

UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE MATEMÁTICA ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA  
Departamento de Estatística

# Detecção de valores aberrantes em modelos de volatilidade estocástica

Pedro Fukui

Orientador: Prof. Dr. Luiz Koodi Hotta

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  
2000

i

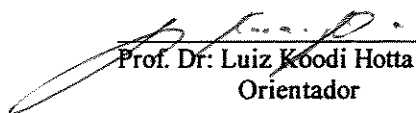
UNICAMP  
BIBLIOTECA CENTRAL  
SECÃO CIRCULANTE

# Detecção de valores aberrantes em modelos de volatilidade estocástica

UNICAMP  
BIBLIOTECA CENTRAL  
SEÇÃO CIRCULANTE

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Pedro Fukui e aprovada pela comissão julgadora.

Campinas, 25 de outubro de 2000

  
Prof. Dr. Luiz Koodi Hotta  
Orientador

Banca Examinadora:

1. Prof. Doutor Pedro Luiz Valls Pereira - IBMECC/SP
2. Prof. Doutor Marinho Gomes de Andrade - ICMC/USP
3. Prof. Doutor Aluísio de Souza Pinheiro - IMECC/UNICAMP

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

|              |                                     |
|--------------|-------------------------------------|
| UNIDADE      | B.C.                                |
| N.º CHAMADA: | 1/Unicamp                           |
|              | F955d                               |
| V.           | Ex.                                 |
| TOMBO BC/    | 43397                               |
| PROC.        | 16-392101                           |
| C            | <input type="checkbox"/>            |
| D            | <input checked="" type="checkbox"/> |
| PREC.        | R\$ 11,00                           |
| DATA         | 09/04/01                            |
| N.º CPD      |                                     |

CM-00153704-9

### FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP

Fukui, Pedro

F955d Detecção de valores aberrantes em modelos de volatilidade estocástica  
/ Pedro Fukui -- Campinas, [S.P. :s.n.], 2000.

Orientador : Luiz Koodi Hotta

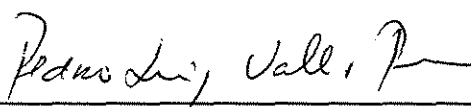
Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de  
Matemática, Estatística e Computação Científica.

1. Kalman, Filtragem de. 2. Derivativos (Finanças). I. Hotta, Luiz Koodi.  
II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e  
Computação Científica. III. Título.

**Dissertação de Mestrado defendida em 29 de setembro de 2000 e aprovada**

**Pela Banca Examinadora composta pelos Profs. Drs.**

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

  
Prof (a). Dr (a). PEDRO LUIZ VALLS PEREIRA

  
Prof (a). Dr (a). MARINHO GOMES DE ANDRADE

UNICAMP  
BIBLIOTECA CENTRAL  
SEÇÃO CIRCULANTE

## **Resumo**

O principal objetivo deste trabalho é propor uma metodologia para detecção e estimação de valores aberrantes em modelos de volatilidade estocástica, sejam aqueles com efeito local ou quebras estruturais do modelo.

Os estudos realizados no desenvolvimento da metodologia foram baseados em métodos computacionais que utilizam simulações estocásticas. As estimativas dos parâmetros foram obtidas através do método de quase-verossimilhança e o critério de BIC (Schwarz) foi utilizado para detecção das observações aberrantes.

A metodologia proposta foi testada em três séries, sendo as duas primeiras simuladas, uma sem outliers e outra com outliers como terceira série foram utilizados dados reais, mais especificamente, dados diários do Índice da Bolsa de Valores de São Paulo, no período de 04 de julho de 1994 - (início do Plano Real) a 7 de abril de 2000, num total de 1424 observações.

## **Abstract**

The main objective of this paper is to propose a methodology for the detection and estimation of outliers caused by local effects or structural breaks in stochastic volatility models.

The studies carried out in the development of the methodology were based on computational methods using stochastic simulation. The estimates of parameters were obtained through the quasi-likelihood method and the Schwarz criteria was used for the detection of outliers.

The proposed methodology was tested in three series. The first two were generated, one with outliers and another without. Real data was used as the third series, more specifically the daily São Paulo Stock Exchange index (BOVESPA) for the period from July 04, 1994 - (beginning of Plano Real) to April 07, 2000, totaling 1424 observations.

Agradeço ao Prof. Dr. Luiz Koodi Hotta pela orientação e amizade.

