UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO
DEPARTAMENTO DE ESTATÍSTICA

Influência Local em Modelos de Séries Temporais

Bruno Reis dos Santos

Orientador: Prof. Dr. Mauricio Enrique Zevallos Herencia

Dissertação apresentada junto ao Departamento de Estatística do Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas, para obtenção do Título de MESTRE em Estatística.

Campinas - SP
2008

Influência Local em Modelos de Séries Temporais

Este exemplar correponde à redação final da dissertação devidamente corrigida e defendida por Bruno Reis dos Santos e aprovada pela comissão julgadora.

Campinas, 25 de abril de 2008.

Prof. Dr. Mauricio Enrique Zevallos Herencia
Orientador

Banca Examinadora:

1 Prof. Dr. Mauricio Enrique Zevallos Herencia (orientador) - IMECC-UNICAMP.

2 Profa. Dra. Glaura da Conceição Franco - ICEx-UFMG.

3 Prof. Dr. Luiz Koodi Hotta - IMECC-UNICAMP.

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requesito parcial para obtenção do Título de MESTRE em Estatística.

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Maria Júlia Milani Rodrigues

	Santos, Bruno Reis dos
Sa59i	Influência local em modelos de séries temporais / Bruno Reis dos Santos -- Campinas, [S.P. :s.n.], 2008.
	Orientador : Maurício Enrique Zevallos Herencia
	Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.
	1. Influência local. 2. Séries temporais. 3. Longa-memória. I. Zevallos Herencia, Maurício Enrique. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Título em inglês: Local influence in time series models.

Palavras-chave em inglês (Keywords): 1. Local influence. 2. Time series. 3. Long memory.

Área de concentração: Séries Temporais

Titulação: Mestre em Estatística

Banca examinadora:

Prof. Dr. Maurício Enrique Zevallos Herencia (IMECC-UNICAMP)
Profa. Dra. Glaura da Conceição Franco (ICEx-UFMG)
Prof. Dr. Luíz Koodi Hotta (IMECC-UNICAMP)

Data da defesa: 25/04/2008

Programa de pós-graduação: Mestrado em Estatística

Dissertação de Mestrado defendida em 25 de abril de 2008 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Zevallos

Prof(a). Dr(a). MAURICIO ENRIQUE ZEVALLOS HERENCIA

Glaura Franco

Prof(a). Dr(a). GLAURA DA CONCEIÇÃO FRANCO

Prof(a). Dr(a). LUIZ KOODI HOTTA

Dedico este trabalho à Fabiana,
à minha mãe Shirlenalda,
ao meu pai Inácio
e aos meus irmãos Luane e Fabiano.

AGRADECIMENTOS

Em primeiro lugar, eu gostaria de agradecer a Deus e a meu mentor por mais uma missão cumprida.

Ao professor Maurício Zevallos, pela orientação, paciência, dedicação e atenção dada até o último momento de nosso trabalho.

Aos amigos que fiz durante todo o percurso do mestrado. Especialmente a: Marley e Patrícia, David, Beth, Daniela, Rodrigo e Mariana, Alejandro, Leme, Márcio, Tsai, Juliana e Isabela. Não menos importantes, a todos os integrantes do laboratório Epifismia.

À minha família, guia de minha jornada e responsáveis por quem sou hoje. Não poderia deixar de citar Fabiana, parte de minha família e companheira de todos os momentos.

Agradeço a Capes pelo apoio financeiro a este projeto.

RESUMO

Nesta dissertação é discutido o uso da metodologia de diagnóstico de Influência Local em modelos de séries temporais. Especificamente, serão estudados os modelos autoregressivos de ordem um, os modelos de regressão com erros autoregressivos de ordem um e modelos de longa-memória. As medidas de influência local consideradas são: Inclinação de Billor e Loynes e Curvatura de Cook. As principais contribuições nesta dissertação são duas. Primeiro, a utilização da metodologia de limiares (*benchmarks*) nos modelos mencionados para determinar se uma observação é influente. Isto permite ter uma ferramenta estatística para identificar pontos influentes a diferença da simples análise exploratória que é o mais comum na literatura. Como segunda contribuição, serão obtidas as expressões para o cálculo das medidas de Inclinação de Billor e Curvatura de Cook nos modelos ARFIMA. Finalmente, as metodologias descritas são ilustradas através de dados simulados e da análise de dados reais.

ABSTRACT

This work is about Time Series Diagnostics using Local Influence. Specifically, first-order autoregressive models, regression models with first-order autoregressive errors and long-memory models are studied. In order to assess Local Influence two statistics are considered: the Slope of Billor and Loynes and Cook's Curvature. The main contributions are two. First, apply a methodology based on benchmarks calculated by simulation on the aforementioned models for determining influential observations. This permits to have a statistical tool to identify influential points instead of the simple exploratory analysis, which is the most common device in the literature. Second, expressions for Billor and Loynes Slope and Cook's Curvature in ARFIMA models are derived. Finally, all methodologies are illustrated using simulated data and the analysis of real data.

Sumário

Lista de Figuras **xvii**

Lista de Tabelas **xxi**

1 Introdução **1**

1.1 Breve descrição da Dissertação 3

2 Medidas de Influência Local e Metodologia de Limiares **5**

2.1 Curvatura de Cook 5

2.2 Inclinação de Billor e Loynes 10

2.3 Metodologia de Diagnóstico para Influência Local: Limiares 13

2.3.1 Limiares - Inclinação de Billor e Loynes 14

2.3.2 Limiares - Curvatura de Cook 16

2.4 Aplicação 17

2.4.1 Inclinação de Billor e Loynes 17

2.4.2 Curvatura de Cook 21

2.4.3 Deleção de uma Observação 23

2.4.4 Conclusões 24

3 Influência Local em Modelos de Regressão Linear com Erros Autoregressivos de Primeira Ordem **27**

3.1 Introdução 27

3.2 Influência Local com a Verossimilhança Exata 29

3.3 Influência Local com a Verossimilhança Perfilada 31

3.4 Influência Local com a Verossimilhança Condicional 32

3.5 Simulações 37

3.5.1 Verossimilhança Exata 37

3.5.2 Conclusões das Simulações 53

3.6 Aplicações 53

3.6.1 Vendas *versus* Despesas 54

3.6.2 Índice de Preços ao Consumidor da Argentina 64

3.7 Modelos de Regressão com Erros AR(1) e Inovações t-Student 70

3.7.1 Verossimilhança Condicional 71

3.7.2 Influência Local 72

3.7.3 Simulações 75

3.7.4 Conclusões das Simulações 91

3.7.5 Aplicação 92

4 Influência Local em Modelos ARFIMA **101**

4.1 Introdução 101

4.2 Processos de Longa Memória 101

4.2.1 Definição do Modelo ARFIMA 102

4.2.2 Aproximação do Modelo ARFIMA(0,d,0) por ARMA(p,q) 104

4.3 Análise de Influência Local 105

4.3.1 Verossimilhança Aproximada 106

4.4 Simulações 109

4.5 Aplicação 127

5 Conclusões e Considerações Finais **133**

Referências **135**

Lista de Figuras

2.1.1 Esquema de uma superfície $\alpha(\boldsymbol{\omega})$ no $\Re^3$. 7

2.2.1 Esquema de uma superfície $\alpha^*(\boldsymbol{\omega})$ no $\Re^3$. 12

2.4.1 Dados de gansos, resíduos em valores absolutos e vetor de inclinação $\mathbf{v}$. . . 19

2.4.2 Limiares para influência local da inclinação. 21

2.4.3 Resíduos em valor absoluto e vetor de curvatura $\mathbf{v}_c$. 22

2.4.4 Limiares para influência local da Curvatura de Cook. 23

2.4.5 Resumo dos limiares para as medidas de influência. 26

3.5.1 Influência local na série 1 ($\rho = 0,5$ e $n = 250$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 42

3.5.2 Influência local na série 1 ($\rho = 0,5$ e $n = 250$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 43

3.5.3 Influência local na série 4 ($\rho = 0,9$ e $n = 250$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 44

3.5.4 Influência local na série 4 ($\rho = 0,9$ e $n = 250$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 45

3.6.1 Gráfico de dispersão, resíduos, autocorrelações e autocorrelações parciais dos resíduos. 55

3.6.2 Influência local com verossimilhança exata e erros AR(1). 59

3.6.3 Influência local com verossimilhança perfilada e erros AR(1). 60

3.6.4 Influência local com verossimilhança condicional e erros AR(1). 61

3.6.5 Influência local com verossimilhança exata e erros $NID(0, \sigma^2)$. 62

3.6.6 Influência local com verossimilhança condicional e erros $NID(0, \sigma^2)$. 63

3.6.7 Retornos de IPCA, autocorrelações e autocorrelações parciais. 66

3.6.8 Influência local com verossimilhança exata. 68

3.7.1 Influência local na série 2 ($\nu = 5$ e $\rho = 0,5$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 80

3.7.2 Influência local na série 2 ($\nu = 5$ e $\rho = 0,5$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 81

3.7.3 Influência local na série 7 ($\nu = 7$ e $\rho = 0,7$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 82

3.7.4 Influência local na série 7 ($\nu = 7$ e $\rho = 0,7$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 83

3.7.5 Influência local na série 12 ($\nu = 12$ e $\rho = 0,9$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 84

3.7.6 Influência local na série 12 ($\nu = 12$ e $\rho = 0,9$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 85

3.7.7 Gráfico de dispersão e autocorrelações e autocorrelações parciais dos resíduos de *CAPM*. 94

3.7.8 Influência local considerando as inovações $\epsilon_t \sim AR(1)$ normais. 98

3.7.9 Influência local considerando as inovações $\epsilon_t \sim AR(1)$ t-Student. 99

4.4.1 Influência local na série 1 ($d = 0,2$ e $n = 250$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 116

4.4.2 Influência local na série 1 ($d = 0,2$ e $n = 250$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 117

4.4.3 Influência local na série 2 ($d = 0,4$ e $n = 250$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 118

4.4.4 Influência local na série 2 ($d = 0,4$ e $n = 250$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 119

4.4.5 Influência local na série 5 ($d = 0,4$ e $n = 500$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 120

4.4.6 Influência local na série 5 ($d = 0,4$ e $n = 500$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 121

4.4.7 Influência local na série 2a ($d = 0,4$ e $n = 250$) com *outlier* positivo em y_{245}. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*. 122

4.4.8 Influência local na série 2a ($d = 0,4$ e $n = 250$) com *outlier* negativo em y_{245}. 123

4.5.1 Gráfico logaritmo da série *varve* e as respectivas autocorrelações. 129

4.5.2 Influência local para $\mathbf{v}$ e $\mathbf{v}_c$. 131

Lista de Tabelas

2.4.1 Parâmetros estimados. 18

2.4.2 Estimativas sem a observação (x_{29}, y_{29}). 23

2.4.3 Estatísticas e limiares para influência sem a observação (x_{29}, y_{29}). 24

2.4.4 Resumo das estatísticas e limiares para influência. 25

3.5.1 Séries Simuladas. 37

3.5.2 Perturbação positiva e negativa no modelo AR(1) - Normal. 38

3.5.3 Estatísticas Globais M_0. 38

3.5.4 Estatísticas Individuais M_1. 39

3.5.5 Estatísticas Individuais M_2. 39

3.5.6 Influência global para S e C_c. 40

3.5.7 Estimativas dos parâmetros ρ e σ^2. 47

3.5.8 Estatísticas Globais M_0. 49

3.5.9 Estatísticas Individuais M_1. 50

3.5.1Œstatísticas Individuais M_2. 51

3.6.1 Resumo do diagnóstico de influência local para $\epsilon_t \sim AR(1)$. 57

3.6.2 Resumo do diagnóstico de influência local para $\epsilon_t \sim NID(0, \sigma^2)$. 57

3.6.3 Resumo do diagnóstico da influência local. 67

3.6.4 Resumo do diagnóstico de influência local. 70

3.7.1 Séries Simuladas. 76

3.7.2 Perturbação positiva e negativa no modelo AR(1) - t-Student. 77

3.7.3 Estatísticas Globais M_0. 77

3.7.4 Estatísticas Individuais M_1. 78

3.7.5 Estatísticas Individuais M_2. 78

3.7.6 Influência global para C_c. 78

3.7.7 Estimativas dos parâmetros ρ e σ^2. 87

3.7.8 Estatísticas Globais M_0. 88

3.7.9 Estatísticas Individuais M_1. 89

3.7.10 Estatísticas Individuais M_2. 90

3.7.11 Parâmetros estimados. 95

3.7.12 Resumo do diagnóstico da influência local. 96

4.4.1 Séries Simuladas. 110

4.4.2 Perturbação Positiva e Negativa no modelo ARFIMA(0,d,0). 111

4.4.3 Estatísticas Globais M_0. 112

4.4.4 Estatísticas Individuais M_1. 112

4.4.5 Estatísticas Individuais M_2. 113

4.4.6 Estimativas de S e C_c para as séries da Tabela 4.4.1. 113

4.4.7 Estimativas dos parâmetros d e σ^2. 126

4.4.8 Estatísticas Globais M_0. 126

4.5.1 Resumo do diagnóstico da influência local. 130

1 *Introdução*

O diagnóstico de modelo é uma das etapas fundamentais no paradigma de análise estatística. Através do diagnóstico podemos avaliar a qualidade do ajuste do modelo para um conjunto de dados. Uma das ferramentas de diagnóstico mais úteis é a análise de Influência Local.

No diagnóstico de influência, destacam-se basicamente duas abordagens. A primeira se baseia na deleção de observações através por exemplo, da distância de Cook [ver Cook e Weisberg, 1982], ou ainda, na medida DFFITS proposta por Belsley, Kuh e Welsh (1980). A segunda metodologia de diagnóstico de influência, chamada de Influência Local, foi proposta por Cook (1986), baseada na medida afastamento da verossimilhança, e conseqüentemente, na avaliação da Curvatura de Cook. Na literatura, através da Curvatura de Cook, pode-se citar: Lesaffre e Verbeke (1998) que utilizaram o enfoque de influência local para avaliar modelos lineares mistos; Liu (2002) onde o método de influência local é aplicado para modelos de regressão linear elípticos multivariados considerando perturbações de casos ponderados na variável explanatória e na variável resposta; Dávila (2000), Paula (2004) e Salgado (2006) que utilizaram a Curvatura de Cook como ferramenta de diagnóstico para os modelos de calibração comparativa, modelos lineares generalizados e modelos elípticos mistos, respectivamente; ou ainda, Galea et al. (2008) que estudou a Curvatura de Cook como diagnóstico de influência para modelos CAPM (*Capital Asset Pricing Model*) sob a suposição de distribuições elípticas

Outras medidas de influência local alternativas foram propostas. Assim, Billor e Loynes (1993) introduzem a máxima inclinação como uma medida de influência local baseada no afastamento da verossimilhança modificado para avaliar modelos de regressão linear, considerando o esquema de perturbação na matriz de variâncias. Baseados nos estudos da Curvatura de Cook (1986) e na Inclinação de Billor e Loynes (1993), Zhang e King (2005) propõem uma curvatura como medida de influência local para avaliar modelos Autoregressivos Generalizados de Heteroscedasticidade Condicional ($GARCH$) em diversos esquemas de perturbação.

O fato da metodologia de influência local não ser baseada na exclusão de observações do específico conjunto de dados a torna uma ferramenta muito útil quando aplicada no contexto de séries temporais. Ou seja, partindo do princípio de que as observações de séries temporais são dependentes, não seria conveniente avaliar pontos influentes a partir de técnicas como distância de Cook ou DFFITS. Atualmente, a literatura é escassa em estudos na área de séries temporais que utilizam a metodologia de influência local como ferramenta de diagnóstico para detectar observações influentes. Alguns dos poucos trabalhos são: Kim e Huggins (1997) que apresentaram o diagnóstico de influência local para o modelo de regressão linear com erros autoregressivos de ordem um, considerando a medida Curvatura de Cook; Tsai e Wu (1992) que investigaram a influência local sobre o modelo de regressão linear com erros autoregressivos de primeira ordem através da verossimilhança perfilada (somente a correlação é parâmetro) usando inclinação como medida de influência; Schall e Dunne (1991) que apresentaram um estudo para investigar a influência local em modelos de regressão com erros ARMA (*Autoregressivos de Média-Móvel*), utilizando diversos esquemas de perturbação; Zhang e King (2005) que propuseram uma medida de curvatura para o diagnóstico de influência local em modelos GARCH; Medeiros (2006) que discutiu a influência local, através da Curvatura de Cook, como método de diagnóstico para modelos autoregressivos simétricos. Recentemente, Zevallos e Hotta (2008) estudam a influência local em modelos de séries temporais com heteroscedasticidade condicional.

Por outro lado, várias séries temporais reais apresentam decaimento hiperbólico para as autocorrelações amostrais, ou seja, a correlação é significativa entre observações bastante distantes. Essa característica, notada por Hurst (1951, 1957), Mandelbrot e Wallis (1968) e McLeod e Hipel (1978), é reproduzida pelos chamados processos de longa memória, ou ainda, diferença fracionária [Hosking, 1981]. A literatura praticamente não apresenta estudos de influência para modelos de longa memória. Os únicos trabalhos do qual temos referência são: Palma et al. (2008), que estudou a influência de observações via distância de *Kullback-Leibler* para o modelo ARFIMA (*Autoregressivo Fracionário Integrado de Média-Móvel*), onde os critérios são influência na estimação e na previsão, e Haldrup e Nielsen (2007) que estudaram a estimação da integração fracionária na presença de perturbações nos dados.

Na literatura, a abordagem predominante na análise de influência local para conjuntos de dados se baseia na simples análise exploratória. Poucos trabalhos fornecem marcas de referência para determinar se, estatisticamente as observações são influentes. Dentre essas exceções, temos Poon e Poon (1999) com modelos de regressão linear e Zhu e Zhang (2004) com o modelo linear generalizado. Adicionalmente, Zhang e King (2005) propuseram a construção de marcas de referência, via simulações de Monte Carlo, para a Inclinação de Billor e Loynes e Curvatura de Cook em modelos GARCH.

Levando-se em consideração, que no trabalho aplicado são necessários critérios estatísticos para caracterizar as observações como sendo influentes, esta dissertação tem dois objetivos. O primeiro é discutir a aplicação da metodologia de marcas de referência (limiares) proposta por Zhang e King (2005) na análise de Influência Local dos modelos autoregressivos de ordem um, com e sem variáveis explanatórias, através da Inclinação de Billor e Loynes (1993) e da Curvatura de Cook (1986). O segundo objetivo é apresentar uma metodologia para realizar a análise de Influência Local em modelos com longa memória através da Inclinação de Billor e Loynes (1993) e da Curvatura de Cook (1986), e discutir a abordagem de Zhang e King (2005) no cálculo das marcas de referência. Para atingir estes objetivos serão discutidos exemplos com dados simulados e dados reais.

1.1 Breve descrição da Dissertação

No Capítulo 2 serão apresentadas as medidas de influência local, e em seguida, será descrita 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) e Curvatura de Cook (1986). A metodologia de limiares será descrita a partir de um modelo simples (de regressão), e sua conseqüente leitura, permitirá entender o que acontece em modelos mais complexos, tais como os estudados nos capítulos seguintes.

O Capítulo 3 trata da influência local para os modelos autoregressivos de primeira ordem e os modelos de regressão com erros autoregressivos de ordem um. Baseado no esquema de perturbação nos dados e nas verossimilhanças exata, perfilada e condicional, a metodologia de limiares descrita no Capítulo 2 será utilizada para construir os níveis de referência das medidas Inclinação de Billor e Loynes e Curvatura de Cook. Esta metodologia é ilustrada através de dados simulados e da análise de dados reais.

O Capítulo 4 visa obter as expressões para as estatísticas relacionadas às medidas de influência local da Inclinação de Billor e Loynes e da Curvatura de Cook, considerando o modelo de longa memória ARFIMA. Por fim, baseado no esquema de perturbação nos dados, a metodologia de limiares será ilustrada através de dados simulados e de dados reais.

O Capítulo 5 apresenta as contribuições, conclusões e futuras linhas de pesquisas para o estudo realizado nesta dissertação.

2 *Medidas de Influência Local e Metodologia de Limiares*

Na literatura, o primeiro trabalho com respeito a influência local foi introduzido por Cook (1986), onde propõe um método de diagnóstico a partir de uma medida chamada afastamento da verossimilhança, aproveitando assim, idéias elementares da geometria diferencial. Existem outras medidas de influência, como a distância de Cook [Cook, 1977] e DFFITS [Besley et al., 1980], que se baseiam na exclusão de observações do conjunto de dados e na posterior investigação das mudanças ocorridas nas estimativas dos parâmetros. Justamente nesse ponto, a proposta de Cook (1986) em relação a metodologia de influência local se faz interessante, pois visa contornar o problema da deleção de observações, principalmente quando se trata de séries temporais. As três medidas de influência local estudadas neste capítulo são: Curvatura de Cook [Cook, 1986] e Inclinação [Billor e Loynes, 1993]. Para tais medidas, será apresentada uma metodologia baseada em níveis de referência para avaliar a significância relacionada a influência local. Por fim, considerando o modelo de regressão linear e o esquema de perturbação na matriz de variâncias e covariâncias, será realizado o diagnóstico de influência local para um conjunto de dados reais estudado por outros autores na literatura.

2.1 Curvatura de Cook

Cook (1986) introduziu um método de diagnóstico denominado influência local. Seja $\mathbf{y} = (y_1, \ldots, y_n)^T$ o vetor de observações geradas por um modelo qualquer denominado modelo *proposto*, definindo-se $L(\boldsymbol{\theta}; \mathbf{y})$, ou simplesmente $L(\boldsymbol{\theta})$, como a log-verossimilhança, onde $\boldsymbol{\theta}_{(p\times 1)}$ é um vetor de parâmetros desconhecidos.

O modelo *proposto* é perturbado através de um vetor de perturbações $\boldsymbol{\omega}_{q\times 1}$, onde $q \leq n$. Como resultado, obtém-se o modelo *perturbado*. Por exemplo, a perturbação pode ser 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 ω_t.

Seja $\boldsymbol{\omega}_{q\times 1}$ o vetor de perturbações e $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ a log-verossimilhança do modelo *perturbado*. Assume-se que exista um $\boldsymbol{\omega}_0 \in \Re^q$ que represente a log-verossimilhança com perturbação nula, ou seja, $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$, para todo $\boldsymbol{\theta}$.

O principal objetivo da metodologia de influência local consiste em variar $\boldsymbol{\omega}$ em $\Re^q$ e avaliar o comportamento da medida denominada afastamento da verossimilhança, definida por

$$LD(\boldsymbol{\omega}) = 2\left\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\right\}, \tag{2.1.1}$$

onde $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ são os estimadores de máxima verossimilhança para $L(\boldsymbol{\theta})$ e $L(\boldsymbol{\theta}|\boldsymbol{\omega})$, respectivamente, assumindo que $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ é duas vezes continuamente diferenciável na vizinhança de $\boldsymbol{\omega}_0$. Portanto, torna-se interessante analisar o gráfico de $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ como uma superfície geométrica formada pelo vetor

$$\alpha(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ LD(\boldsymbol{\omega}) \end{pmatrix}, \tag{2.1.2}$$

de dimensões $((q+1)\times 1)$. Esta superfície $\alpha(\boldsymbol{\omega})$ é chamada de Gráfico de Influência. Assim, deseja-se utilizar $LD(\boldsymbol{\omega})$ através de $\alpha(\boldsymbol{\omega})$ para comparar as verossimilhanças *proposta* e *perturbada*.

Como ilustração, na Figura 2.1.1 é apresentada uma superfície $\alpha(\boldsymbol{\omega})$ no $\Re^3$. Seja a intersecção de $\alpha(\boldsymbol{\omega})$ com o plano contendo o vetor normal e um dos vetores tangentes a um ponto em particular, por exemplo $\boldsymbol{\omega}_0$. A intersecção é uma curva (seção normal) no plano, e essa tem uma curvatura. A curvatura pode ser vista como o inverso do raio de um círculo que se aproxima do ponto $\boldsymbol{\omega}_0$, ou ainda, como a taxa de mudança do ângulo que o vetor tangente faz com o eixo horizontal relacionado ao comprimento do arco ao longo da curva.

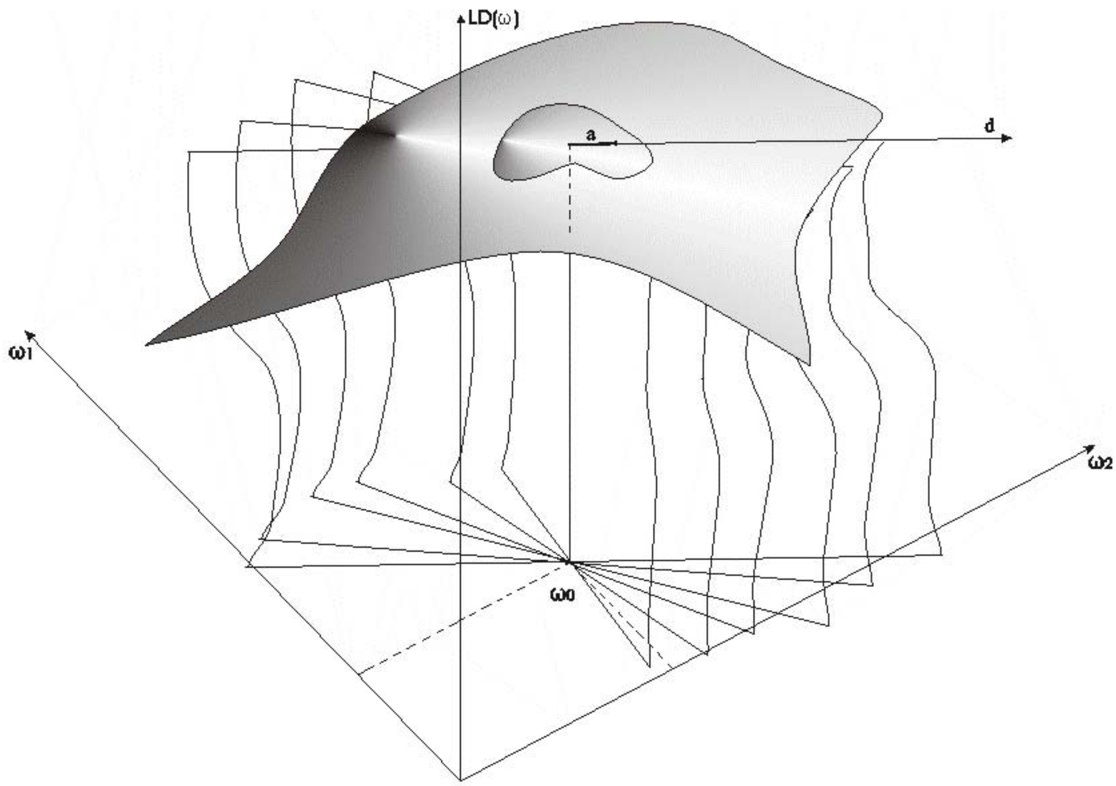

Figura 2.1.1: Esquema de uma superfície $\alpha(\boldsymbol{\omega})$ no $\Re^3$.

Especificamente, deseja-se descrever como a superfície $\alpha(\boldsymbol{\omega})$ desvia de seu plano tangente em $\boldsymbol{\omega}_0$. Essa descrição pode ser realizada utilizando as curvaturas das seções normais da superfície $\alpha(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$. As curvaturas das seções normais ilustradas pela Figura 2.1.1 são chamadas de curvaturas normais.

Considerando uma seção normal projetada perpendicularmente sobre o plano $\Re^q$, obtém-se um vetor $\mathbf{d}$ não-nulo de comprimento unitário em $\Re^q$, ou seja, $\|\mathbf{d}\| = 1$, tal que

$$\boldsymbol{\omega}(a) = \boldsymbol{\omega}_0 + a\mathbf{d}, \tag{2.1.3}$$

onde $a \in \Re$, $\mathbf{d}$ e a seção normal pertencem ao mesmo plano. Portanto, será equivalente considerar uma seção normal a $\alpha(\boldsymbol{\omega})$, em $\boldsymbol{\omega}_0$, a partir de um vetor $\mathbf{d} \in \Re^q$, indicando assim, a direção onde existe maior ou menor influência em torno de $\boldsymbol{\omega}_0$.

Suponha a classe de todos os vetores $\mathbf{d} \in \Re^q$ associados às seções normais da superfície $\alpha(\boldsymbol{\omega})$ (Figura 2.1.1) em $\boldsymbol{\omega}_0$. Se as magnitudes a associadas às curvaturas normais, estiverem "próximas" de zero, os vetores $\boldsymbol{\omega}(a)$ em (2.1.3) terão comprimento "próximo" à magnitude de $\boldsymbol{\omega}_0$. Equivale dizer que, considerando a classe de todas as seções normais da superfície $\alpha(\boldsymbol{\omega})$, $LD(\boldsymbol{\omega})$ está "próximo" de zero, indicando que existe estabilidade do modelo ajustado sob o particular esquema de perturbação que está sendo considerado.

Por outro lado, se uma magnitude a não estiver "próxima" de zero, então ocorrerá maior oscilação da função $LD(\boldsymbol{\omega})$, e portanto, existirá maior influência do esquema de perturbação sob esta seção normal.

Para avaliar a influência local Cook (1986) propõe o uso da curvatura normal, dada por

$$C = 2\left\|\mathbf{d}^T\ddot{\mathbf{F}}\mathbf{d}\right\|, \tag{2.1.4}$$

onde $\ddot{\mathbf{F}}$ é uma matriz $(q \times q)$ definida como

$$\ddot{\mathbf{F}} = \frac{\partial^2 L(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^T},$$

e por sua vez, pode ser expressada como

$$\ddot{\mathbf{F}} = \mathbf{J}^T \ddot{\mathbf{L}} \mathbf{J},$$

onde -$\ddot{\mathbf{L}}$ é a matriz de informação observada para o modelo postulado ($\boldsymbol{\omega} = \boldsymbol{\omega}_0$), $\mathbf{J}$ é uma matrix $(p \times q)$, tal que

$$\mathbf{J} = -(\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta},$$

e $\boldsymbol{\Delta}$ é uma matriz $(p \times q)$ com elementos

$$\Delta_{ij} = \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \theta_i \partial \omega_j},$$

avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, $i = 1, ..., p$, $j = 1, ..., q$. Assim, obtém-se

$$\ddot{\mathbf{F}} = \boldsymbol{\Delta}^T (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta}. \tag{2.1.5}$$

Por fim, através de (2.1.4), a representação da curvatura normal é dada por

$$C = 2 \left\| \mathbf{d}^T \boldsymbol{\Delta}^T (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta} \mathbf{d} \right\|. \tag{2.1.6}$$

A partir de agora, a máxima curvatura normal de Cook será representada por C_c. Portanto, é possível avaliar o comportamento da curvatura relativa a superfície $\alpha(\boldsymbol{\omega})$, através das estatísticas, dadas por

$$C_c \quad = \quad \text{maior autovalor de } 2\ddot{\mathbf{F}}, \tag{2.1.7}$$

$$\mathbf{v}_c \quad = \quad \text{autovetor associado ao maior autovalor de } 2\ddot{\mathbf{F}}. \tag{2.1.8}$$

Assim, a estatística C_c fica definida como medida de influência global e a estatística $\mathbf{v}_c$ como medida individual, onde a i-ésima componente do vetor indica a influência da i-ésima observação. Se a i-ésima componente de $\mathbf{v}_c$ for suficientemente "grande", diz-se que a i-ésima observação do específico conjunto de dados é localmente influente. Adicionalmente, considerando o modelo de regressão linear, Cook propõe relacionar a i-ésima componente de $\mathbf{v}_c$ com a i-ésima componente do vetor de resíduos. Ou seja, "grandes" componentes dos resíduos em conjunção aos "grandes" elementos de $\mathbf{v}_c$ dão sinais de que as referidas observações são influentes.

Considerando o modelo de regressão e o esquema de perturbação de casos ponderados, Cook (1986) propõe o valor 2 como referência para a curvatura (2.1.7). Ou seja, se C_c é suficientemente "grande" em relação à magnitude 2, existe indício de que as observações são globalmente influentes, isto é, existe pelo menos uma observação influente. Pode-se dizer ainda que a C_c é um útil indicador para a existência de observações localmente influentes.

Para avaliar a influência local através de C_c e $\mathbf{v}_c$, Cook (1986) utiliza a pertubação de casos ponderados e a perturbação aditiva na variável exploratória do modelo de regressão linear. Vários esquemas de perturbação podem ser introduzidos através de $\boldsymbol{\omega}$, e esses são divididos em dois grupos, conforme resumido por Billor e Loynes (1993):

- 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, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ é substituída por $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 diag^{-1}(\omega_i))$, onde ω_i são as perturbações especificadas e $diag^{-1}(\omega_i)$ é a inversa da matriz diagonal de componentes ω_i, para $i = 1, \ldots, n$.

- Perturbação nos Dados: Perturbar a variável resposta ou as variáveis explanatórias 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.

2.2 Inclinação de Billor e Loynes

Quatro dificuldades práticas e teóricas que surgem no enfoque de influência local introduzido por Cook (1986) foram apontadas por Billor e Loynes (1993): i) a escolha de uma referência para C_c, ii) a fórmula explícita para C_c é difícil de ser encontrada analiticamente, iii) a falta de invariância da curvatura sobre reparametrizações do esquema de perturbação e a iv) falta de definição dos parâmetros.

Para contornar esses problemas, Billor e Loynes (1993) propõem utilizar o afastamento da verossimilhança modificado

$$LD^*(\boldsymbol{\omega}) = -2\left\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}|\boldsymbol{\omega})\right\}, \tag{2.2.1}$$

onde $L(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}|\boldsymbol{\omega})$ é a verossimilhança do modelo *perturbado* avaliada sobre $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$, que é o estimador de máxima verossimilhança de $\boldsymbol{\theta}_{\boldsymbol{\omega}}$ quando perturbações são introduzidas no modelo *proposto*. Assim, é possível definir uma superfície $\alpha^*(\boldsymbol{\omega}) = (\boldsymbol{\omega}, LD^*(\boldsymbol{\omega}))$ para investigar o comportamento de $LD^*(\boldsymbol{\omega}_0 + a\mathbf{d}^*)$, ou seja, em torno do ponto de perturbação nula $\boldsymbol{\omega}_0$.

Em particular, Billor e Loynes (1993) obtêm através de $\alpha^*(\boldsymbol{\omega})$ a medida de influência local, aqui chamada de máxima Inclinação, dada por

$$S = \|\nabla LD^*(\boldsymbol{\omega})\| = \left\|\frac{\partial}{\partial\boldsymbol{\omega}} LD^*(\boldsymbol{\omega})\right\|.$$

Assim, avaliando S sobre $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, tem-se

$$S = 2\|\mathbf{v}\|, \quad \mathbf{v} = \frac{\partial}{\partial\boldsymbol{\omega}} L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0). \tag{2.2.2}$$

Para julgar se existem observações influentes, Billor e Loynes (1993) propõem as estatísticas

$$S \quad = \quad \text{norma da 1ª derivada do vetor } \frac{\partial}{\partial\boldsymbol{\omega}} L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0), \tag{2.2.3}$$

$$\mathbf{v} \quad = \quad \text{duas vezes as componentes do vetor } \frac{\partial}{\partial\boldsymbol{\omega}} L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0). \tag{2.2.4}$$

A estatística S fica definida como medida global e a estatística $\mathbf{v}$ como medida individual.

Considerando o modelo de regressão linear com esquema de perturbação inserido na matriz de variâncias e covariâncias, Billor e Loynes (1993) propõem uma referência específica para a inclinação, ou seja, se a medida global S for maior do que essa referência existe evidência de que pelo menos uma observação é influente. Propõem ainda, que a

i-ésima observação do conjunto de dados estudado será localmente influente, se a i-ésima componente da medida individual $\mathbf{v}$ é notadamente "maior" do que as demais. Ou seja, através da simples inspeção visual da i-ésima componente de $\mathbf{v}$, considera-se influente a i-ésima observação.

Note que calcular S envolve a primeira derivada, em comparação ao cálculo de C_c que envolve duas derivadas. Neste sentido, a medida global S é mais simples. Em Cook (1986), a primeira derivada do afastamento da verossimilhança (2.1.1) em relação a $\boldsymbol{\omega}$ é igual a zero, e não pode ser utilizada para examinar a influência local.

A Figura 2.2.1 apresenta uma ilustração de S para a superfície $\alpha^*(\boldsymbol{\omega})$ no $\Re^3$.

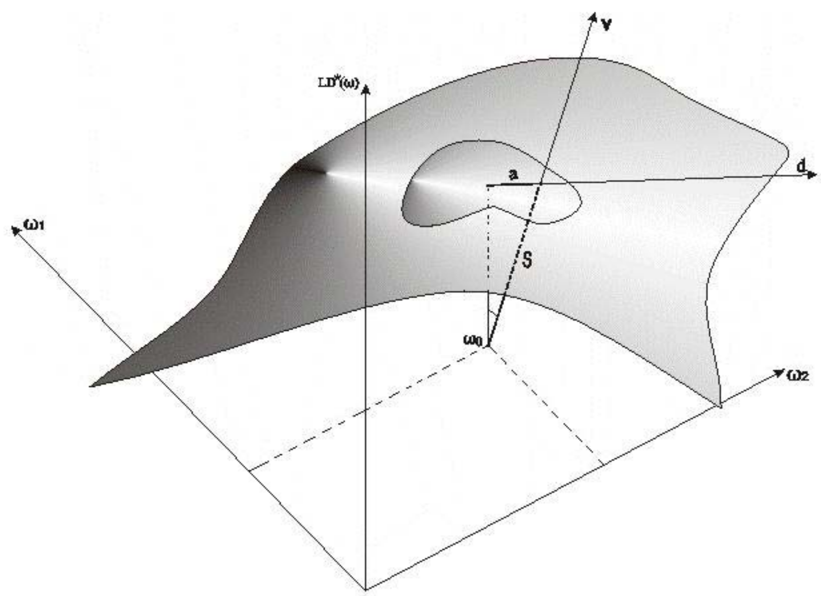

Figura 2.2.1: Esquema de uma superfície $\alpha^*(\boldsymbol{\omega})$ no $\Re^3$.

Supondo o afastamento da verossimilhança modificado, cada seção normal (Figura 2.2.1), ou ainda, cada curvatura C (2.1.6), é associada a uma inclinação S (2.2.2). Assim, percebe-se através da Figura 2.2.1 que diminuindo a magnitude de a, diminui-se o comprimento de S. Dessa maneira, quanto menor S, menor será influência do esquema de perturbação na determinada seção normal, apontando indícios de estabilidade do modelo ajustado sob o particular esquema de perturbação, em torno de $\boldsymbol{\omega}_0$. Ao contrário, aumentando-se a magnitude de a, o comprimento de S aumenta, proporcionando maior influência do esquema de perturbação sob a seção normal em torno de $\boldsymbol{\omega}_0$.

2.3 Metodologia de Diagnóstico para Influência Local: Limiares

Conforme apresentado nas Seções 2.1 e 2.2, a partir de um modelo proposto e um específico esquema de perturbação, é possível calcular as medidas de influência local S e C_c, e os respectivos vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$. Assim, o diagnóstico baseado nas estatísticas S e C_c fica definido como *abordagem global*. Adicionalmente, o diagnóstico baseado nas estatísticas $\mathbf{v}$ e $\mathbf{v}_c$ é definido como *abordagem individual*.

Diferentes critérios de diagnóstico são propostos para avaliar a influência local diante da abordagem global. Billor e Loynes (1993) propuseram uma referência para S considerando o modelo de regressão linear e o esquema de perturbação inserido na matriz de variâncias e covariâncias. Cook (1986) considerou o valor dois, fundamentado em uma amostra aleatória simples, como referência para C_c, investigando assim a influência local no modelo de regressão. Poon e Poon (1999) com modelo de regressão linear e Zhu e Zhang (2004) com o modelo linear generalizado, propuseram a construção de níveis de referência para avaliar as medidas de influência local Curvatura Normal Conformal e C_c, respectivamente. Zhang e King (2005), também baseados na construção de níveis de referência, propuseram a estimação de percentis através de simulações de Monte Carlo para inclinação e curvatura (afastamento da verossimilhança modificado), considerando o modelo GARCH e diversos esquemas de perturbação.

No que diz respeito às estatísticas da abordagem individual, Billor e Loynes (1993) propuseram a simples inspeção visual das maiores componentes de $\mathbf{v}$ para o modelo de regressão linear com esquema de perturbação na matriz de variâncias e covariâncias. Cook (1986) comparou as "grandes" componentes dos resíduos do modelo de regressão com os "grandes" elementos de $\mathbf{v}_c$, considerando o esquema de perturbação de casos ponderados. Zhang e King (2005) apresentaram um diagnóstico baseado em simulações de Monte Carlo para o modelo GARCH e diversos esquemas de perturbação, e assim construíram níveis de referência para os vetores de diagnóstico da inclinação e curvatura.

Em resumo, as técnicas de diagnóstico propostas apresentam dois tipos de problemas. O primeiro trata das aproximações "grosseiras" apresentadas por Billor e Loynes (1993) para S e Cook (1986) para C_c. A segunda é a determinação de que uma observação específica é influente através da simples análise exploratória de gráficos para as estatísticas da abordagem individual.

Como as distribuições das estatísticas globais S e C_c, e individuais $\mathbf{v}$ e $\mathbf{v}_c$ são desconhecidas, Zhang e King (2005) propõem o uso de marcas de referência construídos através de simulações de Monte Carlo, motivando assim, uma metodologia para investigar a significância tanto das medidas de influência local, como dos seus respectivos vetores de diagnóstico. A partir de agora, as marcas de referência também serão chamadas de limiares, como sugerido por Prescott (1986).

Considerando um modelo proposto, ou seja, sobre o ponto de perturbação nula, as simulações de Monte Carlo se tornam uma importante ferramenta, pois permitem estimar as distribuições das medidas de influência local. Assim, através de parâmetros pré-determinados, para cada replicação é gerada uma amostra de tamanho n, e conseguinte são computadas as estatísticas S e C_c e os vetores $\mathbf{v}$ e $\mathbf{v}_c$. A partir dessas replicações, pode-se estimar a distribuição das estatísticas S e C_c. Nessa dissertação, são realizadas 1000 replicações para cada simulação.

A seguir são descritos os procedimentos para o cálculo de limiares, construídos a partir das simulações de Monte Carlo, e o conseqüente diagnóstico de influência local, levando-se em consideração a Inclinação de Billor e Loynes e a Curvatura de Cook.

2.3.1 Limiares - Inclinação de Billor e Loynes

Considere o modelo *proposto*, ou seja, sem perturbação. São simulados 1000 modelos e calculadas as inclinações S em (2.2.3). Seja M_0 o percentil 95% das 1000 inclinações S simuladas. M_0 é uma marca de referência ou *limiar*. Seja agora um conjunto de dados com inclinação calculada por (2.2.3) igual a S_d. Se $S_d > M_0$, então pelo menos uma observação específica do conjunto de dados analisado é influente. A marca de referência M_0 será chamada de critério global para a influência local.

A partir das 1000 simulações do modelo *proposto* são calculados os vetores de inclinação $\mathbf{v}_j$ em (2.2.4), onde $j = 1, \ldots, 1000$. Em cada simulação é calculado o maior valor de $|v_{j_i}|$, onde v_{j_i} é a i-ésima componente do vetor $\mathbf{v}_j$. M_1 é o percentil 95% das maiores componentes η_j para os 1000 vetores $\mathbf{v}_j$ simulados. M_1 é uma marca de referência ou *limiar*. Seja agora um conjunto de dados com vetor de inclinação S_d calculado por (2.2.4) igual a $\mathbf{v}_d$. Se $v_{d_i} > M_1$, onde v_{d_i} é a i-ésima componente do vetor $\mathbf{v}_d$, diz-se portanto, que a i-ésima observação do conjunto de dados é individualmente influente.

Por outro lado, seja $D_j = \{D_1, \ldots, D_{1000}\}$ o conjunto das 1000 inclinações S simuladas. Seja também $D_{j_u} = \{D_{j_1}, \ldots, D_{j_m}\}$ o conjunto das inclinações S simuladas tal que $S > M_0$, onde $u = 1, \ldots, m$ e $\{j_1, \ldots, j_m\} \in \{1, \ldots, 1000\}$. Para cada inclinação S do conjunto D_{j_u}, é calculado um vetor $\mathbf{v}_j$ em (2.2.4), cujos componentes são representadas por η_{j_u}. Define-se M_2 como o percentil 5% das maiores componentes $|\eta_{j_u}|$ dos m vetores $\mathbf{v}_j$ simulados. M_2 é uma marca de referência ou *limiar*. Seja agora um conjunto de dados com vetor de inclinação calculado por (2.2.4) igual a $\mathbf{v}_d$. Se $v_{d_i} > M_2$, então a i-ésima observação do conjunto de dados é individualmente influente. As marcas de referência M_1 e M_2 serão chamadas de critérios individuais para a influência local.

Em resumo, considerando k replicações, os limiares são calculados da seguinte maneira.

I) Global:
$D_j = S$, $j = 1, \ldots, k$
M_0 = Percentil 95% de $\{D_1, \ldots, D_k\}$.

