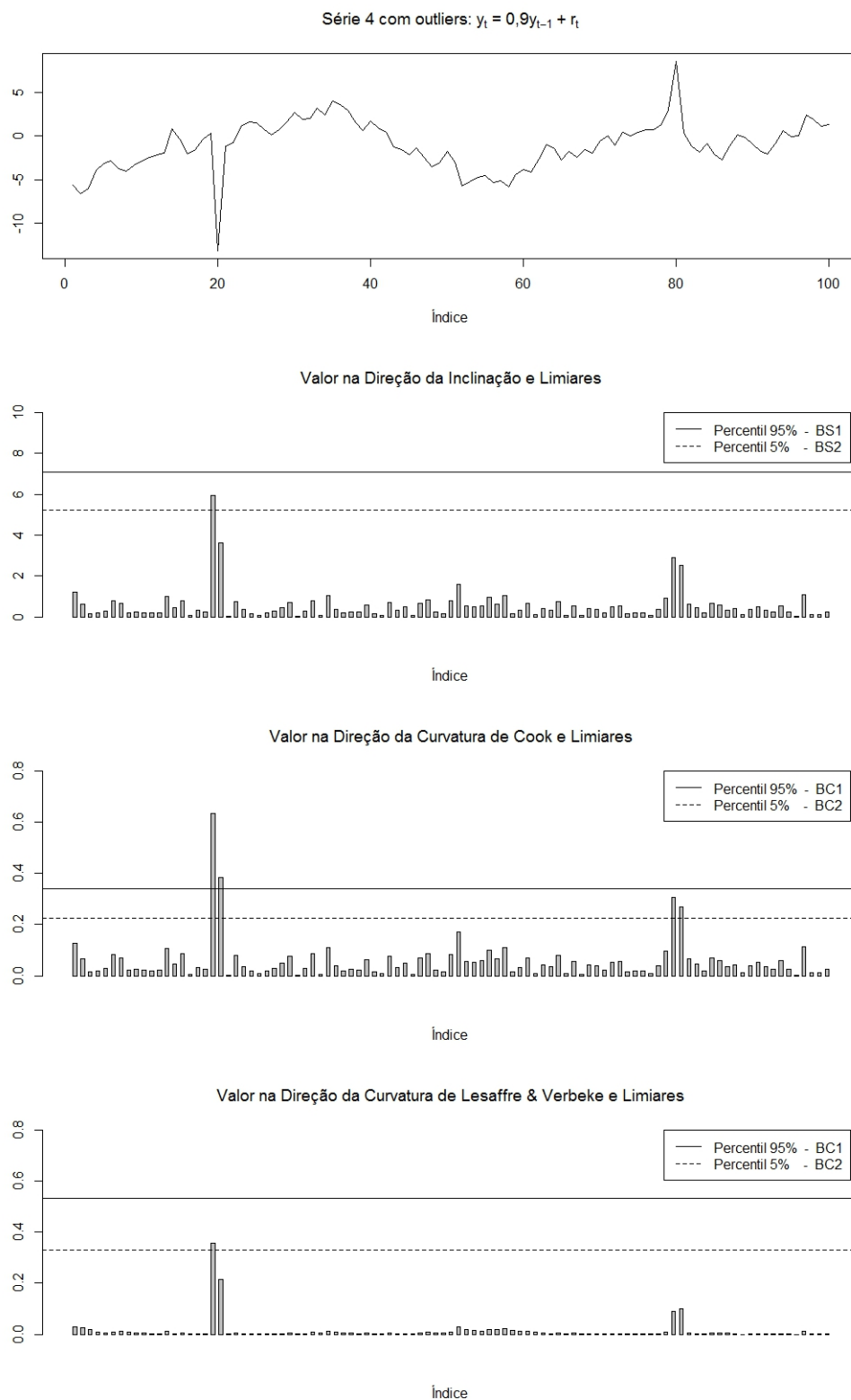
Fonte: Autoria própria.

**Figura 3.7** Diagnóstico de influência local na série 3 ( $\phi = 0,7$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA- $t_4$ .



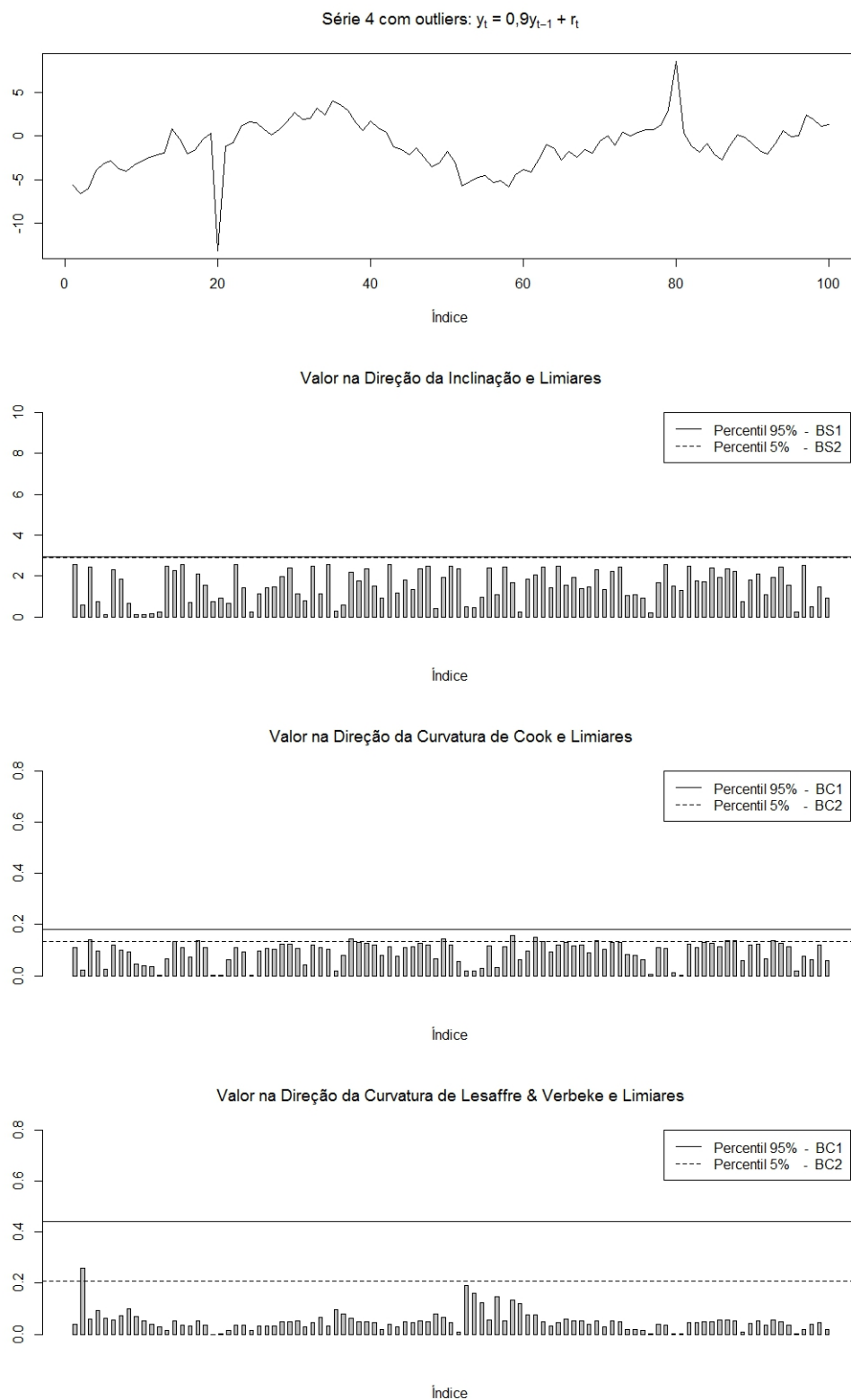
Fonte: Autoria própria.

**Figura 3.8** Diagnóstico de influência local na série 4 ( $\phi = 0,9$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA-normal.



Fonte: Autoria própria.

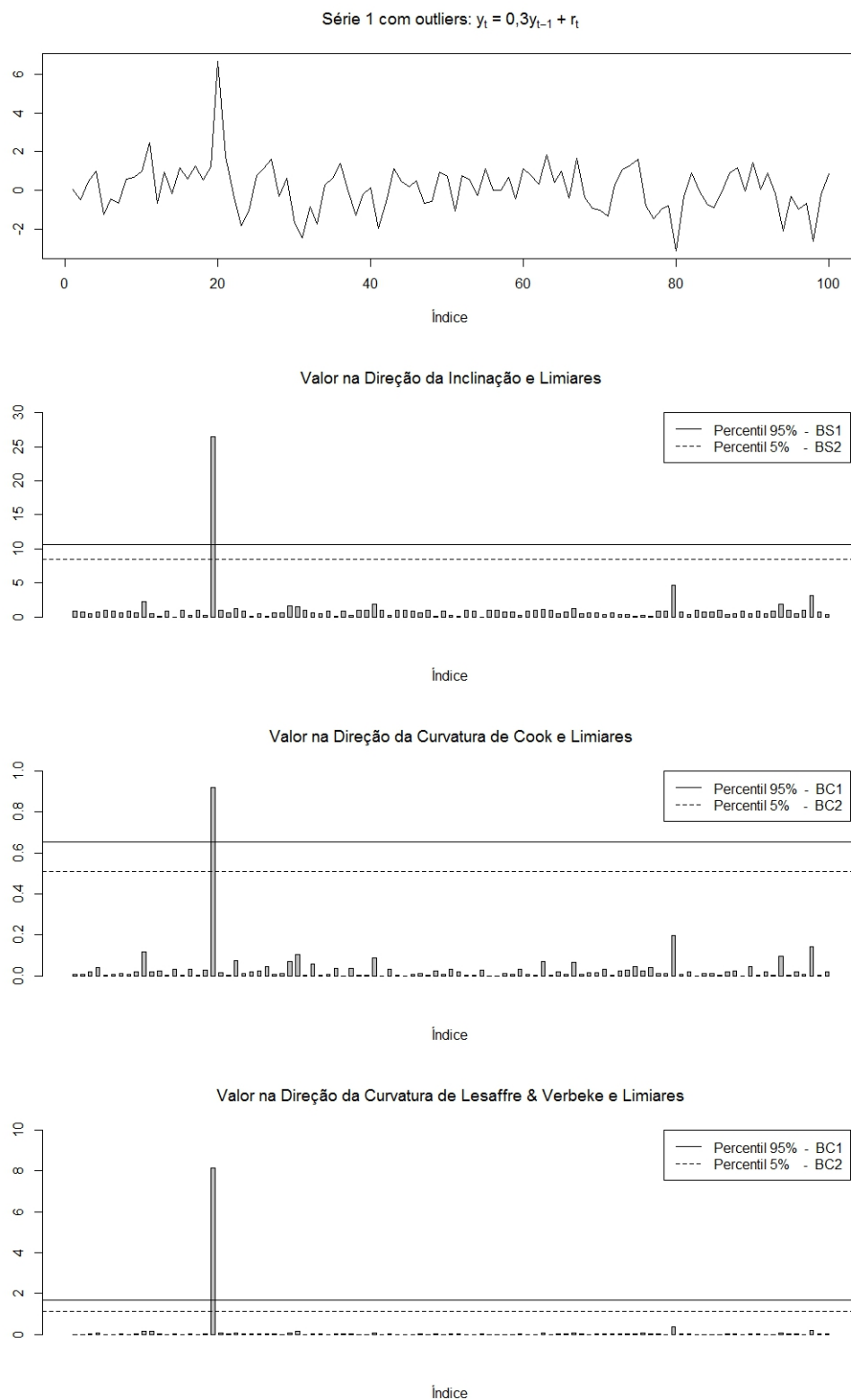
**Figura 3.9** Diagnóstico de influência local na série 4 ( $\phi = 0,9$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação nos dados para o modelo SYMARMA- $t_4$ .



Fonte: Autoria própria.

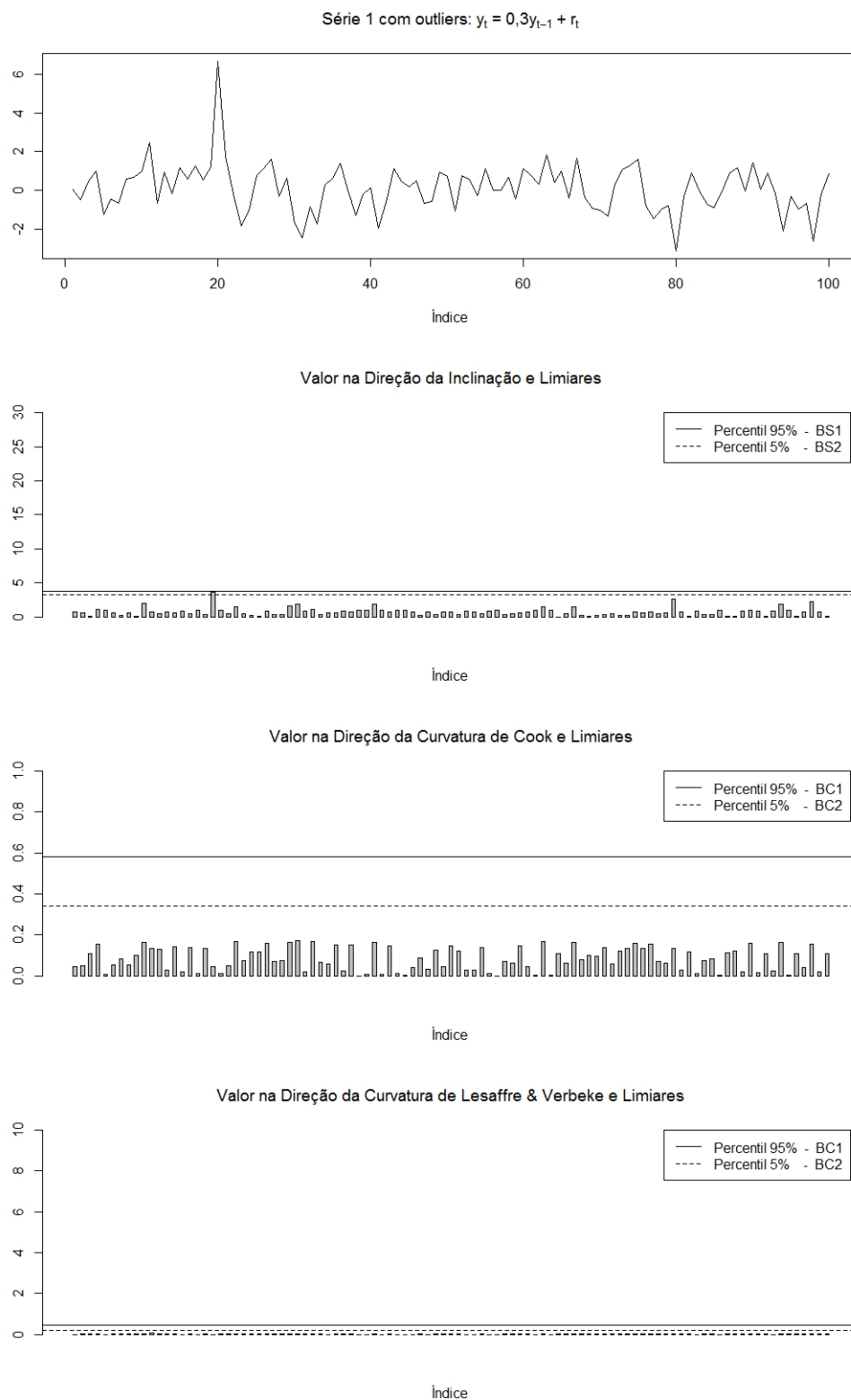


**Figura 3.10** Diagnóstico de influência local na série 1 ( $\phi = 0,3$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA-normal.



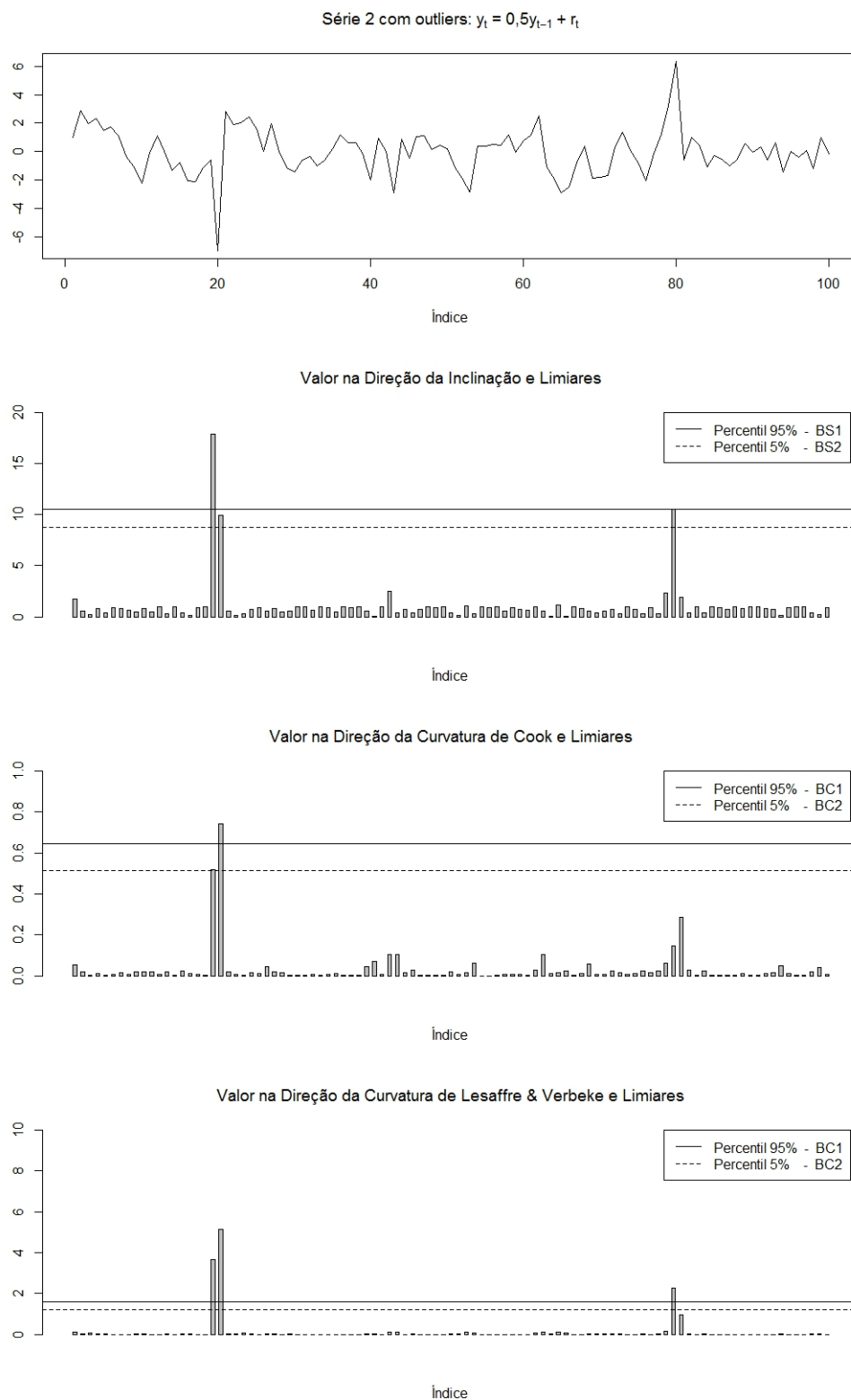
Fonte: Autoria própria.

**Figura 3.11** Diagnóstico de influência local na série 1 ( $\phi = 0,3$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA- $t_4$ .