Agradeço à Prof. Dra. Cicília Yuko Wada pelo incentivo aos estudos.

Agradeço aos meus pais pelo apoio e compreensão.

Agradeço ao CNPq pela bolsa concedida

Dedico esta dissertação à Bia que sempre esteve  
ao meu lado nos momentos difíceis e felizes.

# Sumário

|   |    |
|---|----|
| Capítulo I - Introdução.  | 1  |
| Capítulo II - Volatilidade de ativos financeiros.   | 3  |
| 2.1 Descrição das características empíricas das séries financeiras: Fatos estilizados.            | 3  |
| 2.2 Volatilidade.   | 6  |
| 2.3 Modelo na forma de espaço de estados.   | 7  |
| 2.3.1 Modelo de volatilidade estocástica.   | 9  |
| 2.4 Filtro de Kalman.   | 11 |
| 2.5 Os diferentes tipos de outliers.  | 12 |
| Capítulo III - Estimação dos parâmetros do modelo AR(1)_SV.                                       | 14 |
| 3.1 Principais métodos para estimação dos parâmetros do Modelo AR(1)_SV.                          | 15 |
| 3.2 Aspectos práticos do método da máxima Quase-Verossimilhança                                   | 17 |
| 3.2.1 Maximização numérica e reparametrização.  | 17 |
| 3.2.2 Critérios de convergência.  | 19 |
| 3.3 Experimentos via simulação de Monte Carlo.  | 21 |
| 3.3.1 Estimativas dos parâmetros.   | 22 |
| 3.3.2 Intervalos de Confiança.  | 29 |
| Capítulo IV - Detecção de outliers através dos resíduos auxiliares                                | 32 |
| 4.1 Resíduos auxiliares.  | 34 |
| 4.1.1 Transformação inversa do resíduo auxiliar irregular   | 37 |
| 4.2 Testes assintóticos para detecção de outliers.  | 38 |
| 4.2.1 Comparação das distribuições das estatísticas K e S para diferentes conjuntos de parâmetros | 40 |
| 4.3 Análise dos resíduos auxiliares utilizando séries simuladas.                                  | 44 |
| Capítulo V - Procedimento de varredura para detecção e estimação de outliers.                     | 51 |
| 5.1 Critério BIC para detecção de outliers.   | 52 |
| 5.2 Procedimento de varredura para detecção de outliers.  | 53 |
| 5.2.1 O procedimento de varredura passo-a-passo.  | 54 |
| 5.3 Aplicação do procedimento.  | 55 |
| 5.3.1 Série simulada sem outliers.  | 56 |
| 5.3.2. Série simulada com outliers.   | 59 |
| 5.3.3. Dados reais Índice da bolsa de valores de São Paulo (BOVESPA).                             | 65 |
| Conclusão.  | 76 |
| Referências Bibliográficas.   | 77 |