II) Individual:
$\eta_j = max\,\{|v_{j_1}|, \ldots, |v_{j_n}|\}$, $j = 1, \ldots, k$
M_1 = Percentil 95% de $\{\eta_1, \ldots, \eta_k\}$
M_2 = Percentil 5% de $\{\eta_{j_1}, \ldots, \eta_{j_m}\}$ onde $\{j_1, \ldots, j_m\} \in \{1, \ldots, k\}$ corresponde a série gerada que satisfaz $D_{j_u} > M_0$, $u = 1, \ldots, m$.

2.3.2 Limiares - Curvatura de Cook

Considere o modelo *proposto*, ou seja, sem perturbação. São simulados 1000 modelos e calculadas as curvaturas C em (2.1.7). Seja M_0 o percentil 95% das 1000 curvaturas C simuladas. M_0 é uma marca de referência ou *limiar*. Seja agora um conjunto de dados com curvatura calculada por (2.1.7) igual a C_d. Se $C_d > M_0$, então pelo menos uma observação específica do conjunto de dados analisado é influente. A marca de referência M_0 será chamada de critério global para a influência local.

Com base nas 1000 simulações do modelo *proposto* são calculados os vetores $\mathbf{v}_j$ em (2.1.8) relativo a curvatura, onde $j = 1, \ldots, 1000$. Em cada simulação é calculado o maior valor de $|v_{j_i}|$, onde v_{j_i} é a i-ésima componente do vetor $\mathbf{v}_j$. M_1 é o percentil 95% das maiores componentes λ_j para os 1000 vetores $\mathbf{v}_j$ simulados. Seja agora um conjunto de dados com vetor de curvatura C_d calculada por (2.1.8) igual a $\mathbf{v}_d$. Se $v_{d_i} > M_1$, onde v_{d_i} é a i-ésima componente do vetor de curvatura $\mathbf{v}_d$, então a i-ésima observação do conjunto de dados é individualmente influente.

Seja $G_j = \{G_1, \ldots, G_{1000}\}$ o conjunto de todas as 1000 curvaturas C simuladas. Seja também $G_{j_u} = \{G_{j_1}, \ldots, G_{j_m}\}$ o conjunto das curvaturas C simuladas tal que $C > M_0$, onde $u = 1, \ldots, m$ e $\{j_1, \ldots, j_m\} \in \{1, \ldots, 1000\}$. Para cada curvatura C do conjunto G_{j_u}, é calculado um vetor $\mathbf{v}_j$ em (2.1.8), cujos componentes são representadas por λ_{j_u}. Define-se M_2 como o percentil 5% das maiores componentes $|\lambda_{j_u}|$ dos m vetores $\mathbf{v}_j$ simulados. Seja agora um conjunto de dados com vetor de curvatura calculado por (2.1.8) igual a $\mathbf{v}_d$. Se $v_{d_i} > M_2$, então a i-ésima observação do conjunto de dados é individualmente influente. As marcas de referência M_1 e M_2 serão chamadas de critérios individuais para a influência local.

Em resumo, considerando k replicações, os limiares são calculados da seguinte maneira.

III) Global:

$G_j = C$, $j = 1, \ldots, k$

$M_0 =$ Percentil 95% de $\{G_1, \ldots, G_k\}$.

IV) Individual:

$\lambda_j = max\,\{|v_{j_1}|, \ldots, |v_{j_n}|\}$, $j = 1, \ldots, k$

M_1 = Percentil 95% de $\{\lambda_1, \ldots, \lambda_k\}$

M_2 = Percentil 5% de $\{\lambda_{j_1}, \ldots, \lambda_{j_m}\}$ onde $\{j_1, \ldots, j_m\} \in \{1, \ldots, k\}$ corresponde as séries de dados geradas as quais satisfazem $G_{j_u} > M_0$, $u = 1, \ldots, m$.

2.4 Aplicação

Esta seção tem o objetivo de ilustrar a técnica de limiares descrita na seção anterior através de um exemplo estudado por Billor e Loynes (1993) em um modelo de regressão com esquema de perturbação na matriz de variâncias e covariâncias.

Seja $\mathbf{y} = (y_1, \ldots, y_n)^T$ o vetor resposta de observações geradas através do modelo de regressão linear,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.4.1}$$

onde $\mathbf{X}$ é a variável explanatória, $\boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{W}^{-1})$ e, $\boldsymbol{\beta}$ e σ^2 são os parâmetros, tal que σ^2 é conhecido e $\mathbf{W}$ é a matriz de perturbação dada por

$$\mathbf{W} = diag(1 + \omega_1, ..., 1 + \omega_n). \tag{2.4.2}$$

Segue-se, portanto, que a log-verossimilhança perturbada é dada por

$$L(\boldsymbol{\beta}|\boldsymbol{\omega}) = -\frac{n}{2}log(2\pi) - \frac{1}{2}log\left\{\prod_{i=1}^{n}\frac{\sigma^2}{\omega_i + 1}\right\} - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

A seguir são discutidas as medidas de influência local Inclinação de Billor e Loynes e Curvatura de Cook.

2.4.1 Inclinação de Billor e Loynes

Billor e Loynes (1993) estudaram amostras referentes a um conjunto de dados de gansos [Weisberg, 1985], os quais consistem de 45 observações do verdadeiro tamanho do rebanho realizadas através de fotografias aéreas ($\mathbf{y}$) e a estimação visual do tamanho do rebanho ($\mathbf{X}$) realizada por um observador. Estes dados foram obtidos para determinar quão bom poderia ser a estimação visual durante uma contagem do censo da população de gansos em relação a atual técnica de contagem do censo realizada por fotografias aéreas.

Através do gráfico de dispersão na Figura 2.4.1, percebe-se evidência de heteroscedasticidade. Este fato justifica a adoção do esquema de perturbação para a estrutura de variâncias e covariâncias (2.4.2) no modelo de regressão linear. Ajustando o modelo de regressão linear aos dados através do comando *lm* do *software R*, encontram-se as estimativas de máxima verossimilhança apresentadas na Tabela 2.4.1. As estimativas são altamente significativas segundo os quocientes t.

Tabela 2.4.1: Parâmetros estimados.

Parâmetro	Estimativa	Desvio Padrão	Razão t
β_0	26,6496	8,6145	3,094
β_1	0,8825	0,0776	11,367
σ	43,8981		
$R^2 = 0,7445$			

A i-ésima componente do vetor de inclinação (2.2.4) é

$$\frac{\partial LD^*(\boldsymbol{\omega}_0)}{\partial \omega_i} = \frac{1}{2}\left(1 - \frac{e_i^2}{\hat{\sigma}^2}\right), \tag{2.4.3}$$

onde $\mathbf{e} = \mathbf{y} - \hat{\boldsymbol{\beta}}\mathbf{X}$, para $i = 1, ..., n$.

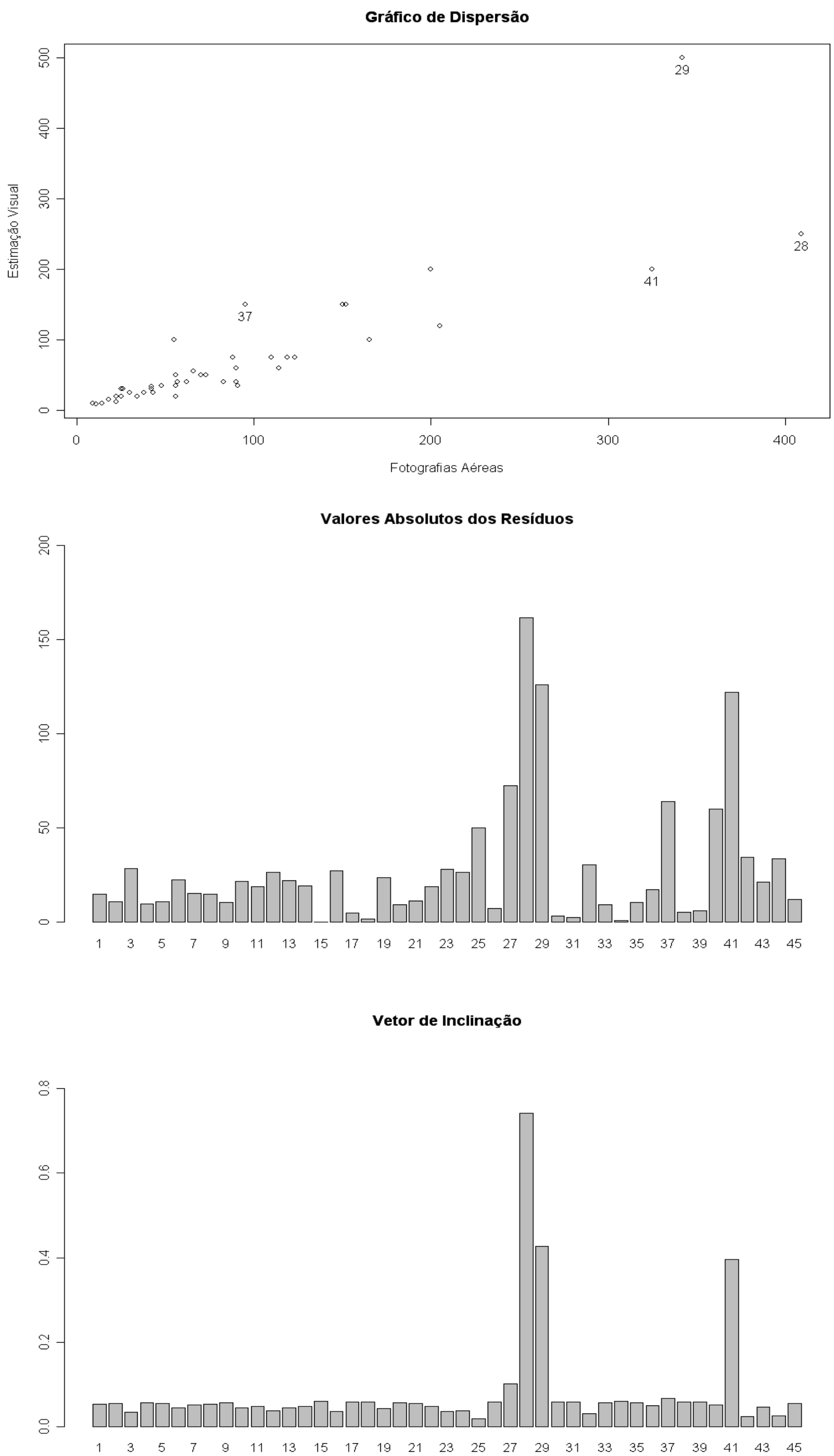

Figura 2.4.1: Dados de gansos, resíduos em valores absolutos e vetor de inclinação **v**.

A Figura 2.4.1 ilustra o vetor de valores absolutos dos resíduos em conjunção ao vetor de inclinação $\mathbf{v}$. Analisando os gráficos da Figura 2.4.1, percebe-se que devido a "grande" magnitude observada através dos resíduos e_{28}, e_{29} e e_{41} e das componentes v_{28}, v_{29} e v_{41}, existem indícios de que as observações y_{28}, y_{29} e y_{41} são influentes..

Billor e Loynes (1993) calcularam a inclinação para o conjunto de dados a partir de (2.2.3) obtendo $S = 17,377$. Eles propuseram uma referência ($\sqrt{2n + 4\sqrt{14n}}$), onde n é o tamanho amostral. Dado que essa referência tem valor 13,798, Billor e Loynes indicam que há influência local. Considerando a abordagem individual, ou seja, a análise das componentes do vetor $\mathbf{v}$, Billor e Loynes realizaram a simples inspeção visual das maiores magnitudes do vetor. Portanto, identificaram como influentes as observações y_{28}, y_{29} e y_{41} (ver Figura 2.4.1).

Com o objetivo de construir níveis de referência para identificar as observações significativamente influentes, o presente trabalho propõe a aplicação da metodologia de limiares, considerando os enfoques global e individual para a inclinação segundo descrito na Seção 2.3. Assim, considerando-se as estimativas dos parâmetros da Tabela 2.4.1, para o critério global encontramos $M_0 = 14,1118$, e como $S = 17,377 > M_0$, onde S é a inclinação calculada em (2.2.3) para o conjunto de dados, então há pelo menos uma observação influente. Adicionalmente, os limiares individuais são $M_1 = 0,7746$ e $M_2 = 0,5098$. A Figura 2.4.2 ilustra a relação do vetor de inclinação $\mathbf{v}$ com os limiares M_1 e M_2. A partir desta figura podemos observar que somente uma componente (y_{28}) sobrepassa a marca M_2. Dessa maneira, a metodologia de limiares identifica a observação y_{28} como influente.

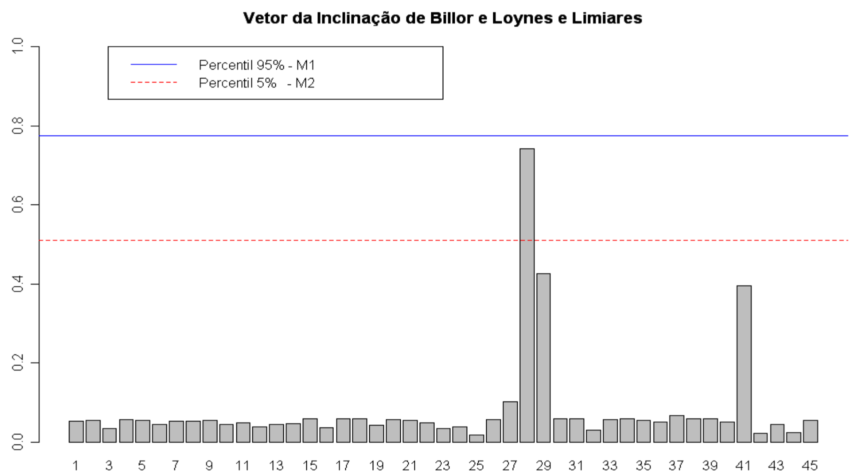

Figura 2.4.2: Limiares para influência local da inclinação.

Billor e Loynes (1993) identificaram três observações como influentes y_{28}, y_{29} e y_{41}. Já a técnica dos limiares identificou somente uma observação como sendo influente (y_{28}). Comparando com a forma de identificação utilizada por Billor e Loynes (1993), a metodologia de limiares proposta no presente trabalho é melhor pelo fato de serem utilizadas medidas de referências estatísticas, e não apenas a simples análise visual das maiores componentes do vetor de diagnóstico.

2.4.2 Curvatura de Cook

Para avaliar a influência local no modelo (2.4.1) considerando o esquema de perturbação (2.4.2), através da Curvatura de Cook são calculadas as seguintes expressões:

$$\frac{\partial^2 L(\hat{\boldsymbol{\beta}}|\boldsymbol{\omega}_0)}{\partial \boldsymbol{\omega} \partial \boldsymbol{\beta}^T} = \frac{1}{\hat{\sigma}^2} D(\mathbf{e})\mathbf{W}\mathbf{X},$$
$$\frac{\partial^2 L(\hat{\boldsymbol{\beta}}|\boldsymbol{\omega}_0)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\frac{1}{\hat{\sigma}^2} \mathbf{X}^T\mathbf{W}\mathbf{X},$$

com $\mathbf{e}$ definido em (2.4.3) e D($\mathbf{e}$) a matriz diagonal dos elementos de $\mathbf{e}$.

Com as estimativas da Tabela 2.4.1 são calculadas a Curvatura de Cook C_c e o vetor da curvatura $\mathbf{v}_c$ a partir de (2.1.7) e (2.1.8), respectivamente.

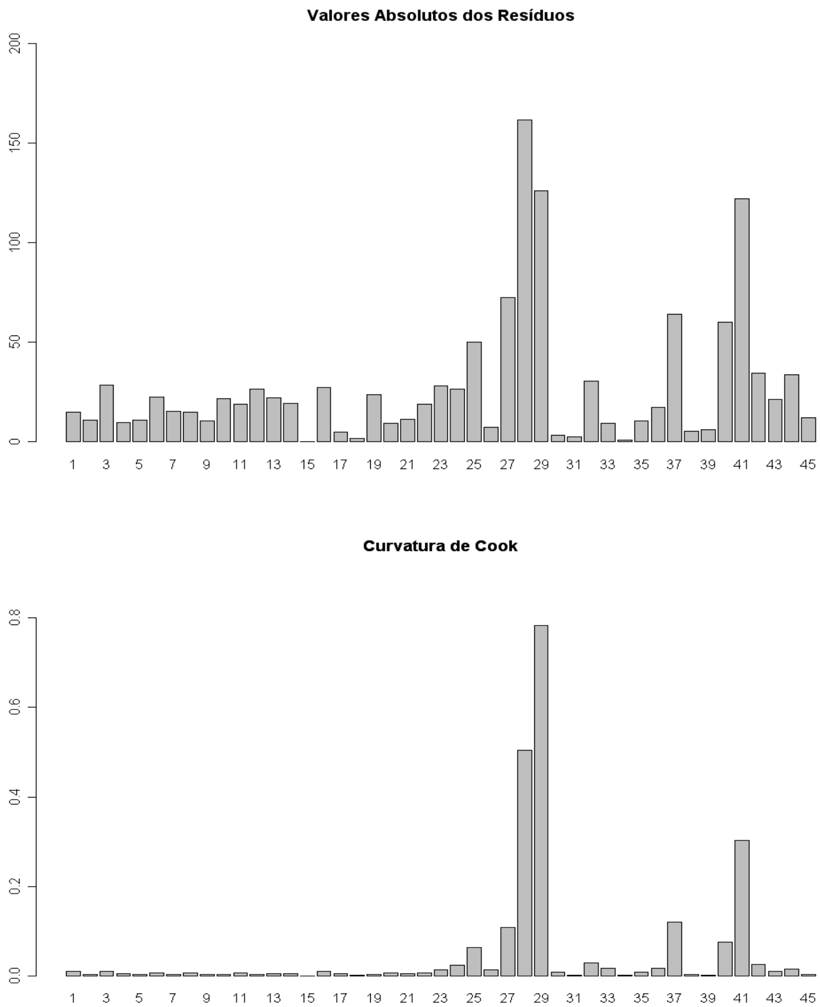

Figura 2.4.3: Resíduos em valor absoluto e vetor de curvatura $\mathbf{v}_c$.

O gráfico dos valores absolutos dos resíduos em conjunção ao vetor da curvatura $\mathbf{v}_c$ apresentado na Figura 2.4.3 aponta indícios de que as observações y_{28}, y_{29} e y_{41} são influentes, pois as componentes v_{28}, v_{29} e v_{41} apresentam magnitudes "grandes".

Com o objetivo de construir níveis de referência para identificar as observações significativamente influentes, o presente trabalho propõe a aplicação da metodologia de limiares, considerando os enfoques global e individual descritos na Seção 2.3. Assim, para o critério global obtemos $M_0 = 5,3285$, e como $C = 12,08 > M_0$. Adicionalmente, os limiares individuais são $M_1 = 0,9247$ e $M_2 = 0,8020$. A Figura 2.4.4 ilustra a relação de $\mathbf{v}_c$ com os limiares M_1 e M_2. A componente 29 do vetor de curvatura $\mathbf{v}_c$ é igual a $0,8233$, ou seja, apresenta valor maior do que M_2. Portanto, y_{29} é uma observação influente.

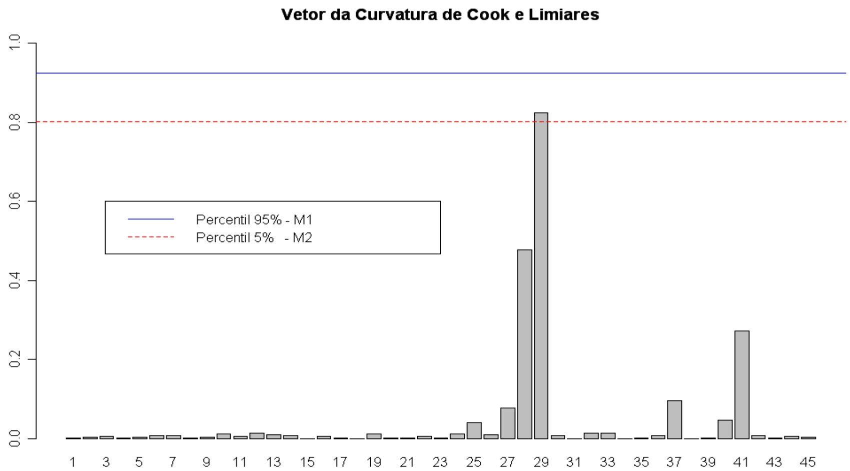

Figura 2.4.4: Limiares para influência local da Curvatura de Cook.

2.4.3 Deleção de uma Observação

Uma outra análise que será realizada é verificar se, após a exclusão de uma observação identificada como influente pelos critérios individuais M_1 e M_2, são identificadas outras observações, isto é, se ocorre mascaramento. A Tabela 2.4.2 apresenta as estimativas dos parâmetros considerando exclusão da observação (x_{29}, y_{29}) que foi identificada como influente com base na Curvatura de Cook.

Tabela 2.4.2: Estimativas sem a observação (x_{29}, y_{29}).

Parâmetro	Estimativa	Desvio Padrão	Razão t
β_0	5,1412	7,4282	0,692
β_1	1,2804	0,0893	14,326
σ	32,9158		
$R^2 = 0,8261$			

Analisando as estimativas apresentadas na Tabela 2.4.2, $\hat{\beta}_0$ não é significativa a 5%.

A Tabela 2.4.3 apresenta os novos limiares sem considerar a observação (x_{29}, y_{29}).

Tabela 2.4.3: Estatísticas e limiares para influência sem a observação (x_{29}, y_{29}).

		Billor e Loynes	Cook
		$\mathbf{v}$	$\mathbf{v}_c$
Observações Influentes			
y_{28}		0,4209	**0,6458**
y_{36}		**0,6646**	0,4202
Estatísticas Globais	S	13,0015	—
	C_c	—	**7,2936**
Limiares	M_0	13,3977	4,3371
	M_1	0,7583	0,7997
	M_2	0,5305	0,4885

Em negrito estão representadas as estatísticas significativas para M_0, M_1 e M_2.

Na Tabela 2.4.3, vemos que $S = 13,0015$ é menor do que $M_0 = 13,3977$, e dessa maneira, não haveria influência global segundo este critério, mas nota-se que os valores de S e M_0 estão muito próximos. A observação y_{36} (ver Figura 2.4.1), que era a observação y_{37} antes do mascaramento de y_{29}, é identificada como individualmente influente pelo critério M_2, considerando o vetor de inclinação de Billor e Loynes. Além disso, a observação y_{28}, que era influente para o vetor de Inclinação, agora é considerada influente através vetor da Curvatura de Cook sob o critério M_2.

Retirando a observação (x_{29}, y_{29}), a observação y_{41} identificada por Billor e Loynes (1993) como influente, continua não sendo identificada pela metodologia de limiares considerando os critérios individuais.

2.4.4 Conclusões

Alguns comentários podem ser destacados com respeito à utilização das duas medidas globais e duas medidas individuais de influência local. O primeiro é que a simples análise exploratória das duas medidas de diagnóstico propostas para avaliar a influência local apresentam as mesmas observações como possíveis candidatas a pontos significativamente influentes, ou seja, y_{28}, y_{29} e y_{41}. O segundo é que utilizar diferentes medidas de influência local pode ocasionar a identificação de diferentes observações influentes. Por exemplo, utilizando a Inclinação de Billor e Loynes, y_{28} é influente, mas através da Curvatura de Cook y_{29} é influente, considerando o enfoque individual M_2. Isto pode ser verificado na Figura 2.4.5 onde é apresentado resumo dos gráficos dos vetores $\mathbf{v}$ e $\mathbf{v}_v$ com os respectivos limiares M_0, M_1, M_2 (Tabela 2.4.4).

Através da deleção do ponto (x_{29}, y_{29}), percebe-se que a observação y_{28} continua a ser influente sob o critério M_2, só que agora essa observação é identificada considerando o vetor da Curvatura de Cook. Observou-se também que a observação y_{37}, que anteriormente não tinha sido identificada como sendo influente, após a exclusão da observação (x_{29}, y_{29}) se torna influente sob o critério individual M_2.

Considerando o modelo de regressão linear (2.4.1) e o esquema de perturbação (2.4.2), através da aplicação nos dados de gansos, percebe-se que a metodologia baseada na construção de níveis de referência é uma poderosa ferramenta para investigar a influência local. A partir da análise sobre os enfoques global e individual para as medidas foi possível identificar as observações y_{28} e y_{29} como individualmente influentes.

Diante dessas conclusões, a análise de influência local indica instabilidade do modelo ajustado para o específico esquema de perturbação. Assim, propõe-se a investigação das observações (x_{28}, y_{28}) e (x_{29}, y_{29}), ou o ajuste de um outro modelo.

Tabela 2.4.4: Resumo das estatísticas e limiares para influência.

		Billor e Loynes	Cook
		$\mathbf{v}$	$\mathbf{v}_c$
Observações Influentes			
y_{28}		**0,7413**	0,4767
y_{29}		0,4263	**0,8233**
Estatísticas Globais	S	**17,377**	—
	C_c	—	**12,08**
Limiares			
	M_0	14,1118	5,3285
	M_1	0,7852	0,9337
	M_2	0,5416	0,7882

Em negrito estão representadas as estatísticas significativas para M_0, M_1 e M_2.

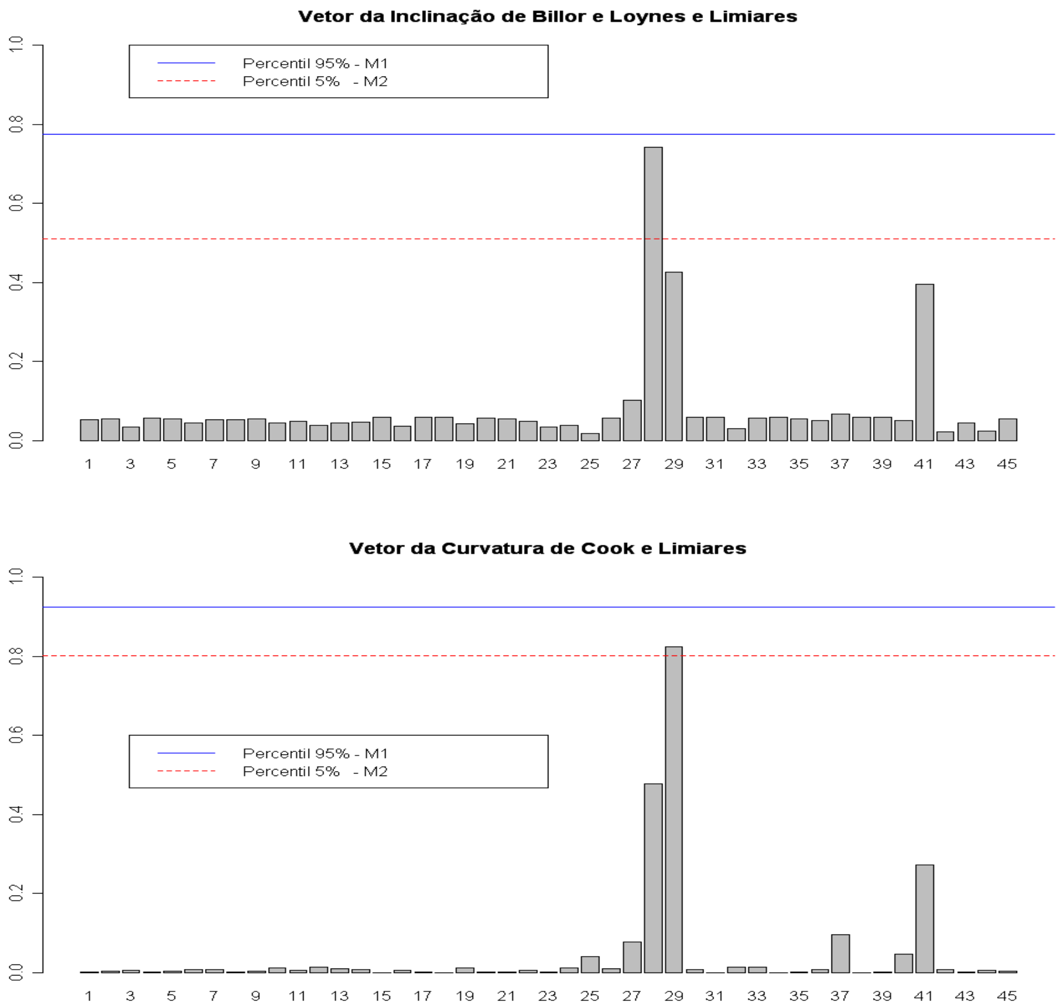

Figura 2.4.5: Resumo dos limiares para as medidas de influência.

3 Influência Local em Modelos de Regressão Linear com Erros Autoregressivos de Primeira Ordem

Este capítulo discute o diagnóstico de influência local em modelos de regressão linear com erros normais Autoregressivos de Primeira Ordem - AR(1), utilizando o esquema de perturbação nos dados. São consideradas três verossimilhanças: exata, perfilada e condicional. A partir das diferentes verossimilhanças são calculadas as medidas Inclinação de Billor e Loynes e Curvatura de Cook, e seus respectivos vetores de diagnóstico. Por fim, a metodologia de limiares apresentada na Seção 2.3, incluindo os enfoques global e individual, é aplicada em quatro conjuntos de dados gerados por simulação e em duas séries de dados reais.

3.1 Introdução

Na literatura, encontram-se alguns estudos sobre influência local para modelos autoregressivos. Assim, Kim e Huggins (1997) investigaram a influência local através da Curvatura de Cook (C_c) para modelos de regressão linear com erros AR(1), utilizando o esquema de perturbação nos dados com a verossimilhança exata. Tsai e Wu (1992) estudaram o modelo de regressão linear com erros autoregressivos de ordem um, através da inclinação e de seu vetor de diagnóstico, considerando o esquema de perturbação nos dados e uma verossimilhança perfilada. Schall e Dunne (1991) estudaram a influência local, para o modelo de regressão com erros ARMA(p,q), através da Curvatura de Cook, onde aplicaram quatro tipos de esquemas de perturbação para uma verossimilhança condicionada. Recentemente, Medeiros (2006) discutiu a influência local, utilizando a Curvatura de Cook, como técnica de diagnóstico para os modelos de regressão com erros pertencentes a classe elíptica e com estrutura autoregressiva de ordem um, considerando os esquemas de perturbações nas variáveis resposta e explicativa.

No entanto, nesses estudos a análise de influência nos conjuntos de dados é realizada de forma exploratória. Neste capítulo será utilizada a metodologia de limiares descrita no Capítulo 2 para avaliar a influência em modelos de regressão com erros autoregressivos de ordem um.

O processo Autoregressivo de Primeira Ordem, ou AR(1), que está definido como

$$\epsilon_t = \rho\epsilon_{t-1} + \eta_t \ , \ t = 1, ..., n \ , \tag{3.1.1}$$

onde $\{\eta_t\}$ é um ruído branco, isto é, as variáveis aleatórias η_t são independentes, com $var[\eta_t] = \sigma^2$, e em conseqüência, $cov[\eta_t, \eta_s] = 0$, $t \neq s$. Notação: $\eta_t \sim RB(0, \sigma^2)$. As variáveis ϵ_t e η_{t+k} são independentes, para $k \geq 1$ e o processo é estacionário se $|\rho| < 1$.

A seguir são apresentadas duas propriedades que permitem identificar o comportamento de uma série temporal como sendo um processo AR(1). A primeira é o decaimento exponencial das autocorrelações, ou seja,

$$\rho(\tau) = \begin{cases} 1 & , \tau = 0 \\ \rho^{\tau} & , \tau \neq 0, \end{cases}$$

onde $\rho(\tau)$ é a função de autocorrelação. A segunda propriedade diz respeito ao comportamento das autocorrelações parciais, dado por

$$\phi_{\tau\tau} = \begin{cases} \rho & , \tau = 1 \\ 0 & , \tau > 1, \end{cases}$$

onde $\phi_{\tau\tau}$ é a função de autocorrelação parcial.

A próxima seção apresenta o estudo da influência local considerando três tipos diferentes de verossimilhanças.

3.2 Influência Local com a Verossimilhança Exata

Esta seção propõe estudar a influência local para o modelo de regressão linear com erros autoregressivos de primeira ordem através da verossimilhança exata, considerando o esquema de perturbação nos dados.

Na literatura, Kim e Huggins (1991) estudaram a influência local para o modelo de regressão linear com erros normais AR(1) através da verossimilhança exata, e assim, investigaram o comportamento do vetor de diagnósico $\mathbf{v}_c$ associado a Curvatura de Cook (1986). No entanto, Kim e Huggins (1991) não apresentaram referências estatísticas para avaliar a significância das componentes dos vetores de diagnóstico. Para contornar este problema, o presente trabalho estuda a influência local no modelo de regressão linear com erros AR(1) através das medidas de Inclinação de Billor e Loynes e Curvatura de Cook utilizando a metodologia de limiares discutida na Seção 2.3. Para tanto, será considerada a verossimilhança exata e o esquema de perturbação nos dados.

Seja $\mathbf{y} = (y_1, \ldots, y_n)^T$ o vetor de observações geradas através do modelo de regressão linear com erros autoregressivos de primeira ordem, isto é,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3.2.1}$$

onde $\mathbf{X}$ é uma matriz $(n \times (k+1))$ de posto completo, $\boldsymbol{\beta}$ é um vetor $((k+1) \times 1)$ de parâmetros desconhecidos e $\boldsymbol{\epsilon}$ é um vetor aleatório $(n \times 1)$, tal que $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2\boldsymbol{\Sigma})$, com

$$\boldsymbol{\Sigma} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \ldots & \rho^{n-1} \\ \rho & 1 & \ldots & \rho^{n-2} \\ \vdots & & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \ldots & 1 \end{pmatrix},$$

sendo ρ coeficiente de correlação desconhecido e σ^2 uma constante desconhecida.

Assim, a verossimilhança exata para o modelo (3.2.1) é dada por

$$L(\boldsymbol{\theta}) \;=\; -\frac{n}{2}log(\sigma^2) + \frac{1}{2}log(1-\rho^2) - \frac{1}{2\sigma^2}(\mathbf{y} - X\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

onde $\boldsymbol{\theta} = (\rho, \sigma^2, \boldsymbol{\beta}^T)^T$ é o vetor de parâmetros.

Considerando o esquema de perturbação nos dados

$$\tilde{\mathbf{y}} = \mathbf{y} + \boldsymbol{\omega}, \tag{3.2.2}$$

onde $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T$, então a verossimilhança perturbada para o modelo (3.2.1) é dada por

$$L(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{n}{2}log(\sigma^2) + \frac{1}{2}log(1-\rho^2) - \frac{1}{2\sigma^2}(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}). \tag{3.2.3}$$

O presente trabalho propõe a investigar a influência local através das medidas S e C_c, e seus respectivos vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$. Dessa maneira, avaliado sobre os estimadores de máxima verossimilhança $\hat{\boldsymbol{\theta}} = (\hat{\rho}, \hat{\sigma}^2, \hat{\boldsymbol{\beta}}^T)^T$ e o vetor de perturbação nula $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = \mathbf{0}$, o vetor na direção da inclinação é dado por

$$\mathbf{v} = 2\frac{L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial \boldsymbol{\omega}} = -\frac{2}{\hat{\sigma}^2}\mathbf{e}^T\boldsymbol{\Sigma}^{-1}.$$

Assim, através de $\mathbf{v}$ é possível encontrar $S = ||\mathbf{v}||$ (vide (2.2.3) e (2.2.4)).

Kim e Huggins (1998) apresentaram os resultados das derivadas necessárias para avaliar a influência local, levando-se em consideração a medida de influência global C_c e individual $\mathbf{v}_c$. Seja portanto,

$$\boldsymbol{\Delta} = \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} = \left(\frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \rho \partial \boldsymbol{\omega}^T} \quad , \quad \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \sigma^2 \partial \boldsymbol{\omega}^T} \quad , \quad \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T}. \right)^T$$

Dessa forma, avaliando-se $\boldsymbol{\Delta}$ sobre $\hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega}_0$, encontra-se:

$$\boldsymbol{\Delta} = \frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} = \frac{1}{\hat{\sigma}^4} \begin{pmatrix} -\hat{\sigma}^2 \mathbf{e}^T \boldsymbol{\Sigma}_{\hat{\rho}}^{-1} \\ \mathbf{e}^T \boldsymbol{\Sigma}^{-1} \\ \hat{\sigma}^2 X^T \boldsymbol{\Sigma}^{-1} \end{pmatrix},$$

onde $\boldsymbol{\Sigma}_{\hat{\rho}}^{-1} = \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \rho}$, $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}$.

A matriz de informação observada, $-\ddot{\mathbf{L}}$, foi calculada tomando negativas as segundas derivadas da função de log-verossimilhança não-perturbada avaliadas em $\hat{\boldsymbol{\theta}}$, ou seja,

$$\ddot{\mathbf{L}} = \frac{\partial^2 L(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{pmatrix} b & c & \mathbf{u}^T \\ c & r & \mathbf{0}^T \\ \mathbf{u} & \mathbf{0} & \mathbf{A} \end{pmatrix},$$

onde

$$b = -\frac{(1+\hat{\rho}^2)}{(1-\hat{\rho}^2)^2} - \frac{1}{\hat{\sigma}^2}\sum_{i=3}^{n}(y_{i-1} - \mathbf{x}_{i-1}^T\hat{\boldsymbol{\beta}})^2, \quad c = -\tfrac{\hat{\rho}}{\hat{\sigma}^2(1-\hat{\rho}^2)},$$

$$r = -\frac{n}{2\hat{\sigma}^4}, \quad \mathbf{u} = \tfrac{1}{\hat{\sigma}^2}\mathbf{X}^T\boldsymbol{\Sigma}_{\hat{\rho}}^{-1}\mathbf{e}, \quad \mathbf{A} = -\frac{1}{\hat{\sigma}^2}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}.$$

Assim, a matriz $\ddot{\mathbf{F}}$ toma a forma

$$\begin{aligned} \ddot{\mathbf{F}} &= \frac{gc}{d\hat{\sigma}^6}\left[\boldsymbol{\Sigma}^{-1}\mathbf{e}\mathbf{e}^T\boldsymbol{\Sigma}_{\hat{\rho}}^{-1}(\mathbf{I} - \mathbf{H}\boldsymbol{\Sigma}^{-1}) + (\mathbf{I} - \boldsymbol{\Sigma}^{-1}\mathbf{H})\boldsymbol{\Sigma}_{\hat{\rho}}^{-1}\mathbf{e}\mathbf{e}^T\boldsymbol{\Sigma}^{-1}\right] + \\ &+ \frac{g}{\hat{\sigma}^4}(\mathbf{I} - \boldsymbol{\Sigma}^{-1}\mathbf{H})\boldsymbol{\Sigma}_{\hat{\rho}}^{-1}\mathbf{e}\mathbf{e}^T\boldsymbol{\Sigma}_{\hat{\rho}}^{-1}(\mathbf{I} - \mathbf{H}\boldsymbol{\Sigma}^{-1}) + \frac{\boldsymbol{\Sigma}^{-1}(\alpha\mathbf{e}\mathbf{e}^T - \mathbf{H}\hat{\sigma}^6)\boldsymbol{\Sigma}^{-1}}{\hat{\sigma}^8}, \end{aligned}$$

onde $\alpha = 1/r + g(c/r)^2$ e $g = \left\{ b - \left(c^2/b - \frac{\mathbf{e}^T\boldsymbol{\Sigma}_{\hat{\rho}}^{-1}\mathbf{H}\boldsymbol{\Sigma}_{\hat{\rho}}^{-1}\mathbf{e}}{\hat{\sigma}^2} \right) \right\}^{-1}$.

A partir de $\ddot{\mathbf{F}}$, é possível calcular o autovetor $\mathbf{v}_c$ associado ao maior autovalor de $\ddot{\mathbf{F}}$ (C_c) tal como em (2.1.7) e (2.1.8).

3.3 Influência Local com a Verossimilhança Perfilada

Tsai e Wu (1992) estudaram a influência local para o modelo de regressão linear com erros AR(1) a partir de uma verossimilhança perfilada, considerando somente ρ como parâmetro, e assim, investigaram o comportamento do vetor de diag nóstico da inclinação. É importante ressaltar que essa inclinação não é baseada nos estudos de Billor e Loynes (1993), mas sim no estimador de máxima verossimilhança $\hat{\rho}(a)$, visto como uma superfície na direção do referido vetor de perturbação. Ou seja, Tsai e Wu (1992) consideram $\mathbf{v}$ calculado a partir da primeira derivada de $\hat{\rho}(a)$. Conforme mostrado no Capítulo 2, Billor e Loynes (1993) consideram os cálculos da inclinação baseados no afastamento da verossimilhança modificado.

O presente trabalho propõe o estudo da influência local através do modelo de regressão com erros AR(1) e uma verossimilhança perfilada, onde o interesse está em torno somente dos parâmetros σ^2 e $\boldsymbol{\beta}$. Ao invés de redefinir a verossimilhança exata (3.2.3), e conseqüentemente, calcular todas as expressões relativas as derivadas em (2.1.7) e (2.1.8), será utilizada a restrição do parâmetro ρ que não é de interesse para o modelo. Ou seja, a curvatura normal de Cook (1986) na direção d é dada por

$$C_c(\sigma^2, \boldsymbol{\beta}) = 2 \left\| \mathbf{d}^T \boldsymbol{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{L}_\rho) \boldsymbol{\Delta} \mathbf{d} \right\|, \tag{3.3.1}$$

em que

$$\ddot{\mathbf{L}} = \frac{\partial^2 L(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{pmatrix} L_{\rho\rho} & L_{21} \\ L_{12} & L_{22} \end{pmatrix},$$

$$\ddot{\mathbf{L}}_\rho = \begin{pmatrix} L_{\rho\rho}^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

A curvatura $C_c(\sigma^2, \boldsymbol{\beta})$ e o vetor de diagnóstico $\mathbf{v}_c$ são obtidos pelo autovalor e autovetor de $\boldsymbol{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{L}_\rho) \boldsymbol{\Delta}$, assim como calculado em (2.1.7) e (2.1.8), respectivamente.

Dados os parâmetros $\boldsymbol{\theta} = (\rho, \sigma^2, \boldsymbol{\beta}^T)^T$ e $\boldsymbol{\theta}_1 = (\sigma^2, \boldsymbol{\beta}^T)^T$ relativos às verossimilhanças exata e perfilada, respectivamente, comparando os resultados da Inclinação de Billor e Loynes S e do vetor de diagnóstico $\mathbf{v}$, os mesmos serão idênticos, visto que, os cálculos baseados em (2.2.3) e (2.2.4) dependem somente da primeira derivada em relação às respectivas verossimilhanças.

3.4 Influência Local com a Verossimilhança Condicional

Schall e Dunne (1991) propuseram o diagnóstico de influência local com a medida Curvatura de Cook para o modelo regressão com erros ARMA, considerando quatro esquemas de perturbação (*outlier* para a mudança na média dos dados, *outlier* para a mudança na variância dos dados, *outlier* para a mudança na média das inovações e *outlier* para a mudança na variância das inovações). Para tanto, utilizaram uma verossimilhança aproximada com o intuito de simplificar a derivação algébrica para influência local, assim como minimizar o esforço computacional. Schall e Dunne (1991) abordaram somente a Curvatura de Cook (C_c) como medida de influência local. Além da Curvatura de Cook, o presente estudo propõe investigar a inclinação S, e os vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$ para detectar os pontos influentes através da metodologia de limiares descrita na Seção 2.3 considerando o modelo AR(1).

Nesta seção será estudada a influência local no modelo de regressão linear com erros AR(1) através das medidas de influência local S e C_c, considerando o esquema de perturbação nos dados e a verossimilhança condicionada na primeira observação.

Seja $\mathbf{y} = (y_1, \ldots, y_n)^T$ o vetor de uma amostra para um modelo qualquer, tal que $\boldsymbol{\theta}$ é o vetor de parâmetos. Então, a verossimilhança é dada por

$$l(\boldsymbol{\theta}) = f(y_1, \ldots, y_n; \boldsymbol{\theta}) = f(y_1)f(y_2|y_1)f(y_3|y_2, y_1)\cdots f(y_n|y_{n-1}, \ldots, y_1) = f(y_1)\prod_{t=2}^{n} f(y_t|F_{t-1}),$$

onde $f(\cdot)$ é a função de densidade de probabilidade e $F_{t-1} = \{y_{t-1}, ..., y_1\}$ representa o passado relativo a y_t. Logo, a verossimilhança é da forma

$$L(\boldsymbol{\theta}) = log\left\{f(y_1)\right\} + \sum_{t=2}^{n} log\left\{f(y_t|F_{t-1})\right\}.$$

Tratando-se de amostras suficientemente grandes a contribuição dada pela primeira observação, através de $log\left\{f(y_1)\right\}$, é desprezível [página 123 de Hamilton, 1994]. Dessa maneira, obtém-se a verossimilhança condicional

$$L(\boldsymbol{\theta}) \cong \sum_{t=2}^{n} log\left\{f(y_t|F_{t-1})\right\}. \tag{3.4.1}$$

Seja o modelo de regressão linear (3.2.1) escrito na forma

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t, \tag{3.4.2}$$

onde os resíduos ϵ_t seguem um processo autoregressivo de ordem um tal como em (3.1.1). Assim, condicionando-se ϵ_t em F_{t-1}, tem-se que

$$\epsilon_t | F_{t-1} \sim N(\rho\epsilon_{t-1}, \sigma^2),$$

e a respectiva verossimilhança condicional de y_t em F_1 é dada por

$$\begin{aligned} L(\boldsymbol{\theta}) = & -\left(\frac{n-1}{2}\right) log(2\pi\sigma^2) - \\ & - \left(\frac{1}{2\sigma^2}\right) \sum_{t=2}^{n} \left[y_t - \beta_0 - \beta_1 x_t - \rho(y_{t-1} - \beta_0 - \beta x_{t-1})\right]^2, \end{aligned} \qquad (3.4.3)$$

onde $\boldsymbol{\theta} = (\rho, \sigma^2, \beta_0, \beta_1)^T$.