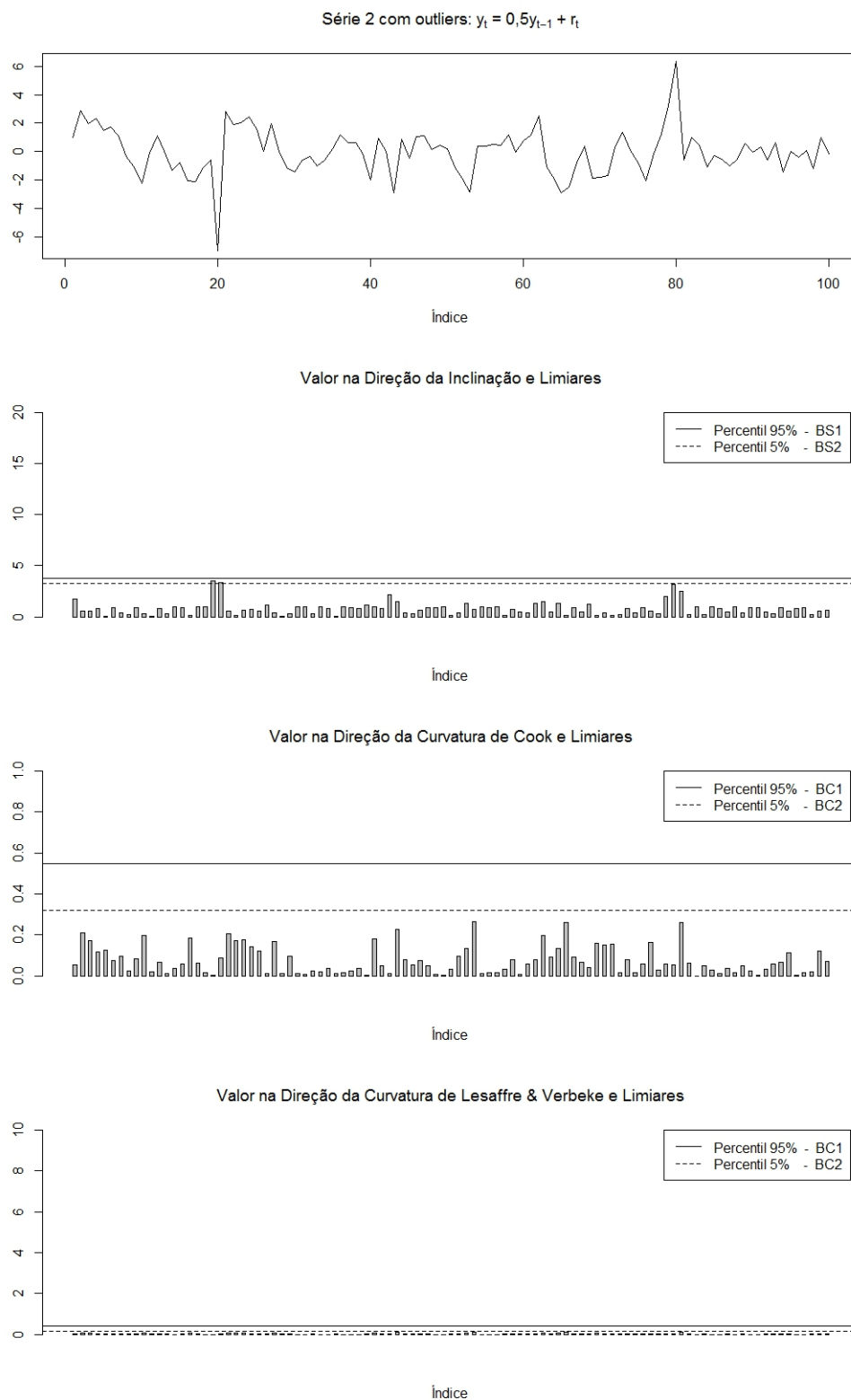
Fonte: Autoria própria.

**Figura 3.12** Diagnóstico de influência local na série 2 ( $\phi = 0,5$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA-normal.



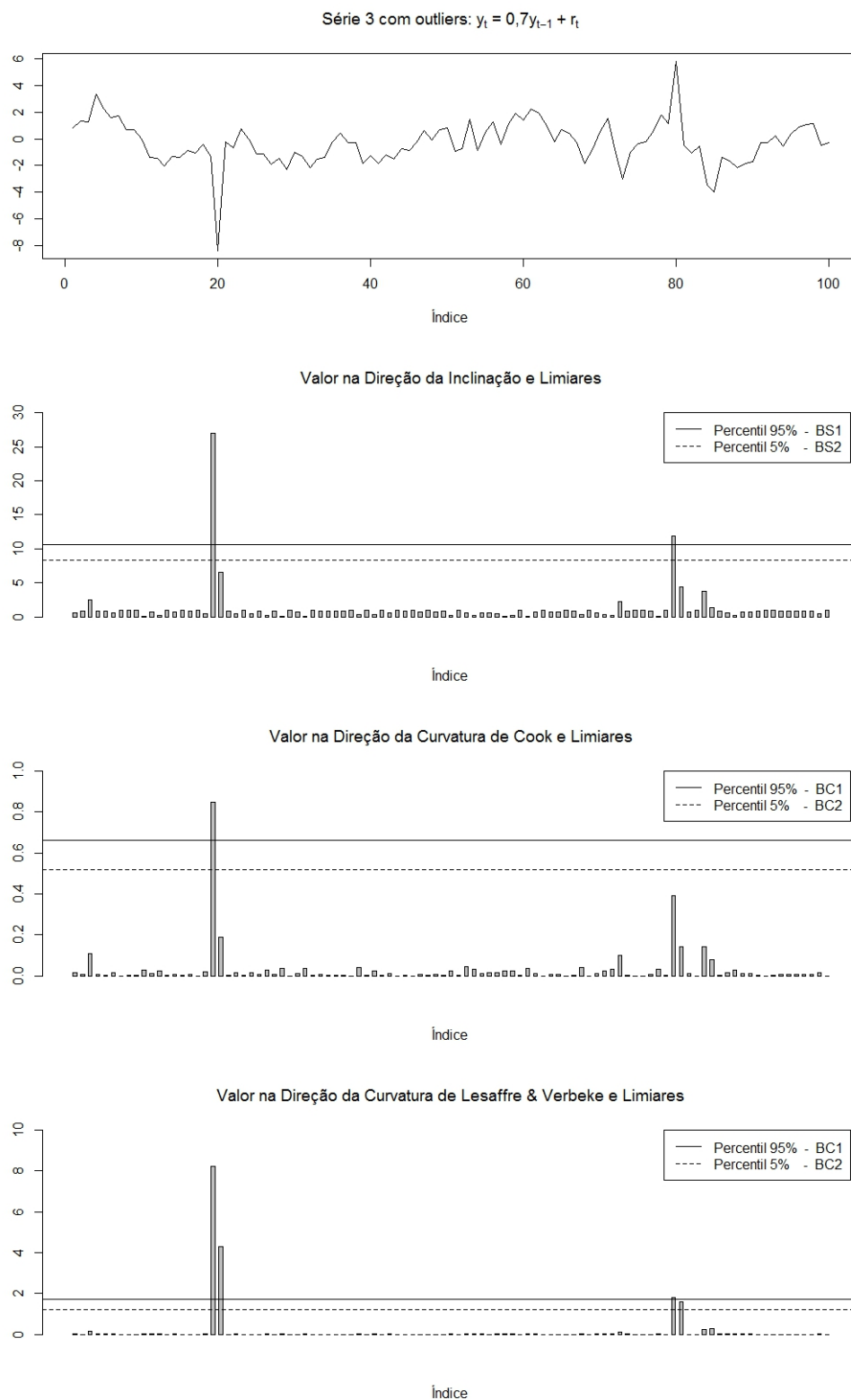
Fonte: Autoria própria.

**Figura 3.13** Diagnóstico de influência local na série 2 ( $\phi = 0,5$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA- $t_4$ .



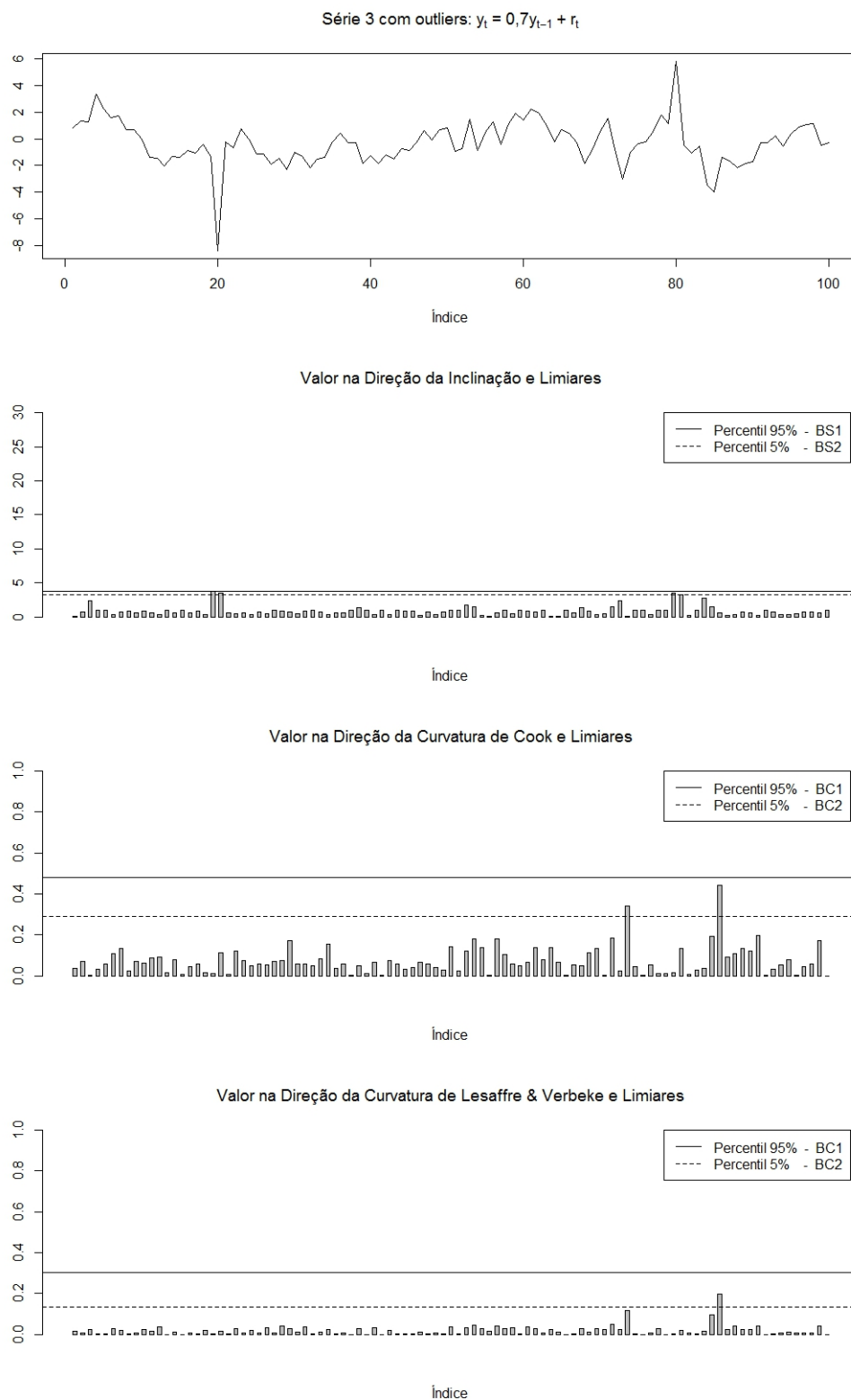
Fonte: Autoria própria.

**Figura 3.14** Diagnóstico de influência local na série 3 ( $\phi = 0,7$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA-normal.



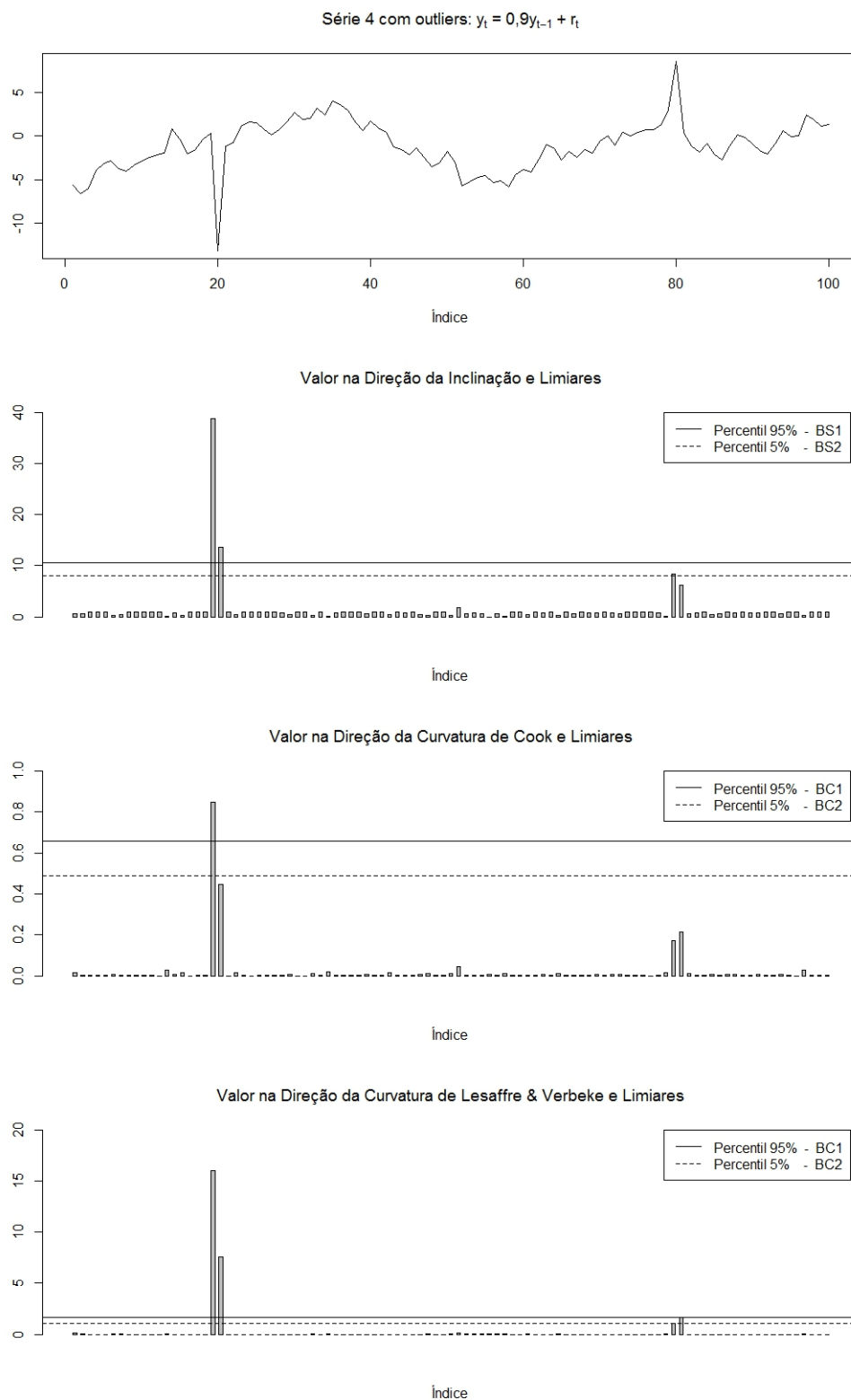
Fonte: Autoria própria.

**Figura 3.15** Diagnóstico de influência local na série 3 ( $\phi = 0,7$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA- $t_4$ .



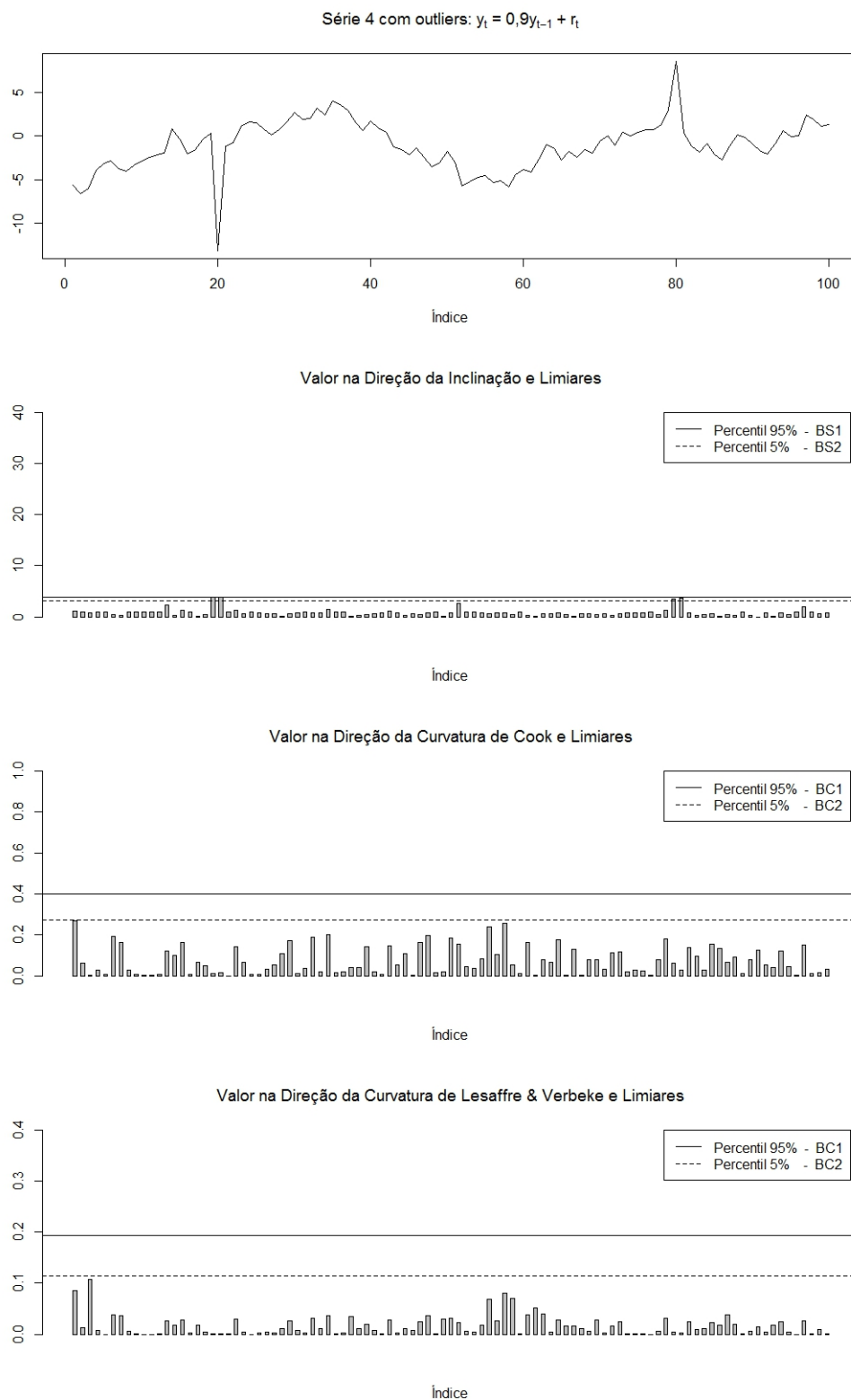
Fonte: Autoria própria.

**Figura 3.16** Diagnóstico de influência local na série 4 ( $\phi = 0,9$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA-normal.



Fonte: Autoria própria.

**Figura 3.17** Diagnóstico de influência local na série 4 ( $\phi = 0,9$ ,  $\varphi = 1$  e  $n = 100$ ) sob o esquema de perturbação no parâmetro de dispersão para o modelo SYMARMA- $t_4$ .



Fonte: Autoria própria.



## Capítulo 4

### Aplicações a Dados Reais

#### 4.1 Introdução

Nessa seção apresentamos dois exemplos práticos com o qual pretendemos avaliar a qualidade do ajuste e a capacidade preditiva do modelo SYMARMA perante séries temporais com *outliers*. O primeiro conjunto de dados é o excesso de retorno nos preços de fechamento diário da *Microsoft*, considerando como covariável o excesso de retorno fornecido pelo mercado, medido pelo índice S&P500, abrangendo o período entre 1 de abril e 5 de setembro de 2002. O segundo conjunto de dados é composto por observações anuais da inflação no Brasil, medida pelo Índice Geral de Preços - Oferta Global (IGP-OG), entre os anos de 1970 e 2014.