# Índice de gráficos e figuras.

|               |   |    |
|---------------|---|----|
| Figura 3.1.   | Distribuição log-qui-quadrado e normal(-1.27,4.93).   | 16 |
| Figura 3.2.   | Função de reparametrização de $\gamma$ .  | 18 |
| Figura 3.3.   | Função de reparametrização $\sigma^2_\eta$  |    |
| Gráfico 3.1.  | Dispersão das estimativas de QML do modelo AR(1)_SV com conjunto de parâmetros $\Theta_1$ e séries com 1000 observações.  | 24 |
| Gráfico 3.2.  | Dispersão das estimativas de QML do modelo AR(1)_SV com conjunto de parâmetros $\Theta_2$ e séries com 1000 observações.  | 24 |
| Gráfico 3.3.  | Densidade de probabilidade das estimativas de QML dos parâmetros $\gamma$ e $\sigma_\eta$ estimada através de 1000 replicações de séries com 1000 e 2000 observações.       | 25 |
| Gráfico 3.4.  | Densidade de probabilidade das estimativas de QML dos parâmetros $\psi$ e $\omega$ estimada através de 1000 replicações de séries com 1000 e 2000 observações.              | 25 |
| Gráfico 3.5.  | Densidade de probabilidade das estimativas de QML dos parâmetros $\alpha$ estimada através de 1000 replicações de séries com 1000 e 2000 observações.                       | 28 |
| Gráfico 4.1.  | Comparação das distribuições empíricas das estatísticas de interesse do resíduo irregular   | 42 |
| Gráfico 4.2.  | Comparação das distribuições empíricas das estatísticas de interesse do resíduo estrutura   | 43 |
| Gráfico 4.3.  | Série dos retornos e dos log-retornos <sup>2</sup> gerados por um modelo AR(1)_SV <b>sem outliers</b> .   | 45 |
| Gráfico 4.4.  | Resíduos auxiliares irregulares transformados. Série dos resíduos, correlograma e histograma com estimativa da função densidade de probabilidade. <b>Série sem outliers</b> | 46 |
| Gráfico 4.5.  | Resíduos auxiliares estruturais. Série dos resíduos, correlograma e histograma com estimativa da função densidade de probabilidade. <b>Série sem outliers</b>               | 46 |
| Gráfico 4.6.  | Série dos retornos e dos log-retornos <sup>2</sup> gerados por um modelo AR(1)_SV <b>com outliers</b> .   | 45 |
| Gráfico 4.7.  | Resíduos auxiliares irregulares transformados. Série dos resíduos, correlograma e histograma com estimativa da função densidade de probabilidade. <b>Série com outliers</b> | 46 |
| Gráfico 4.8.  | Resíduos auxiliares estruturais. Série dos resíduos, correlograma e histograma com estimativa da função densidade de probabilidade. <b>Série com outliers</b>               | 46 |
| Gráfico 5.1.  | Retornos com volatilidade verdadeira e estimada pelo modelo AR1_SV. Série exemplo #1 (sem outliers).  | 58 |
| Gráfico 5.2.  | Resíduos auxiliares irregulares série exemplo #2(com outliers) do modelo.   | 63 |
| Gráfico 5.3.  | Resíduos auxiliares estruturais da série exemplo #2(com outliers) do modelo considerando os outliers detectados pelo procedimento de varredura.                             | 65 |
| Gráfico 5.4.  | Retornos com volatilidade verdadeira e estimada pelo modelo AR1_SV com os outliers detectados pelo procedimento de varredura. Série exemplo #2                              | 65 |
| Gráfico 5.5.  | Densidade de probabilidade estimada dos retornos do IBOVESPA. (de 04/Jul/1994 à 07/Abr/2000) correspondendo à 1424 observações  | 65 |
| Gráfico 5.6.  | Log(Ret <sup>2</sup> ) e retornos da série do IBOVESPA. (de 04/Jul/1994 à 07/Abr/2000) 1424 observações.  | 66 |
| Gráfico 5.7.  | Resíduos auxiliares estruturais sem considerar a presença de outliers. Série IBOVESPA.  | 68 |
| Gráfico 5.8.  | Resíduos auxiliares irregulares sem considerar a presença de outliers. Série IBOVESPA.  | 68 |
| Gráfico 5.9.  | Resíduos auxiliares estruturais considerando os outliers detectados pelo procedimento de varredura. Série IBOVESPA  | 74 |
| Gráfico 5.10. | Resíduos auxiliares irregulares considerando os outliers detectados pelo procedimento de varredura. Série IBOVESPA.   | 74 |
| Gráfico 5.11. | Retornos com volatilidade estimada com e sem os outliers detectados pelo procedimento de varredura. Série IBOVESPA.   | 75 |