Seja uma pequena perturbação ω_t introduzida na variável observada y_t para o modelo de regressão AR(1) (3.1.1) de tal forma que

$$\tilde{y}_t = y_t + \omega_t \quad , t = 2, ..., n \quad , \qquad (3.4.4)$$

onde $\tilde{y}_t$ é a variável perturbada. Assim, a verossimilhança perturbada condicionada em y_1 é dada por

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{\omega}) & = -\left(\frac{n-1}{2}\right) log(2\pi\sigma^2) - \\ & - \left(\frac{1}{2\sigma^2}\right) \sum_{t=2}^{n} \left[y_t + \omega_t - \beta_0 - \beta x_t - \rho(y_{t-1} + \omega_{t-1} - \beta_0 - \beta x_{t-1})\right]^2, \end{aligned} \qquad (3.4.5)$$

onde $\boldsymbol{\omega} = (\omega_2, ..., \omega_n)^T$.

A seguir este trabalho, baseado na verossimilhança condicional (3.4.5), apresenta as expressões que permitem calcular as medidas de influência local Inclinação de Billor e Loynes (S) e Curvatura de Cook (C_c), e seus respectivos vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$. Todos os resultados são avaliados sobre $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} = (\hat{\rho}, \hat{\sigma}^2, \hat{\beta}_0, \hat{\beta}_1)^T$ (estimadores de máxima verossimilhança) e $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = \mathbf{0}$ (perturbação nula).

- S:

$$\begin{aligned}\frac{\partial L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial \omega_i} &= -\frac{1}{2\hat{\sigma}^2}\sum_{t=2}^{n}\left\{2\left(I_{[t=i]} - \hat{\rho} I_{[t=i+1]}\right) A_t\right\} \\ &= -\frac{1}{\hat{\sigma}^2}\begin{cases} A_i - \hat{\rho} A_{i+1} & , \quad i = 2, ..., n-1 \\ A_i & , \quad i = n, \end{cases}\end{aligned}$$

onde $A_i = e_i - \rho e_{i-1}$ e $e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_t$ e $I_{[\,]}$ função indicadora, tal que

$$I_{[t=i]} = \begin{cases} 1 \ , \ t = i \ , \\ 0 \ , \ t \neq i. \end{cases}$$

- C_c:

$$\begin{aligned}\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial \omega_i \partial \rho} &= \frac{1}{\hat{\sigma}^2}\begin{cases} e_{i-1} - \hat{\rho} e_i + A_{i+1} \ , \ i = 2, ..., n-1 \\ e_{i-1} \ , \ i = n \end{cases} \\ \frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial \omega_i \partial \sigma^2} &= \frac{1}{\hat{\sigma}^4}\begin{cases} A_i - \hat{\rho} A_{i+1} \ , \ i = 2, ..., n-1 \\ A_i \ , \ i = n \end{cases} \\ \frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial \omega_i \partial \beta_0} &= \frac{1-\hat{\rho}}{\sigma^2}\begin{cases} 1 - \hat{\rho} \ , \ i = 2, ..., n-1 \\ 1 \ , \ i = n \end{cases} \\ \frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial \omega_i \partial \beta_1} &= \frac{1}{\hat{\sigma}^2}\begin{cases} x_i - \hat{\rho} x_{i-1} - (x_{i+1} - \hat{\rho} x_i) \ , \ i = 2, ..., n-1 \\ x_i - \hat{\rho} x_{i-1} \ , \ i = n \end{cases}\end{aligned}$$

As derivadas cruzadas entre os parâmetros são descritas a seguir.

$$
\begin{aligned}
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\rho\partial\rho} &= -\frac{1}{\hat{\sigma}^2}\sum_{t=2}^{n} e_{i-1}^2 \\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\sigma^2\partial\sigma^2} &= -\frac{(n-1)}{2\hat{\sigma}^4} \\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\beta_0\partial\beta_0} &= -\frac{(n-1)(1-\hat{\rho})^2}{\hat{\sigma}^2} \\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\beta_1\partial\beta_1} &= -\frac{1}{\hat{\sigma}^2}\sum_{t=2}^{n}(x_i-\hat{\rho}x_{i-1})^2 \\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\rho\partial\sigma^2} &= -\frac{1}{\hat{\sigma}^4}\sum_{t=2}^{n} A_t e_{i-1} \\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\rho\partial\beta_0} &= -\frac{1}{\hat{\sigma}^2}\left[\sum_{t=2}^{n} A_t + (1-\hat{\rho})\sum_{t=2}^{n} e_{t-1}\right] \\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\rho\partial\beta_1} &= -\frac{1}{\hat{\sigma}^4}\sum_{t=2}^{n}(x_{t-1}A_t + (x_i-\hat{\rho}x_{i-1})e_{t-1}) \\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\sigma^2\partial\beta_0} &= -\frac{(1-\hat{\rho})}{\hat{\sigma}^4}\sum_{t=2}^{n} A_t \\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\sigma^2\partial\beta_1} &= -\frac{(1-\hat{\rho})}{\hat{\sigma}^4}\sum_{t=2}^{n} A_t \\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\beta_0\partial\beta_1} &= -\frac{(1-\hat{\rho})}{\sigma^2}\sum_{t=2}^{n}(x_i-\hat{\rho}x_{i-1}).
\end{aligned}
$$

Estas expressões podem ser obtidas a partir das equações na Seção 3.2 desprezando a primeira observação. Portanto, quando o tamanho da amostra é razoável as verossimilhanças exata e condicional são praticamente iguais, a menos que ρ seja muito próximo de um.

A partir dos resultados das derivadas calculadas nesta seção e das expressões (2.2.3), (2.2.4), (2.1.7) e (2.1.8), relativas as medidas de influências globais e individuais apresentadas no Capítulo 2, será possível investigar a influência local das séries simuladas e dos conjuntos de dados apresentados nas seguintes seções.

3.5 Simulações

A seguir será apresentado o diagnóstico de influência local, baseado na técnica de limiares proposta na Seção 2.3, para séries temporais AR(1) geradas por simulação. O objetivo das simulações é estudar o comportamento das marcas de nível para os critérios M_0, M_1 e M_2 apresentados na Seção 2.3, para quatro séries com diferentes combinações de correlações e tamanhos amostrais.

3.5.1 Verossimilhança Exata

As simulações são baseadas no modelo autoregressivo de primeira ordem (3.1.1). Quatro séries temporais foram geradas considerando as combinações de tamanho amostral $n = 250$ e parâmetros de correlação $\rho = 0,5$, $\rho = 0,6$, $\rho = 0,7$ e $\rho = 0,9$, com $\sigma^2 = 1$. Essas séries serão denominadas conforme a Tabela 3.5.1.

Considera-se o esquema de perturbação nos dados. Nesse sentido, uma vez simuladas as séries com tamanho amostral $n = 250$, foram introduzidas duas observações "aberrantes" (*outliers* positivos e negativos), com magnitudes de cinco e três desvios padrões (σ_y) nas componentes y_{50} e y_{200}, respectivamente, onde σ_y é o desvio padrão da amostra gerada.

Tabela 3.5.1: Séries Simuladas.

	Tamanho Amostral	Coeficiente de Correlação	Posição do *Outlier*	Tamanho do *Outlier* em Valor Absoluto
Série 1	$n = 250$	$\rho = 0,5$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 2	$n = 250$	$\rho = 0,6$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 3	$n = 250$	$\rho = 0,7$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 4	$n = 250$	$\rho = 0,9$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$

Para as séries simuladas é realizada a análise de influência local utilizando os limiares descritos na Seção 2.3, e assim, espera-se identificar as observações geradas como influentes (*outliers*). As séries apresentadas na Tabela 3.5.1 foram escolhidas de tal maneira que as estimativas de ρ estivessem próximas aos verdadeiros valores (ver Tabela 3.5.7). Os limiares foram calculados em princípio supondo conhecidos os parâmetros do modelo e gerados com um programa escrito nos *softwares MATLAB* e *R* utilizando 1000 replicações.

A Tabela 3.5.2 apresenta o esquema de perturbação inserido para as séries 1 e 4, e auxilia na análise das observações influentes. Adicionalmente, nas Tabelas 3.5.3 - 3.5.5 são apresentadas as marcas de nível correspondentes ao critério global M_0 e aos critérios individuais M_1 e M_2, respectivamente. Além disso, foram calculados os limiares M_0, M_1 e M_2 considerando o tamanho amostral $n = 500$.

Tabela 3.5.2: Perturbação positiva e negativa no modelo AR(1) - Normal.

	y_{48}	y_{49}	y_{50}	y_{51}	y_{52}	ω_{50}	$\tilde{y}_{50} = 5\sigma_y$
Série 1	-0,268	0,001	-0,020	2,169	0,448	5,864	5,843
Série 1	-0,268	0,001	-0,020	2,169	0,448	-5,823	-5,843
Série 4	3,704	2,614	1,187	-0,167	1,499	9,649	10,836
Série 4	3,704	2,614	1,187	-0,167	1,499	-12,024	-10,836
	y_{198}	y_{199}	y_{200}	y_{201}	y_{202}	ω_{200}	$\tilde{y}_{200} = 3\sigma_y$
Série 1	-0,651	0,076	-0,923	-0,024	-1,193	4,429	3,506
Série 1	-0,651	0,076	-0,923	-0,024	-1,193	-2,582	-3,506
Série 4	3,471	6,009	5,695	5,389	3,862	0,806	6,502
Série 4	3,471	6,009	5,695	5,389	3,862	-12,197	-6,502

Tabela 3.5.3: Estatísticas Globais M_0.

Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$ 0,5	0,6	0,7	0,9
		90	37,5450	39,5354	41,1683	45,4813
$n = 250$	Inclinação (S)	95	38,0293	40,2616	41,9053	46,3569
		99	39,5861	41,4635	43,1131	48,0860
		90	7,5710	8,1784	8,5549	9,1938
$n = 250$	Curvatura (C_c)	95	7,8407	8,5217	8,8752	9,6579
		99	8,4221	9,1842	9,5273	10,2811
		90	52,2744	54,3806	57,0814	63,2683
$n = 500$	Inclinação (S)	95	52,9482	55,2024	57,9122	63,9850
		99	53,9177	56,8652	59,1083	65,7524
		90	6,2555	7,7850	8,2063	8,8109
$n = 500$	Curvatura (C_c)	95	6,4357	8,0209	8,4759	9,1114
		99	6,7931	8,6533	8,7486	9,5384

Tabela 3.5.4: Estatísticas Individuais M_1.

Tamanho Amostral	Medida	Percentis (%) $\boldsymbol{\rho}$	0,5	0,6	0,7	0,9
		90	0,2229	0,2219	0,2195	0,2206
$n = 250$	Inclinação (S)	95	0,2334	0,2322	0,2284	0,2339
		99	0,2567	0,2663	0,2588	0,2559
		90	0,2200	0,2199	0,2185	0,2205
$n = 250$	Curvatura (C_c)	95	0,2315	0,2313	0,2289	0,2341
		99	0,2510	0,2575	0,2623	0,2553
		90	0,1648	0,1644	0,1654	0,1655
$n = 500$	Inclinação (S)	95	0,1740	0,1723	0,1719	0,1724
		99	0,1891	0,1862	0,1863	0,1923
		90	0,1653	0,1651	0,1649	0,1653
$n = 500$	Curvatura (C_c)	95	0,1747	0,1717	0,1718	0,1720
		99	0,1880	0,1823	0,1874	0,1926

Tabela 3.5.5: Estatísticas Individuais M_2.

Tamanho Amostral	Medida	Percentis (%) $\boldsymbol{\rho}$	0,5	0,6	0,7	0,9
		Mínimo	0,1533	0,1517	0,1586	0,1470
		Quantil 5%	0,1588	0,1667	0,1633	0,1662
$n = 250$	Inclinação (S)	Máximo	0,2324	0,2596	0,2379	0,2338
		Média	0,1913	0,1924	0,1877	0,1906
		Desvio Padrão	0,0208	0,0210	0,0156	0,0202
		Mínimo	0,1553	0,1408	0,1526	0,1476
		Quantil 5%	0,1637	0,1576	0,1627	0,1647
$n = 250$	Curvatura (C_c)	Máximo	0,2510	0,2414	0,2333	0,2336
		Média	0,1930	0,1888	0,1880	0,1880
		Desvio Padrão	0,0221	0,0225	0,0170	0,0187
		Mínimo	0,1141	0,1226	0,1202	0,1221
		Quantil 5%	0,1285	0,1289	0,1236	0,1294
$n = 500$	Inclinação (S)	Máximo	0,1880	0,1811	0,2073	0,2181
		Média	0,1441	0,1451	0,1458	0,1454
		Desvio Padrão	0,0150	0,0131	0,0164	0,0162
		Mínimo	0,1165	0,1205	0,1190	0,1231
		Quantil 5%	0,1225	0,1261	0,1251	0,1296
$n = 500$	Curvatura (C_c)	Máximo	0,1834	0,1702	0,1797	0,2193
		Média	0,1438	0,1450	0,1448	0,1465
		Desvio Padrão	0,0150	0,0128	0,0148	0,0178

Tabela 3.5.6: Influência global para S e C_c.

Tamanho Amostral	Medida	$\boldsymbol{\rho}$	0,5	0,6	0,7	0,9
$n = 250$	Inclinação (S)		**43,505**	**45,131**	**57,329**	**55,695**
$n = 250$	Curvatura (C_c)		**9,5322**	**10,791**	**15,102**	**13,561**

Em negrito estão representadas as estatísticas significativas para o percentil 5% de M_0.

Na Tabela 3.5.3, se correlação e percentil são mantidos fixos, os limiares de S baseados no critério global M_0 aumentam quando o tamanho amostral aumenta. Ao contrário, para C_c, mantendo-se constantes o percentil e ρ, as marcas de nível M_0 decrescem quando o tamanho amostral aumenta. Percebe-se também que para tamanhos amostrais e percentis fixos, se a correlação aumenta os limiares de S e C_c têm comportamento monótono crescente. Portanto, existem indícios de que os limiares para o critério global M_0 não são robustos dado a variação de ρ.

Na Tabela 3.5.4, fixando ρ e percentil, os limiares de S e C_c baseados no critério individual M_1 diminuem quando o tamanho amostral aumenta. No entanto, fixados n e percentil, os limiares para S e C_c permanecem praticamente constantes quando a correlação ρ aumenta, ou seja, há indícios de que o critério M_1 seja robusto quanto a variação de ρ.

Na Tabela 3.5.5, se o tamanho amostral varia de $n = 250$ para $n = 500$ com ρ constante, tanto os limiares baseados no critério individual M_2 quanto as demais estatísticas, têm suas magnitudes diminuídas. Agora, se a correlação aumenta, para percentil e tamanho amostral fixos, os limiares M_2 e as demais estatíticas de S e C_c permanecem praticamente inalteradas. Assim como M_1, o critério M_2 apresenta indícios de robustez quanto a variação do coeficiente de correlação ρ.

A Tabela 3.5.6 apresenta as estatísticas S e C_c, calculadas para as séries 1 - 4. Utilizando o critério M_0 (Tabela 3.5.3) para avaliar a influência local das quatro séries geradas, a Tabela 3.5.6 mostra que todas as estatísticas S e C_c são significativas. Por exemplo, para a série 1 o valor de S é 43,5 é maior que o limiar 38,0 baseado no critério global M_0 dado na Tabela 3.5.3. Assim, existem indícios de influência global, e isso será confirmado através dos vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$, em conjunção aos limiares M_1 e M_2, ilustrados nas Figuras 3.5.1 - 3.5.4, ou seja, para as séries 1 e 4. Em cada uma dessas figuras o gráfico superior corresponde à série sem *outlier*. O segundo gráfico corresponde à série em valores absolutos com *outlier*.

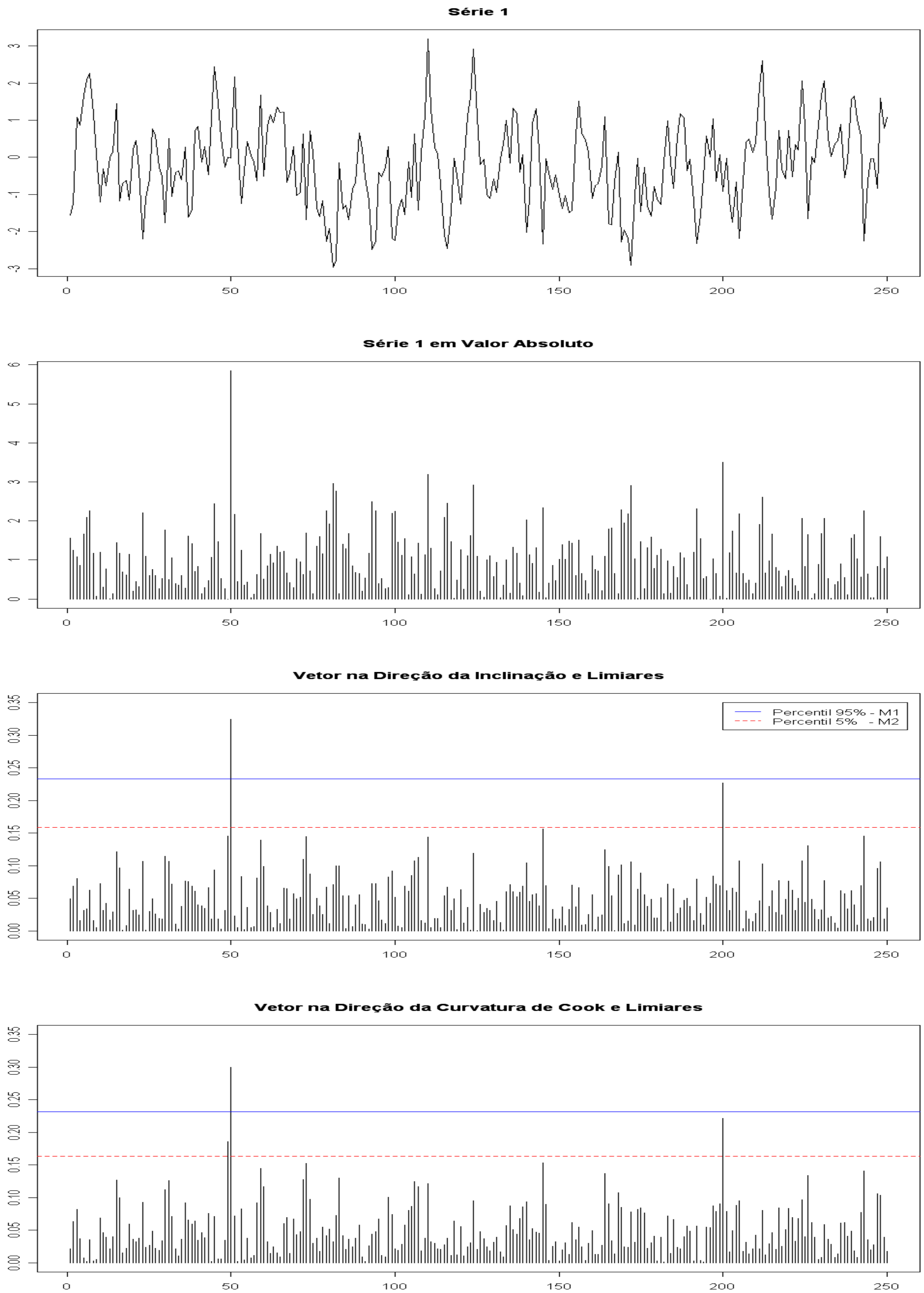

Figura 3.5.1: Influência local na série 1 ($\rho = 0,5$ e $n = 250$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

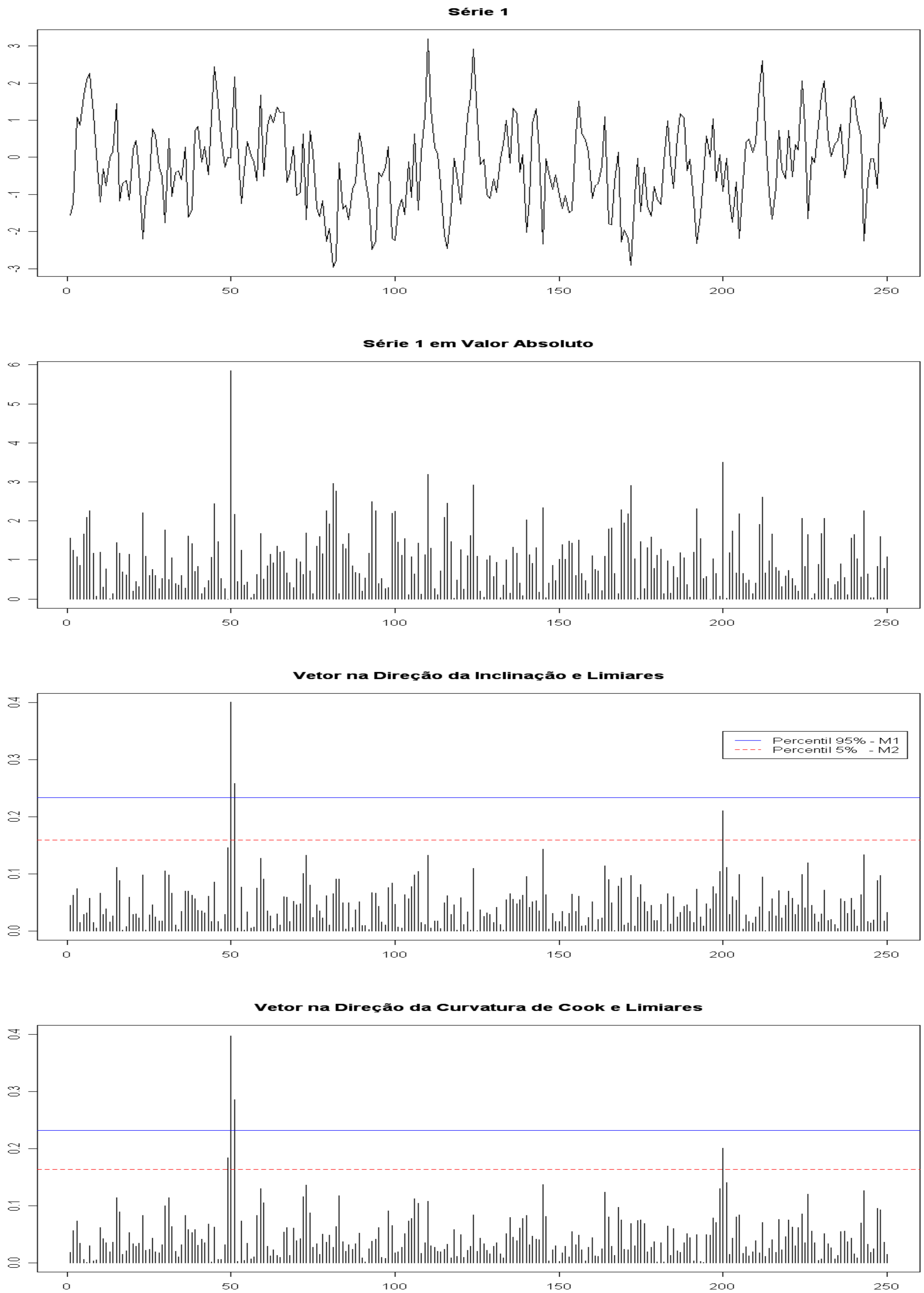

Figura 3.5.2: Influência local na série 1 ($\rho = 0,5$ e $n = 250$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

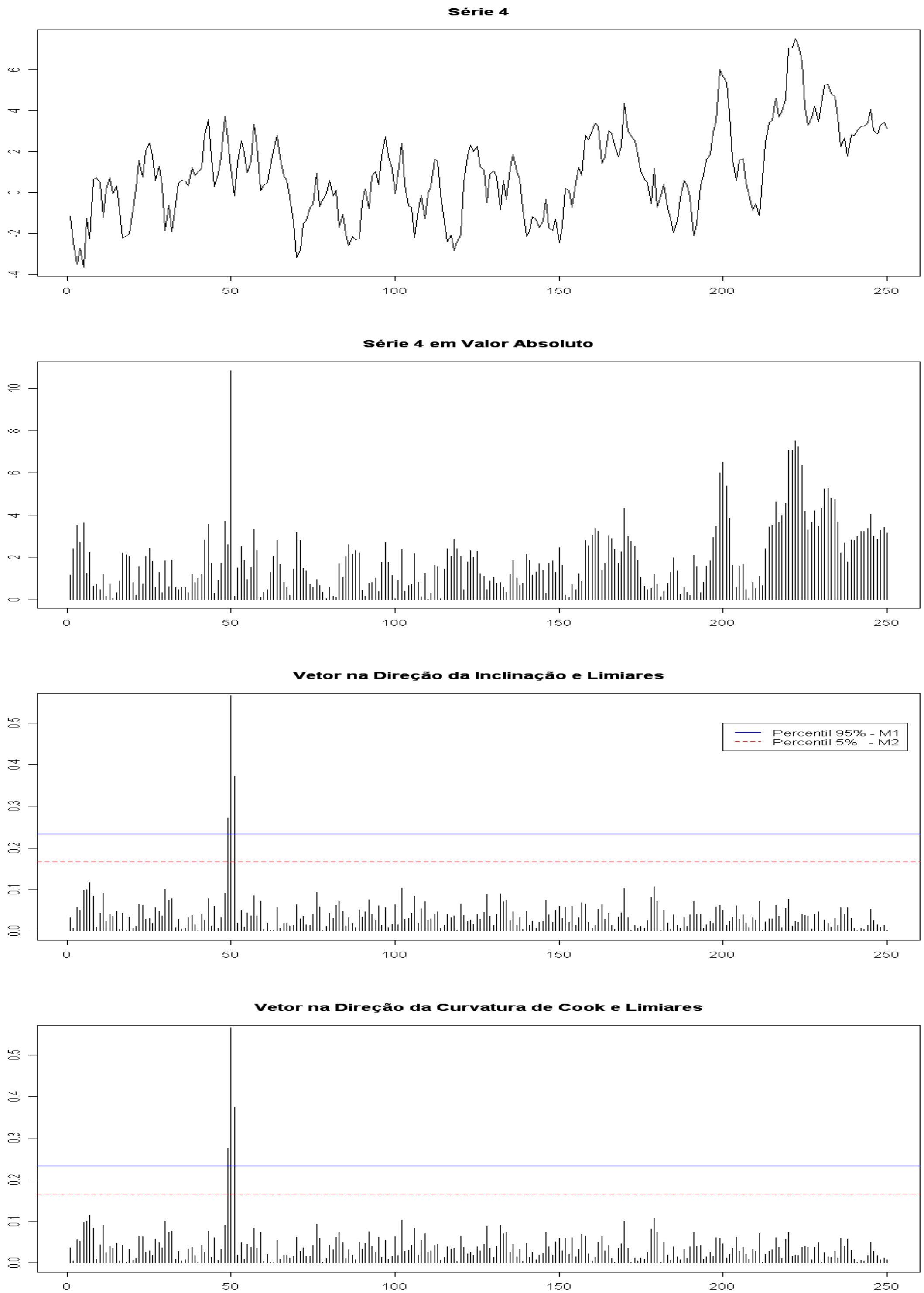

Figura 3.5.3: Influência local na série 4 ($\rho = 0,9$ e $n = 250$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

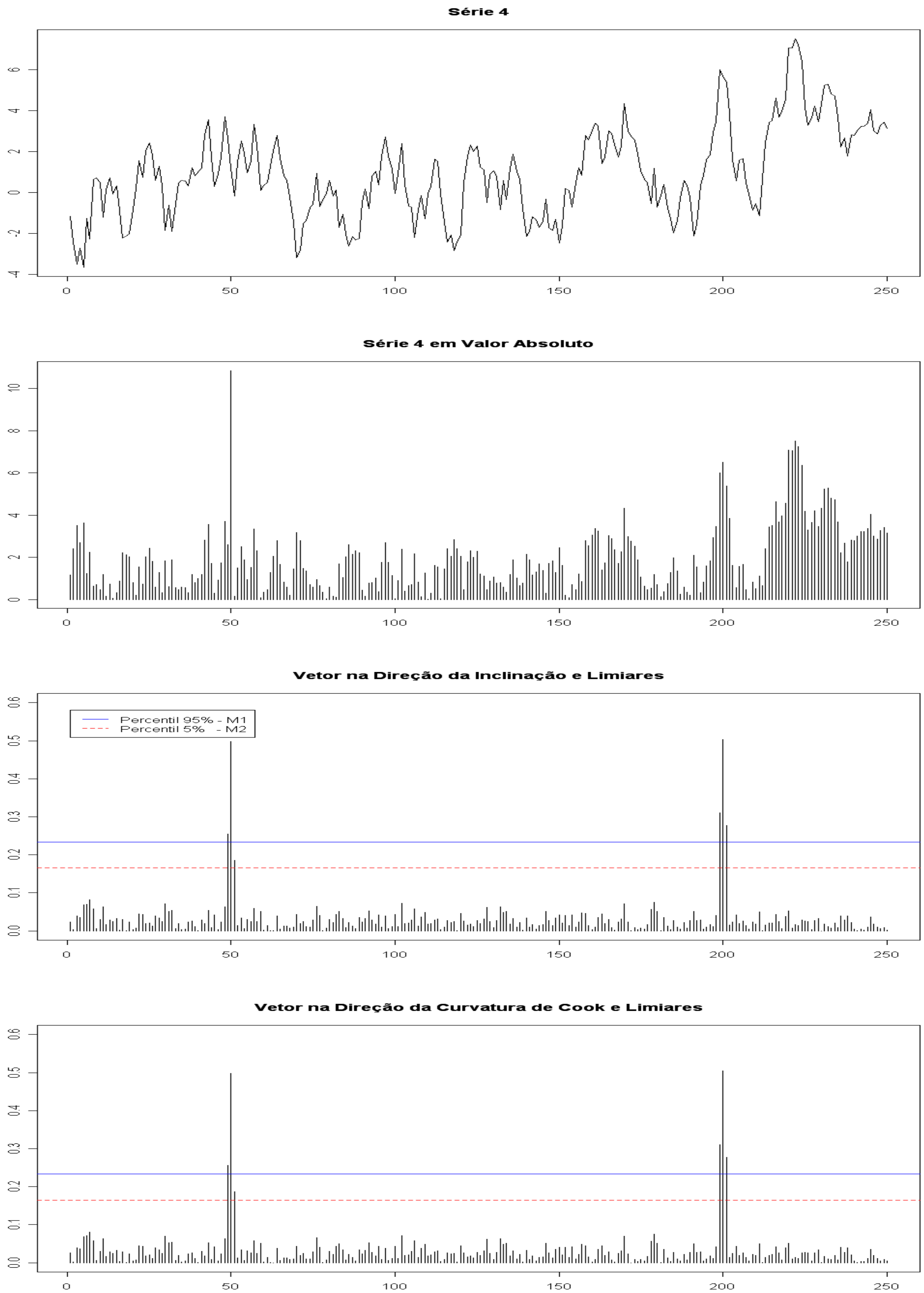

Figura 3.5.4: Influência local na série 4 ($\rho = 0,9$ e $n = 250$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

Para a série 1 ilustrada na Figura 3.5.1, com base nos vetores de diagnóstico de S e C_c e considerando a perturbação positiva $\tilde{y}_{50} = 5,843$, a componente 50 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada corretamente pelos critérios M_1 e M_2. Já componente 200 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada somente pelo critério M_2. Quando a mesma série é analisada com perturbações negativas, o comportamento das componentes 50 e 200 de $\mathbf{v}$ e $\mathbf{v}_c$ é idêntico ao observado na Figura 3.5.1, ou seja, y_{50} é uma observação influente sob os critérios M_1 e M_2, e y_{200} é uma observação influente sob o critério M_2.

Para a série 4 ilustrada na Figura 3.5.3, quando a perturbação é positiva, a componente 50 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada pelos critérios M_1 e M_2. Ao contrário, a componente 200 de $\mathbf{v}$ e $\mathbf{v}_c$ não é identificada por nenhum dos critérios individuais M_1 e M_2. Isso ocorre pois a perturbação $\omega_{200} = 0,806$ é muito pequena (ver Tabela 3.5.2). Para a mesma série, agora com perturbações negativas (Figura 3.5.4), a componente 50 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada pelos critérios M_1 e M_2. Desta vez a componente 200 também é identificada por M_1 e M_2, pois o valor absoluto da diferença entre $\tilde{y}_{200} = -6,502$ e $y_{200} = 5,695$ é maior do que o valor absoluto da diferença entre $\tilde{y}_{200} = 6,502$ (perturbação positiva) e $y_{200} = 5,695$ (ver Tabela 3.5.2).

A técnica de limiares é eficaz no seguinte sentido. Existem observações que não foram geradas como *outliers*, mas apresentam notada magnitude em relação à observação y_{200}, e mesmo assim, não são identificadas como influentes pelos critérios M_1 e M_2. Isto pode ser visto, através das observações y_{220}, y_{221}, y_{222}, y_{223} e y_{224} na série 4, e das respectivas componentes nos vetores $\mathbf{v}$ e $\mathbf{v}_c$ das Figuras 3.5.3 e 3.5.4.

Existe ainda, um "reflexo" no seguinte sentido. A observação anterior e/ou posterior, à observação detectada como influente, apresentam valores para os vetores de diagnóstico de S e C_c que as caracterizam como influentes, quando na verdade não foram geradas como tais. Esse "reflexo" está diretamente ligado a estrutura do modelo autoregressivo de ordem um (3.1.1) e ao sinal das observações inseridas como *outliers*. Assim, considerando a série 1 na Tabela 3.5.2, percebe-se que variando t de 48 a 49 a magnitude da observação y_t aumenta. Se t vai de 49 para 50 y_t praticamente se mantém constante. Se t vai de 50 para 51 y_t aumenta. E finalmente, se t vai de 51 para 52 y_t diminui. Quando se insere o *outlier* positivo $\tilde{y}_{50} = 5,843$, a dinâmica da série é alterada para um valor positivo muito maior do que $y_{50} = -0,020$, ocasionando uma "grande" variação entre $y_{49} = 0,001$ e $\tilde{y}_{50} = 5,843$. Assim, na Figura 3.5.1 a componente 49 de $\mathbf{v}_c$ é identificada pelo critério M_2. Quando se insere o *outlier* negativo $\tilde{y}_{50} = -5,843$ (Figura 3.5.2), a dinâmica da série é alterada para um valor negativo muito menor do que $y_{50} = -0,020$. Adicionalmente, como ocorre uma "grande" variação entre $y_{49} = 0,001$ e $\tilde{y}_{50} = -5,843$, e entre $y_{50} = -5,843$ e $\tilde{y}_{51} = 2,169$, a componente 49 de $\mathbf{v}_c$ é identificada por M_2 e a 51 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada por M_1 e M_2. Esse comportamento é o que entendemos como "reflexo".

O "reflexo" também pode ser observado através das componentes 49 e 51 de $\mathbf{v}$ e $\mathbf{v}_c$ das Figuras 3.5.3 e 3.5.4 e das componentes 199 e 201 de $\mathbf{v}$ e $\mathbf{v}_c$ na Figura 3.5.4. A exceção ocorre em $\mathbf{v}$ e $\mathbf{v}_c$ na Figura 3.5.3, pois como comentado anteriormente a perturbação $\omega_{200} = 0,806$ é muito pequena.

Em relação aos critérios individuais M_1 e M_2, o fato dos limiares apresentados nas Tabelas 3.5.4 e 3.5.5, respectivamente, não variarem tanto para diferentes valores de ρ, aponta indícios de que essas marcas de nível podem ser robustas quanto a estimação dos parâmetros. Deve-se observar que para gerar os limiares foram utilizados os mesmos parâmetros empregados nas simulações das séries temporais 1 - 4. No entanto, na prática não são conhecidos os valores "reais" dos parâmetros. Nesse sentido, faz-se interessante verificar o comportamento das marcas de nível em relação aos parâmetros estimados.

A Tabela 3.5.7 apresenta as estimativas de ρ e σ^2 para as séries 1 - 4.

Tabela 3.5.7: Estimativas dos parâmetros ρ e σ^2.

Séries	Tamanho Amostral	Parâmetros		Estimativas	
Série 1	$n = 250$	$\rho = 0,5$	$\sigma^2 = 1$	$\hat{\rho} = 0,4970(0,0548)$	$\hat{\sigma}^2 = 1,1943(0,1070)$
Série 2	$n = 250$	$\rho = 0,6$	$\sigma^2 = 1$	$\hat{\rho} = 0,4313(0,0572)$	$\hat{\sigma}^2 = 1,2135(0,1087)$
Série 3	$n = 250$	$\rho = 0,7$	$\sigma^2 = 1$	$\hat{\rho} = 0,6865(0,0458)$	$\hat{\sigma}^2 = 1,2689(0,1137)$
Série 4	$n = 250$	$\rho = 0,9$	$\sigma^2 = 1$	$\hat{\rho} = 0,8485(0,0338)$	$\hat{\sigma}^2 = 1,7202(0,1541)$

Entre parênteses se encontram os desvios padrões das estimativas dos parâmetros.

Será interessante verificar se diante da melhor situação em termos de estimação os limiares têm comportamentos similares às marcas de referências calculadas utilizando-se os verdadeiros valores dos parâmetros. Assim, as séries 1 - 4 foram escolhidas de forma que os parâmetros estimados estivessem próximos dos parâmetros. A exceção ocorre para a estimativa de $\rho = 0,6$, que é diferente do verdadeiro valor do parâmetro, e dessa forma, será verificada de que maneira uma estimativa afastada do real valor afeta o comportamento dos limiares.

Nas Tabelas 3.5.8 - 3.5.10 são apresentados os limiares calculados a partir dos verdadeiros parâmetros e os novos limiares gerados a partir das estimativas dos parâmetros apresentadas na Tabela 3.5.7. Adicionalmente foram calculados os limiares para as quatro séries geradas com tamanho amostral $n = 500$.

Tabela 3.5.8: Estatísticas Globais M_0.

Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,5	0,6	0,7	0,9
		90		37,5450	39,5354	41,1683	45,4813
$n = 250$	Inclinação (S)	95		38,0293	40,2616	41,9053	46,3569
		99		39,5861	41,4635	43,1131	48,0860
		90		7,5710	8,1784	8,5549	9,1938
$n = 250$	Curvatura (C_c)	95		7,8407	8,5217	8,8752	9,6579
		99		8,4221	9,1842	9,5273	10,2811
			$\hat{\boldsymbol{\rho}}$	0,4970	0,4313	0,6865	0,8485
			$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	1,1943	1,2135	1,2689	1,7202
		90		34,3105	33,0852	36,3080	33,8174
$n = 250$	Inclinação (S)	95		34,7531	33,6750	36,9651	34,5123
		99		36,1810	34,7355	38,0201	35,5968
		90		6,3256	5,9552	6,6925	5,2679
$n = 250$	Curvatura (C_c)	95		6,5506	6,1920	6,9539	5,5255
		99		7,0392	6,6168	7,4689	5,8998
Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,5	0,6	0,7	0,9
		90		52,2744	54,3806	57,0814	63,2683
$n = 500$	Inclinação (S)	95		52,9482	55,2024	57,9122	63,9850
		99		53,9177	56,8652	59,1083	65,7524
		90		6,2555	7,7850	8,2063	8,8109
$n = 500$	Curvatura (C_c)	95		6,4357	8,0209	8,4759	9,1114
		99		6,7931	8,6533	8,7486	9,5384
			$\hat{\boldsymbol{\rho}}$	0,4889	0,5393	0,6902	0,8500
			$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	1,1539	1,0277	1,1275	1,3422
		90		48,2862	52,4273	53,5607	53,0056
$n = 500$	Inclinação (S)	95		48,9691	53,0991	54,1960	53,7727
		99		50,4364	54,1421	55,6444	54,9023
		90		6,2807	7,3175	7,2476	6,4475
$n = 500$	Curvatura (C_c)	95		6,4752	7,5329	7,4845	6,6647
		99		6,9925	7,9550	7,9232	6,8982

Tabela 3.5.9: Estatísticas Individuais M_1.

Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,5	0,6	0,7	0,9
		90		0,2229	0,2219	0,2195	0,2206
$n = 250$	Inclinação (S)	95		0,2334	0,2322	0,2284	0,2339
		99		0,2567	0,2663	0,2588	0,2559
		90		0,2200	0,2199	0,2185	0,2205
$n = 250$	Curvatura (C_c)	95		**0,2315**	0,2313	0,2289	0,2341
		99		0,2510	0,2575	0,2623	0,2553
			$\hat{\boldsymbol{\rho}}$	0,4970	0,4313	0,6865	0,8485
			$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	1,1943	1,2135	1,2689	1,7202
		90		0,2229	0,2191	0,2191	0,2197
$n = 250$	Inclinação (S)	95		0,2334	0,2288	0,2285	0,2315
		99		0,2569	0,2540	0,2585	0,2564
		90		0,2202	0,2185	0,2184	0,2198
$n = 250$	Curvatura (C_c)	95		**0,2317**	0,2299	0,2293	0,2307
		99		0,2510	0,2608	0,2612	0,2550
Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,5	0,6	0,7	0,9
		90		0,1648	0,1644	0,1654	0,1655
$n = 500$	Inclinação (S)	95		0,1740	0,1723	0,1719	0,1724
		99		0,1891	0,1862	0,1863	0,1923
		90		0,1653	0,1651	0,1649	0,1653
$n = 500$	Curvatura (C_c)	95		0,1747	0,1717	0,1718	0,1720
		99		0,1880	0,1823	0,1874	0,1926
			$\hat{\boldsymbol{\rho}}$	0,4889	0,5393	0,6902	0,8500
			$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	1,1539	1,0277	1,1275	1,3422
		90		0,1643	0,1647	0,1636	0,1653
$n = 500$	Inclinação (S)	95		0,1723	0,1735	0,1706	0,1716
		99		0,1897	0,1872	0,1848	0,1889
		90		0,1653	0,1652	0,1633	0,1646
$n = 500$	Curvatura (C_c)	95		0,1712	0,1744	0,1700	0,1717
		99		0,1825	0,1884	0,1831	0,1896

Em negrito são comparados os limiares M_1 calculados com os parâmetros verdadeiros e estimados.

Tabela 3.5.10: Estatísticas Individuais M_2.

Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,5	0,6	0,7	0,9
		Mínimo		0,1533	0,1517	0,1586	0,1470
		Quantil 5%		0,1588	0,1667	0,1633	0,1662
$n = 250$	Inclinação (S)	Máximo		0,2324	0,2596	0,2379	0,2338
		Média		0,1913	0,1924	0,1877	0,1906
		Desvio Padrão		0,0208	0,0210	0,0156	0,0202
		Mínimo		0,1553	0,1408	0,1526	0,1476
		Quantil 5%		0,1637	0,1576	0,1627	0,1647
$n = 250$	Curvatura (C_c)	Máximo		0,2510	0,2414	0,2333	0,2336
		Média		0,1930	0,1888	0,1880	0,1880
		Desvio Padrão		0,0221	0,0225	0,0170	0,0187
			$\hat{\boldsymbol{\rho}}$	0,4970	0,4313	0,6865	0,8485
			$\hat{\boldsymbol{\sigma}}^2$	1,1943	1,2135	1,2689	1,7202
		Mínimo		0,1532	0,1507	0,1590	0,1604
		Quantil 5%		0,1588	0,1645	0,1624	0,1667
$n = 250$	Inclinação (S)	Máximo		0,2323	0,2348	0,2389	0,2524
		Média		0,1913	0,1924	0,1872	0,1914
		Desvio Padrão		0,0208	0,0172	0,0162	0,0186
		Mínimo		0,1552	0,1427	0,1521	0,1612
		Quantil 5%		0,1637	0,1694	0,1626	0,1679
$n = 250$	Curvatura (C_c)	Máximo		0,2510	0,2335	0,2330	0,2465
		Média		0,1930	0,1901	0,1878	0,1925
		Desvio Padrão		0,0221	0,0175	0,0172	0,0173
Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,5	0,6	0,7	0,9
		Mínimo		0,1141	0,1226	0,1202	0,1221
		Quantil 5%		0,1285	0,1289	0,1236	0,1294
$n = 500$	Inclinação (S)	Máximo		0,1880	0,1811	0,2073	0,2181
		Média		0,1441	0,1451	0,1458	0,1454
		Desvio Padrão		0,0150	0,0131	0,0164	0,0162
		Mínimo		0,1165	0,1205	0,1190	0,1231
		Quantil 5%		0,1225	0,1261	0,1251	0,1296
$n = 500$	Curvatura (C_c)	Máximo		0,1834	0,1702	0,1797	0,2193
		Média		0,1438	0,1450	0,1448	0,1465
		Desvio Padrão		0,0150	0,0128	0,0148	0,0178
			$\hat{\boldsymbol{\rho}}$	0,4889	0,5393	0,6902	0,8500
			$\hat{\boldsymbol{\sigma}}^2$	1,1539	1,0277	1,1275	1,3422
		Mínimo		0,1196	0,1161	0,1213	0,1193
		Quantil 5%		0,1266	0,1282	0,1254	0,1228
$n = 500$	Inclinação (S)	Máximo		0,1843	0,1872	0,1927	0,1728
		Média		0,1447	0,1443	0,1493	0,1446
		Desvio Padrão		0,0146	0,0155	0,0170	0,0134
		Mínimo		0,1188	0,1174	0,1214	0,1199
		Quantil 5%		0,1247	0,1224	0,1241	0,1237
$n = 500$	Curvatura (C_c)	Máximo		0,1702	0,1823	0,1813	0,1752
		Média		0,1443	0,1442	0,1466	0,1456
		Desvio Padrão		0,0135	0,0156	0,0159	0,0136

Na Tabela 3.5.8 no caso de $\hat{\rho}$, para n e percentil fixos não é verificado o comportamento monótono crescente dos limiares sob o critério M_0, a diferença do que acontece quando os limiares são calculados através de ρ. Assim, as estimativas dos parâmetros estão afetando significativamente os cálculos dos limiares sob o critério M_0. Especificamente, considerando que as estimativas de ρ estão próximas aos verdadeiros valores, os limiares dependem bastante da estimação de σ^2.