Consideramos os esquemas de perturbação aditiva nos dados e no parâmetro de dispersão. Três medidas de influência são estudadas: Curvatura de Cook (1986), Curvatura de Lesaffre e Verbeke (1998) e Inclinação de Billor e Loynes (1993). A técnica de limiares, proposta na Seção 3.3 do Capítulo 3, é utilizada para identificar observações influentes através dos critérios globais  $BS_0$  e  $BC_0$  e dos critérios individuais  $BS_i$ ,  $BC_i^{Cook}$  e  $BC_i^{LV}$  para  $i = 1$  e  $2$ .

Todas as funções para estimação, testes de hipóteses e diagnóstico foram desenvolvidas pelo autor a partir de rotinas construídas utilizando o ambiente de programação R em sua versão 3.1.2. Esta linguagem foi criada por Ross Ihak e Robert Gentleman na Universidade de Auckland. O R encontra-se disponível gratuitamente em <http://www.r-project.org>.

A seguir apresentamos os resultados obtidos pelo ajuste dos modelos SYMARMA sob distribuições condicional normal,  $t$ -Student e exponencial potência para os conjuntos de dados

citados acima. A suspeita é que distribuições com caudas pesadas acomodam melhor as observações atípicas presentes nas séries estudadas, fazendo com que elas não exerçam grande influência na modelagem dos dados e, por esse motivo, não são detectadas pelas medidas de influência local propostas. Por outro lado, na modelagem realizada a partir do modelo SYMARMA sob distribuição normal, esses mesmos pontos atípicos devem ser destacados pelos gráficos de influência.

## 4.2 Excesso de retorno da *Microsoft*

O conjunto de dados utilizado nessa seção corresponde a série excesso de retorno da *Microsoft* ( $Y$ ) e o excesso de retorno do índice S&P500 ( $X$ ) entre 1 de abril e 5 de setembro de 2002, totalizando uma amostra de 109 observações. Essa mesma série, para o período correspondente de 2 de novembro de 2001 até 31 de janeiro de 2003, foi utilizada anteriormente por Maior e Cysneiros (2009) tendo sido ajustado o modelo CAPM (*Capital Asset Pricing Model*) para estimar o risco sistemático da *Microsoft*. Os excessos de retornos estudados são definidos como

$$y_t = r_t - r_{ft} \quad \text{e} \quad x_t = r_{mt} - r_{ft},$$

em que  $r_t$  denota o retorno líquido<sup>1</sup> da ação no período  $t$ ,  $r_{ft}$  indica a taxa livre de risco durante o  $t$ -ésimo período avaliada pela taxa T-bill<sup>2</sup> e  $r_{mt}$  é o retorno fornecido pelo mercado medido por algum índice, por exemplo, no caso do Brasil o IBOVESPA e no caso do mercado norteamericano, o índice S&P500. Uma breve descrição do conjunto de dados completo pode ser encontrada em Ruppert (2004, p.239). Os dados podem ser obtidos em [http://www.de.ufpe.br/~cysneiros/elliptical/time\\_series.html](http://www.de.ufpe.br/~cysneiros/elliptical/time_series.html).

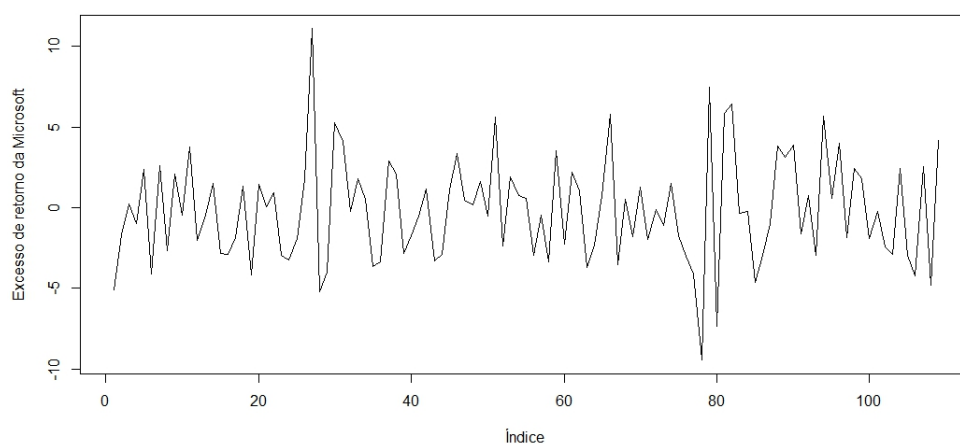
Os valores máximo e mínimo para o excesso de retorno da *Microsoft* são 11,10% e -9,47%, respectivamente. A inspeção direta do gráfico na Figura 4.1 sugere que a série não apresenta tendência e podemos notar a presença de *outliers*. A Figura 4.2 mostra a relação entre o excesso de retorno da S&P500 e o excesso de retorno da *Microsoft*. Podemos notar uma forte tendência

<sup>1</sup>O retorno líquido é definido como  $r_t = (p_t - p_{t-1})/p_{t-1}$ , em que  $p_t$  é o preço da ação no tempo  $t$ .

<sup>2</sup>A taxa T-bill foi dividida por 100 para conversão do valor percentual e, em seguida, por 253 para conversão em uma taxa diária (RUPPERT, 2004).

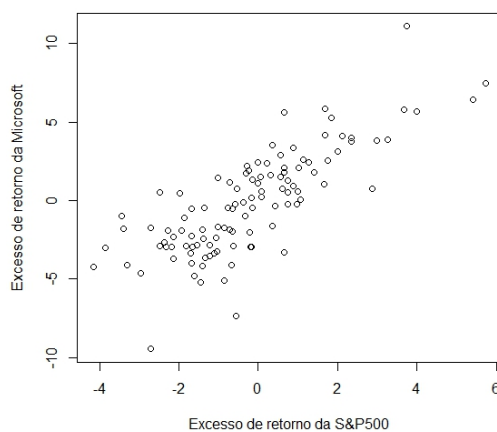
linear entre os excessos de retornos; coeficiente de correlação igual a 0,79.

**Figura 4.1** Série excesso de retorno da *Microsoft*.



Fonte: Autoria própria.

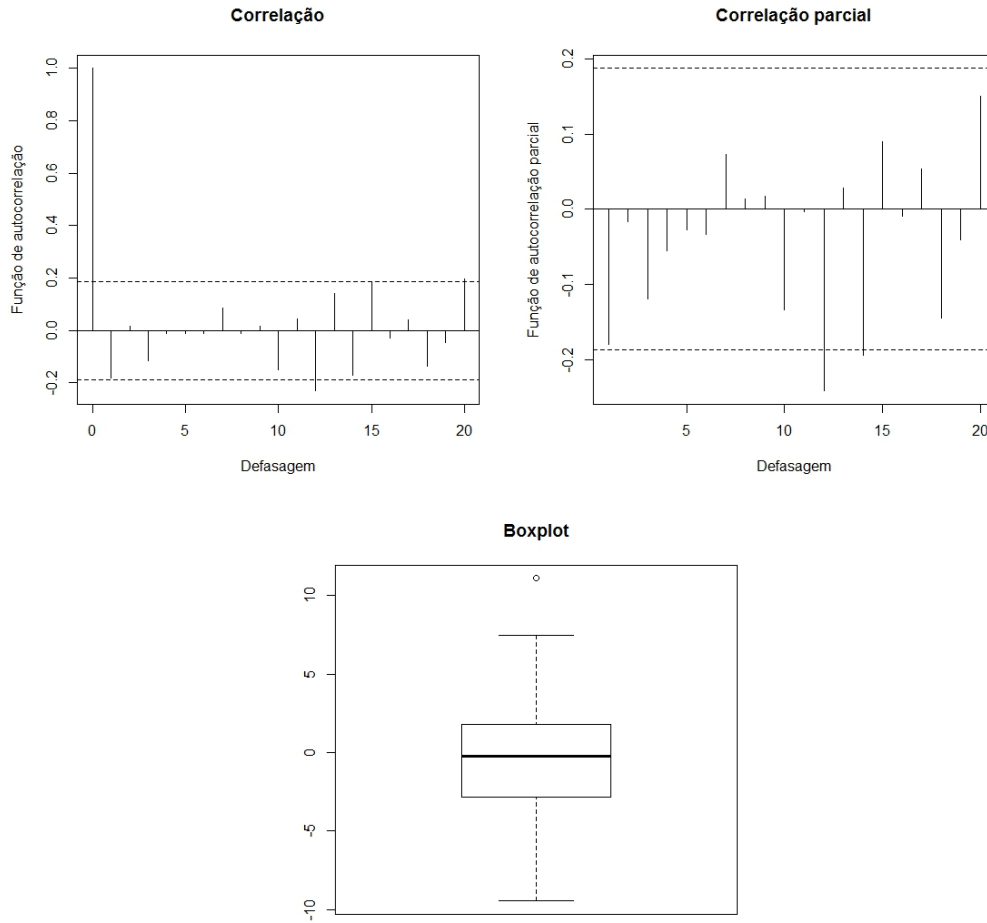
**Figura 4.2** Excesso de retorno da S&P500 versus excesso de retorno da *Microsoft*.



Fonte: Autoria própria.

Observando a Figura 4.3 notamos o decaimento exponencial da correlação, o que sugere a estacionariedade da série. E ainda, a partir do gráfico da correlação parcial, temos indícios que um modelo autorregressivo de ordem 12 pode explicar bem o comportamento dessa série. Por fim, o boxplot destaca um ponto atípico e podemos suspeitar que a distribuição dos dados é simétrica com caudas pesadas.

**Figura 4.3** Função de autocorrelação, autocorrelação parcial e boxplot da série excesso de retorno da *Microsoft*.



Fonte: Autoria própria.

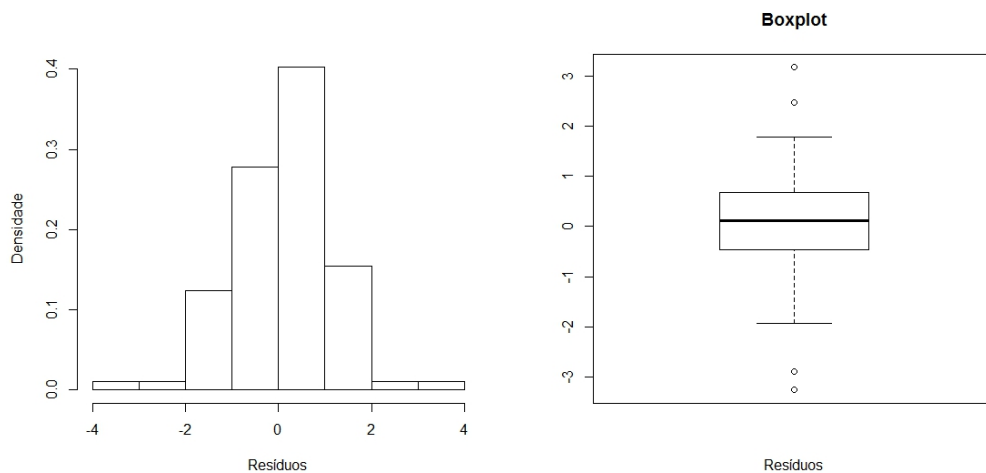
Inicialmente, ajustamos o modelo ARMA, isto é, o modelo SYMARMA-normal, dado por

$$y_t = \beta X_t + \phi \{Y_{t-12} - \beta X_{t-12}\} + r_t, \quad t = 13, \dots, 109. \quad (19)$$

Baseado no procedimento desenvolvido por Chen e Liu (1993) para detecção de *outliers* em modelos da classe ARIMA, os pontos  $y_{27} = 11,10\%$  e  $y_{80} = -7,34\%$  são considerados *outliers* aditivos. O histograma, boxplot e o gráfico normal de probabilidades dos resíduos obtidos pelo ajuste do modelo SYMARMA-normal são apresentados nas Figuras 4.4 e 4.5, respectivamente. A partir dessas figuras, podemos considerar que os dados são simétricos com

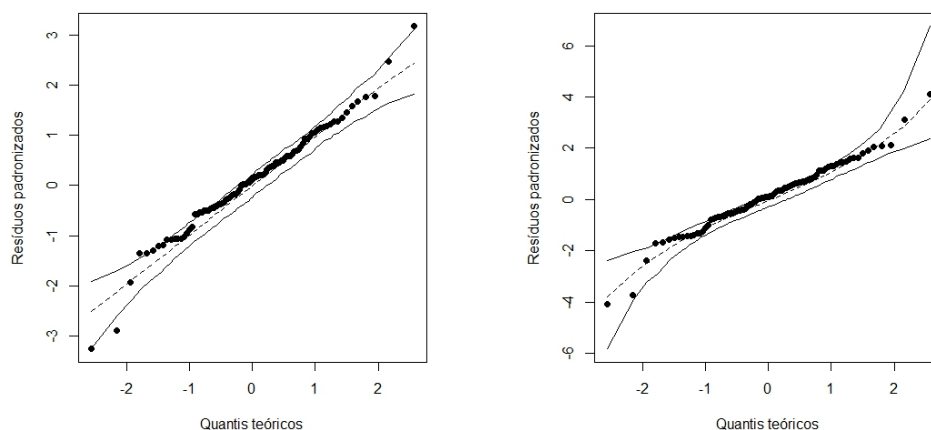
caudas pesadas. Adicionalmente, os coeficientes de assimetria e curtose empíricos são  $-0,227$  e  $4,465$ , respectivamente.

**Figura 4.4** Histograma (a) e boxplot (b) dos resíduos obtidos pelo ajuste do modelo SYMARMA-normal



Fonte: Autoria própria.

**Figura 4.5** Gráfico normal de probabilidades com envelope para o ajuste dos modelos SYMARMA-normal (a) e SYMARMA- $t_5$  (b).



Fonte: Autoria própria.

Como proposta para acomodar melhor as observações atípicas presentes na série, reajustamos o modelo (19) considerando duas outras distribuições de probabilidade com caudas mais pesadas: as distribuições  $t$ -Student e exponencial potência. Utilizamos os critérios AIC, BIC e

a raiz do erro médio quadrático (RMSE) para selecionar os graus de liberdade ( $\nu$ ) do modelo  $t$ -Student, SYMARMA- $t$ , e o parâmetro  $\kappa$  do modelo exponencial potência, SYMARMA-EP. A Tabela 4.1 apresenta os resultados para alguns valores de  $\nu$  entre 3 e 15 sendo selecionado o modelo SYMARMA- $t$  com  $\nu = 5$ . Similarmente, na Tabela 4.2 temos os resultados para alguns valores do parâmetro  $\kappa$  variando entre 0 e 1 sendo selecionado o modelo SYMARMA-EP com  $\kappa = 0,3$ . O parâmetro  $\kappa$  é responsável pela curtose da distribuição exponencial potência; quanto maior o valor de  $\kappa$ , maior a curtose. Além disso, as distribuições normal e Laplace são casos particulares obtidos quando  $\kappa = 0$  e  $\kappa = 1$ , respectivamente.

**Tabela 4.1** Valores para algumas medidas de comparação obtidas do ajuste de modelos SYMARMA- $t$  para a série excesso de retorno da *Microsoft*.

Modelo	AIC	BIC	RMSE
SYMARMA- $t_3$	413,91	419,06	3,427
SYMARMA- $t_4$	412,26	417,41	3,364
<b>SYMARMA-<math>t_5</math></b>	<b>411,68</b>	<b>416,83</b>	<b>3,340</b>
SYMARMA- $t_8$	411,74	416,89	3,345
SYMARMA- $t_{12}$	412,47	417,63	3,393
SYMARMA- $t_{15}$	412,96	418,11	3,429

**Tabela 4.2** Valores para algumas medidas de comparação obtidas do ajuste de modelos SYMARMA-EP para a série excesso de retorno da *Microsoft*.

Modelo	AIC	BIC	RMSE
SYMARMA-EP(0,1)	414,49	419,64	3,657
<b>SYMARMA-EP(0,3)</b>	<b>412,78</b>	<b>417,92</b>	<b>3,622</b>
SYMARMA-EP(0,5)	412,58	417,73	3,643
SYMARMA-EP(0,7)	413,32	418,47	3,681
SYMARMA-EP(0,9)	414,49	419,63	3,979

As estimativas e os erros-padrão assintóticos das estimativas dos parâmetros dos três mode-

los ajustados, juntamente com algumas medidas de comparação do ajuste, são apresentadas na Tabela 4.3. Notamos que o modelo SYMARMA- $t_5$  obteve o menor valor para as medidas AIC, BIC e RMSE. Além disso, as estimativas dos parâmetros  $\beta$ ,  $\phi$  e  $\varphi$  do modelo SYMARMA- $t_5$  têm menores erros-padrão assintóticos quando comparado aos modelos SYMARMA-normal e SYMARMA-EP.

**Tabela 4.3** Estimativas e erros-padrão assintóticos (em parênteses) das estimativas dos parâmetros dos três modelos ajustados e algumas medias de comparação - série excesso de retorno da *Microsoft*.

Modelo	Parâmetro	Estimativas	AIC	BIC	RMSE
SYMARMA-normal	$\beta$	1,327 (0,105)	416,20	421,35	3,728
	$\phi$	-0,132 (0,100)			
	$\varphi$	4,103 (0,589)			
SYMARMA- $t_5$	$\beta$	1,273 (0,097)	411,68	416,83	3,340
	$\phi$	-0,162 (0,093)			
	$\varphi$	2,628 (0,477)			
SYMARMA-EP(0,3)	$\beta$	1,299 (0,100)	412,78	417,92	3,622
	$\phi$	-0,134 (0,096)			
	$\varphi$	2,310 (0,378)			

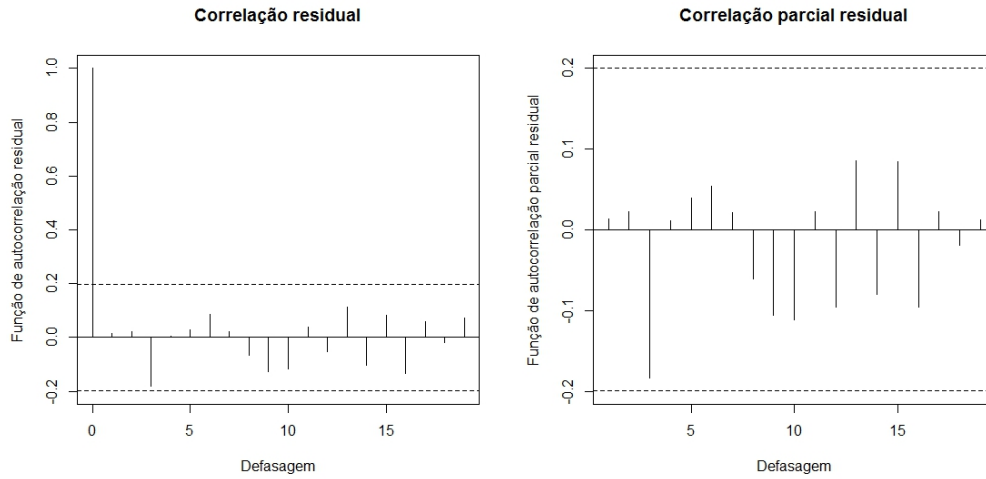
Os gráficos de correlação e correlação parcial dos resíduos obtidos pelo ajuste do modelo SYMARMA- $t_5$  e o gráfico normal de probabilidades com envelope indicam uma boa adequação do modelo (ver Figuras 4.6 e 4.5).

O teste de Ljung-Box (LJUNG; BOX, 1978), utilizado aqui para avaliar a aleatoriedade dos resíduos, obteve valor  $p$  igual a 0,891 indicando que os resíduos do ajuste são não correlacionados ao nível de 5% de significância. Na Figura 4.7 apresentamos a série ajustada pelo modelo SYMARMA- $t_5$  juntamente com a série original indicando um bom ajuste.