# Índice de tabelas.

|              |   |    |
|--------------|---|----|
| tabela 3.1.  | Frequência observada nas estimativas dos parâmetros para os respectivos tamanho das séries e os conjuntos de parâmetros $\Theta_1=\{\alpha=0; \gamma=0.98; \sigma_\eta=0.12\}$ ; $\Theta_2=\{\alpha=0; \gamma=0.95; \sigma_\eta=0.20\}$ . | 22 |
| tabela 3.2.  | Estatísticas descritivas de 1000 estimativas de máxima quase-verossimilhança de séries de tamanho 1000 e 2000. Entre parênteses os parâmetros verdadeiros.  | 26 |
| tabela 3.3.  | Estatísticas descritivas de 1000 estimativas de máxima quase-verossimilhança de séries de tamanho 1000 e 2000.  | 26 |
| tabela 3.4.  | Estatísticas descritivas de 1000 estimativas de máxima quase-verossimilhança de séries de tamanho 1000 e 2000.  | 27 |
| tabela 3.5.  | Estatísticas descritivas de 1000 estimativas de máxima quase-verossimilhança de séries de tamanho 1000 e 2000.  | 27 |
| tabela 3.6.  | Estatísticas descritivas de 1000 estimativas de máxima quase-verossimilhança de séries de tamanho 1000 e 2000.  | 29 |
| tabela 3.7.  | Confiança dos intervalos estimada pelo método de Monte Carlo.   | 30 |
| tabela 4.1.  | Valores de $\phi$ calculados pela equação 4.4 para alguns valores típicos dos parâmetros do modelo AR(1)_SV.  | 36 |
| tabela 4.2.  | Estatísticas descritivas das distribuições das estatísticas de interesse para cada conjunto de parâmetro  | 42 |
| tabela 4.3.  | Estatísticas descritivas das distribuições das estatísticas de interesse para cada conjunto de parâmetro  | 43 |
| tabela 5.1.  | Estimativas do modelo AR1_SV sem considerar a presença de outliers. Série exemplo #1 (sem outliers).  | 56 |
| tabela 5.2.  | Resultados da varredura para série #1.  | 57 |
| tabela 5.3.  | Estimativas do modelo AR1_SV sem considerar a presença de outliers. Série exemplo #2 (com outliers).  | 59 |
| tabela 5.4.  | Resultados da primeira varredura para série #2.   | 60 |
| tabela 5.5.  | Resultados da segunda varredura para série #2.  | 61 |
| tabela 5.6.  | Estimativas do modelo AR1_SV considerando a presença dos outliers detectados pelo procedimento de varredura. Série exemplo #2 (com outliers).   | 62 |
| tabela 5.7.  | Estimativas do modelo AR(1)_SV sem considerar a presença de outliers. Série IBOVESPA.   | 67 |
| tabela 5.8.  | Resultados da primeira varredura para série do IBOVESPA.  | 69 |
| tabela 5.9.  | Resultados da segunda varredura para série do IBOVESPA..  | 70 |
| tabela 5.10. | Resultados da segunda varredura para série do IBOVESPA.   | 71 |
| tabela 5.11. | Estimativas do modelo AR(1)_SV considerando a presença dos outliers detectados pelo procedimento de varredura. Série IBOVESPA.  | 72 |
| tabela 5.12  | Resultado do primeiro procedimento de varredura com os possíveis outliers não detectados  | 73 |

## Capítulo I - Introdução.

A evolução do mercado financeiro e o crescimento do mercado de derivativos têm exigido ferramentas de análise cada vez mais sofisticadas e complexas.

O estudo da volatilidade, nesse contexto, ganhou grande importância, tanto na quantificação do risco, como também na precificação de produtos financeiros. Black (1976) contribuiu para o estudo da volatilidade ao descrever pela primeira vez os efeitos de alavancagem. Taylor (1980) propôs o modelo de volatilidade estocástica. Engle (1982) também colaborou ao propor o modelo da família ARCH para estudar os conglomerados de volatilidade. Esses dois modelos foram desenvolvidos com a mesma finalidade, porém os modelos da família xARCH foram mais difundidos e utilizados, devido a sua simplicidade e facilidade na obtenção das estimativas de máxima verossimilhança.

Ao longo da década de 90, o mercado financeiro vivenciou enormes instabilidades, que influenciaram a volatilidade das séries financeiras, pela incidência de quebras estruturais e observações aberrantes. Os modelos utilizados precisavam ser modificados, para incorporarem o aumento imprevisto da volatilidade. Essas observações aberrantes, também conhecidas como outliers, ganharam importância na modelagem da volatilidade.

A proposta dessa dissertação é desenvolver um algoritmo para a detecção e estimação de outliers utilizando o modelo de volatilidade estocástica.

O modelo de volatilidade estocástica é escrito na forma de espaço de estados. Existem algumas alternativas para a estimação dos parâmetros dos modelos escritos na forma de espaço de estados sendo as mais utilizadas: método dos momentos, método da máxima verossimilhança e a abordagem bayesiana. O método dos momentos apresenta várias desvantagens, sendo a principal o fato de não fornecer a estimativa dos estados (volatilidade). Já a abordagem bayesiana é ineficiente computacionalmente. A alternativa mais adequada seria o uso do método da máxima verossimilhança, pois apresenta eficiência computacional, e fornece as estimativas dos estados (volatilidade). Entretanto, a verossimilhança exata no modelo de volatilidade estocástica não pode ser obtida utilizando o filtro de Kalman, pois um dos componentes aleatórios do modelo de volatilidade estocástica tem distribuição log-qui-quadrado. Por isso, utilizaremos o método da máxima quase-verossimilhança, que aproxima a distribuição log-qui-quadrado por uma distribuição normal com mesma média e variância.