Para as séries 1-4, avaliando-se as medidas globais S e C_c (Tabela 3.5.6) a partir dos limiares M_0 (Tabela 3.5.8) calculados com as estimativas dos parâmetros, percebe-se que em todos os casos existe significância quanto a influência global. Isso ocorre devido ao decréscimo dos limiares calculados a partir da estimativa dos parâmetros (Tabela 3.5.8), quando comparados aos limiares calculados a partir dos verdadeiros valores dos parâmetros. Pode-se inferir que, devido a não robustez de M_0 em relação à variação de ρ e à estimação dos parâmetros, os limiares para critério global considerando a Inclinação de Billor e Loynes e para a Curvatura de Cook não são ferramentas úteis na prática para avaliar a influência local.

Pelo contrário, os limiares para os critérios M_1 e M_2 calculados a partir das estimativas dos parâmetros, os quais são apresentados nas Tabelas 3.5.9 e 3.5.10, são idênticos aos resultados encontrados nas Tabelas 3.5.4 e 3.5.5, respectivamente. Um exemplo disso é apresentado na Tabela 3.5.9, onde o limiar $M_1 = 0,2315$ calculado a partir dos verdadeiros parâmetros (curvatura, $\rho = 0,5$ e $\sigma^2 = 1$) é idêntico ao limiar $M_1 = 0,2317$ calculado a partir dos parâmetros estimados (curvatura, $\hat{\rho} = 0,4970$ e $\hat{\sigma}^2 = 1,1943$). Assim, os limiares baseados nos critérios M_1 e M_2 são considerados robustos quanto a estimação dos parâmetros e portanto, constituem ferramentas muito úteis para avaliar a influência local.

Os gráficos relativos as análises de influência local, considerando os critérios individuais para os vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$ calculados com os parâmetros estimados, são idênticos aos gráficos dos vetores de diagnóstico calculados com os verdadeiros valores dos parâmetros. Em conseqüência, as conclusões de influência individual são as mesmas que no caso dos limiares M_1 e M_2 calculados com parâmetros verdadeiros.

3.5.2 Conclusões das Simulações

A partir dos experimentos de simulação, pode-se concluir que nas quatro séries o critério global M_0 calculdo com parâmetros verdadeiros indica a existência de observações influentes. Mas, quando são utilizadas as estimativas dos parâmetros, mesmo que $\hat{\rho}$ esteja próximo de ρ, os limiares M_0 são bastante afetados pela estimativa de σ^2. Em conseqüência, M_0 não seria uma ferramenta útil na prática.

Por outro lado, para as séries 1-4 se verifica a robustez das marcas de nível quanto à estimação dos parâmetros, considerando os critérios individuais M_1 e M_2. Esse fato é importante, pois permitirá o uso dos limiares M_1 e M_2, os quais são calculados, na prática, com as estimativas dos parâmetros. Portanto, M_1 e M_2 são úteis na prática.

É importante lembrar que devido a estrutura do modelo autoregressivo de primeira ordem (3.1.1) e pela maneira na qual são inseridos os *outliers* (positivos ou negativos), existem observações anteriores e/ou posteriores à observação detectada, que são consideradas influentes segundo os critérios M_1 e/ou M_2, quando na realidade não foram geradas como tais. Baseado nisso, quando a técnica de limiares for aplicada para avaliar a influência local, diante do modelo AR(1) e do esquema de perturbação nos dados, pode-se utilizar essas observações como guias para identificar os pontos realmente influentes.

A técnica não detecta corretamente observações grandes que não foram geradas como *outliers*. Portanto, pôde-se verificar que a análise de influência local através da técnica de limiares se torna uma poderosa ferramenta para identificar observações influentes.

3.6 Aplicações

Nesta seção será ilustrada a metodologia de influência local em dois conjuntos de dados reais. Considerando o esquema de perturbação nos dados e as medidas Inclinação de Billor e Loynes (1993) e Curvatura de Cook (1986), aplica-se a técnica de limiares proposta na Seção 2.3 para identificar observações influentes através do critério global M_0 e dos critérios individuais M_1 e M_2.

3.6.1 Vendas *versus* Despesas

O conjunto de dados desta subseção corresponde ao exemplo utilizado por Kim e Huggins (1997) e Tsai e Wu (1992), nos quais é ajustado um modelo de regressão linear com erros AR(1). Os dados foram obtidos da Tabela 9.1 de Montgomery e Peck (1982) que consistem de uma amostra de $n = 20$ observações consecutivas da concentração de vendas ($\mathbf{y}$) e despesas ($\mathbf{X}$) anuais para um determinado produto e região. O gráfico de dispersão é apresentado na Figura 3.6.1.

É ajustado o modelo $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$ (3.4.2), onde $\epsilon_t \sim AR(1)$, através do comando *arima* do *software R*. Os parâmetros estimados por máxima verossimilhança são: $\hat{\beta}_0 = 1609,743(23,732)$ $\hat{\beta}_1 = 0,0201(0,0003)$, $\hat{\rho} = 0,4322(0,2801)$ e $\hat{\sigma} = 18,3505$, onde as quantidades dentro dos parênteses são os erros padrões. Esses resultados mostram que $\hat{\beta}_0$ e $\hat{\beta}_1$ são altamente significativas, mas $\hat{\rho}$ não, o que é observado na Figura 3.6.1 através do gráfico das autocorrelações dos resíduos. Mesmo que a primeira autocorrelação não seja significativa, em princípio o modelo AR(1) será mantido para os resíduos com o objetivo de comparar os resultados obtidos nas análises de influência local realizadas por Kim e Huggins (1997) e Tsai e Wu (1992). Posteriormente, será realizado o diagnóstico de influência local considerando o modelo de regressão linear simples, ou seja, $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$, onde $\epsilon_t \sim NID(0, \sigma^2)$. Neste caso, as estimativas dos parâmetros são: $\hat{\beta}_0 = 1609(17,02)$ $\hat{\beta}_1 = 0,0201(0,0001)$ e $\hat{\sigma} = 19,984$.

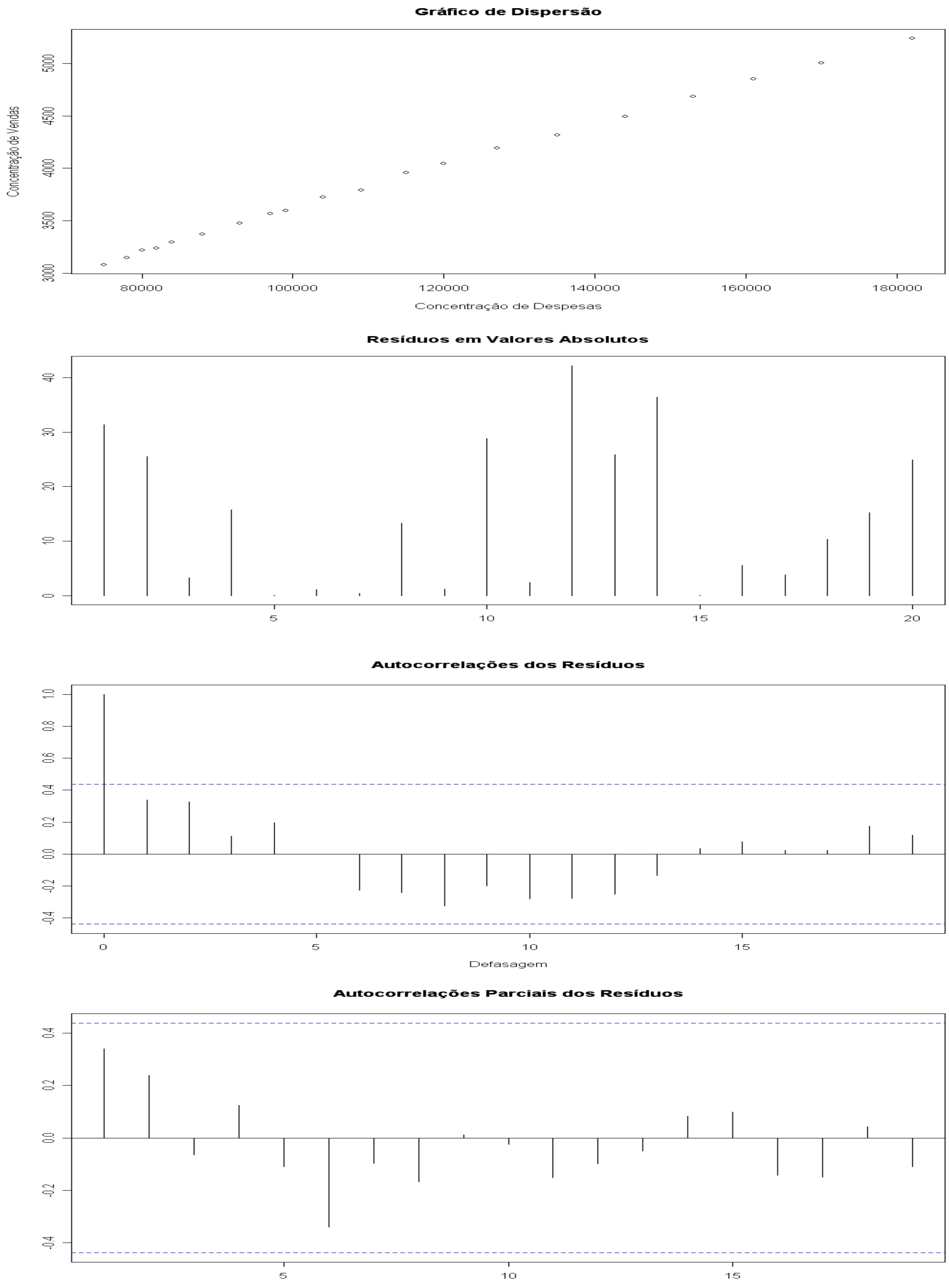

Figura 3.6.1: Gráfico de dispersão, resíduos, autocorrelações e autocorrelações parciais dos resíduos.

Considerando o esquema de perturbação nos dados e a verossimilhança exata (3.2.3) para avaliar a influência local, Kim e Huggins (1997) calculam o vetor de diagnóstico da Curvatura de Cook (1986) e comparam as respectivas componentes de $\mathbf{v}_c$ com as componentes das seguintes medidas: afastamento da verossimilhança perturbada $(2L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}(y_i \pm \hat{\sigma}))$ e influência conjunta $(2L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}(y_i \pm \hat{\sigma}, y_j \pm \hat{\sigma}))$. Comparando as "grandes" componentes dessas medidas com as "grandes" componentes de $\mathbf{v}_c$, o método de afastamento da verossimilhança perturbada identifica como pontos influentes, em ordem decrescente de magnitude, as observações y_{11}, y_{10}, y_3, y_{12} e y_9. O método da influência conjunta identifica como pontos influentes as observações y_{11} e y_{10}.

Por outro lado, através do esquema de perturbação nos dados e da verossimilhança perfilada com $\theta = \rho$, Tsai e Wu (1992) identificam as observações y_3, y_9, y_{11} e y_{13} como influentes comparando as "grandes" componentes obtidas pelo método de influência conjunta com as "grandes" componentes do vetor $\mathbf{v}$ associado à inclinação.

Resumindo, através de simples forma exploratória, Kim e Huggins (1997) e Tsai e Wu (1992) identificam como observações influentes aquelas com componentes dos vetores de inclinação ou curvatura que se destacam em relação as demais.

Nessa dissertação, para o esquema de perturbação nos dados, a metodologia de limiares busca investigar a influência local para o modelo de regressão linear com erros AR(1) e para o modelo de regressão linear simples, através de marcas de referência (limiares) que apresentem significância baseada nos critérios global e individual discutidos na Seção 2.3.

A série contém um número pequeno de observações, e por isso, faz-se interessante avaliar o comportamento dos resultados da análise de influência local, com relação às verossimilhanças exata, perfilada e condicional. Dessa maneira será possível verificar se o tamanho da amostra interfere no diagnóstico de influência para as diferentes verossimilhanças. Lembrando que a Curvatura de Cook e o vetor $\mathbf{v}_c$ calculados para a verossimilhança perfilada são baseados em $C_c(\sigma^2, \boldsymbol{\beta})$ expressado em (3.3.1). Nas Tabelas 3.6.1 e 3.6.2 são apresentados os valores das componentes $\mathbf{v}$ e $\mathbf{v}_c$, assim como os limiares M_0, M_1 e M_2 calculados através de 1000 replicações, conforme discutido na Seção 2.3.

Tabela 3.6.1: Resumo do diagnóstico de influência local para $\epsilon_t \sim AR(1)$.

		Tipos de Verossimilhanças					
		Exata		Perfilada		Condicional	
		$\mathbf{v}$	$\mathbf{v}_c$	$\mathbf{v}$	$\mathbf{v}_c$	$\mathbf{v}$	$\mathbf{v}_c$
Observações	y_{11}	0,3700	**0,5329**	0,3701	**0,5307**	0,3798	**0,5246**
Influentes	y_{12}	**0,4399**	0,3073	**0,4398**	0,3078	**0,4513**	0,3204
Estatísticas	S	0,5392	—	0,5392	—	0,5255	—
Globais	C_c	—	0,0245	—	0,0245	—	0,0221
Limiares	M_0	0,6918	0,0589	0,6884	0,0543	0,6540	0,0501
	M_1	0,6195	0,5976	0,6266	0,6140	0,6235	0,6116
	M_2	0,3957	0,3739	0,4010	0,3714	0,3998	0,3917

Legenda: Em negrito estão representadas as estatísticas significativas para os critérios M_0 e M_1 (95%) e M_2 (5%).

Tabela 3.6.2: Resumo do diagnóstico de influência local para $\epsilon_t \sim NID(0, \sigma^2)$.

		Tipos de Verossimilhanças			
		Exata		Condicional	
		$\mathbf{v}$	$\mathbf{v}_c$	$\mathbf{v}$	$\mathbf{v}_c$
Observações	y_{12}	**0,4595**	**0,4595**	**0,4949**	**0,4778**
Influentes	y_{14}	0,3896	**0,3896**	**0,4196**	**0,4123**
Estatísticas	S	0,4362	—	0,4050	—
Globais	C_c	—	0,0095	—	0,0087
Limiares	M_0	0,5545	0,0163	0,5431	0,0240
	M_1	0,6173	0,6105	0,6344	0,6232
	M_2	0,3979	0,3762	0,4165	0,4000

Legenda: Em negrito estão representadas as estatísticas significativas para os critérios M_0 e M_1 (95%) e M_2 (5%).

Considerando $\epsilon_t \sim AR(1)$ na Tabela 3.6.1, as medidas S e C_c não são significativas sob o critério global M_0. Sob o critério individual M_2 e utilizando as verossimilhanças exata, perfilada e condicional são identificadas as observações y_{11} e y_{12}.

Ainda na Tabela 3.6.1, as marcas de nível M_0, M_1 e M_2 são praticamente as mesmas considerando as diferentes verossimilhanças. Percebe-se também que os valores das componentes dos vetores $\mathbf{v}$ e $\mathbf{v}_c$ para y_{11} e y_{12} são bastante parecidos.

Considerando $\epsilon_t \sim NID(0, \sigma^2)$ na Tabela 3.6.2, as medidas S e C_c também não são significativas sob o critério global M_0. Utilizando as verossimilhanças exata e condicional são identificadas as observações y_{12} e y_{14} sob o critério individual M_2. Nota-se que os limiares M_0, M_1 e M_2 e os vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$, calculados através das verossimilhanças exata e condicional, são muito similares aos obtidos na Tabela 3.6.1.

As Figuras 3.6.2 - 3.6.6 ilustram o comportamento para os vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$ calculados a partir das verossimilhanças exata, perfilada e condicional com erros AR(1) e verossimilhanças exata e condicional com erros $NID(0, \sigma^2)$, respectivamente.

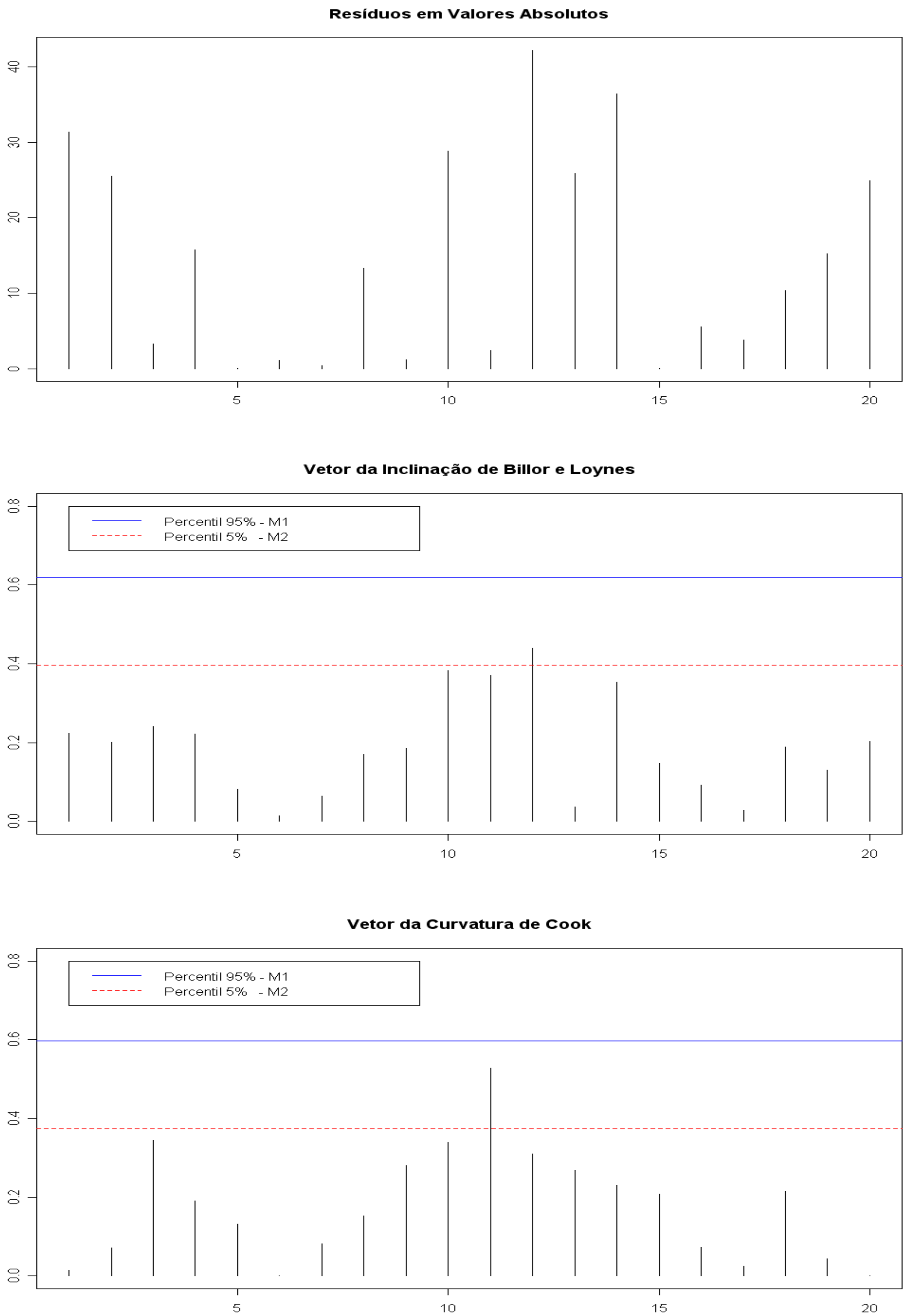

Figura 3.6.2: Influência local com verossimilhança exata e erros AR(1).

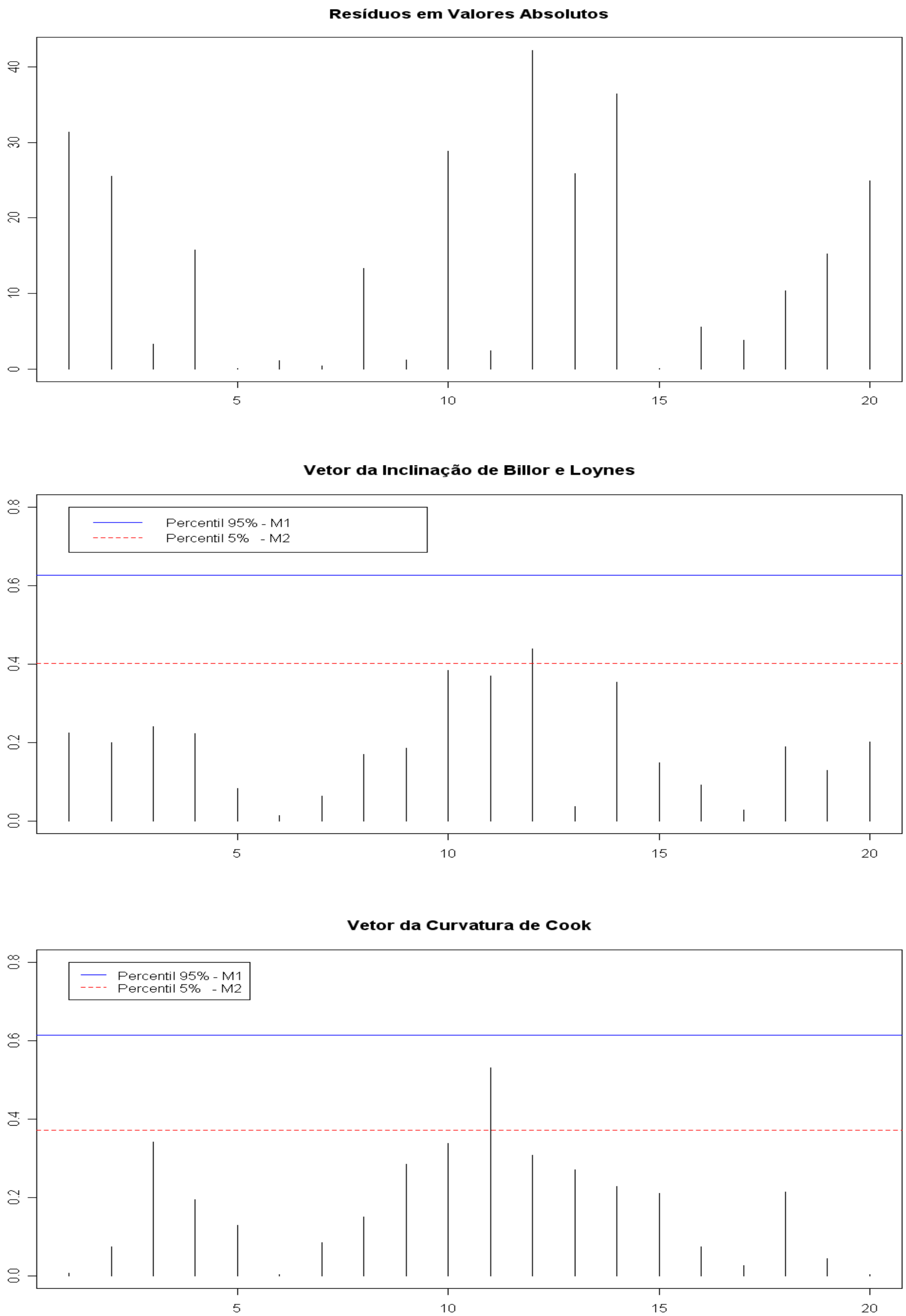

Figura 3.6.3: Influência local com verossimilhança perfilada e erros AR(1).

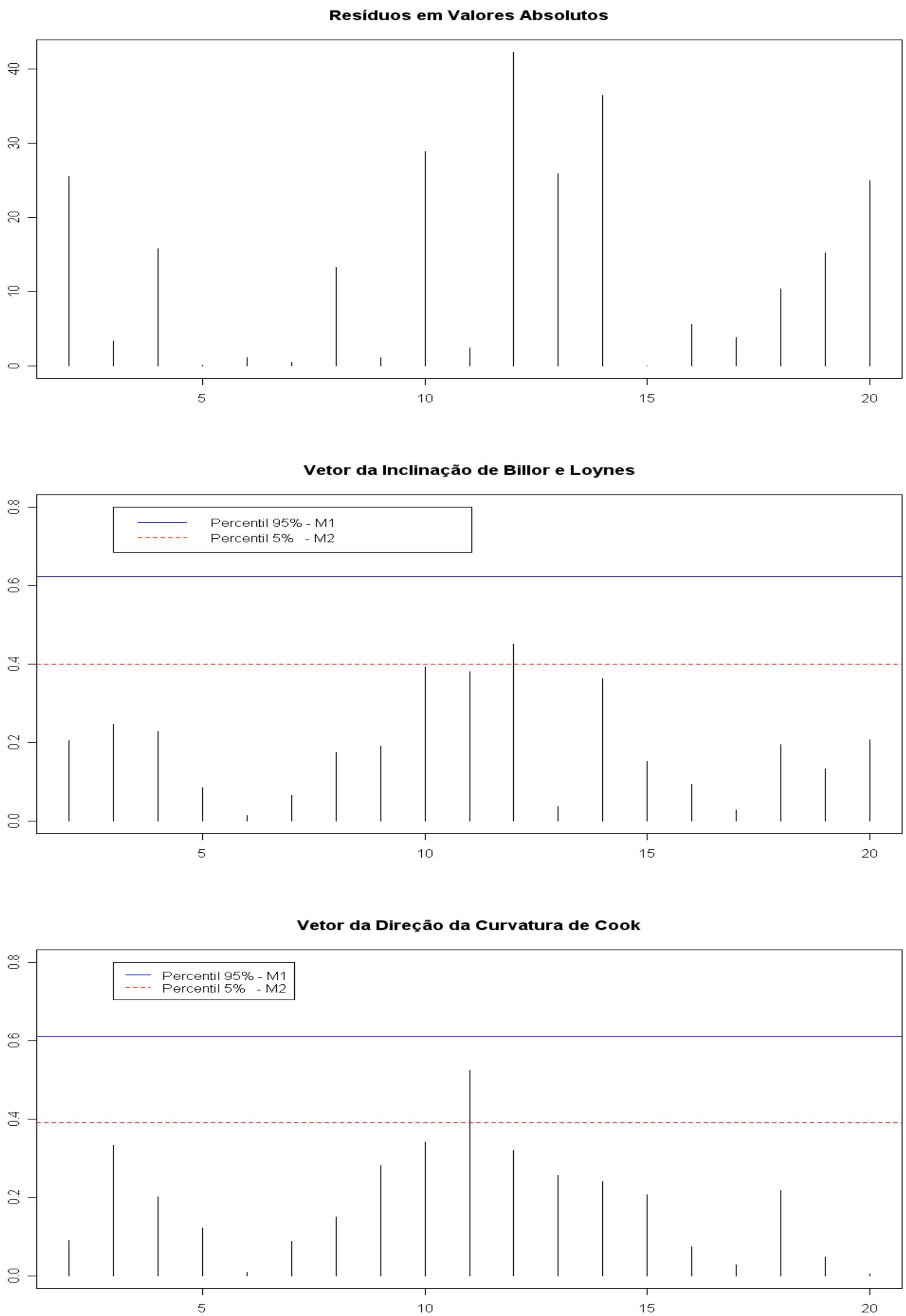

Figura 3.6.4: Influência local com verossimilhança condicional e erros AR(1).

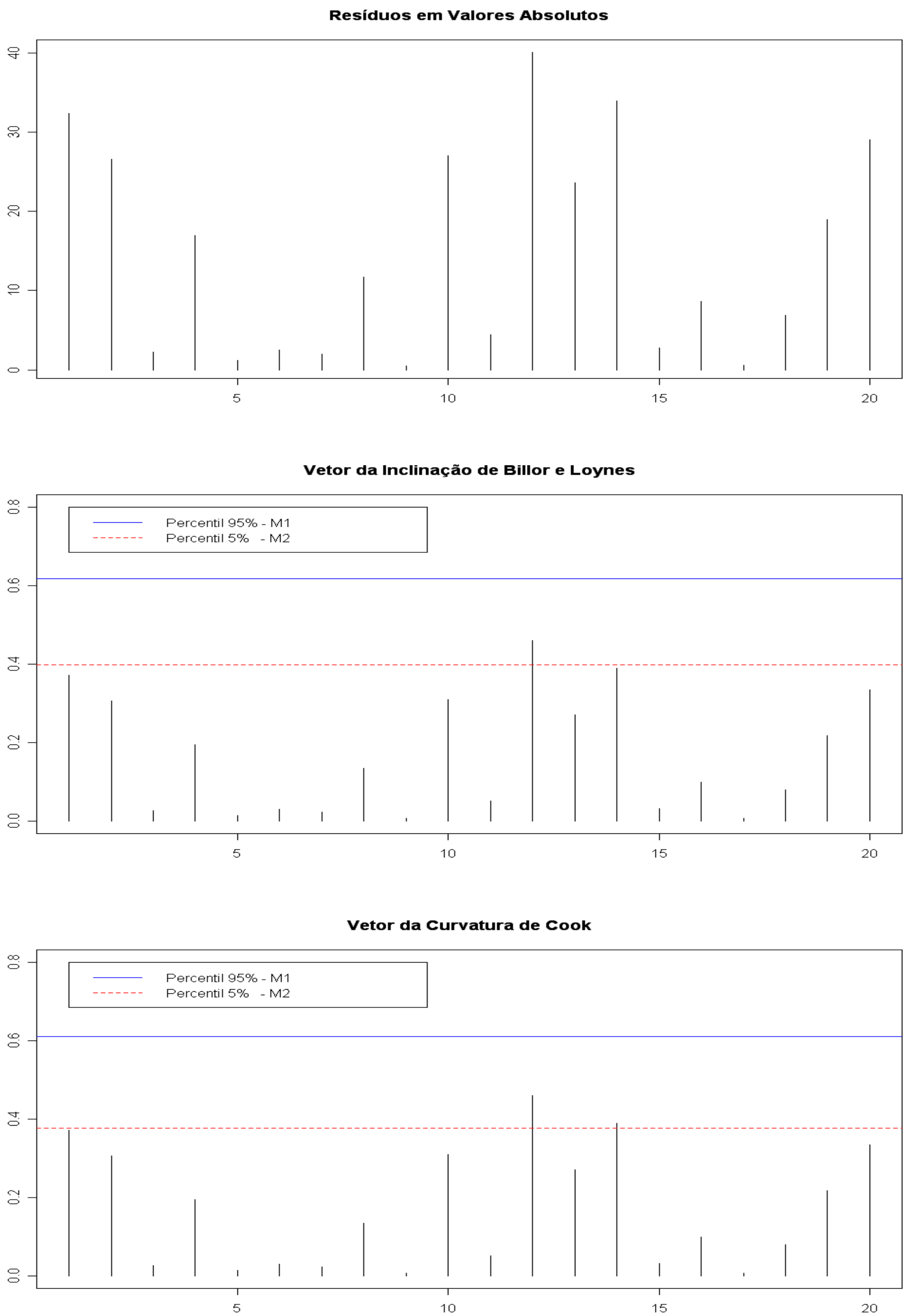

Figura 3.6.5: Influência local com verossimilhança exata e erros $NID(0, \sigma^2)$.

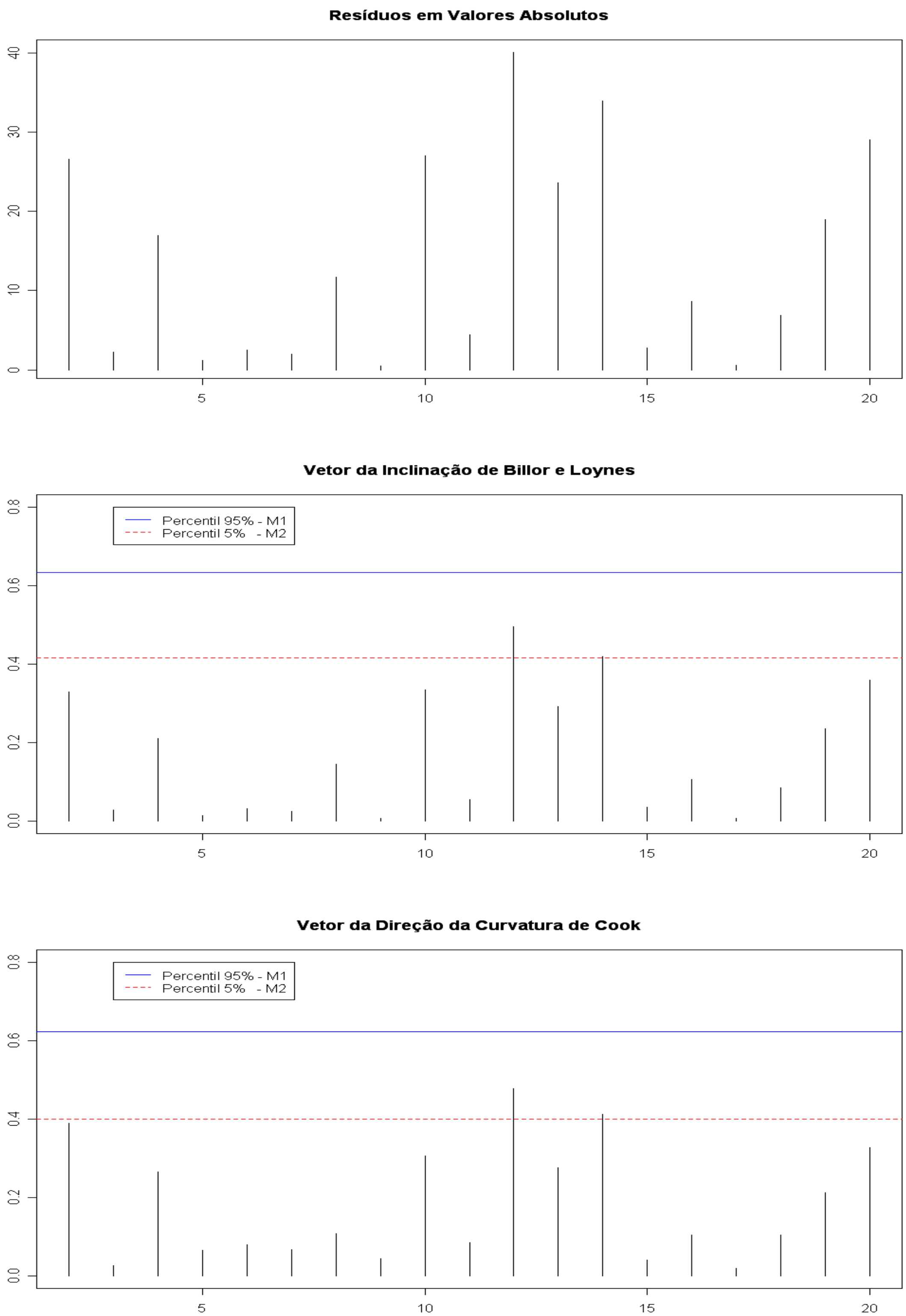

Figura 3.6.6: Influência local com verossimilhança condicional e erros $NID(0, \sigma^2)$.

Na Figura 3.6.2, percebe-se que as componentes 10, 11 e 14 de $\mathbf{v}$ estão próximas do nível M_2. Mas, as observações associadas a esses pontos não são consideradas significativas para influência local. Esse comportamento também é observado para as componentes 10, 11 e 14 dos vetores $\mathbf{v}$ e $\mathbf{v}_c$ nas Figuras 3.6.3 e 3.6.4 e para as componentes 1 e 2 dos vetores $\mathbf{v}_c$ das Figuras 3.6.5 e 3.6.6, respectivamente.

Comparando as magnitudes de $\mathbf{v}$ e $\mathbf{v}_c$ com as componentes dos resíduos na Figura 3.6.2 - 3.6.4, e conforme discutido para as simulações da Seção 3.5, o grande valor observado da componente 11 de $\mathbf{v}$ e $\mathbf{v}_c$ pode ter ocorrido como "reflexo" relativo aos resíduos e_{10} e e_{12}.

A única observação em comum identificada como influente por este trabalho, Kim e Huggins (1997) e Tsai e Wu (1992) é y_{11}. O que torna interessante avaliar a influência local através dos limiares é justamente a identificação das observações através de níveis de referência significativos. Por exemplo, a componente 3 do vetor da inclinação de Tsai e Wu (1992) e as componentes 3 e 10 do vetor da curvatura de Kim e Huggins (1997) apresentam "grandes" magnitudes quando comparadas aos demais pontos, e por esse motivo eles as identificam como influentes. Através das Figuras 3.6.2 a 3.6.4 também é possível perceber que e_{10} tem tamanho superior à maioria dos demais pontos. Mas, avaliadas sobre os critérios individuais M_1 e M_2, as componentes (10) dos vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$ associadas à e_{10}, e conseqüentemente à observação y_{10}, não são significativas para a influência local. Já as componentes 9 e 13 de $\mathbf{v}$ e $\mathbf{v}_c$ citadas pelos dois autores, não apresentam qualquer indícios de influência local pelos critérios M_1 e M_2. Esse trabalho identifica ainda a observação y_{12} como influente através do critério M_2, o que vai de acordo com o diagnóstico realizado por Kim e Huggins.

3.6.2 Índice de Preços ao Consumidor da Argentina

A série histórica estudada nesta subseção contém 80 retornos trimestrais do Índice de Preços ao Consumidor da Argentina (IPCA) e foi estudada por Franses (1998). O período que será estudado se refere ao primeiro quadrimestre (Q.1) de 1970 até o quarto quadrimestre (Q.4) de 1989. Como os retornos são calculados a partir da diferença dos logaritmos da série IPCA, então uma observação é perdida, ou seja, nesta aplicação serão estudados 79 observações referentes ao período do segundo quadrimestre de 1970 até o quarto quadrimestre de 1989. O período em questão se refere ao intervalo de tempo que culmina no auge da crise política e econômica de 1989 ocorrida na Argentina.

O modelo AR(1) utilizado para ajustar os referidos dados não considera a média, e dessa maneira, da série diferenciada dos logaritmos do IPCA é subtraída sua média amostral. Na Figura 3.6.7, observa-se que a série dos retornos para o índice IPCA apresenta indícios de estacionariedade. Observando a FAC (Função de Autocorrelação) e FACP (Função de Autocorrelação Parcial), tem-se que um modelo razoável é o AR(1) descrito por 3.1.1). Os parâmetros estimados pelo método de máxima verossimilhança são $\hat{\sigma} = 0,2188$ e $\hat{\rho} = 0,5583(0,0924)$, onde a quantidade dentro do parêntese é o erro padrão. O coeficiente de correlação é altamente significativo, o que é confirmado pelo gráfico das autocorrelações ilustrado na Figura 3.6.7.

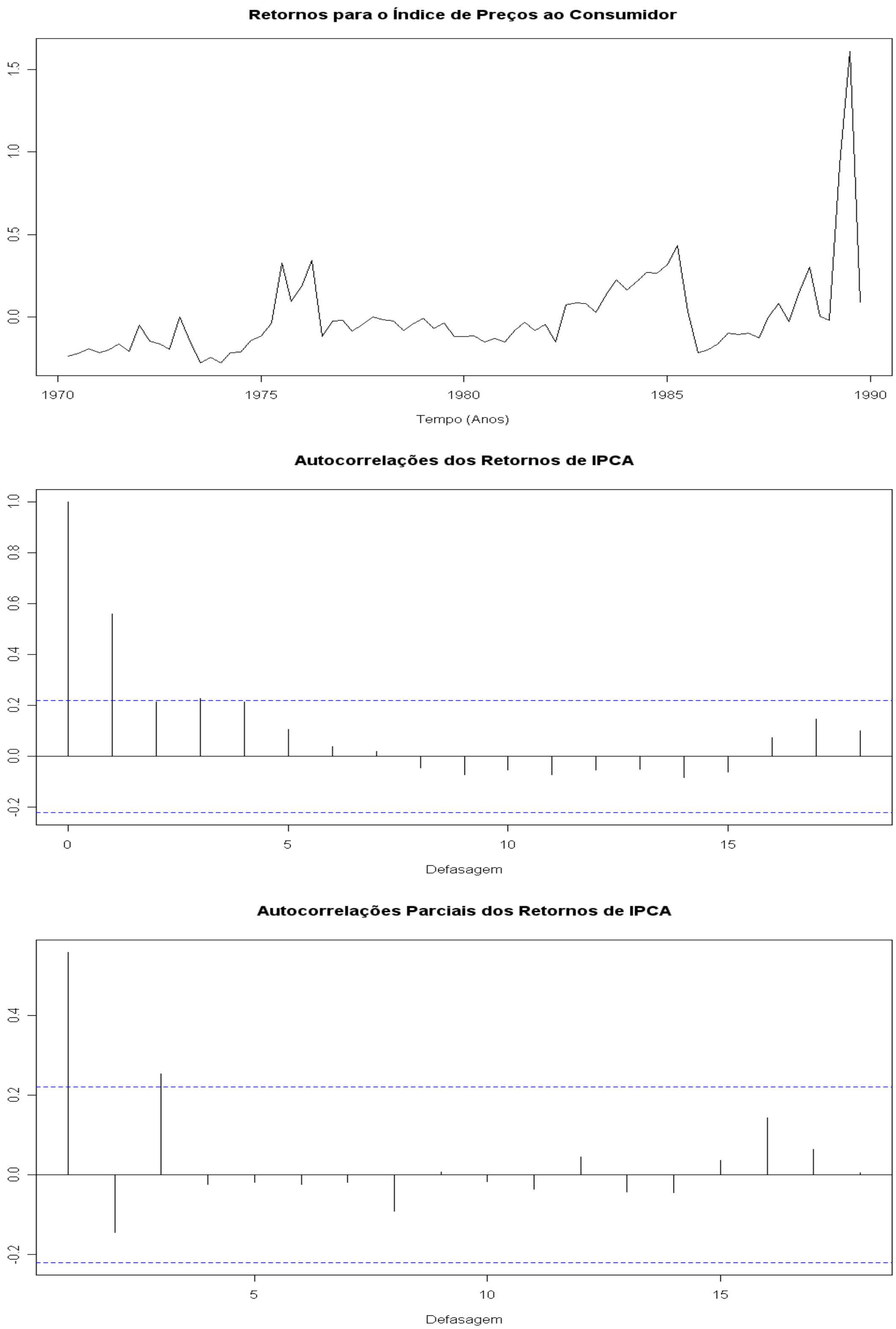

Figura 3.6.7: Retornos de IPCA, autocorrelações e autocorrelações parciais.

Baseado na metodologia proposta por Cook (1986) e Billor e Loynes (1993), na metodologia de limiares estudada na Seção 2.3, considerando o esquema de perturbação (3.4.4) e a verossimilhança exata será realizado o diagnóstico de influência local para o presente conjunto de dados.

A Tabela 3.6.3 apresenta um resumo para o diagnóstico de influência local através dos níveis de referência M_0, M_1 e M_2, considerando a verossimilhança exata.

Tabela 3.6.3: Resumo do diagnóstico da influência local.

Pontos Influentes	Data	$\mathbf{v}$	$\mathbf{v_c}$
76	1989(Q.1)	**0,2526**	**0,3088**
78	1989(Q.3)	**0,7184**	**0,6426**
79	1989(Q.4)	**0,3755**	**0,4837**
Estatísticas Globais	S	90,0002	—
	C_c	—	131,9085
Limiares	M_0	107,3218	198,9195
	M_1	0,3689	0,3621
	M_2	0,2363	0,2487

Legenda: Em negrito estão representadas as estatísticas significativas para os critérios M_0 e M_1 (95%) e M_2 (5%).

A Tabela 3.6.3 não apresenta nenhuma das medidas S ou C_c como significativas para o critério M_0, ou seja não há influência global. Mesmo assim, as componentes 76, 78 e 79 de $\mathbf{v}$ e $\mathbf{v}_c$, avaliadas sobre os critérios individuais M_1 e M_2, são identificadas como influentes, ou seja, as observações referentes ao primeiro, terceiro e quarto quadrimestres do ano de 1989, respectivamente.