A Tabela 4.4 apresenta as previsões 12 passos à frente e a RMSE para os modelos SYMARMA ajustados. Baseado na RMSE, o modelo SYMARMA-normal teve uma melhor desempenho a curto prazo e o modelo SYMARMA- $t_5$  a longo prazo.

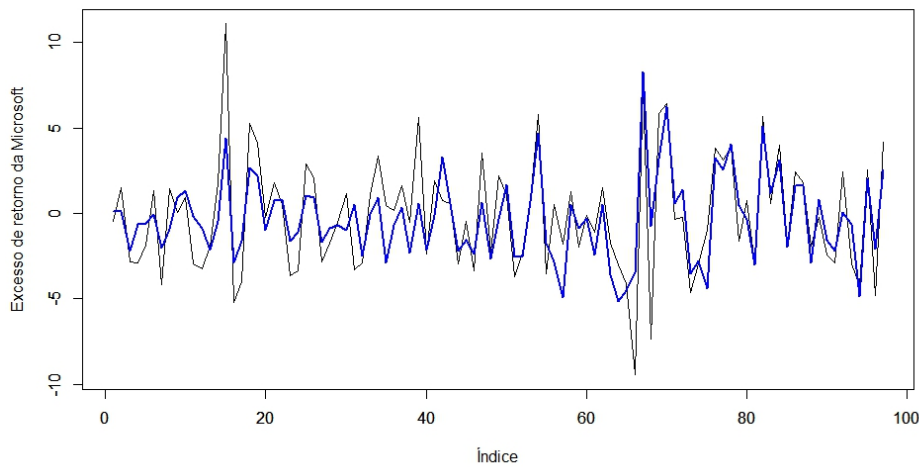
A partir de agora vamos avaliar a influência local das observações considerando dois esque-

**Figura 4.6** Função de autocorrelação e autocorrelação parcial dos resíduos obtidos do ajuste do modelo SYMARMA- $t_5$  - série excesso de retorno da *Microsoft*.



Fonte: Autoria própria.

**Figura 4.7** Série original (linha preta) e série ajustada pelo modelo SYMARMA- $t_5$  (linha azul) - série excesso de retorno da *Microsoft*.



Fonte: Autoria própria.

mas de perturbação: perturbação aditiva nos dados e perturbação no parâmetro de dispersão. Primeiramente, para cada esquema de perturbação, calculamos o vetor de inclinação e os vetores de curvatura a partir das expressões derivadas nas Seções 3.2 e 3.4 do Capítulo 3. Em seguida, estimamos os limiares  $BS_i$ ,  $BC_i^{Cook}$  e  $BC_i^{LV}$  para  $i = 0, 1$  e  $2$ , definidos na Seção 3.3 do Capítulo



3, a partir de 2.000 simulações dos modelos estimados sob distribuição normal e  $t$ -Student. Os resultados são apresentadas nas Tabelas 4.5 e 4.6. Por fim, comparamos os vetores de inclinação e de curvatura com os limiares seguindo o procedimento descrito na Seção 3.3 do Capítulo 3.

**Tabela 4.4** Previsões 12 passos à frente, RMSE e limites do intervalo com 95% de confiança via bootstrap para o ajuste dos modelos SYMARMA-normal e SYMARMA- $t_5$ .

Passos à frente		Excesso de retorno		SYMARMA-normal				SYMARMA- $t_5$			
$\ell$		$y_{n+\ell}$	$\hat{y}_{n+\ell}$	$L_l$	$L_u$	RMSE		$\hat{y}_{n+\ell}$	$L_l$	$L_u$	RMSE
1		1.84	1.24	1.09	1.61	0.33		1.15	1.09	1.52	0.38
2		2.26	0.97	0.84	1.13	0.47		0.92	0.83	1.05	0.50
3		-2.46	-0.18	-0.26	0.19	0.66		-0.19	-0.23	0.15	0.67
4		-2.94	-3.14	-3.92	-2.78	0.57		-2.97	-3.67	-2.77	0.58
5		1.61	0.50	0.30	0.56	0.60		0.52	0.29	0.55	0.60
6		-1.31	-2.38	-2.87	-2.10	0.64		-2.25	-2.69	-2.07	0.62
7		0.97	-0.95	-1.13	-0.14	0.95		-1.00	-1.09	-0.18	0.96
8		-1.18	-3.64	-4.88	-3.21	1.16		-3.40	-4.62	-3.23	1.11
9		0.55	0.15	0.04	0.57	1.17		0.13	0.09	0.51	1.08
10		-4.70	-1.88	-2.16	-1.59	1.08		-1.82	-2.02	-1.57	1.04
11		0.93	-1.95	-2.98	-1.68	1.39		-1.76	-2.82	-1.65	1.33
12		1.88	3.04	2.68	3.98	1.34		2.83	2.69	3.76	1.28

Informações relacionadas ao diagnóstico de influência são apresentadas nas Tabelas 4.7 e 4.8. Os limiares para análise de influência global,  $BS_0$  e  $BC_0$ , foram calculados a partir do percentil 95%. E, os limiares para análise de influência individual  $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$  foram calculados a partir do percentil 5%.

Discutimos inicialmente o caso do modelo SYMARMA-normal (ver Tabela 4.7). Para ambos esquemas de perturbação, as três medidas de influência local adotadas sugerem que a série excesso de retorno da *Microsoft* possui pelo menos uma observação influente ( $O_s > BS_0$  e  $O_c > BC_0$ ). Baseado nos limiares  $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$ , encontramos 12 pontos influentes sob o esquema de perturbação nos dados e 2 pontos influentes sob o esquema de perturbação no parâmetro de dispersão. Notamos que as observações  $y_{27}$  e  $y_{80}$  foram classificadas como influentes por todas as medidas de diagnóstico e esquemas de perturbação considerados. Além disso, os valores ajustados  $\hat{y}_{27}$  e  $\hat{y}_{80}$  estão associados aos maiores resíduos.

Para o modelo SYMARMA- $t_5$  (ver Tabela 4.8), não encontramos nenhuma significância global, ou seja,  $O_s < BS_0$  e  $O_c < BC_0$  para ambos esquemas de perturbação e medidas de diagnóstico consideradas. Sob o esquema de perturbação nos dados, também não foram encontrados indícios de observações individualmente influentes. Por outro lado, duas observações foram destacadas pelo diagnóstico de curvatura sob o esquema de perturbação no parâmetro de dispersão. Essas observações correspondem aos pontos  $y_{28}$  e  $y_{81}$ . Tais pontos correspondem a observações posteriores aos *outliers* detectados pelo procedimento desenvolvido por Chen e Liu (1993). Notamos, ainda, que essas observações não estão associadas a grandes resíduos.

Nas Figuras 4.8 e 4.9, apresentamos (em valores absolutos) os vetores de inclinação e de curvatura para as perturbações nos dados e no parâmetro de dispersão, respectivamente. Os limiares  $BS_i$ ,  $BC_i^{Cook}$  e  $BC_i^{LV}$  estão representados por linhas contínuas (para  $i = 1$ ) e por linhas tracejadas (para  $i = 2$ ). Em quase todos os cenários, o modelo SYMARMA-normal destaca pontos que não estão presentes no diagnóstico baseado no modelo SYMARMA- $t_5$ , evidenciando a robustez do modelo SYMARMA- $t$ .

**Tabela 4.5** Limitares para o modelo SYMARMA-normal calculados para série excesso de retorno da *Microsoft*.

Limiares ( <i>Benchmark</i> )	Estatística	Tipo de perturbação					
		Aditiva			Dispersão		
		Inclinação	Curvatura de Cook	Curvatura de Verbeke	Inclinação	Curvatura de Cook	Curvatura de Verbeke
BS <sub>0</sub>  BC <sub>0</sub>	99,0	3,8938	0,5616	0,5616	9,4549	4,5036	4,5036
	97,5	3,7770	0,5284	0,5284	9,0269	4,1463	4,1464
	95,0	3,7122	0,5103	0,5104	8,6721	3,9087	3,9087
	90,0	3,5997	0,4799	0,4799	8,2142	3,6265	3,6265
BS <sub>1</sub>  BC <sub>1</sub> <sup>Cook</sup>  BC <sub>1</sub> <sup>LV</sup>	99,0	2,4836	0,3671	0,0605	13,5552	0,7215	2,2193
	97,5	2,3410	0,3448	0,0534	11,8425	0,6782	1,8936
	95,0	2,2346	0,3296	0,0479	10,7342	0,6348	1,5634
	90,0	2,1148	0,3139	0,0437	9,6417	0,5886	1,2938
BS <sub>2</sub>  BC <sub>2</sub> <sup>Cook</sup>  BC <sub>2</sub> <sup>LV</sup>	Máx	2,5788	0,3418	0,0711	21,3035	0,8697	5,1049
	Mín	1,4934	0,2001	0,0227	7,4393	0,4354	0,8796
	5.0	1,6919	0,2206	0,0291	8,6217	0,4959	1,1570
	Média	2,0206	0,2654	0,0412	11,7636	0,6321	1,8315
	DP	0,2326	0,0304	0,0094	2,2625	0,0888	0,6202

**Tabela 4.6** Limitares para o modelo SYMARMA- $t_5$  calculados para série excesso de retorno da *Microsoft*.

Limitares ( <i>Benchmark</i> )	Estatística	Tipo de perturbação					
		Aditiva			Dispersão		
		Inclinação	Curvatura de Cook	Curvatura de Verbeke	Inclinação	Curvatura de Cook	Curvatura de Verbeke
BS <sub>0</sub>  BC <sub>0</sub>	99,0	7,4545	1,7248	1,7248	6,6659	1,7047	1,7047
	97,5	7,2853	1,6354	1,6354	6,5219	1,5651	1,5651
	95,0	7,0435	1,5420	1,5420	6,3783	1,4584	1,4584
	90,0	6,8459	1,4631	1,4631	6,2445	1,3697	1,3697
BS <sub>1</sub>  BC <sub>1</sub> <sup>Cook</sup>  BC <sub>1</sub> <sup>LV</sup>	99,0	2,2279	0,1788	0,5422	4,7853	0,6466	0,6154
	97,5	2,1525	0,1675	0,4207	4,6680	0,5847	0,4930
	95,0	2,0959	0,1595	0,3306	4,5600	0,5165	0,3930
	90,0	2,0384	0,1510	0,2699	4,4221	0,4632	0,3069
BS <sub>2</sub>  BC <sub>2</sub> <sup>Cook</sup>  BC <sub>2</sub> <sup>LV</sup>	Máx	2,4778	0,1293	0,7652	4,9342	0,8577	1,4384
	Mín	2,0352	0,1812	0,0775	3,5032	0,2486	0,1157
	5.0	2,0477	0,1493	0,1114	3,6977	0,3141	0,1541
	Média	2,1607	0,1392	0,2407	4,2953	0,4666	0,4083
	DP	0,0854	0,0103	0,1315	0,3222	0,1323	0,2733

**Tabela 4.7** Diagnóstico de influência local para a série excesso de retorno da *Microsoft* considerando o modelo SYMARMA-normal.  $r_t$  são os resíduos.

Data (dia mês ano)	Observação	$r_t$	Tipo de perturbação			
			Aditiva		Dispersão	
			Inclinação	Curvatura de Cook	Inclinação	Curvatura de Cook
07 05 02	27	6,41	<b>-3,3113</b>	<b>0,3135</b>	<b>9,6130</b>	<b>-0,5114</b>
00 00 02	28	-2,38	1,2300	-0,1164	0,4644	-0,0265
30 05 02	43	-3,90	<b>2,0167</b>	-0,1909	2,9367	-0,1538
05 06 02	47	3,42	<b>-1,7643</b>	0,1670	2,0130	-0,1546
11 06 02	51	5,02	<b>-2,5912</b>	<b>0,2453</b>	5,4989	-0,3135
08 07 02	68	3,57	<b>-1,8445</b>	0,1746	2,2929	-0,1290
00 00 02	69	3,27	-1,6872	0,1597	1,7552	-0,1566
22 07 02	78	-5,81	<b>3,0019</b>	<b>-0,2842</b>	7,7224	-0,3852
00 00 02	79	-1,01	0,5228	-0,0495	-0,7354	-0,0277
24 07 02	80	-6,61	<b>3,4134</b>	<b>-0,3231</b>	<b>10,2771</b>	<b>-0,5070</b>
00 00 02	81	2,68	-1,3840	0,1310	0,8539	-0,0356
02 08 02	87	3,65	<b>-1,8877</b>	0,1787	2,4492	-0,1555
$O_s$ e $O_c$			5,2816	1,0332	9,6406	4,5115
BS <sub>0</sub> e BC <sub>0</sub>			3,7122	0,5104	8,6721	3,9087
BS <sub>1</sub> , BC <sub>1</sub> <sup>Cook</sup> e BC <sub>1</sub> <sup>LV</sup>			2,2346	0,3296	10,7342	1,5634
BS <sub>2</sub> , BC <sub>2</sub> <sup>Cook</sup> e BC <sub>2</sub> <sup>LV</sup>			1,6919	0,2206	8,6217	1,1570

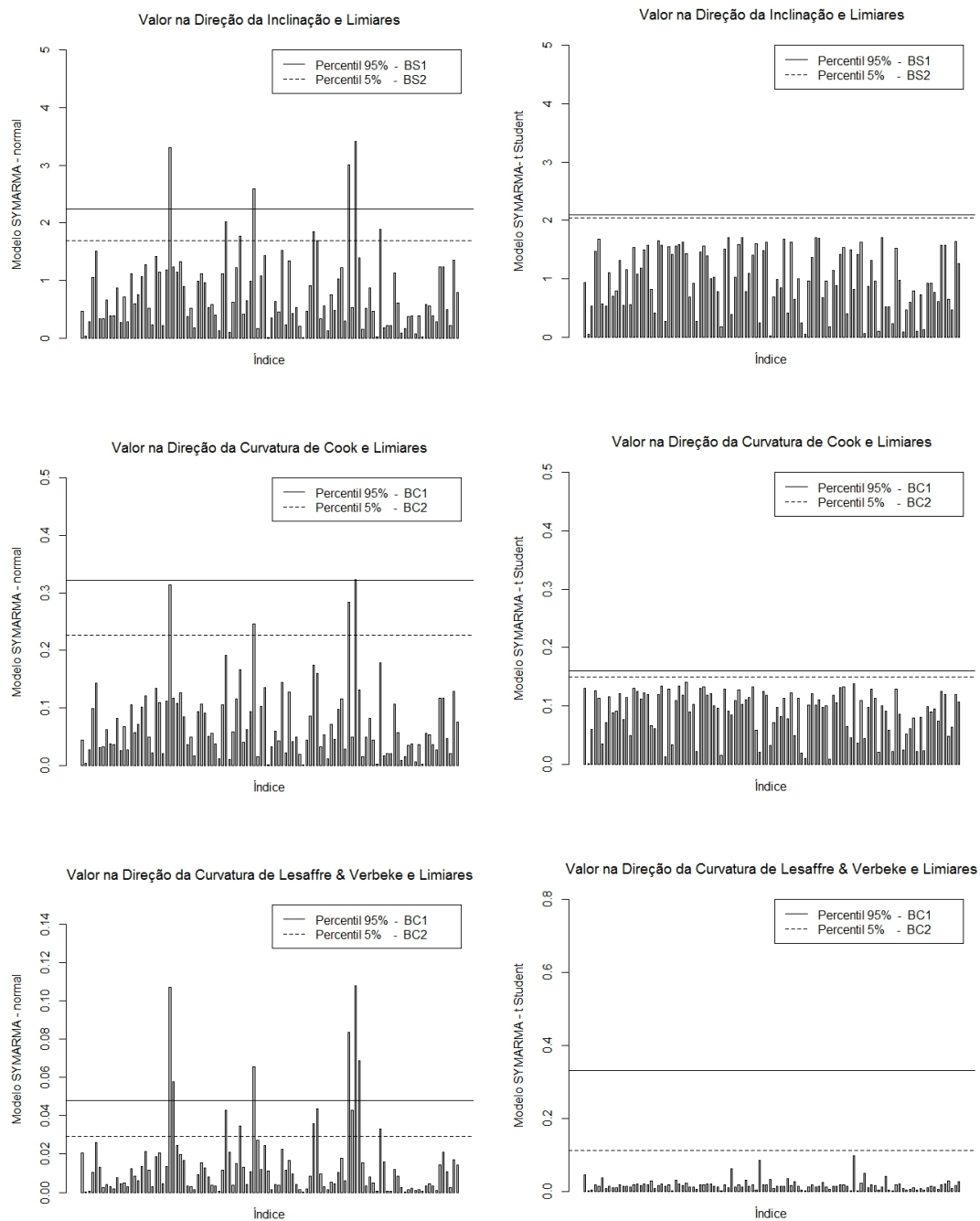
Nota: Pontos influentes sob BS<sub>2</sub>, BC<sub>2</sub><sup>Cook</sup> e BC<sub>2</sub><sup>LV</sup> são apresentados em negrito. Estatísticas globalmente significativas  $O_s$  e  $O_c$  são apresentadas em itálico.

**Tabela 4.8** Diagnóstico de influência local para a série excesso de retorno da *Microsoft* considerando o modelo SYMARMA- $t_5$ .  $r_t$  são os resíduos.

Data (dia mês ano)	Observação	$r_t$	Tipo de perturbação					
			Aditiva			Dispersão		
			Inclinação	Curvatura de Cook	Curvatura de Verbeke	Inclinação	Curvatura de Cook	Curvatura de Verbeke
08 05 02	28	-2,28	1,5550	-0,1084	0,0302	0,7708	<b>-0,4074</b>	<b>0,2128</b>
25 07 02	81	2,60	-1,6294	0,1091	0,0223	1,1184	<b>-0,4174</b>	<b>0,2219</b>
$O_s$ e $O_c$			5,7832	1,0468	1,0468	5,6051	1,2582	1,2582
BS <sub>0</sub> e BC <sub>0</sub>			7,0435	1,5420	1,5420	6,3783	1,4584	1,4584
BS <sub>1</sub> , BC <sub>1</sub> <sup>Cook</sup> e BC <sub>1</sub> <sup>LV</sup>			2,0959	0,1595	0,3306	4,5600	0,5165	0,3930
BS <sub>2</sub> , BC <sub>2</sub> <sup>Cook</sup> e BC <sub>2</sub> <sup>LV</sup>			2,0477	0,1493	0,1114	3,6977	0,3141	0,1541

Nota: Pontos influentes sob BS<sub>2</sub>, BC<sub>2</sub><sup>Cook</sup> e BC<sub>2</sub><sup>LV</sup> são apresentados em negrito. Estatísticas globalmente significativas  $O_s$  e  $O_c$  são apresentadas em itálico.