A Figura 3.6.8 ilustra o comportamento de M_1 e M_2 perante os vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$, para a verossimilhança exata.

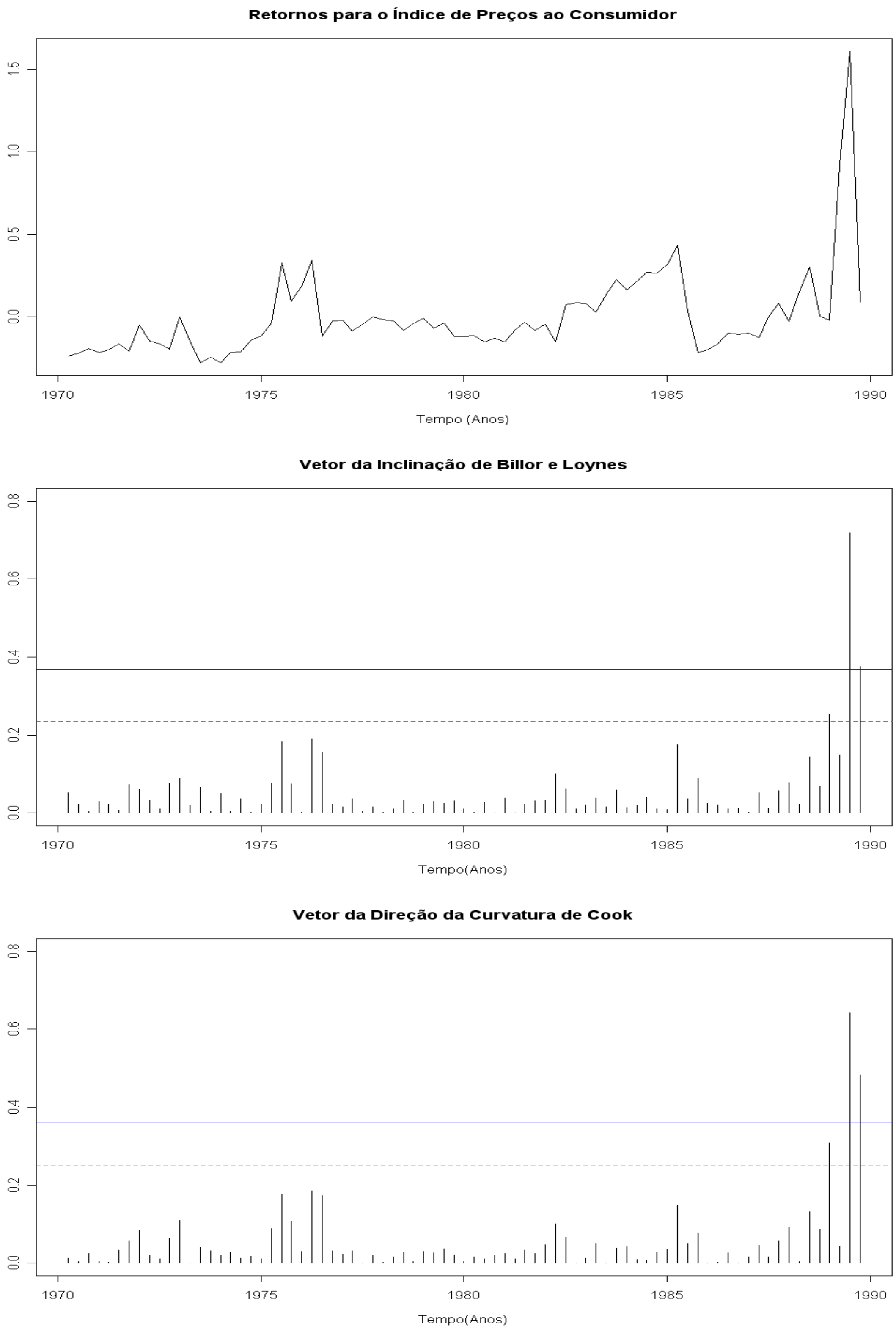

Figura 3.6.8: Influência local com verossimilhança exata.

Na Figura 3.6.8, as observações relativas ao terceiro e quarto quadrimestres de 1989 tem sua significância baseadas nos critérios M_1 e M_2 e a observação relativa ao primeiro quadrimestre de 1989 no critério M_2. Assim, ficam identificadas como influentes as observações referentes ao primeiro, terceiro e quarto quadrimestres do ano de 1989. Percebe-se que a observação relativa ao quarto quadrimestre do ano de 1989 tem magnitude "pequena" quando comparada as demais observações, e poderá ser um "reflexo" do retorno $e_{78} = 1,61$, pois a diferença entre o retornos e_{78} e $e_{79} = 0,09$ é "grande", e dessa maneira a dinâmica da série é quebrada.

Adicionamente, é importante notar que a análise de influência local através dos limiares individuais está identificando observações influentes no extremo da série, o que geralmente não é possível quando são utilizadas técnicas de detecção de *outliers*.

Em resumo, a técnica de limiares para avaliar a influência local para a série dos retornos de IPCA se torna uma ferramenta muito útil, pois detecta justamente os pontos referentes ao ápice da crise econômica e política de 1989 ocorrida na Argentina.

Como forma de avaliação do comportamento dos limiares considerando os critérios global e individual, serão excluídas as observações y_{78} e y_{79} referentes ao terceiro e quarto quadrimestres de 1989. A Tabela 3.6.4 apresenta os referidos resultados para M_0, M_1 e M_2 calculados a partir da verossimilhança exata.

Tabela 3.6.4: Resumo do diagnóstico de influência local.

Pontos		Sem 1989(Q.3) e 1989(Q.4)	
Influentes	Data	$\mathbf{v}$	$\mathbf{v}_c$
23	1975(Q.3)	**0,2648**	**0,2713**
26	1976(Q.2)	**0,2755**	**0,2820**
27	1976(Q.3)	0,2387	**0,2667**
76	1989(Q.1)	**0,4040**	**0,4742**
77	1989(Q.2)	**0,5521**	**0,4199**
Estatísticas Globais	S	139,5394	—
	C_c	—	332,146
Limiares	M_0	157,5214	416,0545
	M_1	0,3743	0,3742
	M_2	0,2498	0,2541

Legenda: Em negrito estão representadas as estatísticas significativas para os critérios M_0 e M_1 (95%) e M_2 (5%).

Assim, na Tabela 3.6.4, nenhuma das estatísticas globais são identificadas pelo critério M_0. Excluindo-se as observações y_{78} e y_{79}, são identificadas como influentes as observações y_{23}(1975 Q.3), y_{26}(1976 Q.2), y_{27}(1976 Q.3), y_{76} e y_{77}. Portanto, as observações 1989(Q.3) e 1989(Q.4) mascaram a influência local de outras observações.

3.7 Modelos de Regressão com Erros AR(1) e Inovações t-Student

Os indícios de caudas pesadas na distribuição das séries temporais em áreas como, por exemplo, hidrologia e finanças sugerem utilizar a distribuição t-Student para as inovações do modelo autoregressivo de primeira ordem. Na recente literatura, Medeiros (2006) discute métodos de diagnóstico (resíduos, pontos de alavanca e influência local) para o modelo de regressão simétrico com erros AR(1), e no que diz respeito a influência local, são propostas as perturbações aditiva na variável resposta e aditiva na i-ésima variável explicativa. Nesta seção, pretende-se estudar a influência local no modelo de regressão AR(1) com inovações t-Student, através do esquema de perturbação nos dados, das medidas de influência Inclinação de Billor e Loynes e Curvatura de Cook. Para tanto, a técnica de limiares discutida na Seção 2.3 será utilizada para avaliar séries temporais simuladas e um conjunto de dados reais.

3.7.1 Verossimilhança Condicional

Seja $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$ o vetor de uma amostra gerada através do modelo AR(1) dado por

$$\epsilon_t = \rho\epsilon_{t-1} + \sigma\eta_t, \tag{3.7.1}$$

onde ρ é a primeira autocorrelação, σ^2 é um parâmetro de dispersão, η_t tem distribuição t-Student padrão $\eta_t \sim t(0,1)$, ou seja, $E[\eta_t] = 0$, $Var[\eta_t] = 1$, considerando o parâmetro ν graus de liberdade.

Seja $\mathbf{y} = (y_1, \ldots, y_n)^T$ um vetor de respostas gerado por um modelo de regressão linear escrito na forma

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t, \tag{3.7.2}$$

onde os resíduos ϵ_t seguem um processo autoregressivo de ordem um, tal como em (3.7.1).

A função de densidade de probabilidade condicionada a $F_{t-1} = y_{t-1}, \ldots, y_1$ é dada por

$$f(y_t|F_{t-1}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\Gamma\left(\frac{\nu}{2}\right)\left(\pi(\nu-2)\right)^{1/2}} \left\{1 + \frac{(y_t - \beta_0 - \beta_1 x_t - \rho\left(y_{t-1} - \beta_0 - \beta_1 x_{t-1}\right))^2}{(\nu-2)\sigma^2}\right\}^{-\frac{\nu+1}{2}},$$

onde ν são os graus de liberdade, $E[y_t|F_{t-1}] = \rho\left(y_{t-1} - \beta_0 - \beta_1 x_{t-1}\right)$, $Var[y_t|F_{t-1}] = \sigma^2$ e $\Gamma(\cdot)$ é a função *Gamma*. Assim, verossimilhança condicionada na primeira observação é dada por

$$\begin{aligned} L(\boldsymbol{\theta}) \;\cong\; & (n-1)log\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\Gamma\left(\frac{\nu}{2}\right)\left(\pi(\nu-2)\right)^{1/2}}\right] - \\ & - \left(\frac{\nu+1}{2}\right)\sum_{t=2}^{n} log\left[1 + \frac{(y_t - \beta_0 - \beta_1 x_t - \rho\left(y_{t-1} - \beta_0 - \beta_1 x_{t-1}\right))^2}{(\nu-2)\sigma^2}\right], \end{aligned} \tag{3.7.3}$$

onde $\boldsymbol{\theta} = (\rho, \sigma^2, \nu, \beta_0, \beta_1)^T$.

3.7.2 Influência Local

Considerando a perturbação nos dados (3.4.4), a log-verossimilhança condicional perturbada é dada por

$$
\begin{aligned}
L(\boldsymbol{\theta}|\boldsymbol{\omega})) &= (n-1)log\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\Gamma\left(\frac{\nu}{2}\right)\left(\pi(\nu-2)\right)^{1/2}}\right]- \qquad (3.7.4)\\
&- \left(\frac{\nu+1}{2}\right)\sum_{t=2}^{n} log\left[1+\frac{(y_t+\omega_t-\beta_0-\beta_1 x_t-\rho\,(y_{t-1}+\omega_{t-1}-\beta_0-\beta_1 x_{t-1}))^2}{(\nu-2)\sigma^2}\right],
\end{aligned}
$$

onde $\boldsymbol{\theta}=(\rho,\sigma^2,\nu,\beta_0,\beta_1)^T$ é o vetor de parâmetros.

Da mesma forma que na Subseção 3.3.3, o presente trabalho introduz as expressões relacionadas às medidas de influência local Inclinação de Billor e Loynes (S) e Curvatura de Cook (C_c), que são avaliadas sobre $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}=(\hat{\rho},\hat{\sigma}^2,\hat{\nu},\hat{\beta}_0,\hat{\beta}_1)^T$ (estimadores de máxima verossimilhança) e $\boldsymbol{\omega}=\boldsymbol{\omega}_0=\mathbf{0}$ (perturbação nula).

- S:

$$
\begin{aligned}
\frac{\partial L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\omega_i} &= -\left(\frac{\hat{\nu}+1}{2}\right)\sum_{t=2}^{n}\left\{\frac{2\frac{A_t}{(\hat{\nu}-2)\hat{\sigma}^2}(I_{[t=i]}-\hat{\rho}I_{[t=i+1]})}{\frac{(\hat{\nu}-2)\hat{\sigma}^2+A_t^2}{(\hat{\nu}-2)\hat{\sigma}^2}}\right\}\\
&= -(\hat{\nu}+1)\begin{cases}\frac{A_i}{(\hat{\nu}-2)\hat{\sigma}^2+A_i^2}-\hat{\rho}\frac{A_{i+1}}{(\hat{\nu}-2)\hat{\sigma}^2+A_{i+1}^2} & , \quad i=2,...,n-1\\ \frac{A_i}{(\hat{\nu}-2)\hat{\sigma}^2+A_i^2} & , \quad i=n,\end{cases}
\end{aligned}
$$

onde $A_i=e_i-\rho e_{i-1}$, onde $e_i=y_i-\hat{\beta}_0-\hat{\beta}_1 x_i$ e $I_{[\,]}$ a função indicadora.

- C_c:

$$\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\omega_i\partial\rho} = (\hat{\nu}+1)\begin{cases} \frac{e_{i-1}\left[(\hat{\nu}-2)\hat{\sigma}^2-A_i^2\right]}{[(\hat{\nu}-2)\hat{\sigma}^2+A_i^2]^2} - \hat{\rho}\frac{e_i\left[(\hat{\nu}-2)\hat{\sigma}^2-A_{i+1}^2\right]}{[(\hat{\nu}-2)\hat{\sigma}^2+A_{i+1}^2]^2} + \frac{A_{i+1}}{(\hat{\nu}-2)\hat{\sigma}^2+A_{i+1}^2}, \\ \frac{e_{i-1}\left[(\hat{\nu}-2)\hat{\sigma}^2-A_i^2\right]}{[(\hat{\nu}-2)\hat{\sigma}^2+A_i^2]^2},\ i=n, \end{cases}$$

$$\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\omega_i\partial\sigma^2} = (\hat{\nu}+1)(\hat{\nu}-2)\begin{cases} \frac{A_i}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_i^2\right]^2} - \hat{\rho}\frac{A_{i+1}}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_{i+1}^2\right]^2}, \\ \frac{A_i}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_i^2\right]^2},\ i=n, \end{cases}$$

$$\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\omega_i\partial\nu} = -\begin{cases} \frac{A_i}{(\hat{\nu}-2)\hat{\sigma}^2+A_i^2} - \hat{\rho}\frac{A_{i+1}}{(\hat{\nu}-2)\hat{\sigma}^2+A_{i+1}^2} - \\ -(\hat{\nu}+1)\hat{\sigma}^2\left\{\frac{A_i}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_i^2\right]^2} - \hat{\rho}\frac{A_{i+1}}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_{i+1}^2\right]^2}\right\}, \\ \\ \frac{A_i}{(\hat{\nu}-2)\hat{\sigma}^2+A_i^2} - (\hat{\nu}+1)\hat{\sigma}^2\frac{A_i}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_i^2\right]^2},\ i=n, \end{cases}$$

$$\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\omega_i\partial\beta_0} = (\hat{\nu}+1)(1-\hat{\rho})\begin{cases} \frac{\left[(\hat{\nu}-2)\hat{\sigma}^2-A_i^2\right]}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_i^2\right]^2} - \hat{\rho}\frac{\left[(\hat{\nu}-2)\hat{\sigma}^2-A_{i+1}^2\right]}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_{i+1}^2\right]^2}, \\ \\ \frac{\left[(\hat{\nu}-2)\hat{\sigma}^2-A_i^2\right]}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_i^2\right]^2},\ i=n, \end{cases}$$

$$\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\omega_i\partial\beta_1} = (\hat{\nu}+1)\begin{cases} \frac{(x_i-\hat{\rho}x_{i-1})\left[(\hat{\nu}-2)\hat{\sigma}^2-A_i^2\right]}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_i^2\right]^2} - \hat{\rho}\frac{(x_{i+1}-\hat{\rho}x_i)\left[(\hat{\nu}-2)\hat{\sigma}^2-A_{i+1}^2\right]}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_{i+1}^2\right]^2}, \\ \\ \frac{(x_i-\hat{\rho}x_{i-1})\left[(\hat{\nu}-2)\hat{\sigma}^2-A_i^2\right]}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_i^2\right]^2},\ i=n. \end{cases}$$

As derivadas cruzadas entre os parâmetros são descritas a seguir.

$$
\begin{aligned}
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\rho\partial\rho} &= -(\hat{\nu}+1)\sum_{t=2}^{n}\frac{e_{t-1}^2\left[(\hat{\nu}-2)\hat{\sigma}^2-A_t^2\right]}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\sigma^2\partial\sigma^2} &= \frac{n-1}{2\hat{\sigma}^4}-\frac{(\hat{\nu}+1)}{2\hat{\sigma}^4}\sum_{t=2}^{n}\frac{A_t^2}{(\hat{\nu}-2)\hat{\sigma}^2+A_t^2}-\\
&- \frac{(\hat{\nu}+1)(\hat{\nu}-2)}{2\hat{\sigma}^2}\sum_{t=2}^{n}\frac{A_t^2}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2\right]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\nu\partial\nu} &= \left(\frac{n-1}{2}\right)\left\{\frac{\Gamma''\left(\frac{\hat{\nu}+1}{2}\right)}{2}-\frac{\Gamma''\left(\frac{\hat{\nu}}{2}\right)}{2}+\frac{1}{(\hat{\nu}-2)^2}\right\}+\\
&+ \frac{1}{2(\hat{\nu}-2)}\sum_{t=2}^{n}\left\{\frac{A_t^2}{(\hat{\nu}-2)\hat{\sigma}^2+A_t^2}\right\}-\frac{3}{2(\hat{\nu}-2)^2}\sum_{t=2}^{n}\left\{\frac{A_t^2}{(\hat{\nu}-2)\hat{\sigma}^2+A_t^2}\right\}-\\
&- \frac{(\hat{\nu}+1)\hat{\sigma}^2}{2(\hat{\nu}-2)}\sum_{t=2}^{n}\left\{\frac{A_t^2}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2}\right\}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\beta_0\partial\beta_0} &= -(\hat{\nu}+1)(1-\hat{\rho})^2\sum_{t=2}^{n}\frac{(\hat{\nu}-2)\hat{\sigma}^2-A_t^2}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2\right]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\beta_1\partial\beta_1} &= -(\hat{\nu}+1)\sum_{t=2}^{n}\frac{\left[x_t-\hat{\rho}x_{t-1}\right]^2\left[(\hat{\nu}-2)\hat{\sigma}^2-A_t^2\right]}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2\right]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\rho\partial\sigma^2} &= -(\hat{\nu}+1)(\hat{\nu}-2)\sum_{t=2}^{n}\frac{e_{t-1}A_t}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\rho\partial\nu} &= \sum_{t=2}^{n}\frac{e_{t-1}A_t}{(\hat{\nu}-2)\hat{\sigma}^2+A_t^2}-(\hat{\nu}+1)\hat{\sigma}^2\sum_{t=2}^{n}\frac{e_{t-1}A_{t-1}}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\rho\partial\beta_0} &= -(\hat{\nu}+1)(1-\hat{\rho})\sum_{t=2}^{n}\frac{e_{t-1}[(\hat{\nu}-2)\hat{\sigma}^2-A_t^2]}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\rho\partial\beta_1} &= -(\hat{\nu}+1)\sum_{t=2}^{n}\frac{e_{t-1}\left(x_t-\hat{\rho}x_{t-1}\right)\left[(\hat{\nu}-2)\hat{\sigma}^2-A_t^2\right]}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\sigma^2\partial\nu} &= \frac{1}{2\hat{\sigma}^2}\sum_{t=2}^{n}\frac{A_t^2}{(\hat{\nu}-2)\hat{\sigma}^2+A_t^2}-\frac{(\hat{\nu}+1)}{2}\sum_{t=2}^{n}\frac{A_t^2}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\sigma^2\partial\beta_0} &= -(\hat{\nu}+1)(1-\hat{\rho})(\hat{\nu}-2)\sum_{t=2}^{n}\frac{A_t}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\sigma^2\partial\beta_1} &= -(\hat{\nu}+1)(\hat{\nu}-2)\sum_{t=2}^{n}\frac{(x_t-\hat{\rho}x_{t-1})A_t}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\nu\partial\beta_0} &= (1-\hat{\rho})\sum_{t=2}^{n}\frac{A_t}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]} - (\hat{\nu}+1)(1-\hat{\rho})\hat{\sigma}^2\sum_{t=2}^{n}\frac{A_t}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\nu\partial\beta_1} &= \sum_{t=2}^{n}\frac{(x_t-\hat{\rho}x_{t-1})A_t}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]} - (\hat{\nu}+1)\hat{\sigma}^2\sum_{t=2}^{n}\frac{(x_t-\hat{\rho}x_{t-1})A_t}{[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2]^2}\\
\frac{\partial^2 L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)}{\partial\beta_0\partial\beta_1} &= -(\hat{\nu}+1)(1-\hat{\rho})\sum_{t=2}^{n}\frac{(x_t-\hat{\rho}x_{t-1})\left[(\hat{\nu}-2)\hat{\sigma}^2-A_t^2\right]}{\left[(\hat{\nu}-2)\hat{\sigma}^2+A_t^2\right]^2},
\end{aligned}
$$

onde $\Gamma''(\cdot) = \frac{\partial^2 log(\Gamma(\cdot))}{\partial\nu\partial\nu}$.

3.7.3 Simulações

A seguir será apresentado o diagnóstico de influência local, baseado na técnica de limiares proposta na Seção 2.3, para séries temporais AR(1) com inovações t-Student geradas por simulação. A partir da verossimilhança (3.7.4), o objetivo das simulações é estudar o comportamento das marcas de referência para os critérios M_0, M_1 e M_2, considerando doze séries com diferentes combinações de correlações e graus de liberdade, com tamanhos amostrais $n = 250$.

As simulações são baseadas no modelo autoregressivo de primeira ordem (3.7.1). Doze séries temporais foram geradas considerando as combinações de parâmetros de correlação $\rho = 0,3$, $\rho = 0,5$, $\rho = 0,7$ e $\rho = 0,9$, e parâmetros graus de liberdade $\nu = 5$, $\nu = 7$ e $\nu = 12$, com $\sigma^2 = 1$. As séries que serão ilustradas pelos gráficos das Figuras 3.7.1 - 3.7.6 serão denominadas conforme a Tabela 3.7.1.

Considera-se o esquema de perturbação nos dados. Nesse sentido, uma vez simuladas as séries, foram introduzidas duas observações "aberrantes" (*outliers* positivos e negativos), com magnitudes de cinco e três desvios padrões (σ_y) nas componentes y_{50} e y_{200}, respectivamente, onde σ_y é o desvio padrão da amostra gerada.

Tabela 3.7.1: Séries Simuladas.

	Graus de Liberdade	Coeficiente de Correlação	Posição do *outlier*	Tamanho do *outlier* em Valor Absoluto
Série 1	$\nu = 5$	$\rho = 0,3$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 2	$\nu = 5$	$\rho = 0,5$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 3	$\nu = 5$	$\rho = 0,7$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 4	$\nu = 5$	$\rho = 0,9$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 5	$\nu = 7$	$\rho = 0,3$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 6	$\nu = 7$	$\rho = 0,5$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 7	$\nu = 7$	$\rho = 0,7$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 8	$\nu = 7$	$\rho = 0,9$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 9	$\nu = 12$	$\rho = 0,3$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 10	$\nu = 12$	$\rho = 0,5$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 11	$\nu = 12$	$\rho = 0,7$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 12	$\nu = 12$	$\rho = 0,9$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$

Para as séries simuladas é realizada a análise de influência local utilizando os limiares descritos na Seção 2.3, e assim, espera-se identificar as observações geradas como influentes (*outliers*). Os limiares foram calculados em princípio supondo conhecidos os parâmetros do modelo e gerados com um programa escrito nos *softwares MATLAB* e *R* utilizando 1000 replicações.

Somente as séries 2, 7 e 12 serão analisadas através da técnica de limiares, e para tanto, a Tabela 3.7.2 apresenta o esquema de perturbação inserido nas respectivas séries, o que auxiliará a análise das observações influentes. Adicionalmente, nas Tabelas 3.7.3 - 3.7.5 são apresentadas as marcas de referência correspondentes ao critério global M_0 e aos critérios individuais M_1 e M_2, respectivamente, considerando a medida de influência local Curvatura de Cook. Como em nenhuma das séries simuladas a metodologia de limiares, realizada através da Inclinação de Billor e Loynes, foi capaz de identificar as observações "aberrantes", serão omitidas as tabelas relativas aos respectivos limiares M_0, M_1 e M_2.

Tabela 3.7.2: Perturbação positiva e negativa no modelo AR(1) - t-Student.

	y_{48}	y_{49}	y_{50}	y_{51}	y_{52}	ω_{50}	$\tilde{y}_{50} = 5\sigma_y$
Série 2	0,558	-0,268	0,663	0,862	0,138	4,295	4,958
Série 2	0,558	-0,268	0,663	0,862	0,138	-5,622	-4,958
Série 7	0,456	-0,397	-0,064	0,609	-0,825	6,496	6,432
Série 7	0,456	-0,397	-0,064	0,609	-0,825	-6,367	-6,432
Série 12	2,889	3,941	4,582	3,882	4,365	8,159	12,741
Série 12	2,889	3,941	4,582	3,882	4,365	-17,324	-12,741
	y_{198}	y_{199}	y_{200}	y_{201}	y_{202}	ω_{200}	$\tilde{y}_{200} = 3\sigma_y$
Série 2	-1,184	-1,079	-0,989	-2,498	-1,052	3,965	2,975
Série 2	-1,184	-1,079	-0,989	-2,498	-1,052	-1,985	-2,975
Série 7	-0,393	-0,988	-0,596	-0,688	-0,963	4,455	3,859
Série 7	-0,393	-0,988	-0,596	-0,688	-0,963	-3,263	-3,859
Série 12	2,099	3,459	3,127	3,347	1,484	4,517	7,645
Série 12	2,099	3,459	3,127	3,347	1,484	-10,772	-7,645

Tabela 3.7.3: Estatísticas Globais M_0.

Graus de Liberdade	Medida	Percentis (%) $\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
		90	18,5136	21,4247	24,4065	37,6396
$\nu = 5$	Curvatura (C_c)	95	36,0685	32,5487	45,8358	83,1595
		99	132,3213	158,5604	208,9140	428,3924
		90	20,0583	27,2735	33,3100	38,1278
$\nu = 7$	Curvatura (C_c)	95	37,8868	45,6139	59,7960	71,6737
		99	185,6951	201,7662	365,2938	397,3722
		90	27,8916	27,2051	35,5074	37,0833
$\nu = 12$	Curvatura (C_c)	95	56,6035	53,8701	64,2573	71,5030
		99	169,5659	230,9542	231,2472	372,9492

Tabela 3.7.4: Estatísticas Individuais M_1.

Graus de Liberdade	Medida	Percentis (%) $\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
$\nu = 5$	Curvatura (C_c)	90	0,2029	0,1950	0,1897	0,1855
		95	0,2169	0,2093	0,2021	0,2108
		99	0,2426	0,2416	0,2507	0,2482
$\nu = 7$	Curvatura (C_c)	90	0,2446	0,2384	0,2269	0,2297
		95	0,2714	0,2631	0,2472	0,2506
		99	0,3165	0,3030	0,2778	0,3046
$\nu = 12$	Curvatura (C_c)	90	0,3648	0,3370	0,3221	0,3054
		95	0,3873	0,3701	0,3491	0,3319
		99	0,4467	0,4250	0,3894	0,3733

Tabela 3.7.5: Estatísticas Individuais M_2.

Graus de Liberdade	Medida	Percentis (%) $\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
$\nu = 5$	Curvatura (C_c)	Mínimo	0,1477	0,1396	0,1430	0,1474
		Quantil 5%	0,1508	0,1469	0,1494	0,1556
		Máximo	0,2435	0,2689	0,2715	0,2586
		Média	0,1878	0,1849	0,1875	0,1826
		Desvio Padrão	0,0232	0,0272	0,0262	0,0240
$\nu = 7$	Curvatura (C_c)	Mínimo	0,1798	0,1880	0,1595	0,1674
		Quantil 5%	0,1987	0,1921	0,1812	0,1820
		Máximo	0,3477	0,3196	0,3361	0,3486
		Média	0,2487	0,2317	0,2269	0,2244
		Desvio Padrão	0,0372	0,0294	0,0323	0,0347
$\nu = 12$	Curvatura (C_c)	Mínimo	0,2339	0,2244	0,2170	0,2104
		Quantil 5%	0,2573	0,2599	0,2304	0,2196
		Máximo	0,5350	0,5237	0,4833	0,3653
		Média	0,3316	0,3218	0,3041	0,2792
		Desvio Padrão	0,0552	0,0578	0,0497	0,0370

Tabela 3.7.6: Influência global para C_c.

Graus de Liberdade	Medida $\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
$\nu = 5$	Curvatura (C_c)	**99,949**	**45,832**	17,233	16,273
$\nu = 7$	Curvatura (C_c)	26,492	14,621	37,86	36,773
$\nu = 9$	Curvatura (C_c)	6,9499	38,76	**94,848**	7,8973

Em negrito estão representadas as estatísticas significativas para o percentil 5% de M_0.

Na Tabela 3.7.3, percebe-se que os limiares baseados no critério global M_0 têm comportamento crescente quando ν e/ou ρ aumentam. Assim como na Tabela 3.5.3 onde a distribuição para as inovações são normais, existem indícios de que os limiares baseados no critério global M_0 não são robustos quanto a variação e/ou estimação dos parâmetros.

Através da Tabela 3.7.4, quando ρ e percentil são constantes os limiares baseados no critério global M_1 aumentam se os graus de liberdade aumentam. Observa-se que ao fixar o percentil e $\nu = 5$ (ou $\nu = 7$), os limiares M_1 permanecem praticamente constantes quando ρ aumenta. Por outro lado, se o percentil estiver fixo e $\nu = 12$ os limiares diminuem quando ρ aumenta. Assim, considerando $\nu = 5$ e $\nu = 7$ existem indícios de que os limiares baseados no critério individual M_1 são robustos quanto a variação e/ou estimação dos parâmetros.

Na Tabela 3.7.5, considerando ρ fixo na Tabela 3.7.5 os limiares (quantil 5%) baseados no critério global M_2 aumentam quando os graus de liberdade aumentam. Fixando os graus de liberdade ($\nu = 5$ ou $\nu = 7$), os limiares (quantil 5%) permanecem praticamente constantes quando ρ aumenta. Por outro lado, se $\nu = 12$ é fixado, os limiares (quantil 5%) diminuem quando ρ aumenta. Assim como no critério M_1 (Tabela 3.5.4), quando $\nu = 5$ ou $\nu = 7$ existem indícios de que os limiares baseados em M_2 são robutos quanto a variação e/ou estimação dos parâmetros.

A Tabela 3.7.6 apresenta as estatísticas C_c, calculadas para as séries 1 - 12. Utilizando-se o critério M_0 (Tabela 3.7.3) para avaliar a influência local das 12 séries geradas, a Tabela 3.7.6 mostra que somente as estatísticas $C_c = 99,949$, $C_c = 45,832$ e $C_c = 94,848$ são significativas sob os respectivos critérios M_0.

A seguir, para as séries 2, 7 e 12, serão apresentados os vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$, em conjunção aos limiares M_1 e M_2, ilustrados nas Figuras 3.7.1 - 3.7.6. Em cada uma dessas figuras o gráfico superior corresponde à série sem *outlier*. O segundo gráfico corresponde à série em valores absolutos com *outlier*.

Figura 3.7.1: Influência local na série 2 ($\nu = 5$ e $\rho = 0,5$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

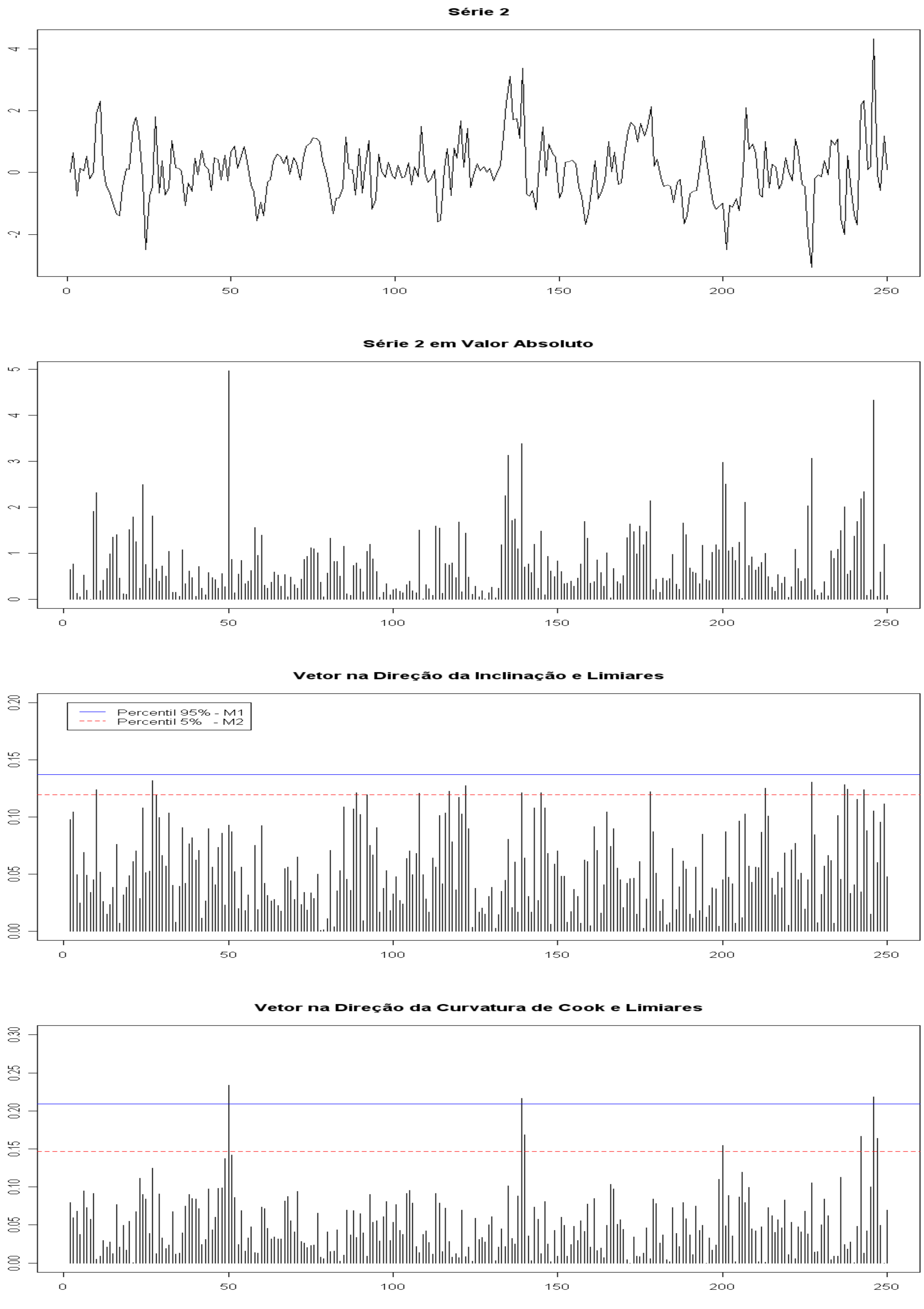

Figura 3.7.2: Influência local na série 2 ($\nu = 5$ e $\rho = 0,5$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

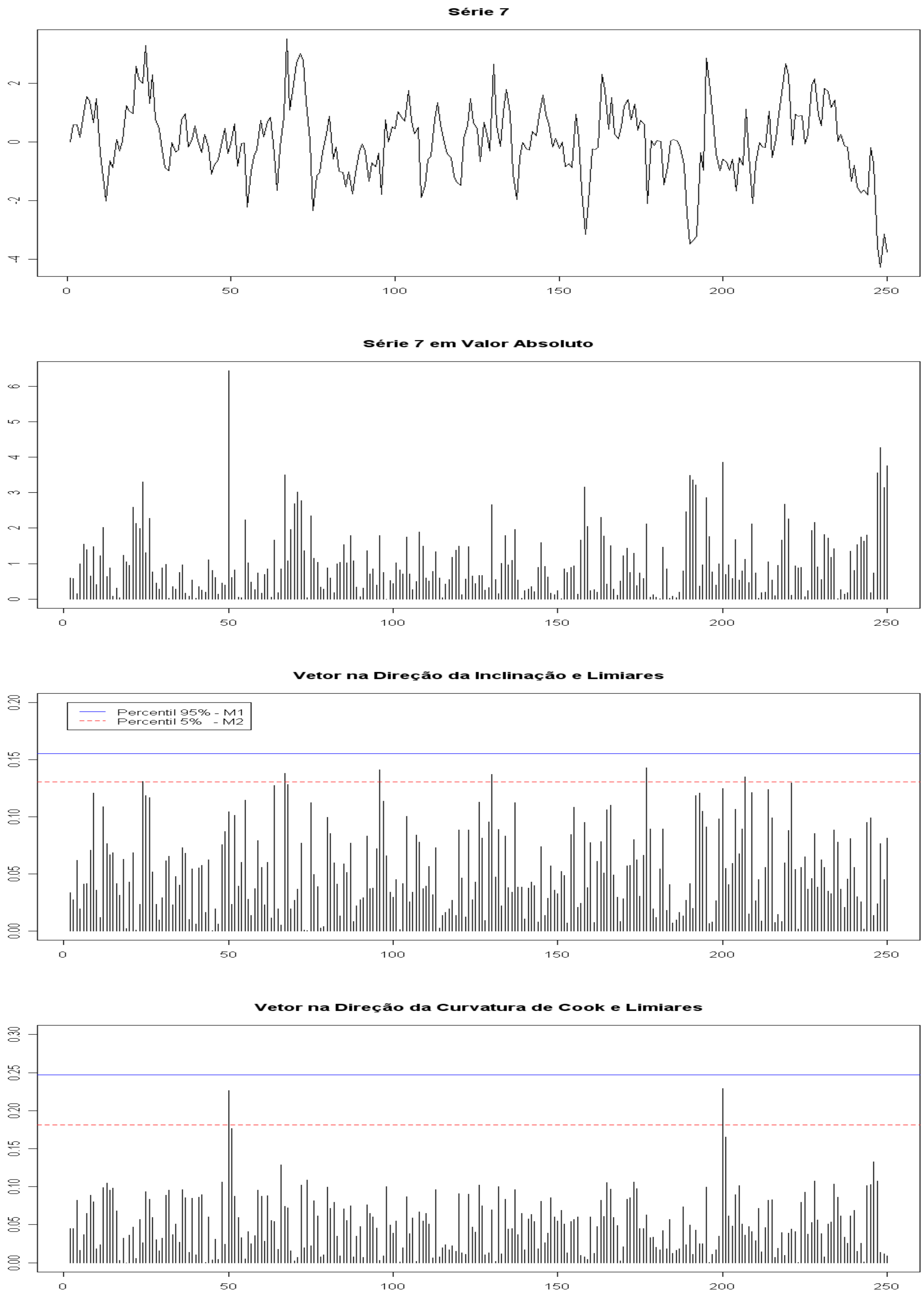

Figura 3.7.3: Influência local na série 7 ($\nu = 7$ e $\rho = 0,7$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

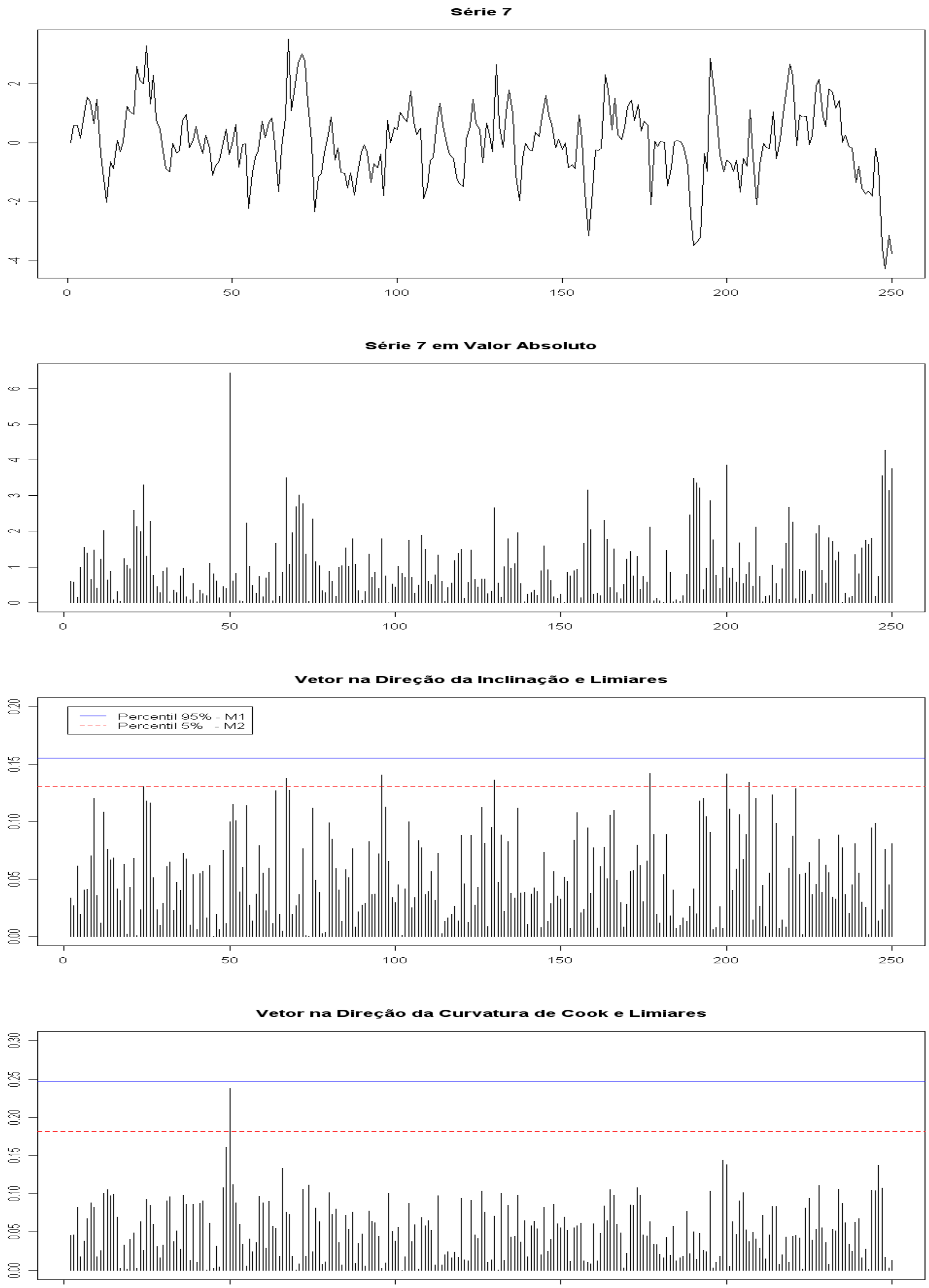

Figura 3.7.4: Influência local na série 7 ($\nu = 7$ e $\rho = 0,7$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

Figura 3.7.5: Influência local na série 12 ($\nu = 12$ e $\rho = 0,9$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

Figura 3.7.6: Influência local na série 12 ($\nu = 12$ e $\rho = 0,9$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

Nas Figuras 3.7.1 - 3.7.6, dados os vetores de diagnóstico da inclinação S e considerando as perturbações positivas e negativas, a maioria das observações geradas como *outliers* não foi identificada pelos limiares baseados nos critérios individuais M_1 e M_2.

Para a série 2 na Figura 3.7.1, dada a perturbação positiva $\tilde{y}_{50} = 4,958$, a componente 50 de $\mathbf{v}_c$ é identificada pelo critério M_2. Isso ocorre pois existe uma "grande" diferença de magnitude em valor absoluto entre $\tilde{y}_{50} = 4,958$ e $y_{50} = 0,663$ (Tabela 3.7.2). Dada a mesma série, se a perturbação é negativa $\tilde{y}_{50} = -4,958$, conforme ilustrado na Figura 3.7.2, a componente 50 de $\mathbf{v}_c$ é identificada pelos dois critérios individuais M_1 e M_2, pois a diferença de magnitude em valor absoluto entre $\tilde{y}_{50} = -4,958$ e $y_{50} = 0,663$ é maior do que no caso onde a perturbação é positiva (ver Tabela 3.7.2). Esse comportamento é o oposto quando se observa a identificação da componente 200 de $\mathbf{v}_c$ nas Figuras 3.7.1 e 3.7.2.

Na Figura 3.7.3, quando a perturbação na série 7 é positiva, as componente 50 e 200 de $\mathbf{v}_c$ são identificadas pelo critério M_2. Isso ocorre em virtude dos valores absolutos das diferenças entre $\tilde{y}_{50} = 6,432$ e $y_{50} = -0,064$ e entre $\tilde{y}_{200} = 3,859$ e $y_{200} = -0,596$ serem "grandes". Ao contrário, quando a perturbação na série 7 é negativa, somente a componente 50 de $\mathbf{v}_c$ é identificada pelo critério M_2 (Figura 3.7.4). Já a componente 200 de $\mathbf{v}_c$ não é identificada por nenhum dos dois critérios individuais.

Para a série 12 ilustrada na Figura 3.7.5, as componentes 50 e 200 de $\mathbf{v}$ e $\mathbf{v}_c$ são identificadas pelos critérios individuais M_1 e M_2 quando a perturbação é positiva. Ao contrário, quando a série 12 é perturbada negativamente as componentes 50 e 200 de $\mathbf{v}$ e $\mathbf{v}_c$ não são identificadas pelos critérios M_1 e M_2 (Figura 3.7.6). Diferentemente do casos anteriores, para a análise de influência local sobre a série 12 não foi encontrada uma justificativa baseada na Tabela 3.7.2.

Quando $\nu = 5$ ou $\nu = 7$, em relação aos critérios individuais M_1 e M_2, o fato dos limiares apresentados nas Tabelas 3.7.4 e 3.7.5, respectivamente, não variarem tanto para diferentes valores de ρ, aponta indícios de que essas marcas de nível podem ser robustas quanto a estimação dos parâmetros. Assim, faz-se interessante verificar o comportamento das marcas de nível em relação aos parâmetros estimados.

A Tabela 3.7.7 apresenta as estimativas de ρ e σ^2 para as séries 1 - 12.

Tabela 3.7.7: Estimativas dos parâmetros ρ e σ^2.

Séries	Parâmetros			Estimativas		
Série 1	$\nu = 5$	$\rho = 0,3$	$\sigma^2 = 1$	$\hat{\nu} = 4,5311(1,0642)$	$\hat{\rho} = 0,2641(0,0580)$	$\hat{\sigma}^2 = 0,9132(0,1289)$
Série 2	$\nu = 5$	$\rho = 0,5$	$\sigma^2 = 1$	$\hat{\nu} = 4,3625(1,0331)$	$\hat{\rho} = 0,3768(0,0609)$	$\hat{\sigma}^2 = 0,9363(0,1458)$
Série 3	$\nu = 5$	$\rho = 0,7$	$\sigma^2 = 1$	$\hat{\nu} = 4,8902(1,1971)$	$\hat{\rho} = 0,6120(0,0467)$	$\hat{\sigma}^2 = 0,9216(0,1206)$
Série 4	$\nu = 5$	$\rho = 0,9$	$\sigma^2 = 1$	$\hat{\nu} = 3,5474(0,4531)$	$\hat{\rho} = 0,9127(0,0227)$	$\hat{\sigma}^2 = 0,9240(0,1262)$
Série 5	$\nu = 7$	$\rho = 0,3$	$\sigma^2 = 1$	$\hat{\nu} = 5,2040(1,1928)$	$\hat{\rho} = 0,2892(0,0537)$	$\hat{\sigma}^2 = 0,9494(0,1100)$
Série 6	$\nu = 7$	$\rho = 0,5$	$\sigma^2 = 1$	$\hat{\nu} = 5,5980(1,5629)$	$\hat{\rho} = 0,4148(0,0542)$	$\hat{\sigma}^2 = 0,9740(0,1190)$
Série 7	$\nu = 7$	$\rho = 0,7$	$\sigma^2 = 1$	$\hat{\nu} = 4,8600(0,9383)$	$\hat{\rho} = 0,6283(0,0483)$	$\hat{\sigma}^2 = 0,9502(0,1062)$
Série 8	$\nu = 7$	$\rho = 0,9$	$\sigma^2 = 1$	$\hat{\nu} = 4,5749(0,9306)$	$\hat{\rho} = 0,8181(0,0361)$	$\hat{\sigma}^2 = 0,9617(0,1179)$
Série 9	$\nu = 12$	$\rho = 0,3$	$\sigma^2 = 1$	$\hat{\nu} = 9,2757(4,2483)$	$\hat{\rho} = 0,2840(0,0579)$	$\hat{\sigma}^2 = 1,0828(0,1162)$
Série 10	$\nu = 12$	$\rho = 0,5$	$\sigma^2 = 1$	$\hat{\nu} = 6,0686(1,5519)$	$\hat{\rho} = 0,5502(0,0520)$	$\hat{\sigma}^2 = 1,1343(0,1190)$
Série 11	$\nu = 12$	$\rho = 0,7$	$\sigma^2 = 1$	$\hat{\nu} = 7,4855(2,2342)$	$\hat{\rho} = 0,6543(0,0462)$	$\hat{\sigma}^2 = 1,1526(0,1086)$
Série 12	$\nu = 12$	$\rho = 0,9$	$\sigma^2 = 1$	$\hat{\nu} = 5,3162(0,9249)$	$\hat{\rho} = 0,9268(0,0232)$	$\hat{\sigma}^2 = 0,9643(0,0989)$

Entre parênteses se encontram os desvios padrões das estimativas dos parâmetros.

Nas Tabelas 3.7.8 - 3.7.10 são apresentados os limiares gerados a partir dos verdadeiros valores dos parâmetros e os novos limiares gerados a partir das estimativas dos parâmetros apresentadas na Tabela 3.7.7.

Tabela 3.7.8: Estatísticas Globais M_0.

Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
		90		18,5136	21,4247	24,4065	37,6396
$\nu = 5$	Curvatura (C_c)	95		36,0685	32,5487	45,8358	83,1595
		99		132,3213	158,5604	208,9140	428,3924
			$\hat{\boldsymbol{\nu}}$	4,5311	4,3625	4,8902	3,5474
			$\hat{\boldsymbol{\rho}}$	0,2641	0,3768	0,6120	0,9127
			$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	0,9132	0,9363	0,9216	0,9240
		90		6,8049	7,0028	7,7053	9,2780
$\nu = 5$	Curvatura (C_c)	95		7,0102	7,2397	8,0759	9,6369
		99		7,4905	7,7097	8,8606	10,2735
Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
		90		20,0583	27,2735	33,3100	38,1278
$\nu = 7$	Curvatura (C_c)	95		37,8868	45,6139	59,7960	71,6737
		99		185,6951	201,7662	365,2938	397,3722
			$\hat{\boldsymbol{\nu}}$	5,2040	5,5980	4,8600	4,5749
			$\hat{\boldsymbol{\rho}}$	0,2892	0,4148	0,6283	0,8181
			$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	0,9494	0,9740	0,9502	0,9617
		90		6,3187	6,5105	7,6859	8,1850
$\nu = 7$	Curvatura (C_c)	95		6,4979	6,7319	7,9833	8,6097
		99		6,8587	7,5768	8,4953	9,3417
Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
		90		27,8916	27,2051	35,5074	37,0833
$\nu = 12$	Curvatura (C_c)	95		56,6035	53,8701	64,2573	71,5030
		99		169,5659	230,9542	231,2472	372,9492
			$\hat{\boldsymbol{\nu}}$	9,2757	6,0686	7,4855	5,3162
			$\hat{\boldsymbol{\rho}}$	0,2840	0,5502	0,6543	0,9268
			$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	1,0828	1,1343	1,1526	0,9643
		90		7,6176	6,2417	7,3273	8,7546
$\nu = 12$	Curvatura (C_c)	95		11,7445	6,6442	8,5077	9,3086
		99		39,9386	8,3876	15,2160	10,8361

Tabela 3.7.9: Estatísticas Individuais M_1.

Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
		90		0,2029	0,1950	0,1897	0,1855
$\nu = 5$	Curvatura (C_c)	95		0,2169	0,2093	**0,2021**	0,2108
		99		0,2426	0,2416	0,2507	0,2482
			$\hat{\boldsymbol{\nu}}$	4,5311	4,3625	4,8902	3,5474
			$\hat{\boldsymbol{\rho}}$	0,2641	0,3768	0,6120	0,9127
			$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	0,9132	0,9363	0,9216	0,9240
		90		0,3305	0,3009	0,1980	0,2096
$\nu = 5$	Curvatura (C_c)	95		0,3720	0,3462	**0,2166**	0,2359
		99		0,4813	0,4299	0,2573	0,2926
Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
		90		0,2446	0,2384	0,2269	0,2297
$\nu = 7$	Curvatura (C_c)	95		0,2714	0,2631	0,2472	0,2506
		99		0,3165	0,3030	0,2778	0,3046
			$\hat{\boldsymbol{\nu}}$	5,2040	5,5980	4,8600	4,5749
			$\hat{\boldsymbol{\rho}}$	0,2892	0,4148	0,6283	0,8181
			$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	0,9494	0,9740	0,9502	0,9617
		90		0,2683	0,2189	0,1938	0,1576
$\nu = 7$	Curvatura (C_c)	95		0,2960	0,2463	0,2129	0,1682
		99		0,3500	0,3018	0,2583	0,2023
Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
		90		0,3648	0,3370	0,3221	0,3054
$\nu = 12$	Curvatura (C_c)	95		0,3873	0,3701	0,3491	0,3319
		99		0,4467	0,4250	0,3894	0,3733
			$\hat{\boldsymbol{\nu}}$	9,2757	6,0686	7,4855	5,3162
			$\hat{\boldsymbol{\rho}}$	0,2840	0,5502	0,6543	0,9268
			$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	1,0828	1,1343	1,1526	0,9643
		90		0,2084	0,1761	0,1524	0,1355
$\nu = 12$	Curvatura (C_c)	95		0,2280	0,1856	0,1601	0,1379
		99		0,2555	0,2232	0,1796	0,1447

Em negrito são comparados os limiares M_1 calculados com os parâmetros verdadeiros e estimados.

Tabela 3.7.10: Estatísticas Individuais M_2.

Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
		Mínimo		0,1477	0,1396	0,1430	0,1474
		Quantil 5%		0,1508	0,1469	0,1494	0,1556
$\nu = 5$	Curvatura (C_c)	Máximo		0,2435	0,2689	0,2715	0,2586
		Média		0,1878	0,1849	0,1875	0,1826
		Desvio Padrão		0,0232	0,0272	0,0262	0,0240
			$\hat{\boldsymbol{\nu}}$	4,5311	4,3625	4,8902	3,5474
			$\hat{\boldsymbol{\rho}}$	0,2641	0,3768	0,6120	0,9127
			$\hat{\boldsymbol{\sigma}}^2$	0,9132	0,9363	0,9216	0,9240
		Mínimo		0,1529	0,1462	0,1228	0,1269
		Quantil 5%		0,1758	0,1585	0,1303	0,1279
$\nu = 5$	Curvatura (C_c)	Máximo		0,3720	0,3273	0,2129	0,4298
		Média		0,2197	0,1970	0,1512	0,1662
		Desvio Padrão		0,0402	0,0347	0,0183	0,0520
Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
		Mínimo		0,1798	0,1880	0,1595	0,1674
		Quantil 5%		0,1987	0,1921	0,1812	0,1820
$\nu = 7$	Curvatura (C_c)	Máximo		0,3477	0,3196	0,3361	0,3486
		Média		0,2487	0,2317	0,2269	0,2244
		Desvio Padrão		0,0372	0,0294	0,0323	0,0347
			$\hat{\boldsymbol{\nu}}$	5,2040	5,5980	4,8600	4,5749
			$\hat{\boldsymbol{\rho}}$	0,2892	0,4148	0,6283	0,8181
			$\hat{\boldsymbol{\sigma}}^2$	0,9494	0,9740	0,9502	0,9617
		Mínimo		0,1333	0,1137	0,1234	0,1142
		Quantil 5%		0,1441	0,1268	0,1279	0,1174
$\nu = 7$	Curvatura (C_c)	Máximo		0,3437	0,2314	0,2696	0,1682
		Média		0,1945	0,1602	0,1521	0,1341
		Desvio Padrão		0,0393	0,0269	0,0281	0,0130
Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{\rho}$	0,3	0,5	0,7	0,9
		Mínimo		0,2339	0,2244	0,2170	0,2104
		Quantil 5%		0,2573	0,2599	0,2304	0,2196
$\nu = 12$	Curvatura (C_c)	Máximo		0,5350	0,5237	0,4833	0,3653
		Média		0,3316	0,3218	0,3041	0,2792
		Desvio Padrão		0,0552	0,0578	0,0497	0,0370
			$\hat{\boldsymbol{\nu}}$	9,2757	6,0686	7,4855	5,3162
			$\hat{\boldsymbol{\rho}}$	0,2840	0,5502	0,6543	0,9268
			$\hat{\boldsymbol{\sigma}}^2$	1,0828	1,1343	1,1526	0,9643
		Mínimo		0,1840	0,1211	0,1219	0,1183
		Quantil 5%		0,2003	0,1232	0,1285	0,1200
$\nu = 12$	Curvatura (C_c)	Máximo		0,2996	0,1560	0,2182	0,1371
		Média		0,2356	0,1362	0,1576	0,1262
		Desvio Padrão		0,0244	0,0089	0,0231	0,0045

Para as séries 1 - 12 na Tabela 3.7.8, percebe-se que os limiares baseados no critério global M_0 são significativamente afetados pelas estimativas dos parâmetros. Especificamente, os limiares baseados em M_0 calculados através de $\hat{\rho}$, $\hat{\nu}$ e $\hat{\sigma}^2$ são menores do que os limiares calculados através dos verdadeiros valores dos parâmetros. Dessa forma, M_0 não é útil na prática.

Na Tabela 3.7.9, quando $\nu = 5$ ($\rho = 0,7$ ou $\rho = 0,9$) os limiares baseados no critério individual M_1 apresentam indícios de robustez quanto à estimação dos parâmetros. Um exemplo disso é ilustrado na Tabela 3.7.9, onde o limiar $M_1 = 0,2021$ calculado com os verdadeiros parâmetros ($\nu = 5$, $\rho = 0,7$ e $\sigma^2 = 1$) está próximo ao limiar $M_1 = 0,2166$ calculado com os parâmetros estimados ($\hat{\nu} = 4,8902$, $\hat{\rho} = 0,6120$ e $\hat{\sigma}^2 = 0,9216$). O mesmo ocorre quando $\nu = 7$ ($\rho = 0,3$ ou $0,5$). Por outro lado, quando essa mesma análise é feita com os limiares M_1 calculados para as séries a partir de $\nu = 12$, e de seus respectivos valores estimados, não existe robustez quanto à estimação dos parâmetros. Em conseqüência, considerando somente essas específicas combinações dos parâmetros, M_1 é pouco útil na prática.

Na Tabela 3.7.10, quando $\nu = 5$ os limiares baseados no critério individual M_2 indicam robustez quanto à estimação dos parâmetros. Assim, M_2 é considerado um limiar pouco útil na prática.

3.7.4 Conclusões das Simulações

A partir dos experimentos de simulação, das doze séries analisadas o critério global M_0 somente indica a existência de observações influentes em três séries (Tabela 3.7.6), quando os limiares são calculados com os verdadeiros valores dos parâmetros. Por outro lado, os limiares M_0 são bastante afetados pelas estimativas dos parâmetros, e em conseqüência, M_0 não seria uma ferramenta útil na prática.

Ao contrário, verifica-se robustez das marcas de nível quanto à estimação dos parâmetros, considerando os critérios individuais M_1 quando $\nu = 5$ ($\rho = 0,7$ ou $\rho = 0,9$) ou $\nu = 7$ ($\rho = 0,3$ ou $\rho = 0,5$), e M_2 quando $\nu = 5$. Assim, como os limiares M_1 e M_2 só identificam corretamente essas específicas combinações de parâmetros, são pouco úteis na prática.

Os critérios M_0, M_1 e M_2 foram aplicados em várias séries simuladas nas quais foram introduzidos *outliers*. Em uma grande porção destas, os critérios M_0, M_1 e M_2 não foram capazes de detectar as observações influentes.

Da mesma maneira que no diagnóstico realizado para as séries simuladas na Seção 3.5 considerando normalidade para a distribuição das inovações, existem observações que não foram geradas como *outliers* as quais são identificadas como influentes segundo os critérios individuais (Figuras 3.7.1 e 3.7.2).

Por outro lado, os limiares para a Curvatura de Cook considerando o critério global M_0 (Tabela 3.5.3), nota-se que supondo normalidade para as inovações do modelo AR(1), os limiares são menores em relação as marcas de referência (Tabela 3.7.3) calculadas sob a distribuição t-Student. À exceção de $\nu = 5$, os limiares baseados nos critérios individuais M_1 e M_2, referentes as Tabelas 3.5.4 e 3.5.5 (normal), também são menores do que os limiares referentes as Tabelas 3.7.4 e 3.7.5 (t-Student), respectivamente.

3.7.5 Aplicação

A série histórica estudada nesta subseção contém 311 preços de fechamento diários da $Microsoft$ e do índice $S\&P500$ no período de 2 de janeiro de 2002 a 28 março 2003. As taxas $T - Bill$ são utilizadas como retornos livres de risco, as quais são divididas por 253 para conversão em taxas diárias. Em particular, será estudado o Modelo de Precificação de Ativos de Capital (*CAPM - Capital Asset Pricing Model*) desenvolvido por Sharpe (1964), Lintner (1965) e Mossin (1966). Uma das formas de se definir o modelo CAPM é dada por

$$R_t - r_{ft} = \beta_0 + \beta_1(R_{Mt} - r_{ft}) + \epsilon_t, \tag{3.7.5}$$

em que R_t é o retorno de um ativo entre os períodos de tempo t e $(t-1)$, R_{Mt} é o retorno de um ativo de mercado entre os períodos t e $(t-1)$, r_{ft} é denotada a taxa de retorno entre os períodos t e $(t-1)$ de um ativo livre de risco e ϵ_t o erro aleatório. A regressão (3.7.7) é definida como $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$, onde $y_t = R_t - r_{ft}$ (Excesso de retorno $Microsoft$)

e $x_t = R_{Mt} - r_{ft}$ (Excesso de retorno $S\&P500$). Considerando o log-retorno do ativo $Microsoft$ (R_t), o log-retorno do índice $S\&P500$ (R_{Mt}) e o log-retorno de T-Bill (r_{ft}), na Figura 3.7.7 é apresentado o gráfico de dispersão, e em destaque a observação y_{87}.

No *CAPM*, o risco de mercado, ou sistemático, é medido pelo coeficiente β_1. Ou seja, um ativo (ação) onde β_1 não é significativamente diferente de 1, deve se mover em concordância com o mercado. Se β_1 é significativamente maior do que 1 a ação é mais agressiva do que o mercado. Por fim, se β_1 é significativamente menor do que 1 esta ação é conservadora em relação ao mercado.

São realizados três ajustes para o modelo de regressão: *i)* $\epsilon_t \sim AR(1)$ (3.7.1) com distribuição t-Student para as inovações, *ii)* $\epsilon_t \sim AR(1)$ (3.1.1) com distribuição Normal para as inovações e *iii)* $\epsilon_t \sim AR(1)$ (3.1.1) com distribuição Normal para as inovações, considerando $\beta_0 = 0$ (ação bem-precificada). Os parâmetros são estimados em dois estágios. No primeiro estágio, estimam-se os parâmetros $\hat{\beta}_0$ e $\hat{\beta}_1$ por mínimos quadrados. No segundo estágio, supondo t-Student para as inovações, estimam-se ρ, σ^2 e ν por máxima verossimilhança. A Tabela 3.7.11 apresenta as estimativas para os respectivos modelos. As autocorrelações e autocorrelações parciais dos resíduos (Figura 3.7.7) motivam o ajuste dos mesmos pelo modelo autoregressivo de primeira ordem (3.7.1), enquanto o gráfico dos quantis dos resíduos pelos quantis de uma normal padrão aponta indícios de caudas pesadas para os resíduos. Existem ainda indícios de que distribuição dos resíduos não seja normal, quando realizado o teste de normalidade $Kolmogorov - Smirnov$ ($p - valor < 2,2.10^{-16}$). O coeficiente de correlação é altamente significativo, o que é confirmado pelo gráfico das autocorrelações ilustrado na Figura 3.7.7.

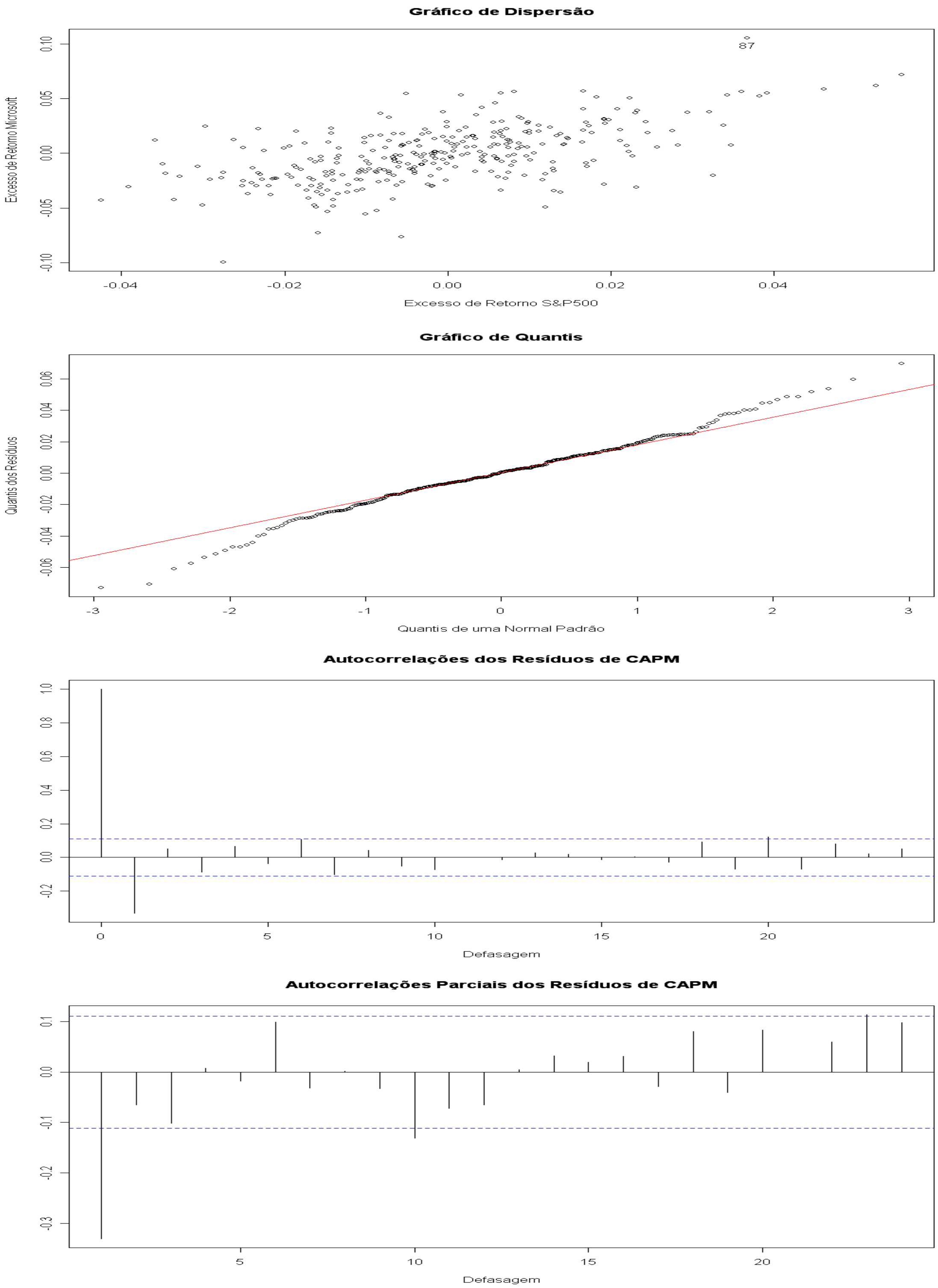

Figura 3.7.7: Gráfico de dispersão e autocorrelações e autocorrelações parciais dos resíduos de *CAPM*.

Tabela 3.7.11: Parâmetros estimados.

	Parâmetro	Estimativa	Desvio Padrão	Razão t	BIC
Normal	β_0	$1,516.10^{-5}$	$1,208.10^{-3}$	0,013	-1534,513
	β_1	0,9681	0,0743	13,027	—
	ρ	-0,3395	0,0541	—	—
	σ^2	$3,985.10^{-4}$	—	—	—
Normal($\beta_0 = 0$)	β_1	0,9680	0,0740	13,07	-1540,246
	ρ	-0,3395	0,0541	—	—
	σ^2	$3,985.10^{-4}$	—	—	—
t-Student	β_0	$1,516.10^{-5}$	$1,208.10^{-3}$	0,013	-1540,814
	β_1	0,9681	0,0743	13,027	—
	ρ	-0,3337	—	—	—
	σ^2	$3,978.10^{-4}$	—	—	—
	ν	7,4194	—	—	—

Através da Tabela 3.7.11, percebe-se que o ativo $Microsoft$ pode ser julgado como, movendo-se com o índice $S\&P500$, pois $\hat{\beta}_1$ não é significativamente diferente de 1. Como a estimativa para β_0 não é significativa, considera-se que os preços das ações da $Microsoft$ se comportam conforme o mercado ($S\&P500$).

Através da metodologia proposta por Cook (1986), será realizado o diagnóstico de influência local considerando o esquema de perturbação nos dados e a metodologia de limiares. A Tabela 3.7.12 apresenta um resumo para o diagnóstico de influência local sobre a medida Curvatura de Cook através das marcas de referência M_0, M_1 e M_2, considerando os modelos $i)$, $ii)$ e $iii)$.

Tabela 3.7.12: Resumo do diagnóstico da influência local.

Pontos Influentes	Data	Normal (ii) $\mathbf{v}$	Normal (ii) $\mathbf{v_c}$	Normal($\beta_0 = 0$) (iii) $\mathbf{v}$	Normal($\beta_0 = 0$) (iii) $\mathbf{v_c}$	t-Student (i) $\mathbf{v}$	t-Student (i) $\mathbf{v_c}$
12	17/01/2002	0,1375	0,1262	0,1374	0,1261	**0,1208**	0,0096
13	18/01/2002	0,1135	**0,1542**	0,1134	**0,1541**	**0,1135**	0,0552
61	01/04/2002	0,1390	0,1328	0,1389	0,1328	**0,1166**	0,0057
86	06/05/2002	0,1259	**0,1692**	0,1260	**0,1653**	0,0958	0,0370
87	07/05/2002	**0,1938**	**0,1652**	**0,1938**	**0,1653**	0,0634	0,1345
106	04/06/2002	0,1069	0,1257	0,1070	0,1258	**0,1116**	0,0634
138	19/07/2002	**0,2129**	**0,1908**	**0,2128**	**0,1907**	0,0908	0,0833
139	22/07/2002	0,0778	**0,1811**	0,0777	**0,1810**	0,0541	0,0593
140	23/07/2002	**0,1599**	0,1035	**0,1598**	0,1034	0,0678	0,1155
209	29/10/2002	0,1268	0,1004	0,1269	0,1005	**0,1136**	0,0213
260	13/01/2003	**0,1698**	**0,1534**	**0,1697**	**0,1533**	0,1052	0,0454
265	21/01/2003	0,1218	0,1474	0,1217	0,1473	**0,1149**	0,0365
268	24/01/2003	0,1208	0,0943	0,1207	0,0942	**0,1170**	0,0318
309	25/03/2003	0,1078	0,1125	0,1077	0,1124	**0,1122**	0,0596
Estatísticas	S	1849,9	—	1849,9	—	1953,6	—
Globais	C_c	—	14157	—	14157	—	19763
Limiares	M_0	2002,12	16291,55	2000,32	16256,30	2056,980	297269,5
	M_1	0,2124	0,2127	0,2093	0,2096	0,1275	0,2502
	M_2	0,1514	0,1502	0,1502	0,1413	0,1111	0,1742

Legenda: Em negrito estão representadas as estatísticas significativas para os critérios M_0 e M_1 (95%) e M_2 (5%).

Na Tabela 3.7.12, observa-se que as medidas globais S ou C_c para os três modelos não são significativas sob o critério M_0 Segundo esse critério não haverá influência global. Porém, as componentes 13, 86, 87, 138, 139, 140 e 260 de $\mathbf{v}$ e $\mathbf{v}_c$, considerando os modelos *ii*) e *iii*), avaliadas sobre os critérios individuais M_1 e/ou M_2, são identificadas como influentes. Percebe-se que utilizando a distribuição t-Student para as inovações do modelo *i*), as componentes 12, 13, 61, 106, 209, 265, 267 e 309 do vetor de diagnóstico para a Inclinação de Billor e Loynes são significativas sob o critério M_2. Por outro lado, considerando M_1 e M_2 nenhuma componente do vetor de diagnóstico da Curvatura de Cook é significativa.

A Figura 3.7.8 ilustra o comportamento de M_1 e M_2 perante os vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$ em conjunção com o vetor de respostas para o modelo *CAPM*, quando a distribuição das inovações é suposta normal (modelo *ii*). O comportamento dos vetores de diagnóstico e dos limiares, quando $\beta_0 = 0$ (modelo *iii*) e as inovações normais, é idêntico ao apresentado na Figura 3.7.8. A Figura 3.7.9 ilustra o comportamento de M_1 e M_2 perante o vetor de diagnóstico $\mathbf{v}_c$ em conjunção com o vetor de respostas e o vetor de resíduos para o modelo $CAPM$ i), considerando a distribuição t-Student para as inovações.

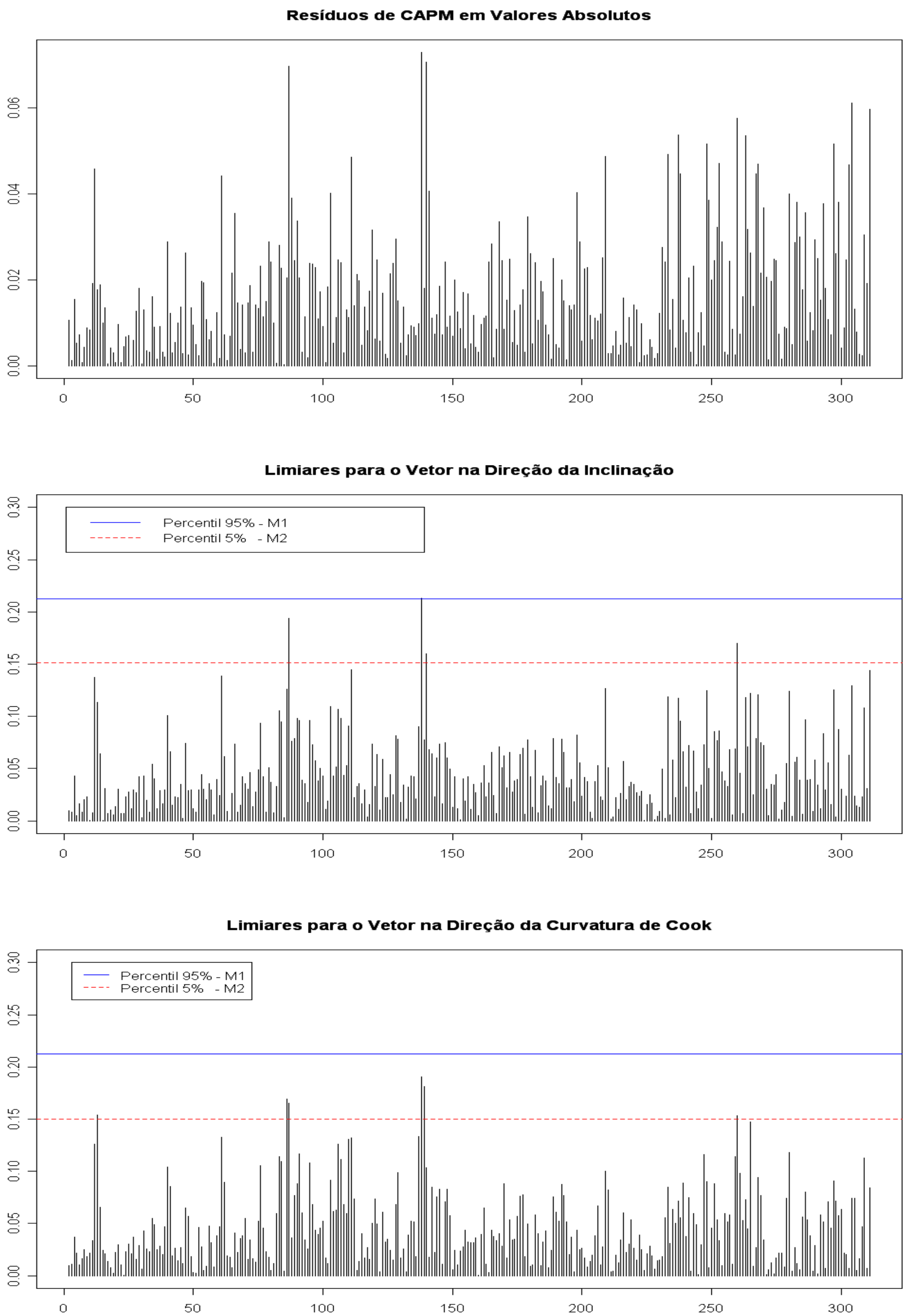

Figura 3.7.8: Influência local considerando as inovações $\epsilon_t \sim AR(1)$ normais.

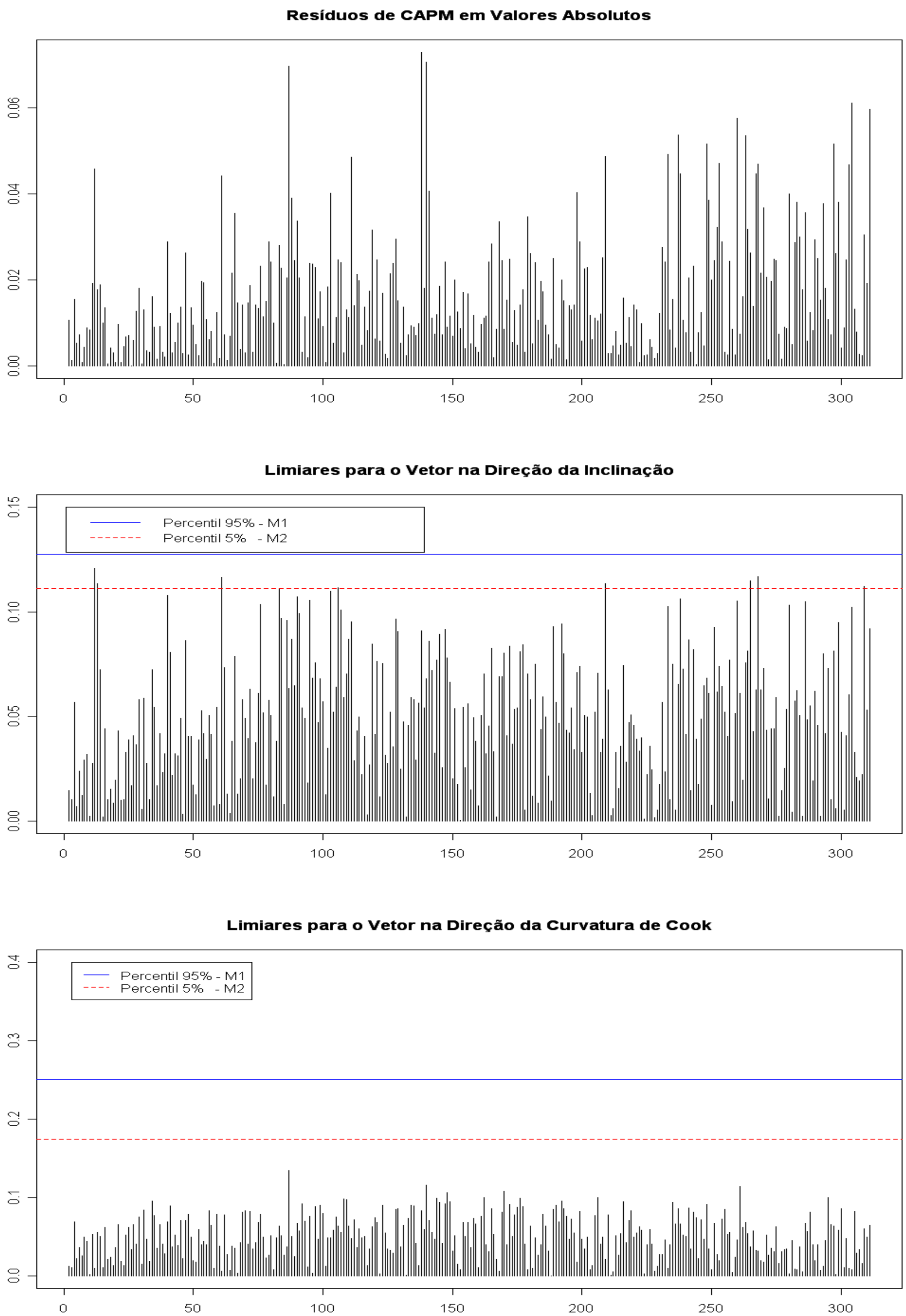

Figura 3.7.9: Influência local considerando as inovações $\epsilon_t \sim AR(1)$ t-Student.

Na Figura 3.7.8, quando as inovações são supostas normais (modelo *ii*), as componentes 87, 138, 140 e 260 de $\mathbf{v}$ e as componentes 13, 86, 87, 138, 139 e 260 de $\mathbf{v}_c$, têm suas significâncias baseadas nos critérios M_1 e M_2, respectivamente. Assim, ficam identificadas como influentes as observações relativas aos dias 18 de janeiro, 6 e 7 de maio, 19, 22 e 23 de julho de 2002, e 13 de janeiro de 2003.

Na Figura 3.7.9, percebe-se que o vetor $\mathbf{v}$ tem suas componentes próximas de zero e compreendidas em uma faixa uniforme quando comparadas ao limiar M_2. Considerando as componentes 12, 13, 61, 106, 209, 265, 267 e 309 do vetor de diagnóstico $\mathbf{v}$, nenhuma tem valor distante da marca M_2. Adicionalmente, tem-se que se o gráfico de influência apresenta significativa não-linearidade, as componentes de $\mathbf{v}$ são próximas de zero [Zhang e King, 2005]. Assim, não é possível identificar observações influentes através da medida de Inclinação de Billor e Loynes.

Ainda na Figura 3.7.9, observa-se que apesar da componente 87 do vetor $\mathbf{v}_c$ se destacar perante as demais em termos de magnitude, não é considerada significativa sob os critérios M_1 e M_2. De maneira análoga ao vetor $\mathbf{v}$, as componentes se encontram próximas de zero em uma faixa uniforme abaixo dos limiares M_1 e M_2. Portanto, considerando a metodologia dos limiares, perturbação nos dados e sob a suposição t-Student para os erros do modelo CAPM, nenhuma observação foi identificada como sendo influente.

Uma comparação importante a ser feita, diz respeito aos vetores de diagnóstico $\mathbf{v}_c$ para a Curvatura de Cook, das Figuras 3.7.8 e 3.7.9, onde são calculados considerando as distribuições normal e t-Student, respectivamente. Considerando o vetor de diagnóstico $\mathbf{v}$, os limiares $M_1 = 0,2127$ e $M_2 = 0,1502$ da Figura 3.7.8 são maiores do que os limiares $M_1 = 0,1275$ e $M_2 = 0,1111$ da Figura 3.7.9. Ao contrário, para o vetor de diagnóstico $\mathbf{v}_c$, percebe-se que os limiares $M_1 = 0,2124$ e $M_2 = 0,1514$ da Figura 3.7.8 são menores do que os limiares $M_1 = 0,2502$ e $M_2 = 0,1742$ da Figura 3.7.9.

4 Influência Local em Modelos ARFIMA

4.1 Introdução

O objetivo deste capítulo é apresentar a metodologia de influência local para o modelo autoregressivo fracionário integrado de média-móvel (ARFIMA). Desta maneira, são apresentadas algumas propriedades dos processos de longa memória importantes para a construção dos modelos ARFIMA(p,d,q). A seguir, uma aproximação do modelo ruído fracionário ARFIMA(0,d,0) por um modelo ARMA(p,q) é discutida. Baseado em uma verossimilhança aproximada, avalia-se a influência local sobre o modelo aproximado ARFIMA(0,d,0) através das medidas propostas por Cook (1986) e Billor e Loynes (1993). Por fim, utiliza-se a técnica de limiares (Seção 2.3) para verificar a significância de potenciais observações influentes, tanto para dados gerados por simulação, quanto para os dados da série histórica $Varve$.

4.2 Processos de Longa Memória

Hosking (1981) e Granger e Joyeux (1980) introduziram uma família de modelos de séries temporais que apresentam correlação significativa entre observações distantes durante um longo período de tempo. A idéia central é construir uma generalização para os modelos ARIMA de Box e Jenkins (1976) que permita o grau de diferenciação d assumir quaisquer valores reais.

Em um processo de longa memória estacionário a função de autocorrelação decresce hiperbolicamente para zero, ou seja,

$$\rho(k) \sim Ck^{2d-1}, \tag{4.2.1}$$

quando $k \to \infty$, com $C \neq 0$ e $0 < d < 0,5$. Neste caso, $\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty$, ao contrário dos processos de curta memória ou memória intermediária, nos quais $\sum_{k=-\infty}^{\infty} |\rho(k)| \leq \infty$.

Beran (1994) aponta dois fatos que têm estimulado o crescimento das pesquisas relacionadas aos processos de longa memória. Primeiro, a inferência estatística baseada em medidas padrões com a média e variância amostrais podem ser substancialmente diferentes para processos de longa memória comparado com processos de memória curta. Segundo, verifica-se que inúmeras séries temporais no mundo real, em áreas como finanças e hidrologia, exibem as características do processo de longa memória. Aplicações recentes são os modelos FIGARCH (*Fracionário Integrado Autoregressivo Generalizado com Heteroscedasticidade Condicional*) introduzido por Baillie et al. (1996), ou ainda, o modelo FIEGARCH (*Fracionário Integrado Exponencial Autoregressivo Generalizado com Heteroscedasticidade Condicional*) em Bollerslev e Mikkelsen (1996).

4.2.1 Definição do Modelo ARFIMA

A seguir é apresentada a definição do modelo ARFIMA, assim como algumas propriedades [ver Hosking, 1981]. O processo $\{y_t\}_{t\in\mathbb{Z}}$ é um processo Autoregressivo Fracionário Integrado de Média-Móvel, ou ARFIMA(p,d,q), com $d \in \Re$, se $\{y_t\}$ for estacionário e satisfizer a equação

$$\phi(B)(1-B)^d y_t = \theta(B)\epsilon_t, \tag{4.2.2}$$

onde $\epsilon_t \sim RB(0,\sigma^2)$ e $\phi(B)$ e $\theta(B)$ são polinômios de forma que

$$\begin{aligned} \phi(B) &= 1 + \phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p, \\ \theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q, \end{aligned}$$

e ∇^d é o operador de diferença fracionária definido pela série binomial como

$$\nabla^d = (1-B)^d = \sum_{k=0}^{\infty} \begin{pmatrix} d \\ k \end{pmatrix} (-B)^k = 1 - dB - \frac{d}{2}(1-d)B^2 - \frac{d}{6}(1-d)(2-d)B^3 - \cdots .$$

Seja $\{y_t\}_{t\in\mathbb{Z}}$ o processo ARFIMA(p,d,q) definido por (4.2.2). Então,

1. $\{y_t\}$ é estacionário sempre que $d < \frac{1}{2}$ e todas as raízes da equação $\phi(B) = 0$ estejam fora do círculo unitário;

2. $\{y_t\}$ é inversível sempre que $d > -1$ e todas as raízes da equação $\theta(B) = 0$ estejam fora do círculo unitário.

Se $p = q = 0$, o processo $\{y_t\}$ é chamado de Ruído Fracionário, ARFIMA(0,d,0), sendo estacionário e inversível quando $-1 < d < \frac{1}{2}$. Assim, o modelo (4.2.2) é escrito como

$$(1-B)^d y_t = \epsilon_t, \tag{4.2.3}$$

onde $\epsilon_t \sim RB(0, \sigma^2)$.

A seguir são descritas importantes propriedades do processo de ruído fracionário as quais serão de fundamental importância na derivação das expressões para as medidas de influência local. Seja $\{y_t\}$ um processo ARFIMA(0,d,0):

a) Dinâmica,

$$\begin{aligned} E[y_t] &= 0; \\ \gamma_\tau &= Cov[y_t, y_{t-\tau}] = \frac{(-1)^\tau(-2d)!}{(\tau-d)!(-\tau-d)!}; \\ \gamma_0 &= Var[y_t] = \frac{(-2d)!}{(-d!)^2}. \end{aligned}$$

b) Quando $d < \frac{1}{2}$, $\{y_t\}$ assume representação média-móvel infinita

$$y_t = \sum_{k=0}^{\infty} \phi_k(d)\epsilon_{t-k},$$

onde

$$\phi_k(d) = \frac{(k+d-1)!}{k!(d-1)!} = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)},$$

sendo $\Gamma(\cdot)$ a função *gamma*.

c) Quando $d > -\frac{1}{2}$, $\{y_t\}$ assume representação autoregressiva infinita

$$\epsilon_t = \sum_{j=0}^{\infty} \theta_j(d) y_{t-j},$$

onde

$$\theta_j(d) = \frac{(j-d-1)!}{j!(-d-1)!} = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}.$$

A partir das representações média-móvel infinita e autoregressiva infinita será possível construir uma aproximação do modelo ARFIMA(p,d,q) por um modelo ARMA(p,q), o que será discutido na seguinte subseção.

4.2.2 Aproximação do Modelo ARFIMA(0,d,0) por ARMA(p,q)

A aproximação do modelo ARFIMA(0,d,0) pelo modelo ARMA(p,q) apresenta inúmeras vantagens práticas. Uma delas é facilitar a utilização de técnicas tradicionais de previsão, as quais são aplicadas em modelos ARMA(p,q) de curta memória, que é um dos desafios encontrados em estudos de processos de longa memória [Chan e Palma, 1998; Tiao e Tsay, 1994].