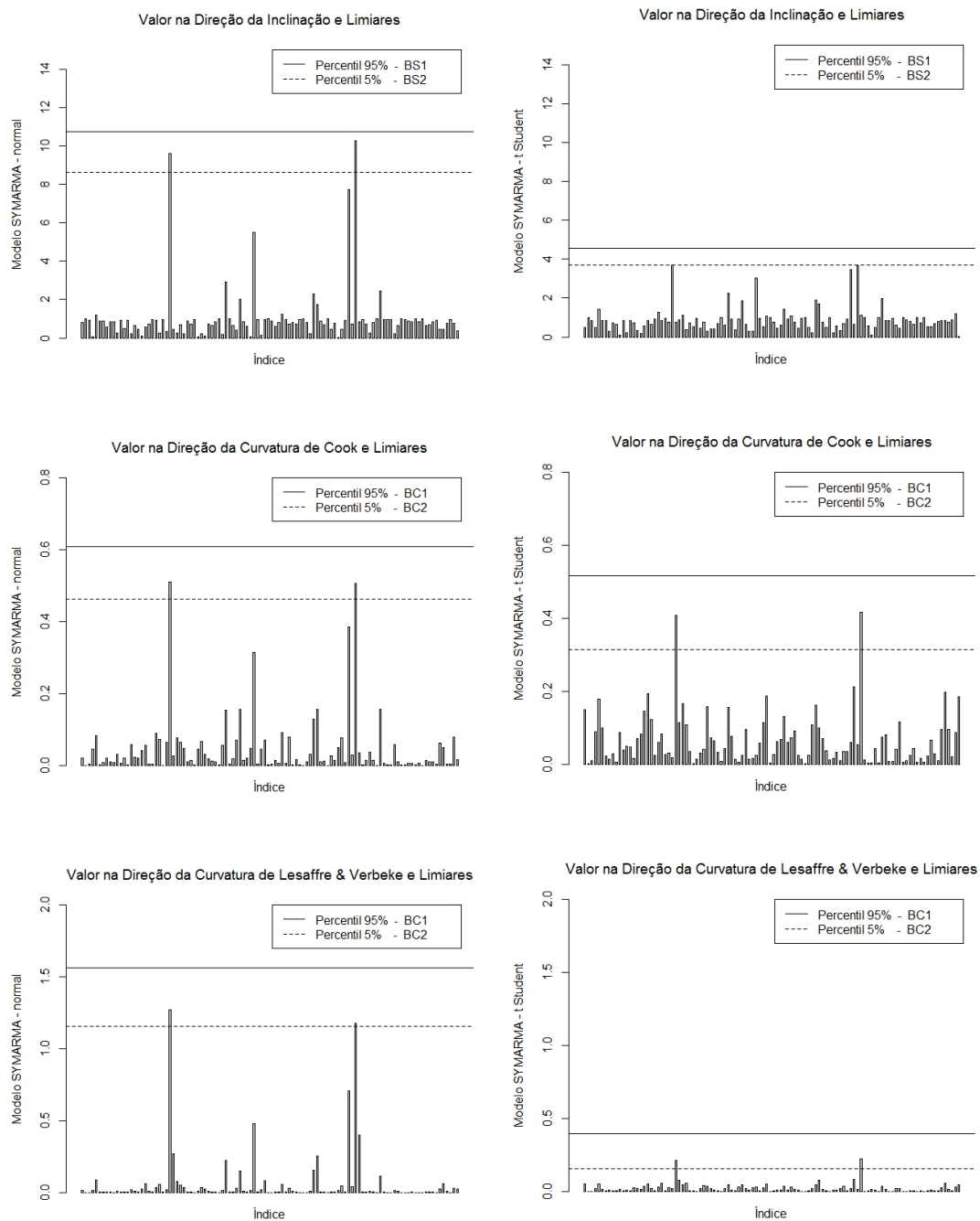
**Figura 4.8** Diagnóstico de influência local na série excesso de retorno da *Microsoft* sob o esquema de perturbação nos dados para os modelos SYMARMA-normal (lado esquerdo) e SYMARMA- $t_5$  (lado direito).



Fonte: Autoria própria.



**Figura 4.9** Diagnóstico de influência local na série excesso de retorno da *Microsoft* sob o esquema de perturbação no parâmetro de dispersão para os modelos SYMARMA-normal (lado esquerdo) e SYMARMA- $t_5$  (lado direito).



Fonte: Autoria própria.

### 4.3 Índice Geral de Preços - Oferta Global

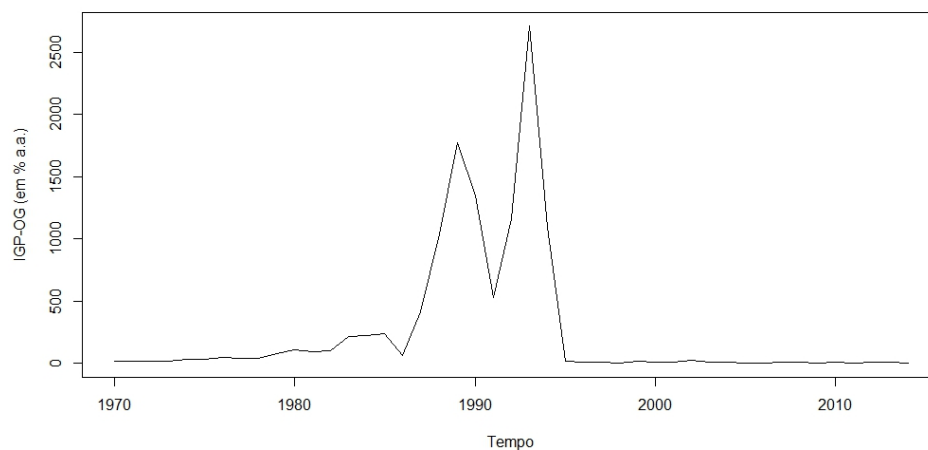
A série histórica estudada nessa seção contém 45 observações anuais do Índice Geral de Preços - Oferta Global (IGP-OG). Este é um dos índices inflacionários utilizados no Brasil, calculado pela Fundação Getúlio Vargas (FGV). O período considerado corresponde aos anos de 1970 até 2014. O IGP-OG é obtido pela média ponderada das parcelas Índice de Preços por Atacado (IPA), Índice de Preços ao Consumidor (IPC) e Índice Nacional de Custo da Construção (INCC), com pesos iguais a 6, 3 e 1, respectivamente.

A série inflacionária em estudo é apresentada na Figura 4.10. Dados de inflação avaliados por meio do IGP também foram estudados por Cribari-Neto e Cassiano (2005) que propuseram quatro medidas robustas para checar impacto de choques econômicos no longo prazo. Da Figura 4.10 observamos que a partir dos anos 80 a série é caracterizada por crescimentos acen- tuados, que levaram o país à hiperinflação em 1989 e 1993. Conforme relatam Cribari-Neto e Cassiano (2005): “Desde 1979 a história econômica brasileira foi marcada por muitas interven- ções governamentais repentinas destinadas a controlar a inflação. Porém, alguns destes planos de choque não obtiveram o efeito esperado, nem mesmo a curto prazo, e por isso não intro- duziram *inliers* significativos na série, uma vez que não conseguiram baixar o nível das taxas inflacionárias. Este foi o caso dos planos de choque Delfim I, Delfim II, Delfim III, Dornelles, Arroz com Feijão, Éris, Marcílio e o Programa de Ação Imediata”.

Dentre as diversas observações atípicas presentes na série, podemos destacar os anos de 1986 (plano Cruzado), 1987 (plano Bresser), 1989 (plano Verão), 1990 (plano Collor I), 1991 (plano Collor II) e 1994 (plano Real). No período considerado, a maior taxa de inflação regis- trada foi de 2.710,13% ao ano que ocorreu em 1993. Uma deflação de -1,40% também pôde ser observada no ano de 2009.

O correlograma e o correlograma parcial da série são apresentados na Figura 4.11. Intui- tivamente é razoável considerar que a inflação no Brasil segue um processo estacionário; a inflação tende a flutuar em torno de algum patamar ao longo do tempo. Esse fato, somado à evidência apresentada no correlograma, permite considerar que a série inflação no Brasil se- gue um processo estacionário. A inspeção direta do correlograma sugere a presença de um parâmetro de médias móveis e o correlograma parcial dá indícios de um possível parâmetro

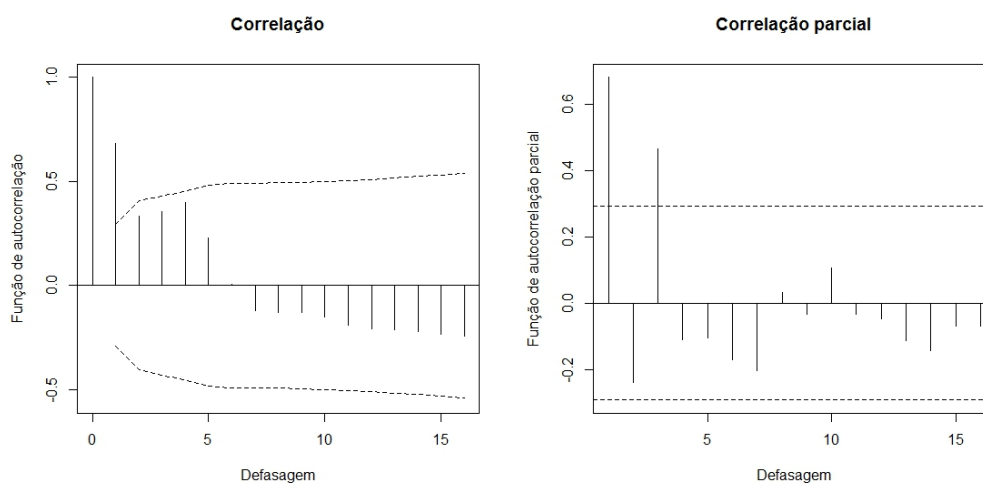
**Figura 4.10** Série inflacionária no Brasil de acordo com IGP-OG.



Fonte: Autoria própria.

autorregressivo no modelo.

**Figura 4.11** Função de autocorrelação e autocorrelação parcial da série IGP-OG.



Fonte: Autoria própria.

Considerando que a série segue um processo estacionário, variantes do modelo ARMA, isto é, SYMARMA-normal, foram estimados. O critério de seleção utilizado para selecionar o modelo final, dentre os modelos estimados, foi o BIC. Os resultados são apresentados na Tabela 4.9 e, como podemos notar, dentre os modelos estimados o que melhor descreveu a dinâmica

dos dados foi o modelo SYMARMA(1,1),

$$y_t = \beta + \phi\{Y_{t-1} - \beta\} + \theta r_{t-1} + r_t, \quad t = 2, \dots, 45, \quad (20)$$

pois apresentou menor BIC e também maior valor para o logaritmo da função de verossimilhança.

**Tabela 4.9** Modelos SYMARMA-normal ajustados a série inflacionária no Brasil, para diferentes configurações, e seus respectivos valores do critério BIC e o logaritmo da função de verossimilhança.

Modelo	BIC	$\ell(\hat{\delta}, \hat{\phi})$
SYMARMA(0,1)-normal	676,46	-332,52
<b>SYMARMA(1,1)-normal</b>	<b>674,51</b>	<b>-329,64</b>
SYMARMA(2,0)-normal	680,37	-332,57
SYMARMA(2,1)-normal	678,31	-329,64

A partir de agora vamos considerar distribuições mais flexível para ajustar os dados com o objetivo de acomodar melhor as observações atípicas presentes na série. Foram ajustados modelos SYMARMA- $t$  com os graus de liberdade variando entre 3 e 12, e modelos SYMARMA-EP com o parâmetro  $\kappa$  variando entre 0 e 1; ver Tabelas 4.10 e 4.11, respectivamente. De acordo com as medidas de comparação adotadas, os modelos SYMARMA- $t$  com 3 graus de liberdade e SYMARMA-EP com o parâmetro  $\kappa = 0,5$  obtiverem os melhores ajustes dentre os modelos considerados.

As estimativas e os erros-padrão assintóticos das estimativas dos parâmetros dos três modelos ajustados, juntamente com algumas medidas de comparação do ajuste, são apresentadas na Tabela 4.12. Notamos que, para todas as medidas utilizadas, o modelo SYMARMA- $t_3$  supera a modelagem usual do modelo ARMA. Além disso, o modelo SYMARMA- $t_3$  possui menores erros padrão assintóticos para as estimativas dos parâmetros do modelo. Por fim, a dispersão da série é melhor acomodada pelo modelo SYMARMA- $t_3$ .

A Figura 4.12 apresenta os gráficos de correlação e correlação parcial dos resíduos obtidos do ajuste do modelo SYMARMA- $t_3$ . O teste de Ljung-Box, utilizado para checar a aleatorie-

**Tabela 4.10** Valores para algumas medidas de comparação obtidas do ajuste de modelos SYMARMA- $t$  para a série IGP-OG.

Modelo	AIC	BIC	RMSE
<b>SYMARMA-<math>t_3</math></b>	<b>574,22</b>	<b>581,36</b>	<b>3,1405</b>
SYMARMA- $t_4$	592,54	599,68	3,5842
SYMARMA- $t_5$	606,06	613,20	4,1795
SYMARMA- $t_8$	627,62	634,76	7,2037
SYMARMA- $t_{12}$	637,70	644,84	10,7221

**Tabela 4.11** Valores para algumas medidas de comparação obtidas do ajuste de modelos SYMARMA-EP para a série IGP-OG.

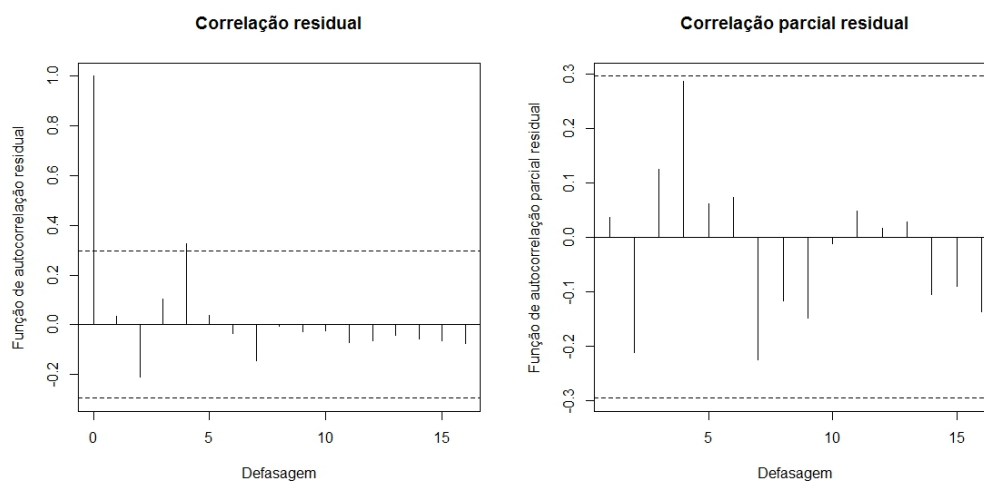
Modelo	AIC	BIC	RMSE
SYMARMA-EP(0,1)	647,83	654,97	14,5577
SYMARMA-EP(0,3)	636,20	643,34	6,5921
<b>SYMARMA-EP(0,5)</b>	<b>621,78</b>	<b>628,92</b>	<b>5,3017</b>
SYMARMA-EP(0,7)	621,48	628,62	8,3588

dade dos resíduos, sugere a não correlação dos resíduos ao nível de 5% de significância (valor  $p$  igual a 0,8038). As séries ajustadas pelos modelos SYMARMA-normal e SYMARMA- $t_3$ , juntamente com a série original, são apresentadas na Figura 4.13. Os resultados mostram os valores ajustados pela distribuição normal foram bastante influenciados pelos *outliers* presentes na série, enquanto que, os valores ajustados pelo modelo SYMARMA- $t_3$  permanecem estáveis ao longo de todo período.

A previsão fornecida pelo modelo SYMARMA- $t_3$  para inflação no ano de 2015, baseada na série histórica do índice IGP-OG, foi de 4,97%.

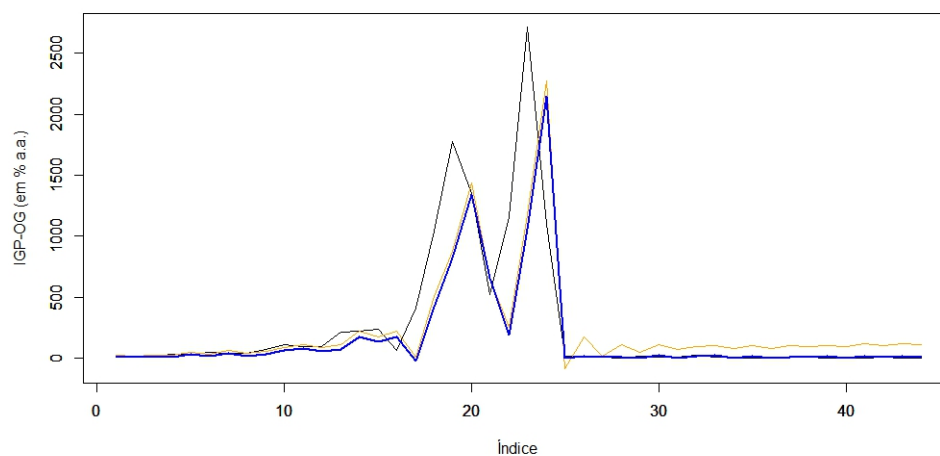
A partir de agora vamos avaliar a influência local das observações através das medidas de influência propostas na Seção 3.2 do Capítulo 3 e o ajuste dos modelos SYMARMA-normal e SYMARMA- $t_3$ . Utilizamos dois esquemas de perturbação: perturbação aditiva nos dados

**Figura 4.12** Função de autocorrelação e autocorrelação parcial dos resíduos obtidos do ajuste do modelo SYMARMA- $t_3$  - série IGP-OG.



Fonte: Autoria própria.

**Figura 4.13** Série original (linha preta), série ajustada pelo modelo SYMARMA-normal (linha laranja) e série ajustada pelo modelo SYMARMA- $t_3$  (linha azul) - série IGP-OG.



Fonte: Autoria própria.

**Tabela 4.12** Estimativas e erros-padrão assintóticos (em parênteses) das estimativas dos parâmetros dos três modelos ajustados e algumas medias de comparação - série IGP-OG.

Modelo	Parâmetro	Estimativas	AIC	BIC	RMSE
SYMARMA-normal	$\beta$	7,158 (3,818)			
	$\phi$	0,465 (0,127)	667,28	674,51	21,45
	$\theta$	0,594 (0,197)			
	$\varphi$	137.729,6 (29.372,6)			
SYMARMA- $t_3$	$\beta$	0,535 (0,559)			
	$\phi$	0,482 (0,017)	574,22	581,36	3,14
	$\theta$	0,503 (0,027)			
	$\varphi$	1.638,64 (494,1)			
SYMARMA-EP(0,5)	$\beta$	-0,420 (3,110)			
	$\phi$	0,455 (0,098)	621,78	628,92	5,30
	$\theta$	0,612 (0,151)			
	$\varphi$	37.907,45 (9.898,3)			

e perturbação no parâmetro de dispersão. Os limiares  $BS_i$ ,  $BC_i^{Cook}$  e  $BC_i^{LS}$  para  $i = 0, 1$  e  $2$ , definidos na Seção 3.3 do Capítulo 3, são estimados a partir de 2.000 simulações de Monte Carlo.

Nas Tabelas 4.13 e 4.14 apresentamos, respectivamente, para os modelos SYMARMA-normal e SYMARMA- $t_3$ , as marcas de referência obtidas para a análise de diagnóstico do vetor de inclinação e do vetor de curvatura. Informações relacionadas ao diagnóstico de influência são apresentadas nas Tabelas 4.15 e 4.16. Os limiares para análise de influência global,  $BS_0$  e  $BC_0$ , foram calculados a partir do percentil 95%. Os limiares para análise de influência individual,  $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$ , foram calculados a partir do percentil 5%.

De acordo com os resultados apresentados para o modelo SYMARMA-normal, as medidas  $O_s$  e  $O_c$  são significativas sob o critério global para a perturbação no parâmetro de dispersão. Sob o critério individual, as observações  $y_{24}$  e  $y_{25}$  são identificadas como influentes para ambos esquemas de perturbação e medidas de influência. Notamos também que essas observações estão associadas a grandes resíduos. As observações  $y_{20}$  e  $y_{23}$ , que também estão associadas a grandes resíduos, foram destacadas pela curvatura de Cook sob o esquema de perturbação nos dados.

Analisando os resultados obtidos para o modelo SYMARMA- $t_3$ , identificamos influência global ao nível de 95% apenas para o diagnóstico de inclinação sob o esquema de perturbação no parâmetro de dispersão. Alguns pontos, como, por exemplo, as observações  $y_{24}$  e  $y_{25}$ , são classificados como individualmente influentes.

As Figuras 4.14 e 4.15 ilustram o comportamento dos vetores de diagnóstico  $\mathbf{d}$ ,  $\mathbf{C}$  e  $\mathbf{C}_\ell$  (em valores absolutos), calculados a partir do ajuste dos modelos SYMARMA-normal e SYMARMA- $t_3$ , para a perturbação aditiva na variável resposta e a perturbação no parâmetro de dispersão, respectivamente. Os limiares  $BS_i$  e  $BC_i$  estão representados por linhas contínuas (para  $i = 1$ ) e por linhas tracejadas (para  $i = 2$ ). Notamos que os valores de  $d_i$ ,  $c_i$  e  $c_{\ell_i}$  para o modelo SYMARMA- $t_3$  são muito próximos as margens de referência  $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$ . Dessa forma, recalculamos os limiares para análise de influência individual ( $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$ ), considerando agora as séries que foram classificadas como globalmente influentes a partir do percentil 99%. Os resultados mostraram que as observações, anteriormente classificadas como influentes no ajuste do modelo SYMARMA- $t_3$ , não são mas significativas sob o critério individual; com excessão apenas para o diagnóstico de curvatura de Lesaffre e Verbeke sob o esquema de perturbação nos dados. Para o ajuste do modelo SYMARMA-normal, as conclusões permaneceram inalteradas.