Aproximações do modelo de longa memória por modelos de curta memória foram estudadas por exemplo em Basak et al. (2001). No referido trabalho, eles propõem a aproximação, tanto do modelo de ruído fracionário quanto do modelo ARFIMA(p,d,q), pelo modelo ARMA(1,1), comparando as melhores previsões dos dois modelos através do critério de quadrados médios de previsão.

Nessa dissertação, através da decomposição do modelo de ruído fracionário (4.2.3) escrita na forma

$$(1-B)^{d/2}y_t = (1-B)^{-d/2}\epsilon_t, \tag{4.2.4}$$

podemos encontrar a aproximação ARMA(p,q)

$$y_t = -\sum_{k=1}^{p}\phi_k(d/2)y_{t-k} + \epsilon_t + \sum_{j=1}^{q}\theta_j(-d/2)\epsilon_{t-j}, \tag{4.2.5}$$

utilizando o truncamento da representação $AR(\infty)$ e $MA(\infty)$. Assim, os coeficientes $\phi_k(d/2)$ e $\theta_j(-d/2)$ são obtidos recursivamente através de

$$\begin{aligned}
\phi_{k+1}(d/2) &= \frac{(k-d/2)}{k+1}\phi_k(d/2)\ ,\ \phi_0(d/2)=1,\\
\theta_{j+1}(-d/2) &= \frac{(j+d/2)}{j+1}\theta_j(-d/2)\ ,\ \theta_0(-d/2)=1.
\end{aligned}$$

O modelo aproximado ARMA(p,q) em (4.2.5) é baseado na definição das ordens p e q para as representações de média-móvel infinita e autoregressiva infinita. A aproximação do modelo ARFIMA(0,d,0) pelo modelo ARMA(p,q) permitirá uma melhor manipulação algébrica no que diz respeito à análise de influência local. Essa vantagem está associada à quantidade de coeficientes presente no modelo. Assim, nesta dissertação será utilizado $p=q=10$, ou $p+q=20$ coeficientes no lugar de utilizar aproximações AR(40) ou MA(40) que são as opções usualmente encontradas na literatura.

4.3 Análise de Influência Local

Schall e Dunne (1991) utilizaram uma verossimilhança aproximada [Anderson, 1977] para o modelo ARMA(p,q), e assim, através da Curvatura de Cook investigaram a influência local, considerando quatro tipos de esquemas de perturbação.

Baseado na verossimilhança aproximada estudada por Schall e Dunne (1991) e no modelo ARMA(p,q) em (4.2.5), o presente trabalho investiga a influência local através das medidas Inclinação de Billor e Loynes (S) e da Curvatura de Cook (C_c), e de seus respectivos vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$. Será considerado o esquema de perturbação nos dados.

4.3.1 Verossimilhança Aproximada

A idéia é utilizar a aproximação (4.2.5) do modelo ARFIMA(0,d,0), para desta forma, aproveitar os resultados de Schall e Dunne (1991) em modelos ARMA. Supondo que a média do modelo ARMA(p,q) é zero, condicionam-se os primeiros p valores de y_t e os primeiros q valores de ϵ_t, da seguinte maneira,

$$\begin{cases} y_{-p+1} = 0, ..., y_0 = 0 \\ \epsilon_{-q+1} = 0, ..., \epsilon_0 = 0. \end{cases}$$

Seja a matriz $\mathbf{L}^t_{(n\times n)}$ dada por

$$\begin{bmatrix} 0 & 0 \\ \mathbf{I}_{n-t} & 0 \end{bmatrix}, \qquad t = 1, ..., n-1,$$

onde $\mathbf{L}^0 = \mathbf{I}_n$ e $\mathbf{L}^n = \mathbf{O}_n$, sendo $\mathbf{I}_j$ a matriz identidade de ordem $(j \times j)$, $j = 1, \ldots, n$ e $\mathbf{O}_n$ a matriz nula. Assim, são definidas as matrizes $\mathbf{\Phi}_{(n\times n)}$ e $\mathbf{\Theta}_{(n\times n)}$, tal que

$$\mathbf{\Phi} = \sum_{k=0}^{p} \phi_k(d/2)\mathbf{L}^k \quad e \quad \mathbf{\Theta} = \sum_{j=0}^{q} \theta_j(d/2)\mathbf{L}^j,$$

onde $\phi_0(d/2) = 1$, $\theta_0(d/2) = 1$ e $max(p, q) \leq n$. Através de (4.2.5), constrói-se a representação matricial do modelo aproximado ARMA(p,q) para o modelo ARFIMA(0,d,0)

$$\mathbf{\Phi y} = \mathbf{\Theta} \boldsymbol{\epsilon}, \tag{4.3.1}$$

onde $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$, $\mathbf{\Phi} = \mathbf{\Phi}(d)$ e $\mathbf{\Theta} = \mathbf{\Theta}(d)$.

Assumindo que $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$, a matriz de covariância para $\mathbf{y}$ é dada por $Cov[\mathbf{y}] = \sigma^2\mathbf{V}$, onde $\mathbf{V}$ é uma função $V(\mathbf{\Phi}, \mathbf{\Theta})$ tal que $\mathbf{V} = \mathbf{\Phi}^{-1}\mathbf{\Theta}\ \mathbf{\Theta}^T\mathbf{\Phi}^{T^{-1}}$.

Notando que $|\mathbf{\Phi}| = |\mathbf{\Theta}| = 1$, a log-verossimilhança para o modelo (4.2.5) é dada por

$$L(d, \sigma^2) = -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{V}^{-1}\mathbf{y}. \tag{4.3.2}$$

A verossimilhança (4.3.2) se torna uma ferramenta de fundamental importância para a análise de influência local, considerando que o modelo ARFIMA(0,d,0) está sendo aproximado pelo modelo ARMA(p,q), em (4.2.5).

Considerando o esquema de perturbação nos dados, ou seja,

$$\tilde{\mathbf{y}} = \mathbf{y} + \boldsymbol{\omega}, \tag{4.3.3}$$

onde $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^T$, obtém-se a log-verossimilhança perturbada

$$\begin{aligned} L(d, \sigma^2|\boldsymbol{\omega}) &= \frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y}+\boldsymbol{\omega})^T \mathbf{V}^{-1}(\mathbf{y}+\boldsymbol{\omega}) \\ &= \frac{n}{2} log(2\pi) - \frac{n}{2} log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y}+\boldsymbol{\omega})^T \boldsymbol{\Phi}^T(\boldsymbol{\Theta}^T)^{-1}\boldsymbol{\Theta}^{-1}\boldsymbol{\Phi}(\mathbf{y}+\boldsymbol{\omega}) \end{aligned} \tag{4.3.4}$$

Shall e Dunne (1991) apresentaram o resultado do cálculo para a primeira derivada da log-verossimilhança aproximada do modelo ARMA(p,q) em relação a perturbação (4.3.3).

Seguem as expressões necessárias para obter os vetores relativos às medidas de influência local Inclinação de Billor (S) e Curvatura de Cook (C_c). Todos os resultados são avaliados sobre os estimadores de máxima verossimilhança $d = \hat{d}$ e $\sigma^2 = \hat{\sigma}^2$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = \mathbf{0}$ (perturbação nula).

- Inclinação de Billor e Loynes (S):

$$\frac{\partial L(\hat{d}, \hat{\sigma}^2|\boldsymbol{\omega}_0)}{\partial \boldsymbol{\omega}} = -\frac{1}{\hat{\sigma}^2}\mathbf{V}^{-1}\mathbf{y}.$$

- Curvatura de Cook (C_C):

$$\begin{aligned} \frac{\partial^2 L(\hat{d}, \hat{\sigma}^2|\boldsymbol{\omega}_0)}{\partial \boldsymbol{\omega} \partial d} &= \frac{1}{\hat{\sigma}^2}\mathbf{V}^{-1}\left[\frac{\partial \mathbf{V}}{\partial d}\right]\mathbf{V}^{-1}\mathbf{y}, \\ \frac{\partial^2 L(\hat{d}, \hat{\sigma}^2|\boldsymbol{\omega}_0)}{\partial \boldsymbol{\omega} \partial \sigma^2} &= \frac{1}{\hat{\sigma}^4}\mathbf{V}^{-1}\mathbf{y}. \end{aligned}$$

As segundas derivadas em relação aos parâmetros são:

$$\begin{aligned}
\frac{\partial^2 L(\hat{d},\hat{\sigma}^2|\boldsymbol{\omega}_0)}{\partial d\partial d} &= \frac{1}{2\hat{\sigma}^2}\mathbf{y}^T\left\{-2\mathbf{V}^{-1}\left[\frac{\partial \mathbf{V}}{\partial d}\right]\mathbf{V}^{-1}\left[\frac{\partial \mathbf{V}}{\partial d}\right]\mathbf{V}^{-1}+\mathbf{V}^{-1}\left[\frac{\partial^2 \mathbf{V}}{\partial d\partial d}\right]\mathbf{V}^{-1}\right\}\mathbf{y},\\
\frac{\partial^2 L(\hat{d},\hat{\sigma}^2|\boldsymbol{\omega}_0)}{\partial \sigma^2\partial \sigma^2} &= -\frac{n}{2\hat{\sigma}^4},\\
\frac{\partial^2 L(\hat{d},\hat{\sigma}^2|\boldsymbol{\omega}_0)}{\partial d\partial \sigma^2} &= -\frac{1}{2\hat{\sigma}^4}\mathbf{y}^T\mathbf{V}^{-1}\left[\frac{\partial \mathbf{V}}{\partial d}\right]\mathbf{V}^{-1}\mathbf{y},
\end{aligned}$$

tal que

$$\begin{aligned}
\frac{\partial \mathbf{V}}{\partial d} &= \frac{\partial}{\partial d}\left\{\boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\ \boldsymbol{\Theta}^T(\boldsymbol{\Phi}^T)^{-1}\right\} = -\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\ \boldsymbol{\Theta}^T(\boldsymbol{\Phi}^T)^{-1}+\\
&+ \boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]\boldsymbol{\Theta}^T(\boldsymbol{\Phi}^T)^{-1}+\boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]^T(\boldsymbol{\Phi}^T)^{-1}-\boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\ \boldsymbol{\Theta}^T\left[\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\right]^T,
\end{aligned}$$

que pode ser escrito como

$$\frac{\partial^2 \mathbf{V}}{\partial d\partial d} = \frac{\partial \mathbf{V}_1}{\partial d}+\frac{\partial \mathbf{V}_2}{\partial d}+\frac{\partial \mathbf{V}_3}{\partial d}+\frac{\partial \mathbf{V}_4}{\partial d},$$

onde

$$\begin{aligned}
\frac{\partial \mathbf{V}_1}{\partial d} &= 2\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\ \boldsymbol{\Theta}^T(\boldsymbol{\Phi}^T)^{-1}-\boldsymbol{\Phi}^{-1}\left[\frac{\partial^2 \boldsymbol{\Phi}}{\partial d\partial d}\right]\boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\ \boldsymbol{\Theta}^T(\boldsymbol{\Phi}^T)^{-1}-\\
&- \boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]\boldsymbol{\Theta}^T(\boldsymbol{\Phi}^T)^{-1}-\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]^T(\boldsymbol{\Phi}^T)^{-1}+\\
&+ \boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\ \boldsymbol{\Theta}^T\left[\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\right]^T,\\
\frac{\partial \mathbf{V}_2}{\partial d} &= -\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]\boldsymbol{\Theta}^T(\boldsymbol{\Phi}^T)^{-1}+\boldsymbol{\Phi}^{-1}\left[\frac{\partial^2 \boldsymbol{\Theta}}{\partial d\partial d}\right]\boldsymbol{\Theta}^T(\boldsymbol{\Phi}^T)^{-1}+\\
&+ \boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]^T(\boldsymbol{\Phi}^T)^{-1}-\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]\boldsymbol{\Theta}^T\left[\Phi^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\Phi^{-1}\right]^T,\\
\frac{\partial \mathbf{V}_3}{\partial d} &= -\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]^T(\boldsymbol{\Phi}^T)^{-1}+\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]^T(\boldsymbol{\Phi}^T)^{-1}+\\
&+ \boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\left[\frac{\partial^2 \boldsymbol{\Theta}}{\partial d\partial d}\right]^T(\boldsymbol{\Phi}^T)^{-1}-\boldsymbol{\Phi}^{-1}\boldsymbol{\Theta}\left[\frac{\partial \boldsymbol{\Theta}}{\partial d}\right]^T\left[\boldsymbol{\Phi}^{-1}\left[\frac{\partial \boldsymbol{\Phi}}{\partial d}\right]\boldsymbol{\Phi}^{-1}\right]^T,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathbf{V}_4}{\partial d} &= \mathbf{\Phi}^{-1}\left[\frac{\partial \mathbf{\Phi}}{\partial d}\right]\mathbf{\Phi}^{-1}\mathbf{\Theta}\ \mathbf{\Theta}^T\left[\mathbf{\Phi}^{-1}\left[\frac{\partial \mathbf{\Phi}}{\partial d}\right]\mathbf{\Phi}^{-1}\right]^T - \\
&- \mathbf{\Phi}^{-1}\left[\frac{\partial \mathbf{\Theta}}{\partial d}\right]\mathbf{\Theta}^T\left[\mathbf{\Phi}^{-1}\left[\frac{\partial \mathbf{\Phi}}{\partial d}\right]\mathbf{\Phi}^{-1}\right]^T - \mathbf{\Phi}^{-1}\mathbf{\Theta}\left[\frac{\partial \mathbf{\Theta}}{\partial d}\right]^T\left[\mathbf{\Phi}^{-1}\left[\frac{\partial \mathbf{\Phi}}{\partial d}\right]\mathbf{\Phi}^{-1}\right]^T + \\
&+ 2\mathbf{\Phi}^{-1}\mathbf{\Theta}\ \mathbf{\Theta}^T\left[\mathbf{\Phi}^{-1}\left[\frac{\partial \mathbf{\Phi}}{\partial d}\right]\mathbf{\Phi}^{-1}\left[\frac{\partial \mathbf{\Phi}}{\partial d}\right]\mathbf{\Phi}^{-1}\right]^T - \\
&- \mathbf{\Phi}^{-1}\mathbf{\Theta}\ \mathbf{\Theta}^T\left[\mathbf{\Phi}^{-1}\left[\frac{\partial \mathbf{\Phi}^2}{\partial d\partial d}\right]\mathbf{\Phi}^{-1}\right]^T.
\end{aligned}$$

Para calcular $\frac{\partial \mathbf{V}_1}{\partial d}$, $\frac{\partial \mathbf{V}_2}{\partial d}$, $\frac{\partial \mathbf{V}_3}{\partial d}$, e $\frac{\partial \mathbf{V}_4}{\partial d}$, serão utilizadas as seguintes expressões recursivas:

$$\begin{aligned}
\frac{\partial \phi_{k+1}(d/2)}{\partial d} &= -\frac{1}{2(k+1)}\phi_k(d/2) + \left(\frac{k-d/2}{k+1}\right)\left[\frac{\partial \phi_k(d/2)}{\partial d}\right], \\
\frac{\partial \theta_{j+1}(d/2)}{\partial d} &= \frac{1}{2(j+1)}\theta_j(d/2) + \left(\frac{j+d/2}{j+1}\right)\left[\frac{\partial \theta_j(d/2)}{\partial d}\right], \\
\frac{\partial^2 \phi_{k+1}(d/2)}{\partial d\partial d} &= -\frac{1}{k+1}\left[\frac{\partial \phi_k(d/2)}{\partial d}\right] + \left(\frac{k-d/2}{k+1}\right)\left[\frac{\partial^2 \phi_k(d/2)}{\partial d\partial d}\right], \\
\frac{\partial^2 \theta_{j+1}(d/2)}{\partial d\partial d} &= \frac{1}{j+1}\left[\frac{\partial \theta_j(d/2)}{\partial d}\right] + \left(\frac{j+d/2}{j+1}\right)\left[\frac{\partial^2 \theta_j(d/2)}{\partial d\partial d}\right],
\end{aligned}$$

onde $\frac{\partial \phi_0(d/2)}{\partial d} = 0$, $\frac{\partial \theta_0(d/2)}{\partial d} = 0$, $\frac{\partial^2 \phi_0(d/2)}{\partial d\partial d} = 0$ e $\frac{\partial^2 \theta_0(d/2)}{\partial d\partial d} = 0$.

A partir dos resultados derivados nesta seção, será possível realizar o diagnóstico de influência local para o modelo ARMA(p,q) aproximado, ou seja, para o modelo ARFIMA(0,d,0), considerando as medidas Inclinação de Billor e Loynes e Curvatura de Cook calculadas em (2.2.3) e (2.1.7), seus respectivos vetores de diagnóstico obtidos em (2.2.4) e (2.1.8), e o esquema de perturbação nos dados.

4.4 Simulações

A seguir será apresentado o diagnóstico de influência local, baseado na técnica de limiares apresentada na Seção 2.3, para séries temporais ARFIMA(0,d,0), considerando o esquema de perturbação nos dados (4.3.3). O objetivo das simulações é estudar o comportamento das marcas de referência para as estatísticas Inclinação de Billor e Loynes

(S), a Curvatura de Cook (C_c) e os vetores $\mathbf{v}$ e $\mathbf{v}_c$, através dos critérios M_0, M_1 e M_2, para diferentes coeficientes de diferença fracionária d e diferentes tamanhos amostrais n.

As simulações são baseadas no modelo de ruído fracionário (4.2.3). Sete séries temporais foram geradas considerando os tamanhos amostrais $n = 250$ e $n = 500$, e parâmetros de diferença fracionária $d = 0,2$, $d = 0,4$ e $d = 0,45$, com $\sigma^2 = 1$. Essas séries serão denominadas conforme a Tabela 4.4.1.

Considerando cada uma das séries com tamanho amostral $n = 250$, foram introduzidas duas observações "aberrantes" (*outliers* positivos e negativos), com magnitudes de cinco e três desvios padrões (σ_y), nas componentes y_{50} e y_{200}, respectivamente. De maneira análoga, para as séries com tamanho amostral $n = 500$, as observações y_{150} e y_{450} são inseridas como pontos "aberrantes" (*outliers* positivos e negativos) com magnitudes de cinco e três desvios padrões, respectivamente. Adicionalmente, a série 2a representa a série 2 com um *outlier* (positivo e negativo) de magnitude $3\sigma_y$ na posição y_{245}, de forma a verificar se a técnica de limiares (M_0, M_1 e M_2) identifica essa observação nos extremos da série.

Tabela 4.4.1: Séries Simuladas.

	Tamanho Amostral	Grau de Liberdade	Posição do *Outlier*	Tamanho do *Outlier* em Valor Absoluto
Série 1	$n = 250$	$d = 0,2$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 2	$n = 250$	$d = 0,4$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 2a	$n = 250$	$d = 0,4$	y_{245}	$3\sigma_y$
Série 3	$n = 250$	$d = 0,45$	y_{50} e y_{200}	$5\sigma_y$ e $3\sigma_y$
Série 4	$n = 500$	$d = 0,2$	y_{150} e y_{450}	$5\sigma_y$ e $3\sigma_y$
Série 5	$n = 500$	$d = 0,4$	y_{150} e y_{450}	$5\sigma_y$ e $3\sigma_y$
Série 6	$n = 500$	$d = 0,45$	y_{150} e y_{450}	$5\sigma_y$ e $3\sigma_y$

Legenda: σ_y é o desvio padrão da amostra gerada.

Para as séries simuladas é realizada a análise de influência local utilizando os limiares descritos na Seção 2.3 e os cálculos das estatísticas Inclinação de Billor e Loynes (S) e Curvatura de Cook (C_c) na Seção 4.3 com $p = q = 10$, para assim, identificar as observações influentes, ou seja, os *outliers*. As séries apresentadas na Tabela 4.4.1 foram escolhidas de tal maneira que as estimativas de d estivessem próximas aos verdadeiros valores (ver Tabela 4.4.7). Os limiares foram calculados em princípio supondo conhecidos os parâmetros do modelo e gerados com um programa escrito no *softwares MATLAB* utilizando 1000 replicações.

A Tabela 4.4.2 apresenta o esquema de perturbação inserido para as séries 1, 2, 5 e 2a, e auxilia na análise das observações influentes. Adicionalmente, nas Tabelas 4.4.3 - 4.4.5 são apresentadas as marcas de referência correspondentes ao critério global M_0 e aos critérios individuais M_1 e M_2, respectivamente.

Tabela 4.4.2: Perturbação Positiva e Negativa no modelo ARFIMA(0,d,0).

	y_{48}	y_{49}	y_{50}	y_{51}	y_{52}	ω_{50}	$\tilde{y}_{50} = 5\sigma_y$
Série 1	0,445	-0,101	-0,004	0,518	0,656	4,864	4,859
Série 1	0,445	-0,101	-0,004	0,518	0,656	-4,854	-4,859
Série 2	-0,599	0,694	-0,354	0,612	-1,213	6,346	5,992
Série 2	-0,599	0,694	-0,354	0,612	-1,213	-5,638	-5,992
	y_{198}	y_{199}	y_{200}	y_{201}	y_{202}	ω_{200}	$\tilde{y}_{200} = 3\sigma_y$
Série 1	-0,738	-1,288	0,927	1,060	-0,153	1,987	2,915
Série 1	-0,738	-1,288	0,927	1,060	-0,153	-3,843	-2,915
Série 2	0,471	-0,103	0,876	-0,652	1,776	2,718	3,595
Série 2	0,471	-0,103	0,876	-0,652	1,776	-4,472	-3,595
	y_{144}	y_{149}	y_{150}	y_{151}	y_{152}	ω_{150}	$\tilde{y}_{150} = 5\sigma_y$
Série 5	0,672	0,336	-0,863	-0,884	-0,309	6,620	5,756
Série 5	0,672	0,336	-0,863	-0,884	-0,309	-4,892	-5,756
	y_{448}	y_{449}	y_{450}	y_{451}	y_{452}	ω_{450}	$\tilde{y}_{450} = 3\sigma_y$
Série 5	0,683	-1,814	-1,619	-1,024	-1,005	5,073	3,453
Série 5	0,683	-1,814	-1,619	-1,024	-1,005	-1,834	-3,453
	y_{243}	y_{244}	y_{245}	y_{246}	y_{247}	ω_{245}	$\tilde{y}_{245} = 3\sigma_y$
Série 2a	0,843	2,263	1,106	0,087	0,167	2,488	3,595
Série 2a	0,843	2,263	1,106	0,087	0,167	-4,702	-3,595

Tabela 4.4.3: Estatísticas Globais M_0.

Tamanho Amostral	Medida	Percentis (%) \ $\boldsymbol{d}$	0,2	0,4	0,45
		90	34,4460	36,7992	37,4881
$n = 250$	Inclinação (S)	95	34,9235	37,3530	38,0493
		99	35,8870	38,6087	39,3201
		90	5,2443	6,9455	8,7955
$n = 250$	Curvatura (C_c)	95	5,3472	7,4042	11,5663
		99	5,6540	9,3863	30,7706
		90	47,8253	50,9277	52,1464
$n = 500$	Inclinação (S)	95	48,4592	51,6505	52,7918
		99	49,6662	52,9162	53,8638
		90	5,0822	6,6583	8,8893
$n = 500$	Curvatura (C_c)	95	5,1845	7,1779	12,4837
		99	5,4471	8,4883	30,1185

Tabela 4.4.4: Estatísticas Individuais M_1.

Tamanho Amostral	Medida	Percentis (%) \ $\boldsymbol{d}$	0,2	0,4	0,45
		90	0,2247	0,2240	0,2231
$n = 250$	Inclinação (S)	95	0,2345	0,2347	0,2357
		99	0,2539	0,2506	0,2554
		90	0,2221	0,2215	0,2220
$n = 250$	Curvatura (C_c)	95	0,2347	0,2321	0,2334
		99	0,2563	0,2507	0,2542
		90	0,1651	0,1633	0,1666
$n = 500$	Inclinação (S)	95	0,1724	0,1709	0,1728
		99	0,1937	0,1859	0,1889
		90	0,1648	0,1638	0,1669
$n = 500$	Curvatura (C_c)	95	0,1722	0,1706	0,1735
		99	0,1906	0,1834	0,1864

Tabela 4.4.5: Estatísticas Individuais M_2.

Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{d}$	0,2	0,4	0,45
		Mínimo		0,1541	0,1546	0,1525
		Quantil 5%		0,1544	0,1567	0,1555
$n = 250$	Inclinação (S)	Máximo		0,2485	0,2365	0,2356
		Média		0,1889	0,1916	0,1916
		Desvio Padrão		0,0194	0,0205	0,0193
		Mínimo		0,1598	0,1440	0,1466
		Quantil 5%		0,1608	0,1536	0,1499
$n = 250$	Curvatura (C_c)	Máximo		0,2615	0,2263	0,2582
		Média		0,1902	0,1894	0,1954
		Desvio Padrão		0,0195	0,0177	0,0216
		Mínimo		0,1194	0,1207	0,1213
		Quantil 5%		0,1211	0,1218	0,1223
$n = 500$	Inclinação (S)	Máximo		0,1723	0,1796	0,1735
		Média		0,1427	0,1453	0,1466
		Desvio Padrão		0,0135	0,0146	0,0138
		Mínimo		0,1249	0,1178	0,1223
		Quantil 5%		0,1250	0,1190	0,1232
$n = 500$	Curvatura (C_c)	Máximo		0,1789	0,1736	0,1892
		Média		0,1432	0,1452	0,1469
		Desvio Padrão		0,0122	0,0126	0,0160

Tabela 4.4.6: Estimativas de S e C_c para as séries da Tabela 4.4.1.

Tamanho Amostral	Medida (%)	$\boldsymbol{d}$	0,2	0,4	0,45
$n = 250$	Inclinação (S)		**35,5390**	**38,8270**	**39,8750**
$n = 250$	Curvatura (C_c)		**5,5698**	6,7229	7,3491
$n = 500$	Inclinação (S)		**50,0100**	**52,6140**	**53,1750**
$n = 500$	Curvatura (C_c)		**5,4338**	6,5229	6,9159

Legenda: Em negrito estão representadas as estatísticas significativas para M_0 (95%).

Na Tabela 4.4.3, se o parâmetro de diferença fracionária d e percentil são mantidos fixos, os limiares de S baseados no critério M_0 aumentam quando o tamanho amostral aumenta. Ao contrário, para C_c, mantendo-se constantes o percentil e d, as marcas de nível decrescem quando o tamanho amostral aumenta. Percebe-se também que para tamanhos amostrais e percentis fixos, se d aumenta os limiares de S e C_c aumentam pouco. Assim, existem indícios de que os limiares baseados no critério global M_0 não são robustos quanto a variação e/ou estimação de d.

Na Tabela 4.4.4, fixando d e percentil, os limiares M_1 de S e C_c diminuem quando o tamanho amostral aumenta. No entanto, fixados n e percentil, os limiares para S e C_c permanecem constantes quando d aumenta, ou seja, existe indício de que o critério M_1 é robusto quanto a variação e/ou estimação do parâmetro de diferença fracionária.

Na Tabela 4.4.5, se o tamanho amostral varia de $n = 250$ para $n = 500$ com d constante, tanto M_2 (quantil 5%) quanto as demais estatísticas, têm suas magnitudes diminuídas. Agora, se d aumenta, para um fixo tamanho amostral, M_2 e as demais estatísticas de S e C_c permanecem praticamente inalteradas. Assim como M_1, o critério M_2 apresenta indícios de robustez, dada a variação e/ou estimação de d.

A Tabela 4.4.6 apresenta as estatísticas S e C_c das sete séries. Comparando com o critério M_0 (Tabela 4.4.3) para avaliar a influência local das 7 séries geradas, todas as estatísticas S são consideradas significativas. Já as estatísticas C_c são consideradas significativas sob o critério M_0 (Tabela 4.4.3), somente quando $d = 0,2$. Ou seja, quando $d = 0,4$ e $d = 0,45$ o critério global M_0 não acusa influência.

Os vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$, relativos as séries são ilustrados nas Figuras 4.4.1 - 4.4.8. Essas figuras mostram as séries geradas 1, 2, 5 e 2a em conjunção aos limiares M_1 e M_2 para avaliar a influência de cada observação a partir dos vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$. Em cada uma dessas Figuras o gráfico superior corresponde à série sem *outlier*. O segundo gráfico corresponde à série em valores absolutos com *outlier*.

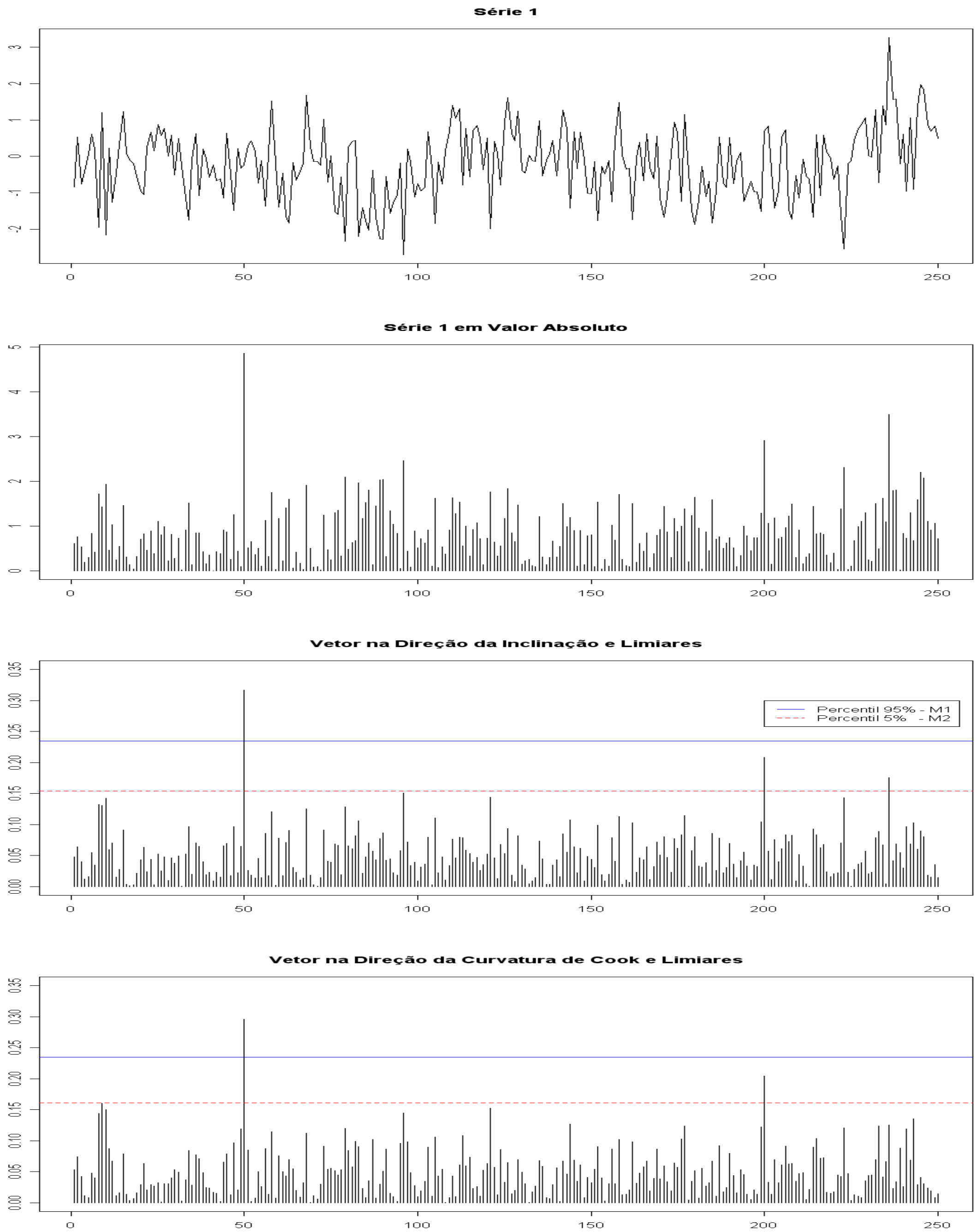

Figura 4.4.1: Influência local na série 1 ($d = 0,2$ e $n = 250$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

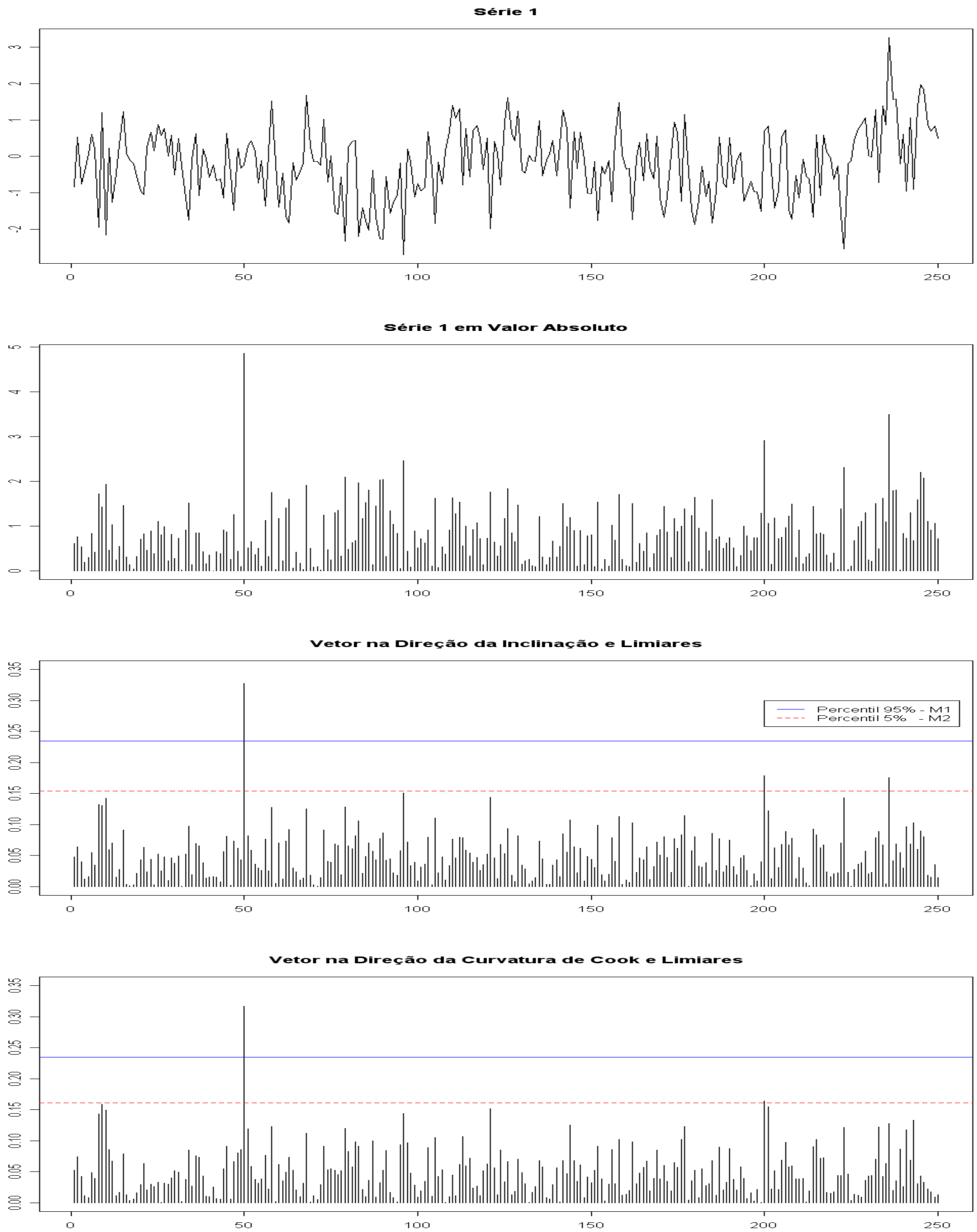

Figura 4.4.2: Influência local na série 1 ($d = 0,2$ e $n = 250$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

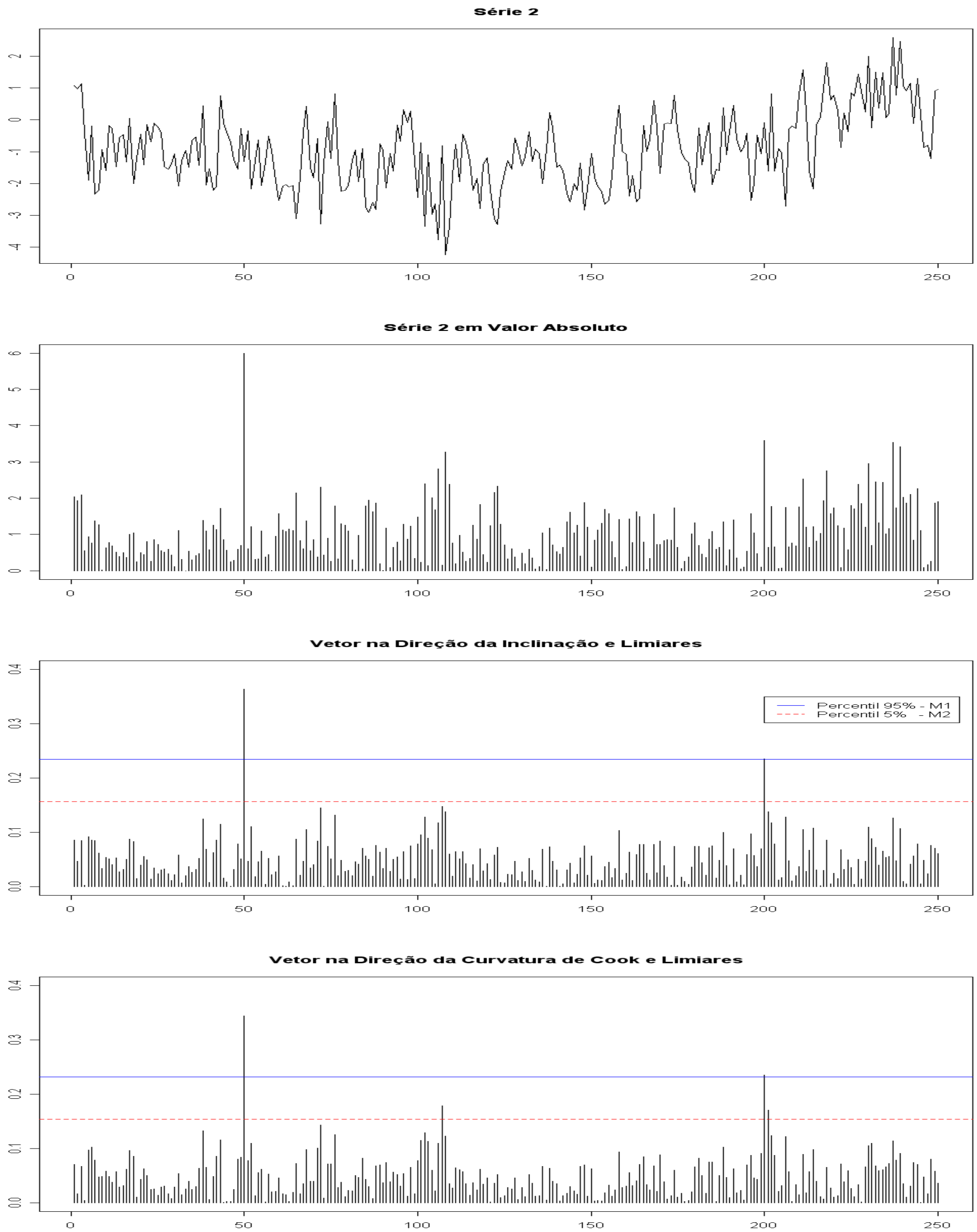

Figura 4.4.3: Influência local na série 2 ($d = 0,4$ e $n = 250$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

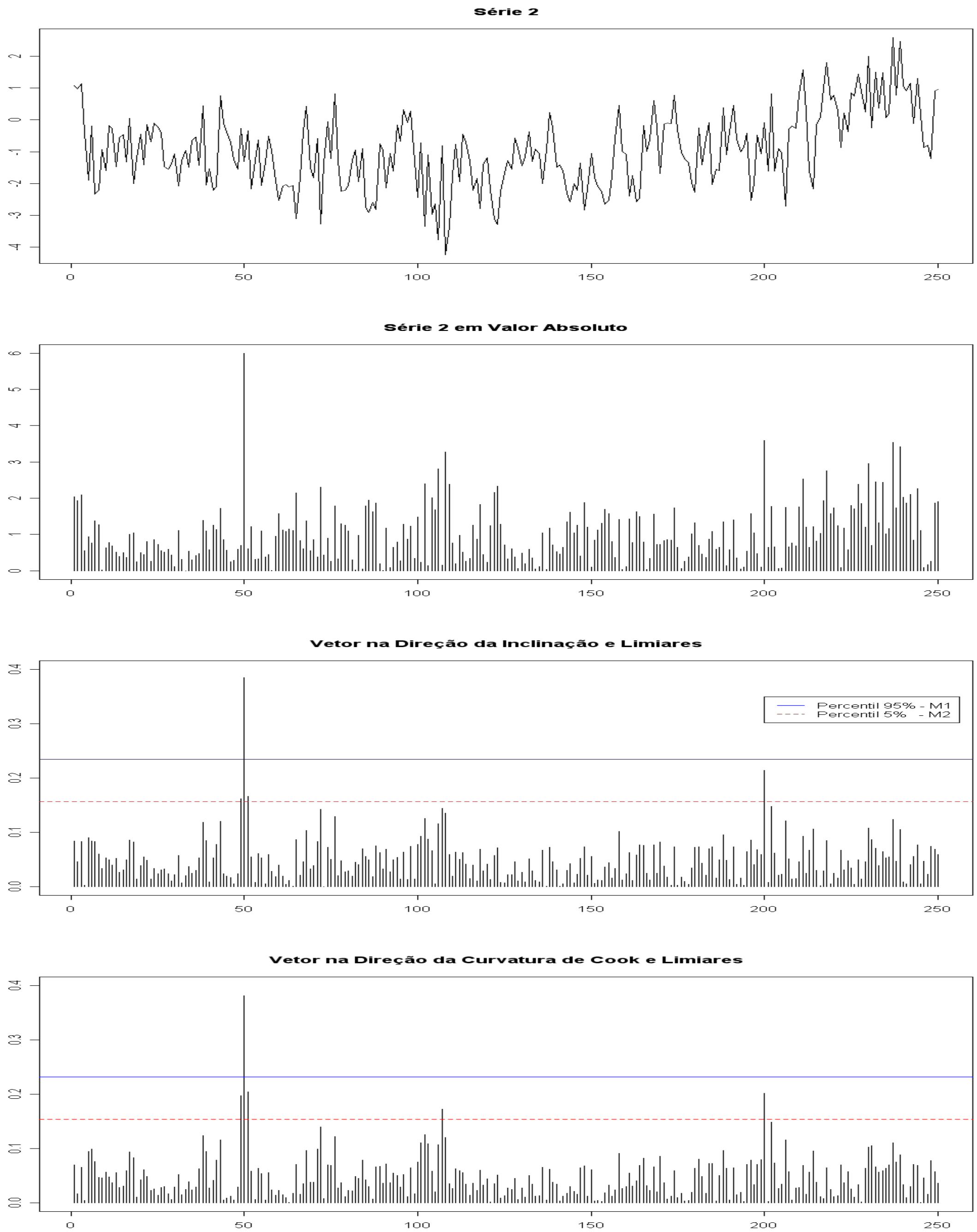

Figura 4.4.4: Influência local na série 2 ($d = 0,4$ e $n = 250$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

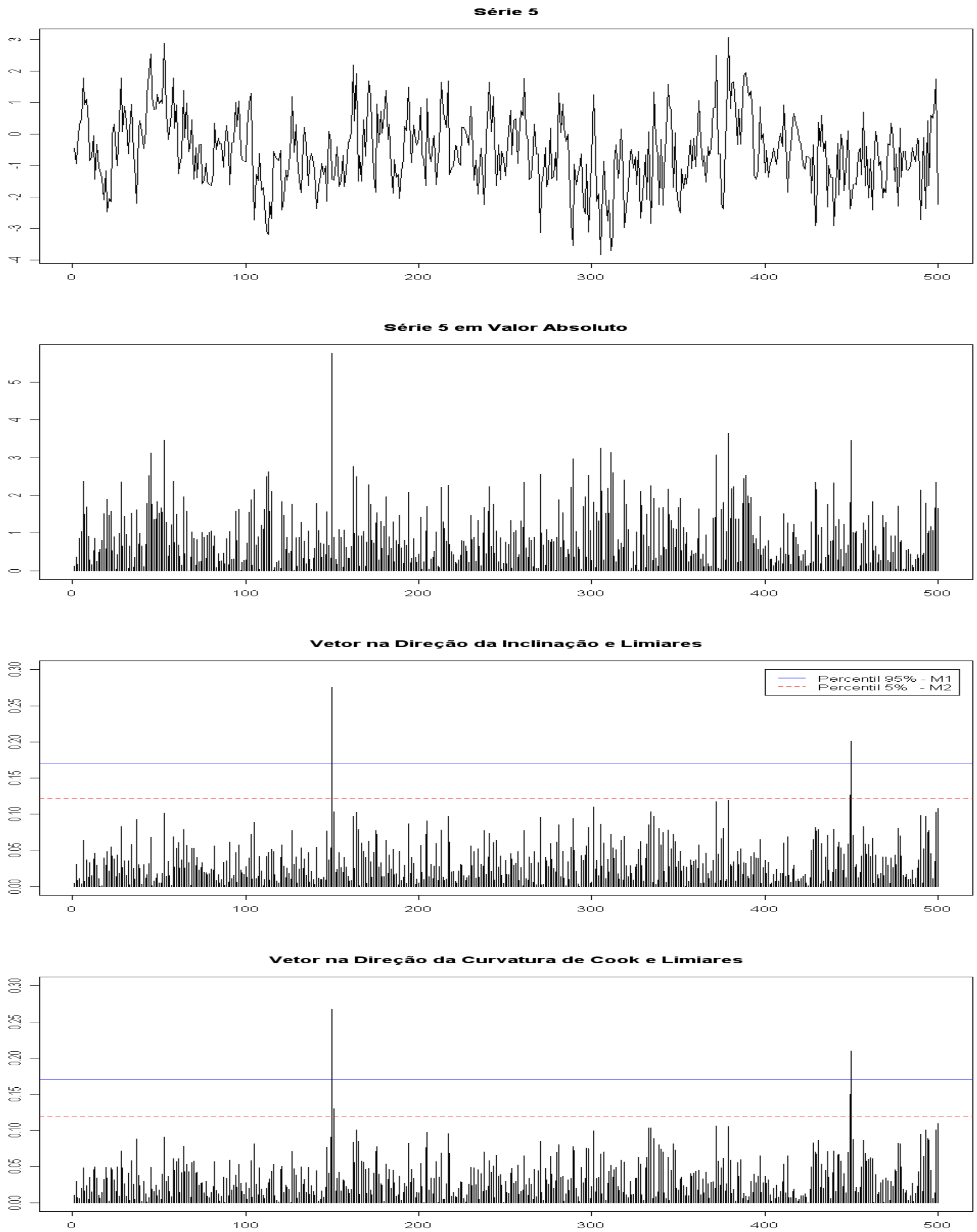

Figura 4.4.5: Influência local na série 5 ($d = 0,4$ e $n = 500$) com *outliers* positivos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

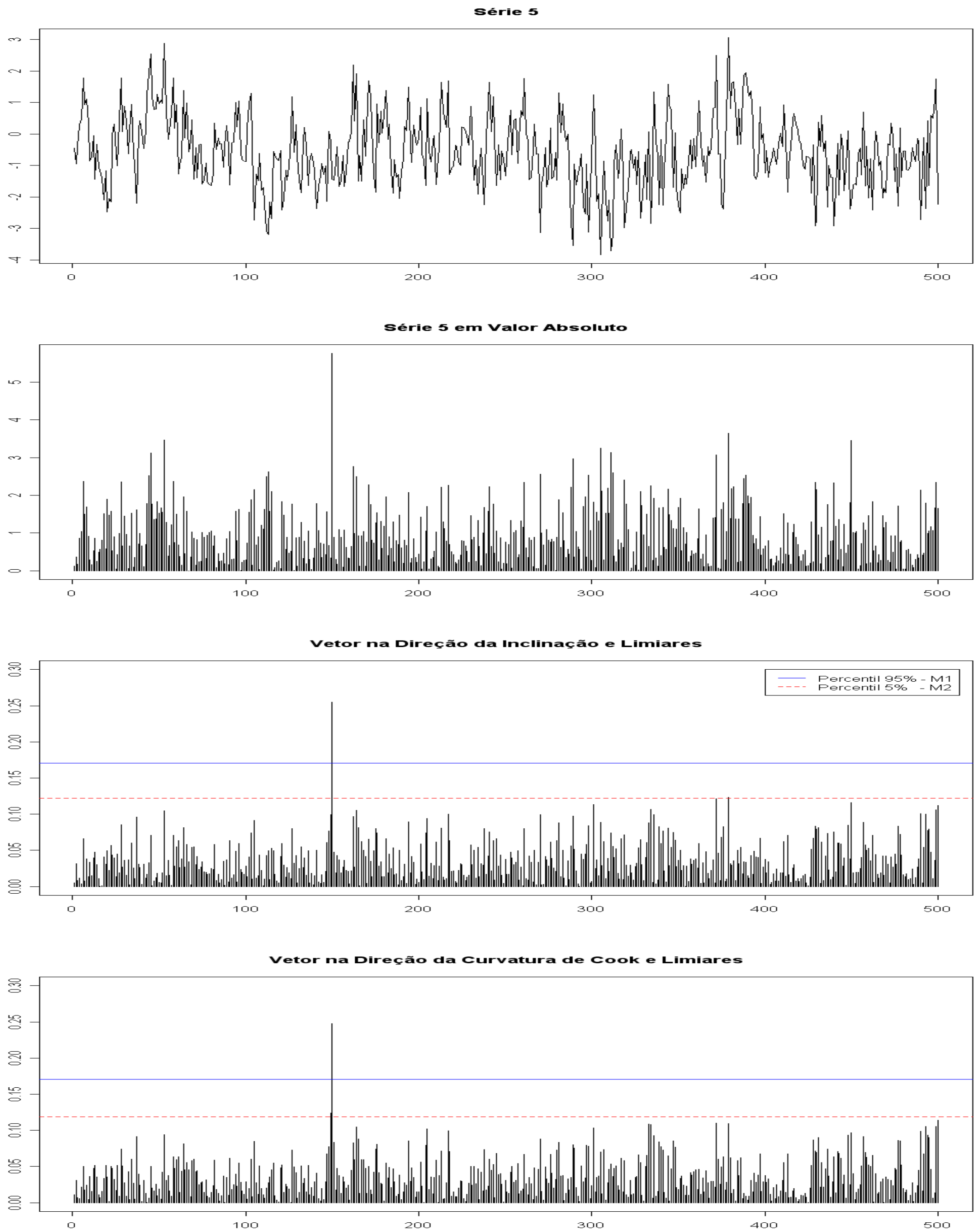

Figura 4.4.6: Influência local na série 5 ($d = 0,4$ e $n = 500$) com *outliers* negativos. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

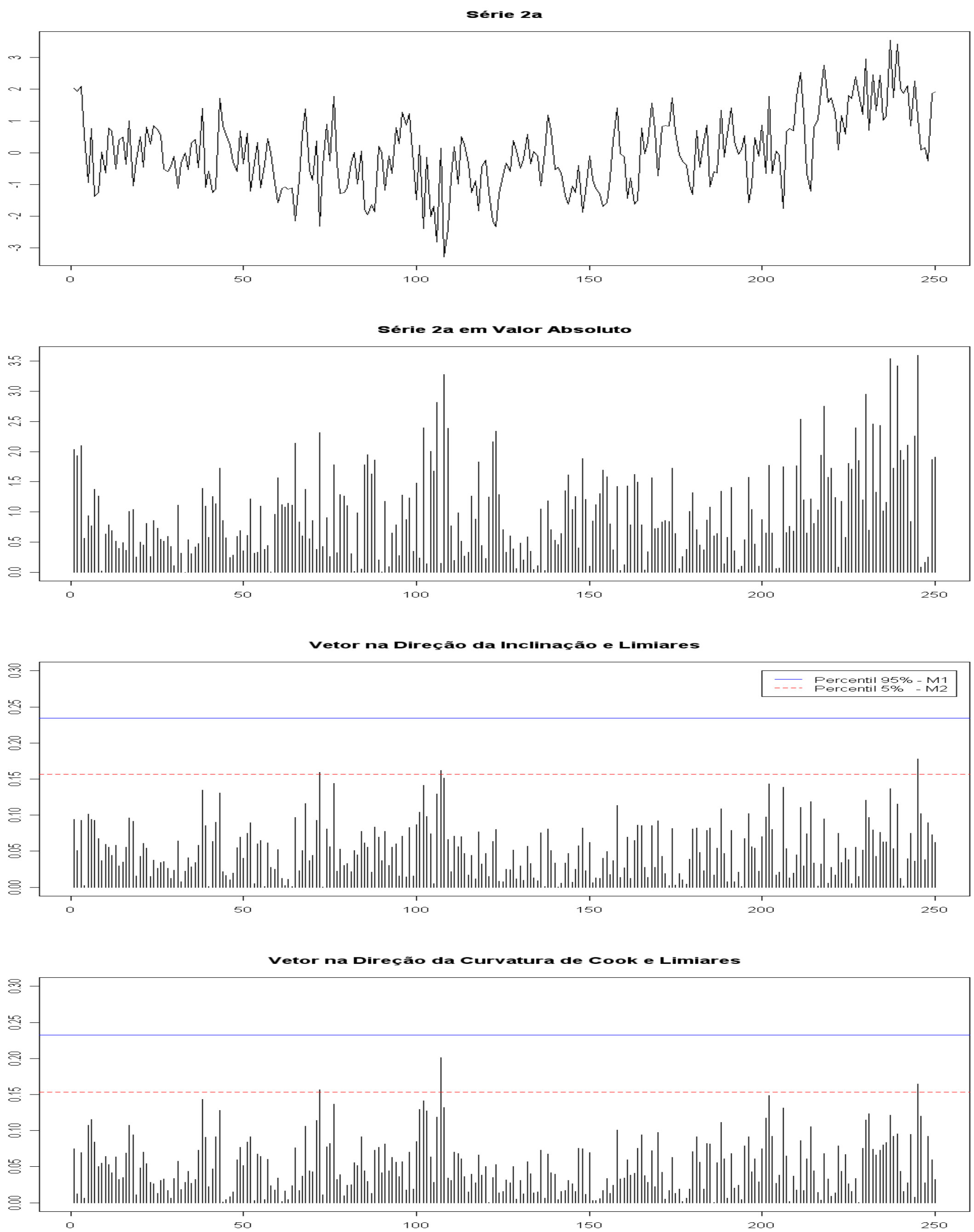

Figura 4.4.7: Influência local na série 2a ($d = 0,4$ e $n = 250$) com *outlier* positivo em y_{245}. O primeiro gráfico corresponde a série sem *outlier*. O segundo gráfico corresponde a série em valor absoluto com *outlier*.

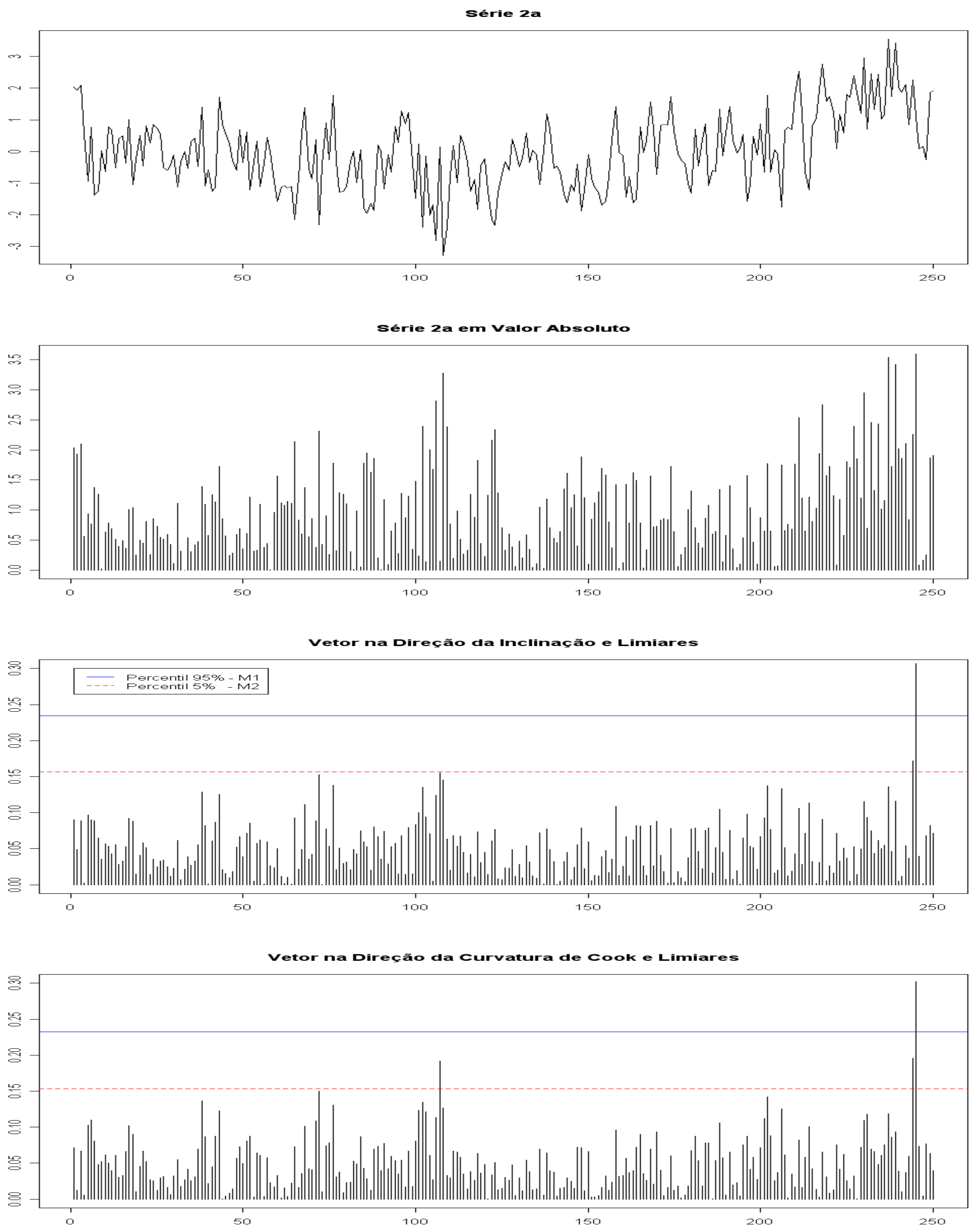

Figura 4.4.8: Influência local na série 2a ($d = 0,4$ e $n = 250$) com *outlier* negativo em y_{245}.

Na Figura 4.4.1 para a série 1, dados os vetores de diagnóstico de S e C_c e considerando a perturbação positiva $\tilde{y}_{50} = 4,859$, a componente 50 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada pelos critérios M_1 e M_2. Isso ocorre devido a magnitude de $\tilde{y}_{50} = 4,859$ ser muito maior do que $y_{50} = -0,004$ (Tabela 4.4.2). A componente 200 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada somente pelo critério M_2. Para a mesma série se a perturbação é negativa $\tilde{y}_{50} = -4,859$ (Figura 4.4.2), o comportamento da componente 50 de $\mathbf{v}$ e $\mathbf{v}_c$ é idêntico ao observado na Figura 4.4.1. Apesar do valor absoluto da diferença entre $\tilde{y}_{200} = -2,915$ e $y_{200} = 0,927$ ser maior quando comparado ao valor absoluto da diferença entre $\tilde{y}_{200} = 2,915$ e $y_{200} = 0,927$ (ver Tabela 4.4.2), a dinâmica da série permanece a mesma, no sentido que se t aumenta de 198 a 200 y_t diminui, e dessa maneira, a componente 200 de $\mathbf{v}$ e $\mathbf{v}_c$ diminui (Figura 4.4.2).

Para a série 2 na Figura 4.4.3, quando as perturbações são positivas, tem-se que as componentes 50 e 200 de $\mathbf{v}$ e $\mathbf{v}_c$ são identificadas pelos critérios M_1 e M_2. Na Figura 4.4.4 considerando a mesma série, quando as perturbações são negativas, a componente 50 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada pelos critérios M_1 e M_2, enquanto a componente 200 é identificada somente por M_2. Novamente, a identificação das componentes dos vetores de diagnóstico se baseia na magnitude do valor absoluto da diferença entre $\tilde{y}_{50}$ e y_{50}, e entre $\tilde{y}_{200}$ e y_{200} (Tabela 4.4.2).

Para a série 5 na Figura 4.4.5, em virtude da perturbação positiva e do "grande" valor absoluto da diferença existente entre $\tilde{y}_{150}$ e y_{150}, e entre $\tilde{y}_{450}$ e y_{450} (Tabela 4.4.2), as componentes 150 e 450 de $\mathbf{v}$ e $\mathbf{v}_c$ são identificadas pelos critérios M_1 e M_2. Para a mesma série quando a perturbação é negativa (Figura 4.4.6), somente a componente 150 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada pelos critérios M_1 e M_2. Visto que o valor absoluto da diferença entre $\tilde{y}_{450} = -3,453$ e $y_{450} = -1,619$ é menor quando comparado ao valor absoluto da diferença entre $\tilde{y}_{450} = 3,453$ e $y_{450} = -1,619$ (ver Tabela 4.4.2), então a componente 450 não é identificada por nenhum dos critérios individuais M_1 ou M_2(Figura 4.4.6).

Na Figura 4.4.7 para a série 2a, considerando a perturbação positiva $\tilde{y}_{245} = 3,595$, a componente 245 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada somente pelo critério M_2. Para a mesma série, quando a perturbação é negativa a componente 245 de $\mathbf{v}$ e $\mathbf{v}_c$ é identificada pelos critérios M_1 e M_2. Isso ocorre devido ao valor absoluto da diferença entre $\tilde{y}_{245} = -3,595$ e $y_{245} = 1,106$ ser maior do que o valor absoluto da diferença entre $\tilde{y}_{245} = 3,595$ e

$y_{245} = 1,106$ (ver Tabela 4.4.2).

Nas Figuras 4.4.1 e 4.4.2, percebe-se que a observação y_{236} apresenta magnitude da mesma ordem que a observação y_{200}, e também é identificada como influente pelo critério M_2 considerando o vetor de diagnóstico $\mathbf{v}$. Esse comportamento também é verificado através das componentes 107 de $\mathbf{v}_c$ nas Figuras 4.4.3, 4.4.4, 4.4.7 e 4.4.8. Portanto, com M_2 seria possível reconhecer observações que realmente não foram geradas como influentes.

Assim como na Seção 3.5, o "reflexo" foi verificado através das componentes 49 e 51 de $\mathbf{v}$ e $\mathbf{v}_c$ na Figura 4.4.4, onde o *outlier* inserido é negativo $\tilde{y}_{50} = -5,992$. Isso ocorre porque a dinâmica da série 2 é alterada (Tabela 4.4.2). Por outro lado, considerando o *outlier* positivo $\tilde{y}_{50} = 5,992$ não ocorre o "reflexo" (Figura 4.4.3), pois mesmo quando inserido o *outlier* a dinâmica crescente da série 2 é respeitada.

Considerando a análise de influência local para a série 5 quando o *outlier* é positivo (Figura 4.4.5), percebe-se o "reflexo" na componente 449 de $\mathbf{v}_c$ devido a quebra da dinâmica da série entre $y_{449} = -1,814$ e $\tilde{y}_{450} = 3,453$ (Tabela 4.4.2). Ao contrário, quando se insere um *outlier* negativo $\tilde{y}_{450} = -3,453$ a componente 450 de $\mathbf{v}$ e $\mathbf{v}_c$ não é identificada (Figura 4.4.6), pois a dinâmica da série 5 é respeitada.

Considerando a Figura 4.4.8, se o *outlier* é negativo a diferença existente entre $\tilde{y}_{245}$ e y_{244} (Tabela 4.4.2) altera a dinâmica da série 2a, fazendo com que a componente 244 de $\mathbf{v}$ e $\mathbf{v}_c$ seja identificada pelo critério M_2.

Na prática, os limiares M_0, M_1 e M_2 são calculados com base nos parâmetros estimados. Assim, são apresentadas na Tabela 4.4.7 as estimativas de d e σ^2 para as séries 1 - 6 perturbadas pelas observações aberrantes. Utilizando o estimador de $Whittle$, foram calculados os estimadores de máxima verossimilhança $\hat{d}$ e $\hat{\sigma}^2$.

A Tabela 4.4.8 apresenta os limiares M_0 calculados a partir dos verdadeiros valores dos parâmetros e a partir dos valores estimados dos parâmetros. As tabelas referentes as marcas de nível M_1 e M_2 serão omitidas, visto que os resultados obtidos são similares aos valores das tabelas 4.4.4 e 4.4.5.

Tabela 4.4.7: Estimativas dos parâmetros d e σ^2.

Séries	Parâmetros			Estimativas	
Série 1	$n = 250$	$d = 0,2$	$\sigma^2 = 1$	$\hat{d} = 0,2230$	$\hat{\sigma}^2 = 0,9863$
Série 2	$n = 250$	$d = 0,4$	$\sigma^2 = 1$	$\hat{d} = 0,3811$	$\hat{\sigma}^2 = 1,1711$
Série 3	$n = 250$	$d = 0,45$	$\sigma^2 = 1$	$\hat{d} = 0,4002$	$\hat{\sigma}^2 = 1,2454$
Série 4	$n = 500$	$d = 0,2$	$\sigma^2 = 1$	$\hat{d} = 0,2052$	$\hat{\sigma}^2 = 1,0665$
Série 5	$n = 500$	$d = 0,4$	$\sigma^2 = 1$	$\hat{d} = 0,3477$	$\hat{\sigma}^2 = 1,1299$
Série 6	$n = 500$	$d = 0,45$	$\sigma^2 = 1$	$\hat{d} = 0,4101$	$\hat{\sigma}^2 = 1,0843$

Tabela 4.4.8: Estatísticas Globais M_0.

Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{d}$	0,2	0,4	0,45
		90		34,4460	36,7992	37,4881
$n = 250$	Inclinação (S)	95		34,9235	37,3530	38,0493
		99		35,8870	38,6087	39,3201
		90		5,2443	6,9455	8,7955
$n = 250$	Curvatura (C_c)	95		5,3472	7,4042	11,5663
		99		5,6540	9,3863	30,7706
			$\boldsymbol{\hat{d}}$	0,2230	0,3811	0,4002
			$\boldsymbol{\hat{\sigma}^2}$	0,9863	1,1711	1,2454
		90		34,9399	33,5990	32,9939
$n = 250$	Inclinação (S)	95		35,6155	34,1654	33,4655
		99		36,4946	35,1540	34,7156
		90		5,4464	5,5778	5,5508
$n = 250$	Curvatura (C_c)	95		5,6013	5,8716	5,9125
		99		5,9160	7,0556	7,6967
Tamanho Amostral	Medida	Percentis (%)	$\boldsymbol{d}$	0,2	0,4	0,45
		90		47,8253	50,9277	52,1464
$n = 500$	Inclinação (S)	95		48,4592	51,6505	52,7918
		99		49,6662	52,9162	53,8638
		90		5,0822	6,6583	8,8893
$n = 500$	Curvatura (C_c)	95		5,1845	7,1779	12,4837
		99		5,4471	8,4883	30,1185
			$\boldsymbol{\hat{d}}$	0,2052	0,3477	0,4101
			$\boldsymbol{\hat{\sigma}^2}$	1,0665	1,1299	1,0843
		90		46,4381	47,1566	49,0362
$n = 500$	Inclinação (S)	95		46,8931	47,7843	49,5821
		99		47,7705	48,8398	50,6800
		90		4,7880	5,2955	8,8893
$n = 500$	Curvatura (C_c)	95		4,8714	5,4369	12,4837
		99		5,0396	5,7337	30,1185

Na Tabela 4.4.8, os valores dos limiares M_0 calculados para a Inclinação de Billor e Loynes e para a Curvatura de Cook, apresentam forte dependência das estimativas dos parâmetros d e σ^2. Por exemplo, percebe-se que para a inclinação S quando $n = 250$ e percentil é fixo os limiares M_0 diminuem quando d aumenta, ao contrário do que ocorre com os limiares M_0 calculados a partir dos verdadeiros valores dos parâmetros. Assim, a variação dos limiares quanto a mudança de $\hat{d}$ confirma os indícios de não robustez dos mesmos considerando o critério global M_0. Estas verificações são muito importantes pois na prática os limiares são gerados a partir das estimativas dos parâmetros. Em conseqüência, M_0 não será uma ferramenta útil na prática.

Por outro lado, a técnica de limiares considerando os critérios M_1 e M_2 é uma ferramenta útil no diagnóstico de observações localmente influentes, dado que os limiares são robustos quanto a estimação de d e σ^2. Em conseqüência, nas análises das séries 1 - 6 encontramos os mesmos pontos influentes utilizando os critérios M_1 e M_2.

Assim, considerando o esquema de perturbação nos dados, através do modelo ARFIMA(0,d,0) e das medidas S e C_c nas séries 1 - 6 geradas, pôde-se verificar que a análise de influência local através da técnica de limiares, sob os critérios M_1 e M_2, torna-se uma poderosa ferramenta para identificar observações influentes.

4.5 Aplicação

Nesta seção será ilustrada a metodoloia de influência local em um conjunto de dados reais. Considerando o esquema de perturbação nos dados e as medidas Inclinação de Billor e Loynes (1993) e Curvatura de Cook (1986), aplica-se a técnica de limiares proposta na Seção 2.3 para identificar observações influentes através do critério global M_0 e dos critérios individuais M_1 e M_2.

Entre a primavera e o verão, as águas de degelo carregam grandes quantidades de silte (resíduos orgânicos) e areia fina para lagoas situadas nas porções frontais das geleiras. A areia fina juntamente com o silte grosso sedimenta-se mais rapidamente, enquanto que o silte mais fino e argila, contendo matéria orgânica, permanecem em suspensão e decantam-se somente no outono e inverno, quando em geral ocorre o congelamento da água de superfície e morte de muitos organismos. Com isso, forma-se um par de lâminas, uma clara e outra mais escura, que pode ter desde alguns milímetros até alguns centímetros de espessura. Este estrato ou camada resultante é denominada varve. Esse sedimento pode ser utilizado como parâmetro, tal como a temperatura, em estudos paleoclimáticos, pois em períodos quentes, uma maior quantidade de sedimentos é depositada quando ocorre um grande descongelamento das geleiras.

Nesta subseção será estudada a conhecida série $Varve$ correspondente a um conjunto de dados de 634 observações [página 36 de Shumway e Stoffer, 2000] que têm início 11384 anos atrás em $Massachusetts$. Foi utilizada a diferença logaritmica para tornar a série estacionária, e posteriormente foi retirada a média.

Na Figura 4.5.1, observa-se que as autocorrelações do logaritmo da série $Varve$, desconsiderando a média, apresentam decaimento hiperbólico, e motivam o ajuste dos dados pelo modelo ARFIMA(0,d,0) (4.2.3). Assim, utilizando a verossimilhança de $Whittle$ foram calculados os estimadores de máxima verossimilhança $\hat{d} = 0,4010$ e $\hat{\sigma}^2 = 0,2353$.

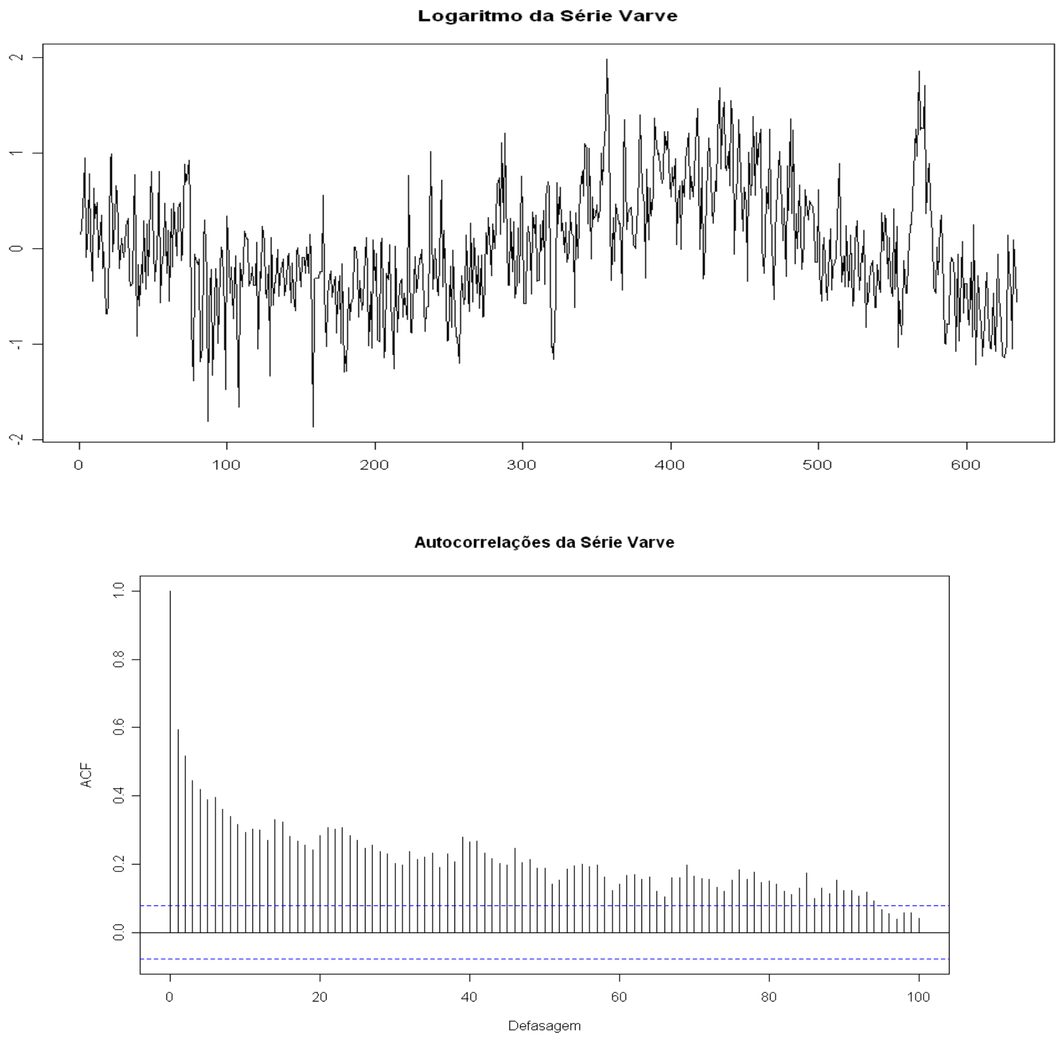

Figura 4.5.1: Gráfico logaritmo da série *varve* e as respectivas autocorrelações.

O diagnóstico de influência local será efetuado segundo o exposto na Seção 4.3, considerando a aproximação ARMA (4.2.5) com p=q=10. A Tabela 4.5.1 apresenta um resumo para o diagnóstico de influência local através dos níveis de referência M_0, M_1 e M_2.

Tabela 4.5.1: Resumo do diagnóstico da influência local.

		Estatísticas Individuais	
		$\mathbf{v}$	$\mathbf{v}_c$
Observações Influentes			
y_{87}		**0,1270**	**0,1205**
y_{99}		**0,1136**	**0,1120**
y_{158}		**0,1495**	**0,1428**
y_{223}		**0,1250**	**0,1309**
y_{238}		**0,1082**	0,1052
Estatísticas Globais	S	114,0700	—
	C_c	—	24,669
Limiares	M_0	119,3820	29,8932
	M_1	0,1553	0,1562
	M_2	0,1061	0,1079

Legenda: Em negrito estão representadas as estatísticas significativas para os critérios M_0 e M_1 (95%) e M_2 (5%).

A Tabela 4.5.1 não apresenta nenhuma medida global S ou C_c significativa para o critério M_0, mas os valores observados são próximos aos M_0 correspondentes. Embora isso aconteça, várias observações são identificadas como significativas para o critério individual M_2. As componentes de $\mathbf{v}$ e $\mathbf{v}_c$ identificadas como influentes são 87, 99, 158, 223 e 238. Percebe-se que o critério M_2 identifica a observação y_{238} como influente considerando somente o vetor de diagnóstico $\mathbf{v}_c$.

A Figura 4.5.2 ilustra o comportamento de M_1 e M_2 perante os vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$.

Através da simples análise exploratória dos dados, as observações y_{357} e y_{568} poderiam ser consideradas influentes quando avaliadas somente pelo gráfico de valores absolutos da série $Varve$, pois exibem notada magnitude quando comparadas as demais. Mas, como pode ser visto através dos vetores de diagnóstico $\mathbf{v}$ e $\mathbf{v}_c$ e dos limiares M_1 e M_2 na Figura 4.5.2, esses pontos não são considerados influentes. Uma explicação para a não identificação das componentes 357 e 568, seria que a série alcança esses pontos de forma gradual, enquanto a observação y_{158} surge na série de forma abrupta.

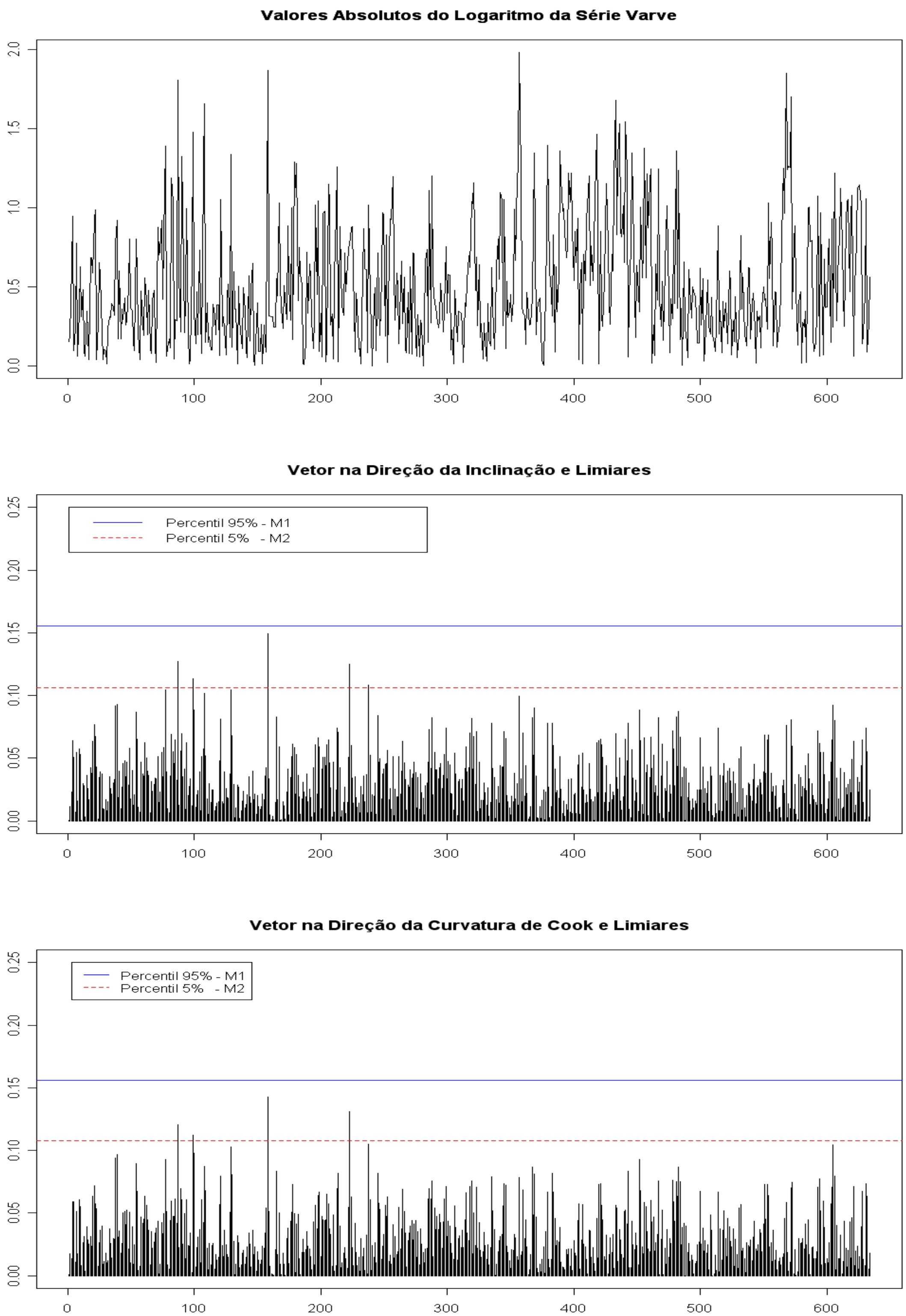

Figura 4.5.2: Influência local para $\mathbf{v}$ e $\mathbf{v}_c$.

5 *Conclusões e Considerações Finais*

Esta dissertação tem duas contribuições. A primeira é ilustrar o uso da metodologia de limiares no diagnóstico de influência local para modelos de séries temporais. Zhang e King (2005) propuseram uma abordagem que permite obter marcas de referência (limiares) para as estatísticas da inclinação e curvatura. A partir dessa metodologia, do esquema de perturbação nos dados e das medidas Inclinação de Billor e Loynes e Curvatura de Cook foi possível investigar os modelos de regressão linear com erros AR(1), modelos autoregressivos de ordem um e modelos de longa memória quanto ao comportamento dos limiares e a eficácia dessa ferramenta no que diz respeito a identificação de observações influentes.

Foram realizadas simulações para os modelos autoregressivos de ordem um, e dessa forma foi possível verificar que, baseados nos parâmetros estimados, a técnica de limiares utilizando o critério global M_0 não é uma ferramenta útil na identificação de observações influentes. Por outro lado, quando os limiares baseados nos critérios individuais M_1 e M_2 são construídos a partir das estimativas dos parâmetros, a técnica é eficaz na identificação dos *outliers* como pontos influentes. Esse mesmo comportamento é observado quando a técnica de limiares é empregada nos modelos ARFIMA(0,d,0). Adicionalmente, considerando os modelos ARFIMA(0,d,0), verificou-se a identificação de pontos "aberrantes" no final da série avaliada, o que dificilmente ocorre nas técnicas usualmente utilizadas na literatura com respeito ao diagnóstico de *outliers*. Verifica-se também nas simulações para AR(1) e ARFIMA(0,d,0) a presença de "reflexos", comportamento esse diretamente ligado à estrutura do modelo em questão e à dinâmica da série investigada. Por outro lado, nos estudos de simulação foi encontrado que o método é eficaz no seguinte sentido: não identifica observações "grandes" que não foram geradas como influentes. A técnica foi ilustrada com sucesso na análise das séries reais Índice de Preços ao Consumidor da Argentina e $Varve$ ajustados sobre os modelos AR(1) e ARFIMA(0,d,0), respectivamente, onde foram identificadas observações influentes.

A segunda contribuição dessa dissertação é o cálculo das expressões das medidas Inclinação de Billor e Loynes e Curvatura de Cook para avaliar a influência local considerando o modelo de longa memória ARFIMA(0,d,0).

Quando são identificadas observações influentes usando a metodologia descrita neste trabalho, e dado o esquema de perturbação nos dados, existe indício de instabilidade do modelo ajustado. Portanto, propõe-se a investigação da observação detectada e/ou o ajuste de um outro modelo.

Finalmente, algumas linhas de futuros trabalhos são descritas a seguir:

1. Utilização de outros esquemas de perturbação para os modelos de séries temporais estudados nessa dissertação;

2. Utilização de blocos de *outliers* e outros tipos de esquemas de perturbação;

3. Investigar a metodologia de limiares para outros modelos de séries temporais como ARFIMA(p,d,q) e ARFIMA-GARCH; e

4. Utilizar a metodologia de limiares considerando a Curvatura de Zhang e King.

Referências

Anderson (1977) Estimation for Autoregressive Moving Average Models in the Time and Frequency Domains. *Annals of Statistics*, Vol. 5, 413-427.

Baillie, R. T., Bollerslev, T. e Mikkelsen, H. O. (1996) Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, Vol. 74, 3-30.

Basak G. K., Chan N. H. e Palma, W. (2001) The Approximation of Long-Memory Process by an ARMA Model. *Journal of Forecasting*, Vol. 20, 367-389.

Besley, D. A., Kuh, E. e Welsch, R. E. (1980) *Regression Diagnostics.* John Wiley. New York.

Billor, N. e Loynes, R. M. (1993) Local Influence: A New Approach. *Communications in Statistics Theory and Methods*, Vol 22, Nº. 6, pp.1595-1611.

Bollerslev, T. e Mikkelsen, H. O. (1996) Modeling and Pricing Long Memory in Stock Market Volatility. *Journal of Econometrics*, Vol. 73, 151-184.

Box, G. E. P. e Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Cadigan, N.G. e Farrel, P. J. (2002) Generalized Local Influence with Applications to Fish Stock Cohort Analysis. *Journal of the Royal Statistical Society*, Series C, Vol. 51, 469-483.

Chan N. H. e Palma, W. (1998) State Space Modeling of Long-Memory Process. *Annals of Statistics*, Vol. 26, 719-740.

Cook, R. D. e Weisberg, S. (1982) *Residuals and Influence in Regression.* Chapman and Hall.

Cook R. D. (1986) Assessment of Local Influence. *Journal of Royal Statistical Society*, Series B, Vol. 42, 133-169.

Dávila, O. L. S. (2000) Influência Local em Modelos de Calibração Comparativa. Dissertação (Mestrado em Estatística) - Instituto de Matemática, Estatística e Computação Científica da Universidade de Campinas, Campinas.

Franses P. H. (1998) Time Series Models for Business and Economic Forecasting. Cambridge University Press.

Galea M., Díaz-García J., Vilca F. (2008) Influence Diagnostics in the Capital asset Pricing Model under Elliptical Distributions. *Journal of Applied Statistics*, Vol. 35, 179-192.

Granger, C. M. G. e Joyeux, R. (1980) An Introduction to Long Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*, Vol. 1, 15-29.

Haldrup N. e Nielsen M. O. (2007) Estimation of Fractional Integration in the Presence of Data Noise. *Computational Statistics & Data Analysis* Vol. 51, 3100-3114.

Hamilton, J. D. (1994) *Time Series Analysis.* Princeton University Press.

Hosking, J. R. M. (1981) Fractional Differencing. *Biometrika*, Vol. 68, 165-176.

Hurst, H. E. (1951) Long-Term Storage Capacity of Reservoirs. *Transaction of the Americn Society of Civil Engineers*, Vol. 16, 770-799.

Hurst, H. E. (1957) A Suggested Statistical Model of Time Series that Occur in Nature. *Nature*, Vol. 180, 494.

Kim, S. K. e Huggins, R. (1997) Dianostics for Autocorrelated Regression Models. *Australian & New Zealand Journal Statistics*, N°. 40, 65-71.

Lintner, J. (1965) The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, Vol. 41, 13-37.

Mossin, J. (1966) Equilibrium in a Capital Asset Market. *Econometrica*, Vol. 35, 768-783.

Liu, S. (2002) Local Influence in Multivariate Elliptical Linear Regression Models. *Linear Algebra an Its Applications*, Vol. 354, 159-174.

Lesaffre, E. e Verbeke, G. (1998) Local Influence in Linear Mixed Models. *Biometrics*, Vol. 54, 570-582.

Mandelbrot, B. B. e Wallis, J. (1968) Noah, Joseph and Operational Hydrology. *Water Resources Research*, Vol. 4, 909-918.

McLeod, A. I. e Hipel, K. W. (1978) Preservation of the Rescaled Adjusted Range, 1: A Reassessment of the Hurst Phenomenon. *Water Resources Research*, Vol. 14, 491-508.

Medeiros M. J. (2006) Métodos de Diagnóstico em Modelos Autoregressivos Simétricos. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo.

Montgomery, D. C. e Peck, E. A. (1982) *Introduction to Linear Regression Analisys.* Wiley, New York.

Palma W. A., Bondon P. B., Tapia, J. C. (2008) *Assessing Influence in Gaussian Long-memory Models.* Computational Statistics & Data Analysis, 52, 4487-4501.

Paula, G. A. (2004) Modelos de Regressão com Apoio Computacional. Instituto de Matemática e Estatística da Universidade de São Paulo.

Poon, W. Y. e Poon, Y. S. (1999) Conformal Normal Curvature and Assessment of Local Influence. *Journal of the Royal Statistical Society*, Series B, Vol. 61, 51-61.

Prescott, P. (1986) Comment on "Assessement of Local Influence", by R. D. Cook. *Journal of the Royal Statistical Society*, Series B, Vol. 48, 161.

Salgado A. O. (2006) Diagnóstico de Influência em Modelos Elípticos com Efeitos Mistos. Tese (Doutorado em Estatística) - Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo.

Schall, R. e Dunne, T. T. (1991) *Diagnostics for Regression-ARMA Time Series.* In Directions in Robust Statistics and Diagnostics, Springer, New York.

Sharpe, W. F. (1964) Capital Asset Prices: a Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance*, Vol. 19, 425-442.

Shumway R. H. e Stoffer D. S. (2000) *Time Series Analysis and Its Applications.* Springer.

Tiao, G. C. e Tsay, R. S. (1994) Some Advances in Non-Linear and Adaptative Modeling in Time Series. *Journal of Forecasting*, Vol. 13, 109-131.

Tsai, C. L. e Wu, X. (1992) Assessing Local Influence in Linear Regression Models with First-Order Autoregressive or Heteroscedastic Error Structure. *Statistics* & *Probability*, Letters 14, 247-252.

Weisberg, D. (1985) *Applied Linear Regression.* Wiley, New York.

Wu, X. e Luo, Z. (1993) Second-Order Approach to Local Inflence. *Journal of the Royal Statistical Society*, Series B, Vol. 55, 929-936.

Zevallos M. e Hotta L. K. (2008) Influential Observations in Conditional Heteroskedastic Time Series. Working Paper.

Zhang, X. e King, M. (2005) Autoregressive Conditional Heteroscedasticity Process. *Journal of Business* & *Economic Statistics*, Vol. 23, Nº 1.

Zhu, H. Zhang, H. (2004) A diagnostic Procedure Based on Local Influence. *Biometrika*, Vol. 91, 3, 579-589.