**Tabela 4.13** Limiares para o modelo SYMARMA-normal calculados para série inflacionária IGP-OG.

Limiares ( <i>Benchmark</i> )	Estatística	Tipo de perturbação					
		Aditiva			Dispersão		
		Inclinação	Curvatura de Cook	Curvatura de Verbeke	Inclinação	Curvatura de Cook	Curvatura de Verbeke
BS <sub>0</sub>  BC <sub>0</sub>	99,0	0,0245	6,6372 e <sup>-05</sup>	6,6372 e <sup>-05</sup>	8,7086	7,0250	7,0250
	97,5	0,0236	5,4438 e <sup>-05</sup>	5,4438 e <sup>-05</sup>	6,4128	5,7229	5,7229
	95,0	0,0230	4,9854 e <sup>-05</sup>	4,9854 e <sup>-05</sup>	5,9799	5,0184	5,0184
	90,0	0,0217	4,4459 e <sup>-05</sup>	4,4659 e <sup>-05</sup>	5,5344	4,5130	4,5130
BS <sub>1</sub>  BC <sub>1</sub> <sup>Cook</sup>  BC <sub>1</sub> <sup>LS</sup>	99,0	0,0213	0,5285	1,3508 e <sup>-05</sup>	13,5105	0,9039	5,1518
	97,5	0,0196	0,4935	1,1116 e <sup>-05</sup>	10,4619	0,8716	3,8830
	95,0	0,0187	0,4673	0,9713 e <sup>-05</sup>	9,3989	0,8422	3,3512
	90,0	0,0175	0,4441	0,8648 e <sup>-05</sup>	7,8790	0,8001	2,8448
BS <sub>2</sub>  BC <sub>2</sub> <sup>Cook</sup>  BC <sub>2</sub> <sup>LS</sup>	Máx	0,0256	0,5073	2,5156 e <sup>-05</sup>	20,3053	0,9619	9,3400
	Mín	0,0128	0,2683	0,5856 e <sup>-05</sup>	7,2252	0,3974	0,7961
	5.0	0,0145	0,3086	0,6658 e <sup>-05</sup>	7,8174	0,4558	2,7160
	Média	0,0178	0,3756	1,0493 e <sup>-05</sup>	11,2983	0,7665	4,0085
	DP	0,0025	0,0464	0,3683 e <sup>-05</sup>	2,9956	0,1427	1,4586

**Tabela 4.14** Limitares para o modelo SYMARMA- $t_3$  calculados para série inflacionária IGP-OG.

Limitares ( <i>Benchmark</i> )	Estatística	Tipo de perturbação					
		Aditiva			Dispersão		
		Inclinação	Curvatura de Cook	Curvatura de Verbeke	Inclinação	Curvatura de Cook	Curvatura de Verbeke
BS <sub>0</sub>  BC <sub>0</sub>	99,0	0,2117	0,0086	0,0086	4,6582	9,8976	9,8976
	97,5	0,2019	0,0045	0,0045	4,0281	3,5743	3,5743
	95,0	0,1910	0,0036	0,0036	3,8340	2,3179	2,3179
	90,0	0,1803	0,0030	0,0030	3,7172	1,8662	1,8662
BS <sub>1</sub>  BC <sub>1</sub> <sup>Cook</sup>  BC <sub>1</sub> <sup>LS</sup>	99,0	0,1112	0,9064	0,0032	5,7257	0,8484	4,1708
	97,5	0,0953	0,8149	0,0024	4,2668	0,7928	1,4792
	95,0	0,0872	0,7266	0,0019	3,1471	0,7248	1,0568
	90,0	0,0810	0,6327	0,0015	2,9292	0,6645	0,7959
BS <sub>2</sub>  BC <sub>2</sub> <sup>Cook</sup>  BC <sub>2</sub> <sup>LS</sup>	Máx	0,1314	0,9762	0,0703	12,3131	0,9997	345,6807
	Mín	0,0765	0,2251	0,0004	2,4348	0,3279	0,3274
	5.0	0,0793	0,2609	0,0008	2,6541	0,3677	0,5585
	Média	0,0874	0,4487	0,0037	4,5539	0,6105	12,0027
	DP	0,0085	0,1685	0,0080	1,9352	0,1665	48,9433

**Tabela 4.15** Diagnóstico de influência local para a série inflacionária IGP-OG considerando o modelo SYMARMA-normal.  $r_t$  são os resíduos.

Data (ano)	Observação	$r_t$	Tipo de perturbação					
			Aditiva			Dispersão		
			Inclinação	Curvatura de Cook	Curvatura de Verbeke	Inclinação	Curvatura de Cook	Curvatura de Verbeke
1988	20	882,38	-0,0131	<b>0,3631</b>	$0,5038 \text{ e}^{-05}$	4,8005	0,2860	1,6288
1991	23	872,27	-0,0130	<b>0,3589</b>	$0,4516 \text{ e}^{-05}$	4,6682	0,1133	1,2450
1992	24	1.532,69	<b>-0,0228</b>	<b>0,6307</b>	<b><math>1,4111 \text{ e}^{-05}</math></b>	<b>16,5008</b>	<b>0,7581</b>	<b>12,2653</b>
1993	25	-1.130,00	<b>0,0168</b>	<b>-0,4650</b>	<b><math>1,4321 \text{ e}^{-05}</math></b>	<b>8,5127</b>	<b>-0,5654</b>	<b>12,1172</b>
1994	26	107,24	-0,0016	0,0441	<b><math>1,1058 \text{ e}^{-05}</math></b>	-0,9143	-0,0152	0,1267
$O_s$ e $O_c$			0,0181	$2,9799 \text{ e}^{-05}$	$2,9799 \text{ e}^{-05}$	<i>10,2859</i>	<i>16,4671</i>	<i>16,4671</i>
$BS_0$ e $BC_0$			0,0230	$4,9854 \text{ e}^{-05}$	$4,9854 \text{ e}^{-05}$	5,9799	5,0184	5,0184
$BS_2$ , $BC_1^{Cook}$ e $BC_1^{LV}$			0,0187	0,4673	$0,9713 \text{ e}^{-05}$	9,3989	0,8422	3,3512
$BS_2$ , $BC_2^{Cook}$ e $BC_2^{LV}$			0,0145	0,3086	$0,6658 \text{ e}^{-05}$	7,8174	0,4558	2,7160

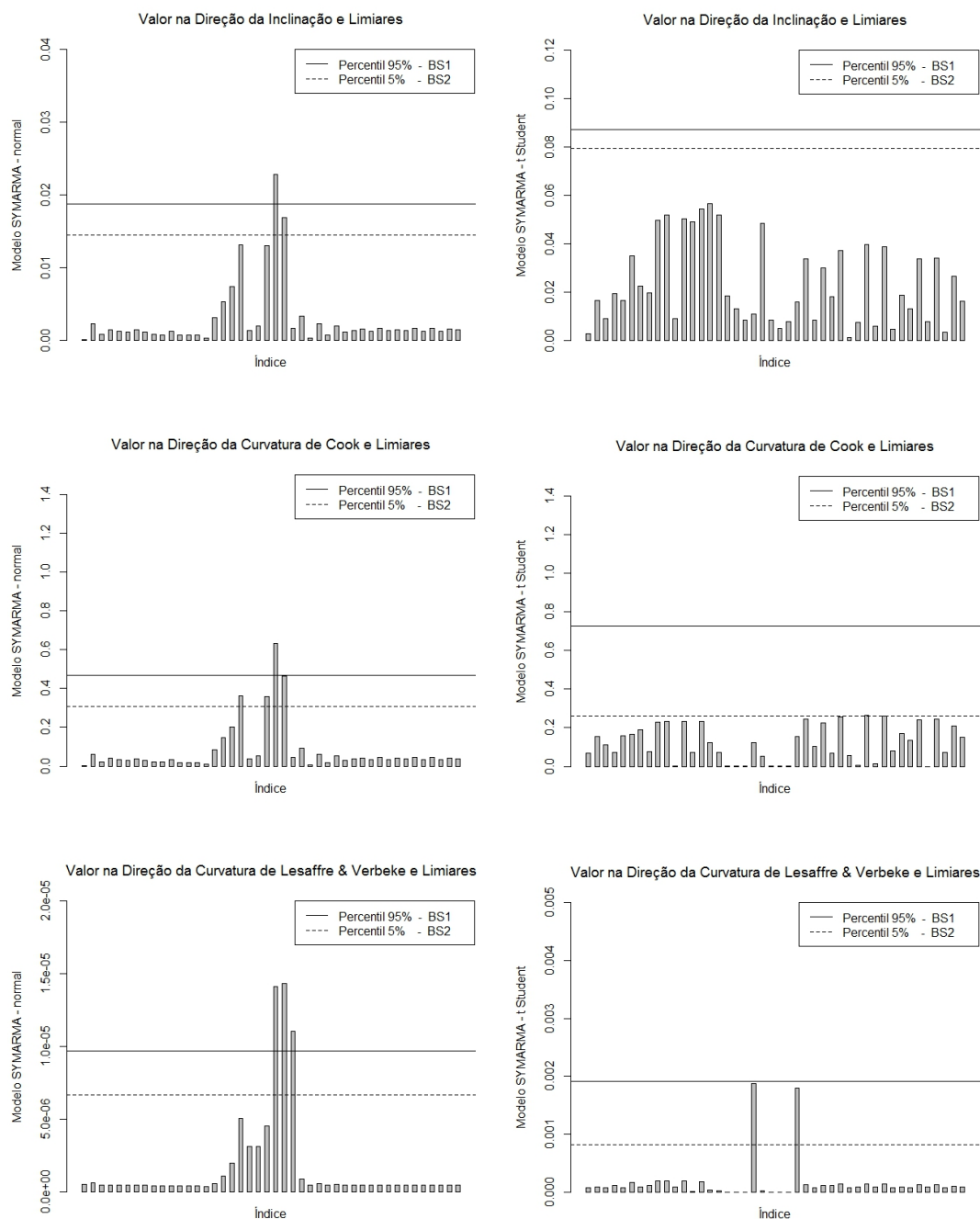
Nota: Pontos influentes sob  $BS_2$ ,  $BC_2^{Cook}$  e  $BC_2^{LV}$  são apresentados em negrito. Estatísticas globalmente significativas  $O_s$  e  $O_c$  são apresentadas em itálico.

**Tabela 4.16** Diagnóstico de influência local para a série inflacionária IGP-OG considerando o modelo SYMARMA- $t_3$ .  $r_t$  são os resíduos.

Data (ano)	Observação	$r_t$	Tipo de perturbação					
			Aditiva			Dispersão		
			Inclinação	Curvatura de Cook	Curvatura de Verbeke	Inclinação	Curvatura de Cook	Curvatura de Verbeke
1979	11	37,07	-0,0519	0,2335	1,8607 e <sup>-04</sup>	-0,0382	<b>-0,3871</b>	0,0983
1983	15	40,93	-0,0544	0,2326	1,7708 e <sup>-04</sup>	0,1138	<b>-0,4034</b>	0,1296
1986	18	423,88	-0,0184	0,0037	5,4153 e <sup>-08</sup>	<b>2,9057</b>	-0,0352	9,1656 e <sup>-04</sup>
1987	19	612,82	-0,0129	0,0022	3,4551 e <sup>-08</sup>	<b>2,9543</b>	-0,0171	2,1825 e <sup>-04</sup>
1988	20	952,77	-0,0084	0,0009	1,6804 e <sup>-08</sup>	<b>2,9810</b>	-0,0070	3,7652 e <sup>-05</sup>
1989	21	5,84	-0,0107	-0,1240	<b>0,0019</b>	-0,9688	0,0010	0,0652
1991	23	965,12	-0,0082	0,0005	4,3319 e <sup>-09</sup>	<b>2,9815</b>	-0,0068	3,5828 e <sup>-05</sup>
1992	24	1.652,12	-0,0048	0,0003	3,2574 e <sup>-09</sup>	<b>2,9937</b>	-0,0023	4,1989 e <sup>-06</sup>
1993	25	-1.044,01	0,0076	0,0010	7,9055 e <sup>-08</sup>	<b>2,9842</b>	-0,0059	2,9134 e <sup>-05</sup>
1994	26	8,82	-0,0160	0,1530	<b>0,0018</b>	-0,9296	0,0451	0,1449
2002	34	-24,51	0,0397	<b>-0,2636</b>	1,4212 e <sup>-04</sup>	-0,5135	-0,0510	0,0385
$O_s$ e $O_c$								
BS <sub>0</sub> e BC <sub>0</sub>			0,0954	0,0020	0,0020	4,8091	0,6402	0,6402
BS <sub>2</sub> , BC <sub>1</sub> <sup>Cook</sup> e BC <sub>1</sub> <sup>LV</sup>			0,1910	0,0036	0,0036	3,8340	2,3179	2,3179
BS <sub>2</sub> , BC <sub>2</sub> <sup>Cook</sup> e BC <sub>2</sub> <sup>LV</sup>			0,0872	0,7266	0,0019	3,1471	0,7248	1,0568
			0,0793	0,2609	0,0008	2,6541	0,3677	0,5585

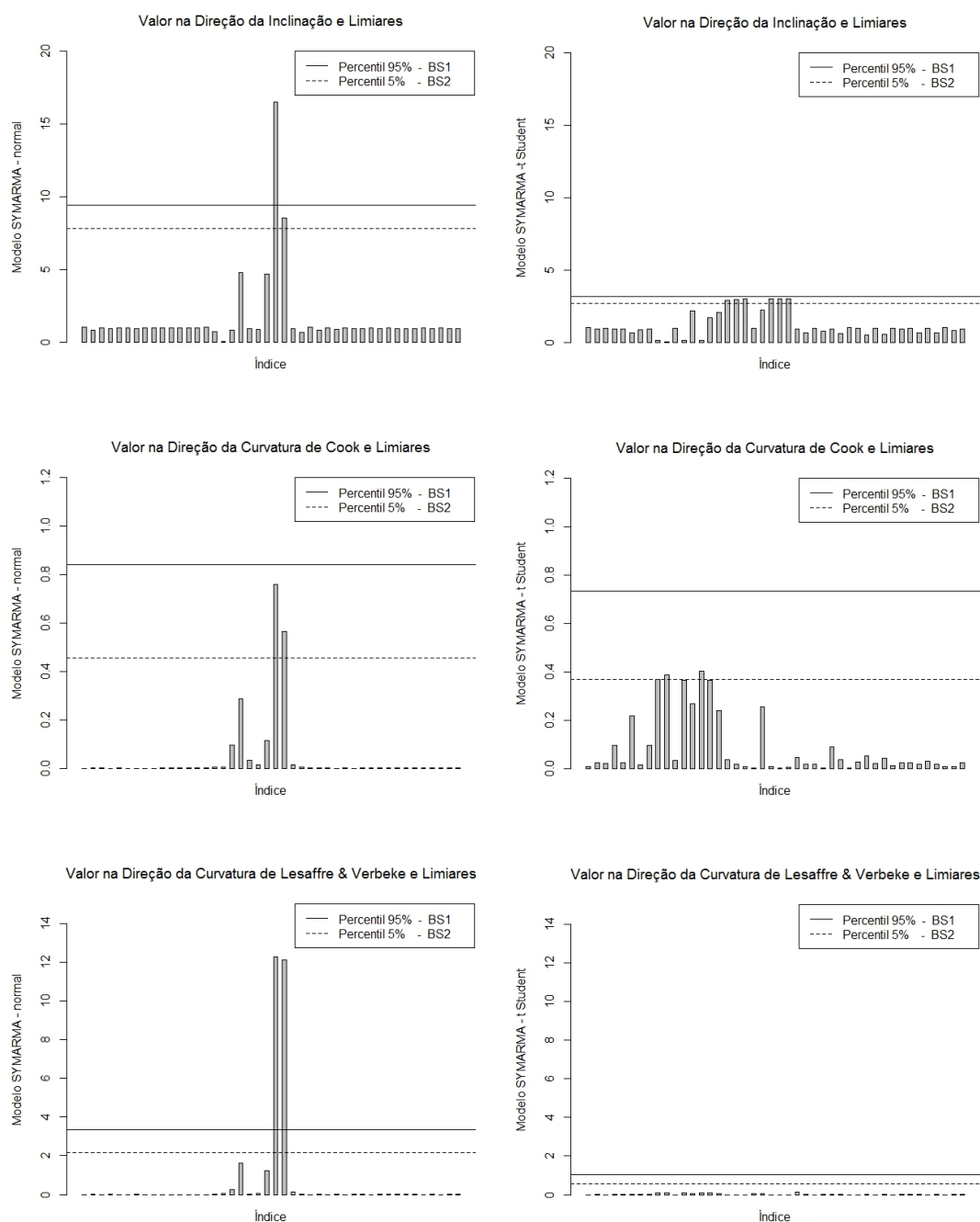
Nota: Pontos influentes sob BS<sub>2</sub>, BC<sub>2</sub><sup>Cook</sup> e BC<sub>2</sub><sup>LV</sup> são apresentados em negrito. Estatísticas globalmente significativas  $O_s$  e  $O_c$  são apresentadas em itálico.

**Figura 4.14** Diagnóstico de influência local na série inflacionária IGP-OG sob o esquema de perturbação nos dados para os modelos SYMARMA-normal (lado esquerdo) e SYMARMA- $t_3$  (lado direito).



Fonte: Autoria própria.

**Figura 4.15** Diagnóstico de influência local na série inflacionária IGP-OG sob o esquema de perturbação no parâmetro de dispersão para os modelos SYMARMA-normal (lado esquerdo) e SYMARMA- $t_3$  (lado direito).



Fonte: Autoria própria.

## Capítulo 5

### Rotinas *elliptical.ts* e *influence.ts*

É importante que toda a teoria desenvolvida nesta tese esteja disponível em algum software. Com esta preocupação foram desenvolvidos macros no software R (IHAKA; GENTLEMAN, 1996). O R é um ambiente integrado que possui grandes facilidades para a manipulação de dados, a geração de gráficos e a modelagem estatística. Devido ao seu código fonte ser aberto, o mesmo tem recebido inúmeras contribuições de várias comunidades científicas. O R encontra-se disponível em <http://www.r-project.org>, bem como diversos macros, que são implementações das mais variadas áreas de estudo.

Com a ideia de difundir a modelagem estatística para dados temporais, desenvolvemos as funções *elliptical.ts*, *QQplot* e *influence.ts*. A função *elliptical.ts* consiste em um conjunto de rotinas computacionais que permitem a definição de distribuições pertencentes a classe simétrica e o ajuste dos parâmetros autoregressivos, de médias móveis e de dispersão do modelo SYMARMA pelo método de máxima verossimilhança condicional. A função *QQplot* permite gerar o gráfico normal de probabilidade com envelope e, por fim, a função *influence.ts* possui rotinas que permitem o diagnóstico de influência local em séries temporais a partir da metodologia de limiares proposta por Zhang e King (2005). Estes conjuntos de rotinas encontram-se disponíveis gratuitamente para uso acadêmico em [http://www.de.ufpe.br/~cysneiros/elliptical/time\\_series.html](http://www.de.ufpe.br/~cysneiros/elliptical/time_series.html).

Vamos agora apresentar a sintaxe do comando para o ajuste de um modelo SYMARMA sob uma particular distribuição simétrica condicional.

```
elliptical.ts <- function(Y, X=cbind(seq(1,1,length=length(Y))), family="Normal",
epsilon=0.0001, maxit=100, trace="F", np, nq, df=3, epw=0.3, dfs=1, alpha=1, mp=1,
epsi=1, sigmap=1, fixed=NULL, ...)
```

A função *elliptical.ts* depende do pacote *Matrix*. Covariáveis precisam ser listadas em uma matriz de dimensão  $n \times k$ , em que  $n$  é o número de observações e  $k$  é o número de covariáveis. O intercepto é considerado uma covariável de uns. Após o ajuste do modelo utilizando a função *elliptical.ts* ficará disponível uma lista de objetos gerados, tais como:

**coefficients:** coeficientes autorregressivos e de médias móveis do modelo ajustado;

**dispersion:** coeficiente de dispersão do modelo ajustado;

**residuals:** resíduo  $(y - \mu) / \sqrt{\hat{\varphi}}$ ;

**fitted.values:** valores ajustados;

**loglik:** o logaritmo da função de verossimilhança maximizada do modelo ajustado;

**Wg:** os valores da função  $W_g(u)$ ;

**Wgder:** os valores da função  $W'_g(u)$ ;

**iter:** número de iterações;

**scale:**  $4d_g$ ;

**scaledispersion:**  $4f_g - 1$ ;

**scalevariance:**  $\xi$ ;

**DesP.C:** desvio-padrão assintótico das estimativas dos coeficientes autoregressivos e de médias móveis do modelo ajustado;

**DesP.D:** desvio-padrão assintótico da estimativa do parâmetro de dispersão do modelo ajustado;



**rmse:** raiz do erro médio quadrático.

Na opção *family*, definimos a família de distribuição a ser ajustada. Esta função, até o presente momento, está definida para as distribuições abaixo:

**Normal:** *family* = Normal();

**t de Student:** *family* = Student(*df*);

**t de Student Generalizada:** *family* = Gstudent(*c(df,dfs)*);

**Logística-I:** *family* = LogisticI();

**Logística-II:** *family* = LogisticII();

**Logística Generalizada:** *family* = Glogistic(*c( $\alpha$ ,mp)*);

**Exponencial Potência:** *family* = ExpPower(*epw*).

As opções *epsilon* e *maxit* definem o erro e o número máximo de iterações para convergência, respectivamente. As opções *np* e *nq* definem as ordens dos parâmetros autorregressivos e de médias móveis, respectivamente. A opção *trace* permite a impressão dos resultados do ajuste. Por fim, a opção *fixed=c(ar,ma,X)* permite selecionar apenas alguns parâmetros para serem estimados, por exemplo, para estimar apenas o parâmetro  $\phi_3$  em um modelo SYMARMA(3,0) utilizaremos *fixed=c(0,0,NA)*. O gráfico normal de probabilidade com envelope é gerado pela função *QQplot*.

```
QQplot<- function(fit, B=300)
```

A opção *fit* define o modelo SYMARMA ajustado e *B* o número de iterações.

Vamos agora apresentar a sintaxe do comando para análise de diagnóstico sob o enfoque de influência local considerando o ajuste de um modelo SYMARMA sob uma particular distribuição simétrica condicional.

```
influence.ts <- function(adjust.model, esq.pert="aditive", diag="incl", seed=2015,
benchmarks.rep=2000, perc.ind1=95, perc.ind2=5, plot="TRUE", trace="TRUE")
```

Na opção *adjust.model* definimos o modelo SYMARMA ajustado pela função *elliptical.ts*, na opção *esq.pert* definimos o esquema de perturbação, na opção *diag* temos a medida de influência local adotada, o comando *plot* permite gerar o gráfico sob o diagnóstico de influência individual. Por fim, *trace* fornece alguns resultados gerais e as opção *benchmarks.rep*, *perc.ind1*, *perc.ind2* e *seed* definem os parâmetros para o cálculo dos limiares.

Até o presente momento, as funções *QQplot* e *influence.ts* estão definidas para as distribuições normal e *t*-Student, a opção *esq.pert* para as perturbações: perturbação aditiva na variável resposta ("aditive") e perturbação no parâmetro de dispersão ("dispersion"), e a opção *diag* para as medidas de influência local propostas por Billor e Loynes (1993), Cook (1986) e Lesaffre e Verbeke (1989), através dos comandos "incl", "cook" e "lv", respectivamente.

Após a análise de diagnóstico utilizando a função *influence.ts* ficará disponível uma lista de objetos gerados, tais como:

**B0:** marca de referência para a medida de influência global;

**B1:** marca de referência para a medida de influência individual 1;

**B2:** marca de referência para a medida de influência individual 2;

**glob:** medida de influência para série em estudo - critério global;

**indiv:** vetor de influência para série em estudo - critério individual.

Vamos agora utilizar a série dos excessos de retorno nas ações da *Microsoft* (ver Seção 4.2, Capítulo 4), para ilustrar algumas saídas disponibilizadas pelas funções implementadas considerando o ajuste do modelo SYMARMA- $t_5$ . Os comandos utilizados são apresentados a seguir.

```
library(Matrix)
dat <- scan(what=list(msf=0,sp500=0,tbill=0))
30.09 1146.54 1.76
```

28.55 1136.76 1.76

28.07 1125.40 1.75

⋮

23.83 893.92 1.62

*attach(dat)*

*.tbill <- tbill/253*

*N <- length(tbill)*

*.sp500 <- ((sp500[2:N]-sp500[1:(N-1)])/sp500[1:(N-1)])\*100*

*.msf <- ((msf[2:N]-msf[1:(N-1)])/msf[1:(N-1)])\*100*

*X <- .sp500-.tbill[1:(N-1)]*

*Y <- .msf-.tbill[1:(N-1)]*

*X <- cbind(X)*

*fit2 <- elliptical.ts(Y, X, family="Student", df=5, np=12, nq=0,  
trace=TRUE, fixed=c(0,0,0,0,0,0,0,0,0,0,0,NA,NA))*

*Call:*

*symarma(12,0) - family: Student*

*Coefficients:*

.	<i>Estimative</i>	<i>s.d.</i>
<i>intercept</i>	1.2732547	0.09657443
<i>ar1</i>	0.0000000	0.00000000
<i>ar2</i>	0.0000000	0.00000000
<i>ar3</i>	0.0000000	0.00000000
<i>ar4</i>	0.0000000	0.00000000
<i>ar5</i>	0.0000000	0.00000000
<i>ar6</i>	0.0000000	0.00000000
<i>ar7</i>	0.0000000	0.00000000

```

ar8      0.0000000  0.00000000
ar9      0.0000000  0.00000000
ar10     0.0000000  0.00000000
ar11     0.0000000  0.00000000
ar12     -0.1619578  0.09284748

```

*varphi* estimated as 2.6282 (s.d. = 0.4774)

log likelihooh = -203.8393

rmse = 3.34

```
rstand.t <- fit2$residuals
```

```
QQplot(fit2)
```

```
influence.ts(fit2,esq.pert="aditive",diag="incl")
```

"Benchmarks - 00% ..."

"Benchmarks - 25% ..."

"Benchmarks - 50% ..."

"Benchmarks - 75% ..."

"Benchmarks - 100%"

Call:

*symarma*(12,0) - family: Student

Billor & Loynes

Measures of local influence:

Inclination 5.320019

Benchmarks:

*BS0 6.309945*

*BS1 1.968821*

*BS2 1.937233*

*influence.ts(fit2,esq.pert="dispersion",diag="cook")*

*"Benchmarks - 00% ..."*

*"Benchmarks - 25% ..."*

*"Benchmarks - 50% ..."*

*"Benchmarks - 75% ..."*

*"Benchmarks - 100%"*

*Call:*

*symarma(12,0) - family: Student*

*Cook*

*Measures of local influence:*

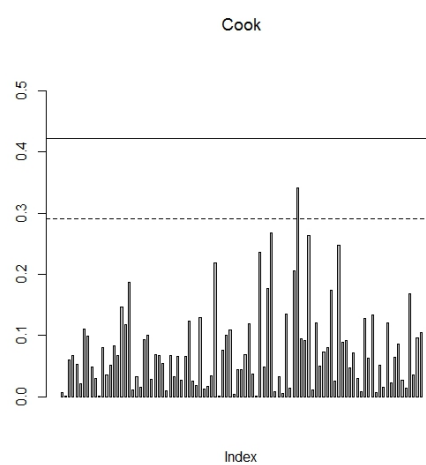
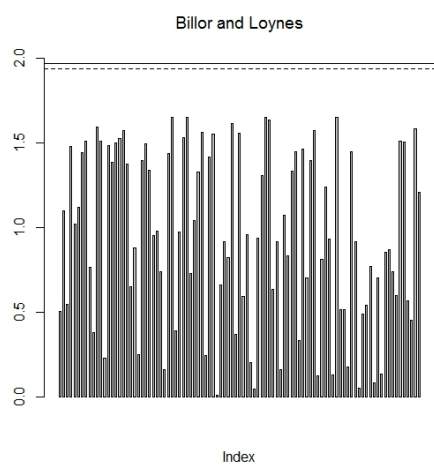
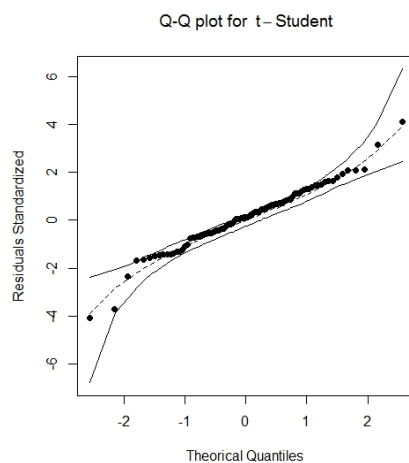
*Curvature 21.07709*

*Benchmarks:*

*BC0 37.1777248*

*BC1 0.4220557*

*BC2 0.2913184*



## Capítulo 6

### Conclusões e Considerações Finais

Nessa tese trabalhamos com a classe de modelos SYMARMA que estende a classe de modelos autorregressivos de médias móveis gaussiano (ARMA) para séries temporais pertencentes à classe de distribuições simétricas. Foram apresentadas expressões mais simples para o vetor escore e para matriz informação esperada de Fisher. Estudos de simulação apontaram que os estimadores de máxima verossimilhança condicional para os parâmetros do modelo são consistentes em erro quadrático médio e normalmente distribuídos. Resultados de simulação baseados na função de influência dão indícios de que o procedimento para estimação dos parâmetros do modelo SYMARMA-normal não é robusto, podendo fornecer estimativas bastantes viesadas para séries perante *outliers*, enquanto que, o procedimento para estimação dos parâmetros no modelo SYMARMA- $t$  é considerado qualitativamente robusto. Adicionalmente, acrescentamos componentes sazonais aos modelos SYMARMA dando origem aos modelos SYMARMA sazonais.

A maior contribuição dessa tese é a utilização do uso da metodologia de limiares, proposta por Zhang e King (2005), no diagnóstico de influência local para modelos SYMARMA. A partir dessa metodologia foi possível estabelecer marcas de referência que determinam, estatisticamente, que observações da série podem ser caracterizadas como influentes. Na maioria das vezes essa caracterização é feita por meio de uma simples análise exploratória. Foram derivadas as expressões das medidas de influência local propostas por Cook (1986), Billor e Loynes (1993) e Lesaffre e Verbeke (1998), considerando os esquemas de perturbação nos dados, perturbação no parâmetro de dispersão, perturbação no vetor de parâmetros autorregressivos e

perturbação no vetor de parâmetros de médias móveis.

Realizamos estudos de simulação em modelos SYMARMA-normal e SYMARMA- $t$ , e dessa forma foi possível verificar que, baseados na técnica de limiares, o modelo SYMARMA- $t$  tende a acomodar melhor *outliers* em séries temporais não destacando-os como influentes. Esse fato evidencia a robustez dos modelos SYMARMA com caudas mais pesadas no ajuste de séries temporais que apresentem *outliers*.

Desenvolvemos as bibliotecas *elliptical.ts* e *influence.ts* que consistem em um conjunto de rotinas computacionais que permitem a definição de distribuições pertencentes a classe simétrica e o ajuste dos parâmetros autorregressivos, de médias móveis e de dispersão do modelo SYMARMA pelo método de máxima verossimilhança condicional. Além disso, permitem o diagnóstico de influência local em séries temporais a partir da metodologia de limiares.

Através de dois exemplos práticos, notamos que os modelos SYMARMA- $t$  e SYMARMA-EP acomodaram melhor as observações atípicas presentes nas séries estudadas, quando comparados aos resultados fornecidos pelo modelo SYMARMA-normal.

Vários trabalhos podem ser desenvolvidos a partir dos resultados apresentados nessa tese. Dentre eles podemos citar:

1. Dar continuidade aos estudos acerca dos modelos SYMARMA. Como por exemplo, estudar o comportamento dos limiares e, conseqüentemente, das medidas de influência local em modelos SYMARMA sazonais considerando também a presença de *outliers* de inovação;
2. Desenvolver a classe de modelos SYMARMA para distribuições condicionais simétricas contínuas com heteroscedasticidade e estudar medidas de diagnóstico para esta nova classe de modelos;
3. Incorporar a ideia de variável resposta transformada a classe de modelos SYMARMA e desenvolver procedimentos para estimação dos parâmetros e análise de diagnóstico. A ideia de variável resposta transformada para dados temporais foi utilizada por Cordeiro e Andrade (2009) no desenvolvimento do modelo TGARMA, autorregressivo de médias móveis generalizado transformado.



Por fim, esperamos que este estudo possa ampliar os conhecimentos a respeito de séries temporais não-gaussianas, inspirando novas pesquisas nesse campo.

## Referências

- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, p. 267-281, 1973.
- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on automatic control*, V. 19, p. 716-723, 1974.
- BENJAMIN, M. A.; RIGBY, R.A.; STASINOPOULOS, M. Generalized autoregressive moving average models. *Journal of the American Statistical Association*, v. 98, p. 214-223, 2003.
- BESLEY, D. A.; KUH, E.; WELSCH, R. E. *Regression diagnostics*. New York: John Wiley, 1980.
- BILLOR, N.; LOYNES, R. M. Local influence: A new approach. *Communications in Statistics Theory and Methods*, v. 22, p. 1595-1611, 1993.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society*, v. 26, p. 211-252, 1964.
- CHEN, C.; LUI, L. M. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, v. 88, p. 284-297, 1993.
- COOK, R. D.; WEISBERG, S. *Residuals and influence in regression*. New York: Chapman and Hall, 1982.
- COOK, R. D. Assessment of local influence (with discussion). *Journal of the Royal Statistical Society*, v. 48, p. 133-169, 1986.
- CORDEIRO G. M.; DE ANDRADE, M. G. Transformed generalized linear models. *Journal of Statistical Planning and Inference*, v. 139, p. 2970-2987, 2009.
- COX, D. R.; HINKLEY, D. V. *Theoretical statistics*. London: Chapman and Hall, 1974.
- COX, D. R. Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, v. 8, p. 93-115, 1981.
- CREAL, D.; KOOPMAN, S. J.; LUCAS, A. Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, v. 28, p. 777-795, 2013.

- CRIBARI-NETO, F.; CASSIANO, K. M. Uma análise da dinâmica inflacionária brasileira. *Revista Brasileira de Economia*, v. 59, p. 535-566, 2005.
- CYSNEIROS, F. J. A. *Métodos restritos e validação de modelos simétricos de regressão*. Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, Brasil, 2004.
- CYSNEIROS, F. J. A.; PAULA, G. A. Restricted methods in symmetrical linear regression models. *Computation Statistics and Data Analysis*, v. 49, p. 689-708, 2005.
- EFRON, B.; TIBSHIRANI, R. J. *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.
- FANG, K. T.; KOTZ, S.; NG, K. W. *Symmetric multivariate and related distributions*. London: Chapman and Hall, 1990.
- FOX, A. J. Outliers in times series. *Journal of the Royal Statistical Society*, v. 34, p. 350-363, 1972.
- GALEA, M.; PAULA, G. A.; URIBE-OPAZO, M. On influence diagnostic in univariate elliptical linear regression models. *Statistical Papers*, v. 44, p. 23-45, 2003.
- HEYDE, C. C.; FEIGIN, P. D. On efficiency and exponential families in stochastic process estimation. *Statistical Distributions in Scientific Work*, v. 1, p. 227-240, 1975.
- IHAKA, R.; GENTLEMAN, R. R. A language for data analysis and graphics. *Journal of Computational Graphical and Statistics*, v. 5, p. 299-314, 1996.
- KELKER, D. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā*, v. 32, p. 419-430, 1970.
- LESAFFRE, F.; VERBEKE, G. Local influence in linear mixed models. *Biometrics*, v. 38, p. 963-974, 1998.
- LI, W. K. Testing model adequacy for some Markov regression models for time series. *Biometrika*, v. 78, p. 83-89, 1991.
- LI, W. K. Time series models based on generalized linear models: some further results. *Biometrics*, v. 50, p. 506-511, 1994.
- LIU, S. Local influence in multivariate elliptical linear regression models. *Linear Algebra and Its Applications*, v. 354, p. 159-174, 2002.
- Ljung, G.M. e Box, G.E.P. (1978). On a measure of a lack of fit in time series models. *Biometrika*, **65**, 297-303.
- LUCAS, A. Robustness of the Student-t based M-estimator. *Communications in Statistics Theory and Methods*, v. 26, p. 1165-1182, 1997.
- MAIOR, V. Q. S.; CYSNEIROS, F. J. A. Estimação do risco sistemático em modelos CAPM com erros normais assimétricos. *Revista Brasileira de Biometria*, v. 27, p. 197-209, 2009.

MAIOR, V. Q. S. *SYMARMA: Um modelo dinâmico para dados temporais sob distribuição simétrica condicional*. Dissertação de mestrado, Departamento de Estatística Universidade Federal de Pernambuco, Brasil, 2012.

MARTIN, R. D.; YOHAI, V. J. Robustness in time series and estimating ARMA models. *Handbook of Statistics*, v. 5, p. 119-155, 1985.

NOCEDAL, J.; WRIGHT, S. J. *Numerical optimization*. New York: Springer, 1999.

OTA, R. *Valores aberrantes em séries temporais: teste de detecção e efeito na previsão de valores agregados*. Dissertação de mestrado, Departamento de Estatística, UNICAMP, Brasil, 1996.

PAULA, G. A.; MEDEIROS, M.; VILCA-LABRAB, F. E. Influence diagnostics for linear models with first-order autoregressive elliptical errors. *Statistics and Probability Letters*, v. 79, p. 339-346, 2009.

PAULA, G. A.; CYSNEIROS, F. J. A. Systematic risk estimation in symmetric models. *Applied Economics Letters*, v. 16, p. 217-221, 2009.

ROCHA, A. V.; CRIBARI-NETO, F. Beta autoregressive moving average models. *Test (Madrid)*, v. 18, p. 529-545, 2009.

RUPPERT, D. *Statistics and Finance*. New York: Springer, 2004.

SCHALL, R.; DUNNE, T. T. *Diagnostics for Regression-ARMA Time Series*. New York: Springer, 1991.

SCHWARZ, G. E. Estimating the dimension of a model. *Annals of Statistics*, v. 6, p. 461-464, 1978.

SHEPHARD, N. *Generalized linear autoregressions*. Technical report, Nuffield College, Oxford University, 1995.

ZEGER, S. L.; QAQISH, B. Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, v. 44, p. 1019-1031, 1988.

ZEVALLOS, M.; HOTTA, L. K. Influential observations in GARCH models. *Journal of Statistical Computation and Simulation*, v. 82, p. 1571-1589, 2012.

ZHANG, X.; KING, M. L. Influence diagnostics in generalized autoregressive conditional heteroscedasticity processes. *Journal of Business and Economic Statistics*, v. 23, p. 118-129, 2005.

## Apêndice A

### Prova do Teorema 1

*Prova:* Denote  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  o polinômio autorregressivo,  $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  o polinômio de médias móveis e  $B^k y_t = y_{t-k}$  o operador de defasagens. Podemos reescrever o modelo SYMARMA como

$$\Phi(B)(Y_t - \mathbf{x}_t^\top \beta) = \Theta(B)r_t,$$

e, desde que  $\Phi(B)$  seja invertível,

$$Y_t = \mathbf{x}_t^\top \beta + \Psi(B)r_t,$$

em que  $r_t$ 's são resíduos tais que  $E(r_t) = 0$  e  $E(r_t r_j) = 0$ , para cada  $t \neq j$ , e  $\Psi(B) = \Theta(B)\Phi(B)^{-1}$ .

Desta forma, assumindo que  $\Phi(B)$  é invertível, a média marginal de  $Y_t$  é dada por

$$E(Y_t) = \mathbf{x}_t^\top \beta.$$

## Apêndice B

### Prova do Teorema 2

*Prova:* Seja  $Y_t = \mu_t + r_t$  em que  $r_t$ 's são resíduos não correlacionados com média zero. Temos que

$$\begin{aligned}\text{Var}(r_t) &= E(r_t^2) = E(E(r_t^2 | H_{t-1})) = E(\text{Var}(r_t | H_{t-1})) = E(\text{Var}(Y_t - \mu_t | H_{t-1})) \\ &= E(\text{Var}(y_t | H_{t-1})) = E(\xi \varphi) = \xi \varphi.\end{aligned}$$

Note que  $\mu_t$ , definido em (6), é  $H_{t-1}$ -mensurável. Portanto, desde que  $\Phi(B)$  seja invertível, a variância marginal de  $Y_t$ ,  $\text{Var}(Y_t)$ , é dada por

$$\begin{aligned}\text{Var}(Y_t) &= \text{Var}(\mathbf{x}_t^\top \beta + \Psi(B)r_t) = \text{Var}(\Psi(B)r_t) = E[(\Psi(B)r_t)^2] \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j E(r_{t-i} r_{t-j}) = \sum_{i=0}^{\infty} \psi_i^2 E(r_{t-i}^2) = \sum_{i=0}^{\infty} \psi_i^2 \text{Var}(r_{t-i}) \\ &= \xi \varphi \sum_{i=0}^{\infty} \psi_i^2.\end{aligned}$$

## Apêndice C

### Prova do Teorema 3

*Prova:* Pelos Teoremas 1 e 2, obtemos que

$$E(Y_t) = \mathbf{x}_t^\top \boldsymbol{\beta} \text{ e } \text{Var}(Y_t) = \xi \varphi \sum_{i=0}^{\infty} \psi_i^2.$$

Utilizando o modelo SYMARMA expresso por meio de um processo linear geral,

$$Y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \Psi(B)r_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=0}^{\infty} \psi_i r_{t-i},$$

temos que a função de covariância fica definida como

$$\begin{aligned} \text{Cov}(Y_t, Y_{t-k}) &= \text{Cov}(\mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=0}^{\infty} \psi_i r_{t-i}, \mathbf{x}_{t-k}^\top \boldsymbol{\beta} + \sum_{i=0}^{\infty} \psi_i r_{t-k-i}) \\ &= \text{Cov}(\psi_0 r_t + \psi_1 r_{t-1} + \cdots, \psi_0 r_{t-k} + \psi_1 r_{t-k-1} + \cdots) \\ &= \text{Var}(r_t) \sum_{i=0}^{\infty} \psi_i \psi_{i+k} \\ &= \xi \varphi \sum_{i=0}^{\infty} \psi_i \psi_{i+k}, \end{aligned}$$

em que  $\psi_0 = 1$ . Adicionalmente,

$$\begin{aligned}
 \text{Corr}(Y_t, Y_{t-k}) &= \frac{\text{Cov}(Y_t, Y_{t-k})}{\sqrt{\text{Cov}(Y_t, Y_t)\text{Cov}(Y_{t-k}, Y_{t-k})}} \\
 &= \frac{\xi \varphi \sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\xi \varphi \sum_{i=0}^{\infty} \psi_i^2} \\
 &= \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}.
 \end{aligned}$$



## Apêndice D

### Matriz de informação condicional esperada de Fisher para modelos SYMARMA( $p, q$ )

Apresentamos nesse apêndice os elementos que compõem a matriz informação condicional esperada de Fisher,  $\mathbf{K}$ , isto é, as expressões resultantes de

$$\mathbf{K}_{\omega_r \omega_s} = -\mathbb{E} \left[ \frac{\partial^2 \ell(\delta, \varphi)}{\partial \omega_r \partial \omega_s} | \mathbf{H}_{t-1} \right] = \mathbb{E} \left[ \frac{\partial \ell(\delta, \varphi)}{\partial \omega_r} \frac{\partial \ell(\delta, \varphi)}{\partial \omega_s} | \mathbf{H}_{t-1} \right],$$

em que  $\omega_r$  e  $\omega_s$  representam parâmetros do modelo e  $\ell$  é o logaritmo da função de verossimilhança condicional expresso por

$$\ell(\delta, \varphi) = \sum_{t=m+1}^n \log f(y_t | \mathbf{H}_{t-1}),$$

sendo  $m = \max\{p, q\}$  em que  $p$  e  $q$  são as ordens dos parâmetros autoregressivos e de médias móveis, respectivamente. Dessa forma,

$$\ell(\delta, \varphi) = -\frac{(n-m)}{2} \log(\varphi) + \sum_{t=m+1}^n \log(g(u_t)), \quad (21)$$

em que  $u_t = (y_t - \mu_t)^2 / \varphi$  e

$$\mu_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i \left\{ y_{t-i} - \mathbf{x}_{t-i}^\top \boldsymbol{\beta} \right\} + \sum_{j=1}^q \theta_j (y_{t-j} - \mu_{t-j}).$$

## D.1 Resultados gerais

Sob condições de regularidade usuais

$$\begin{aligned}
 E\left(\frac{\partial \ell_t(\delta, \varphi)}{\partial \mu_t} \mid \mathbf{H}_{t-1}\right) &= E\left(\frac{\partial \log f(y_t \mid \mathbf{H}_{t-1})}{\partial \mu_t}\right) = \int_{-\infty}^{\infty} \frac{\partial \log f(y_t \mid \mathbf{H}_{t-1})}{\partial \mu_t} f(y_t \mid \mathbf{H}_{t-1}) d\mu_t \\
 &= \int_{-\infty}^{\infty} \left(\frac{1}{f(y_t \mid \mathbf{H}_{t-1})} \frac{\partial f(y_t \mid \mathbf{H}_{t-1})}{\partial \mu_t}\right) f(y_t \mid \mathbf{H}_{t-1}) d\mu_t \\
 &= \int_{-\infty}^{\infty} \frac{\partial f(y_t \mid \mathbf{H}_{t-1})}{\partial \mu_t} d\mu_t = \frac{\partial}{\partial \mu_t} \int_{-\infty}^{\infty} f(y_t \mid \mathbf{H}_{t-1}) d\mu_t = 0.
 \end{aligned} \tag{22}$$

Após de algumas manipulações algébricas, temos que

$$\frac{\partial \ell_t(\delta, \varphi)}{\partial \mu_t} = -\frac{2}{\sqrt{\varphi}} W_g(u_t) z_t,$$

em que  $W_g(u_t) = \partial \log(g(u_t)) / \partial u_t$  e  $z_t = \sqrt{u_t} = (y_t - \mu_t) / \sqrt{\varphi}$ . Portanto, utilizando o resultado em (22), temos que

$$E(W_g(u_t) z_t \mid \mathbf{H}_{t-1}) = 0. \tag{23}$$

Temos ainda que as expressões

$$\frac{\partial \mu_t}{\partial \beta_l} = x_{tl} - \sum_{i=1}^p \phi_i x_{(t-i)l}, \quad \frac{\partial \mu_t}{\partial \phi_i} = y_{t-i} - \mathbf{x}_{t-i}^\top \beta \quad \text{e} \quad \frac{\partial \mu_t}{\partial \theta_j} = y_{t-j} - \mu_{t-j} \tag{24}$$

são  $\mathbf{H}_{t-1}$ -mensuráveis.

## D.2 Elementos da matriz $\mathbf{K}_{\delta\delta}$

Temos que

$$\begin{aligned}
 E\left(\frac{\partial \ell_t(\delta, \varphi)}{\partial \delta_i} \frac{\partial \ell_t(\delta, \varphi)}{\partial \delta_j} \mid \mathbf{H}_{t-1}\right) &= E\left[\left(\frac{-2W_g(u_t)}{\sqrt{\varphi}} \frac{\partial \mu_t}{\partial \delta_i} z_t\right) \left(\frac{-2W_g(u_t)}{\sqrt{\varphi}} \frac{\partial \mu_t}{\partial \delta_j} z_t\right) \mid \mathbf{H}_{t-1}\right] \\
 &= \frac{4}{\varphi} E\left[W_g^2(u_t) z_t^2 \frac{\partial \mu_t}{\partial \delta_i} \frac{\partial \mu_t}{\partial \delta_j} \mid \mathbf{H}_{t-1}\right] \\
 &= \frac{4}{\varphi} E\left[W_g^2(u_t) z_t^2 \mid \mathbf{H}_{t-1}\right] \frac{\partial \mu_t}{\partial \delta_i} \frac{\partial \mu_t}{\partial \delta_j} \\
 &= \frac{4}{\varphi} d_g \frac{\partial \mu_t}{\partial \delta_i} \frac{\partial \mu_t}{\partial \delta_j},
 \end{aligned}$$

em que  $d_g = E[W_g^2(u_t) z_t^2 \mid \mathbf{H}_{t-1}]$ , ou seja,  $d_g = E[W_g^2(U^2) U^2 \mid \mathbf{H}_{t-1}]$ , sendo  $U \sim S(0, 1, g)$ .

Utilizando os resultados em (24) podemos facilmente encontrar as expressões para os elementos de  $\mathbf{K}_{\delta\delta}$ .

### D.3 Elementos da matriz $\mathbf{K}_{\varphi\varphi}$

Da equação (21) obtemos que

$$\begin{aligned}
 E\left(\frac{\partial \ell_t(\delta, \varphi)}{\partial \varphi} \frac{\partial \ell_t(\delta, \varphi)}{\partial \varphi} | \mathbf{H}_{t-1}\right) &= E\left[\left(-\frac{1}{2\varphi} - \frac{W_g(u_t)}{\varphi} u_t\right) \left(-\frac{1}{2\varphi} - \frac{W_g(u_t)}{\varphi} u_t\right) | \mathbf{H}_{t-1}\right] \\
 &= E\left[\frac{1}{4\varphi^2} + \frac{W_g(u_t)u_t}{\varphi^2} + \frac{W_g^2(u_t)u_t^2}{\varphi^2} | \mathbf{H}_{t-1}\right] \\
 &= \frac{1}{4\varphi^2} + \frac{1}{\varphi^2} E[W_g(u_t)u_t | \mathbf{H}_{t-1}] + \frac{1}{\varphi^2} E[W_g^2(u_t)u_t^2 | \mathbf{H}_{t-1}] \\
 &= \frac{1}{4\varphi^2} + \frac{1}{\varphi^2} \left(-\frac{1}{2}\right) + \frac{1}{\varphi^2} f_g \\
 &= \frac{1}{\varphi^2} f_g - \frac{1}{4\varphi^2} = \frac{1}{4\varphi^2} (4f_g - 1),
 \end{aligned}$$

em que  $f_g = E[W_g^2(u_t)u_t^2 | \mathbf{H}_{t-1}]$ , ou seja,  $f_g = E[W_g^2(U^2)U^4 | \mathbf{H}_{t-1}]$ , sendo  $U \sim S(0, 1, g)$ . O resultado  $E[W_g(u_t)u_t | \mathbf{H}_{t-1}] = -1/2$  foi retirado de Fang *et al.* (1990, p. 94).

Utilizando o resultado acima obtemos que

$$\mathbf{K}_{\varphi\varphi} = \sum_{t=m+1}^n \frac{1}{4\varphi^2} (4f_g - 1) = \frac{(n-m)}{4\varphi^2} (4f_g - 1).$$

### D.4 Elementos da matriz $\mathbf{K}_{\delta\varphi}$

Temos que

$$\begin{aligned}
 E\left(\frac{\partial \ell_t(\delta, \varphi)}{\partial \delta_i} \frac{\partial \ell_t(\delta, \varphi)}{\partial \varphi} | \mathbf{H}_{t-1}\right) &= E\left[\left(\frac{-2W_g(u_t)}{\sqrt{\varphi}} \frac{\partial \mu_t}{\partial \delta_i} z_t\right) \left(-\frac{1}{2\varphi} - \frac{W_g(u_t)}{\varphi} u_t\right) | \mathbf{H}_{t-1}\right] \\
 &= E\left[\frac{W_g(u_t)}{\varphi\sqrt{\varphi}} z_t | \mathbf{H}_{t-1}\right] \frac{\partial \mu_t}{\partial \delta_i} + E\left[\frac{2W_g^2(u_t)}{\varphi\sqrt{\varphi}} z_t u_t | \mathbf{H}_{t-1}\right] \frac{\partial \mu_t}{\partial \delta_i} \\
 &= \frac{1}{\varphi\sqrt{\varphi}} \{E[W_g(u_t)z_t | \mathbf{H}_{t-1}] + 2E[W_g^2(u_t)z_t u_t | \mathbf{H}_{t-1}]\} \frac{\partial \mu_t}{\partial \delta_i} \\
 &= 0.
 \end{aligned}$$

O resultado  $E[W_g^2(u_t)z_t u_t | \mathbf{H}_{t-1}] = 0$  foi retirado de Fang *et al.* (1990, p. 94) e de (23), temos que  $E[W_g(u_t)z_t | \mathbf{H}_{t-1}] = 0$ .

## Apêndice E

### Matriz de informação condicional observada de Fisher para modelos SYMARMA( $p, q$ )

Esse apêndice contém as expressões que compõem a matriz de derivadas parciais de segunda ordem do logaritmo da função de verossimilhança condicional do modelo SYMARMA, dado por

$$\ell(\delta, \varphi) = -\frac{(n-m)}{2} \log(\varphi) + \sum_{t=m+1}^n \log(g(u_t)), \quad (25)$$

em que  $u_t = (y_t - \mu_t)^2 / \varphi$  e

$$\mu_t = \mathbf{x}_t^\top \beta + \sum_{i=1}^p \phi_i \left\{ y_{t-i} - \mathbf{x}_{t-i}^\top \beta \right\} + \sum_{j=1}^q \theta_j (y_{t-j} - \mu_{t-j}).$$

#### E.1 Resultados gerais

As derivadas de primeira e segunda ordem do logaritmo da função de verossimilhança condicional  $\ell_t(\delta, \varphi)$  com relação a  $\mu_t$  são expressas, respectivamente, por

$$\frac{\partial \ell_t(\delta, \varphi)}{\partial \mu_t} = -\frac{2}{\varphi} W_g(u_t)(y_t - \mu_t) \quad \text{e} \quad \frac{\partial^2 \ell_t(\delta, \varphi)}{\partial \mu_t^2} = \frac{2}{\varphi} (W_g(u_t) + 2W'_g(u_t)u_t),$$

em que  $W_g(u_t) = \partial \log(g(u_t)) / \partial u_t$  e  $W'_g(u_t) = \partial W_g(u_t) / \partial u_t$ .

Após algumas manipulações algébricas temos que

$$\begin{aligned} \frac{\partial \mu_t}{\partial \beta_l} &= x_{tl} - \sum_{i=1}^p \phi_i x_{(t-i)l}, & \frac{\partial \mu_t}{\partial \phi_i} &= y_{t-i} - \mathbf{x}'_{t-i} \beta, & \frac{\partial \mu_t}{\partial \theta_j} &= y_{t-j} - \mu_{t-j}, \\ \frac{\partial^2 \mu_t}{\partial \beta_l \partial \phi_i} &= -x_{(t-i)l}, & \frac{\partial^2 \mu_t}{\partial \phi_i \partial \theta_j} &= 0 & \text{e} & \frac{\partial^2 \mu_t}{\partial \beta_l \partial \theta_j} = 0. \end{aligned}$$

## E.2 Derivadas de segunda ordem entre elementos do vetor $\delta$

Da equação (25) obtemos que

$$\begin{aligned}
 \frac{\partial^2 \ell(\delta, \varphi)}{\partial \delta_i \partial \delta_j} &= \sum_{t=m+1}^n \frac{\partial}{\partial \delta_j} \left( \frac{\partial \ell_t(\delta, \varphi)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \delta_i} \right) \\
 &= \sum_{t=m+1}^n \frac{\partial^2 \ell_t(\delta, \varphi)}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \delta_j} \frac{\partial \mu_t}{\partial \delta_i} + \frac{\partial \ell_t(\delta, \varphi)}{\partial \mu_t} \frac{\partial^2 \mu_t}{\partial \delta_i \partial \delta_j} \\
 &= \sum_{t=m+1}^n \frac{2}{\varphi} (W_g(u_t) + 2W'_g(u_t)u_t) \frac{\partial \mu_t}{\partial \delta_j} \frac{\partial \mu_t}{\partial \delta_i} - \frac{2}{\varphi} W_g(u_t)(y_t - \mu_t) \frac{\partial^2 \mu_t}{\partial \delta_i \partial \delta_j} \\
 &= \sum_{t=m+1}^n -\frac{1}{\varphi} \left\{ 2s_t \frac{\partial^2 \mu_t}{\partial \delta_i \partial \delta_j} + a_t \frac{\partial \mu_t}{\partial \delta_j} \frac{\partial \mu_t}{\partial \delta_i} \right\},
 \end{aligned}$$

em que  $s_t = W_g(u_t)(y_t - \mu_t)$  e  $a_t = -2 \{W_g(u_t) + 2W'_g(u_t)u_t\}$ .

Utilizando os resultados da Seção E.1 podemos facilmente encontrar as expressões para as derivadas de segunda ordem entre os elementos do vetor  $\delta$ .

## E.3 Derivadas de segunda ordem entre elementos do vetor $\delta$ e $\varphi$

Nesse caso, temos que

$$\begin{aligned}
 \frac{\partial^2 \ell(\delta, \varphi)}{\partial \delta_i \partial \varphi} &= \sum_{t=m+1}^n \frac{\partial}{\partial \delta_j} \left( -\frac{1}{2\varphi} - \frac{W_g(u_t)}{\varphi} u_t \right) \\
 &= \sum_{t=m+1}^n \frac{1}{\varphi} \left[ 2W'_g(u_t) \frac{(y_t - \mu_t)}{\varphi} \frac{\partial \mu_t}{\partial \delta_i} u_t + 2W_g(u_t) \frac{(y_t - \mu_t)}{\varphi} \frac{\partial \mu_t}{\partial \delta_i} \right] \\
 &= \sum_{t=m+1}^n \frac{2}{\varphi^2} [W'_g(u_t)u_t + W_g(u_t)] (y_t - \mu_t) \frac{\partial \mu_t}{\partial \delta_i} \\
 &= \sum_{t=m+1}^n \frac{2}{\varphi^2} b_t \frac{\partial \mu_t}{\partial \delta_i},
 \end{aligned}$$

em que  $b_t = \{W_g(u_t) + W'_g(u_t)u_t\} (y_t - \mu_t)$ .

Dessa forma, com auxílio dos resultados apresentados na Seção E.1, podemos encontrar as expressões para as derivadas de segunda ordem entre o vetor de parâmetros  $\delta$  e  $\varphi$ .

## E.4 Derivadas de segunda ordem do parâmetro $\varphi$

Temos que

$$\begin{aligned}
 \frac{\partial^2 \ell(\delta, \varphi)}{\partial \varphi^2} &= \sum_{t=m+1}^n \frac{\partial}{\partial \varphi} \left( -\frac{1}{2\varphi} - \frac{W_g(u_t)}{\varphi^2} (y_t - \mu_t)^2 \right) \\
 &= \sum_{t=m+1}^n \frac{1}{2\varphi^2} + \frac{2W_g(u_t)}{\varphi^3} (y_t - \mu_t)^2 + \frac{1}{\varphi^2} W'_g(u_t) \frac{(y_t - \mu_t)^2}{\varphi^2} (y_t - \mu_t)^2 \\
 &= \sum_{t=m+1}^n \frac{1}{\varphi^2} \left[ \frac{1}{2} + 2W_g(u_t)u_t + W'_g(u_t)u_t^2 \right].
 \end{aligned}$$