No primeiro capítulo, definiremos os principais conceitos utilizados ao longo da dissertação: as principais características de séries financeiras (fatos estilizados), a volatilidade e modelos na forma de espaço de estados, modelos de volatilidade estocástica, filtro de Kalman e, por fim, os diferentes tipos de outliers.

No capítulo II, avaliaremos a estimação dos parâmetros do modelo de volatilidade estocástica, examinaremos as metodologias utilizada - método da quase-verossimilhança - e discutiremos os principais aspectos práticos para implementação deste método. Ainda neste capítulo, realizaremos simulações para avaliação dos estimadores e dos intervalos de confiança. Todas as simulações, inclusive as do capítulo II, serão realizadas com a utilização do programa OX (Object Oriented Matrix Language) desenvolvido por Doornik e Koopman<sup>1</sup>.

No capítulo III, examinaremos o processo de detecção de outliers nos modelos volatilidade estocástica através dos resíduos auxiliares e as estatísticas de assimetria e curtose dos resíduos auxiliares para a detecção de outliers. Mostraremos que, mesmo com a correta especificação do modelo e condicionado aos parâmetros verdadeiros, estes resíduos são serialmente correlacionados e correlacionados entre si. Ainda nesse capítulo, apresentaremos duas séries simuladas e seus respectivos resíduos auxiliares estimados pelo filtro de Kalman..

No IV, apresentaremos uma metodologia para detecção e estimação dos efeitos dos diferentes tipos de outliers, apresentaremos o critério BIC (Bayes Information Criteriation) que será utilizado para detecção dos outliers. descreveremos o procedimento de varredura para detecção de outliers. aplicaremos o procedimento nas séries exemplo #1(sem outliers), série exemplo #2 (com outliers) e em dados reais (IBOVESPA de 04/Jul/1994 à 07/Abr/2000 correspondendo a 1424 observações). Os outliers na série exemplo #2 foram escolhidos para ressaltar os efeitos dos outliers em um modelo AR(1)\_SV e conseqüentemente teremos bem claros os efeitos dos outliers nas estatísticas de assimetria (S) e curtose (K), nos resíduos auxiliares e o procedimento de varredura será preciso na identificação e localização dos outliers.

---

<sup>1</sup> Toda documentação e versões de demonstração do OX podem ser encontradas nos seguintes endereços (<http://hicks.nuff.ox.ac.uk/Users/Doornik/doornik.html#ox> e <http://www.econ.vu.nl/koopman/#sjme>). A primeira referência é do programa básico e a segundo é das bibliotecas para modelos na forma de espaço de estado e filtro de Kalman.

## Capítulo II - Modelos de Volatilidade.

### 2.1. Descrição das características empíricas das séries financeiras: Fatos Estilizados.

Séries financeiras geralmente não são estacionárias na média e, por este motivo, são necessárias transformações . Uma transformação muito utilizada em séries financeiras é a taxa de retorno ou simplesmente *retorno*, que definimos a seguir.

Seja  $P_t$  o preço de um determinado ativo no instante  $t$ . Os retornos podem ser definidos de várias maneiras, sendo a seguinte forma a mais utilizada:

$$Y_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln(P_{t-1}) \quad (2.1)$$

As variações de um dia para outro costumam ser pequenas e, por isso, a razão  $P_t/P_{t-1}$  oscila em torno de 1. Tomando-se o logaritmo desta razão, em geral, obtemos uma distribuição aproximadamente simétrica em torno de zero. Se considerarmos que  $\ln(1+\Delta) \cong \Delta$  para  $\Delta$  pequeno, então o retorno definido desta forma é aproximadamente a variação relativa do preço dos ativos analisados. Outro detalhe importante é que, sendo os retornos função das razões, eles não dependem de suas unidades originais mas de suas variações relativas, facilitando a comparação direta de duas séries distintas. Além disso, o retorno em  $p$  períodos é igual a soma dos retornos em cada período.

Os retornos de séries financeiras apresentam características peculiares, chamadas de fatos estilizados, descritas a seguir:

### 1. Caudas Pesadas.

Os retornos  $Y_t$  apresentam uma distribuição simétrica leptocúrtica, ou seja, os valores longe da média ocorrem com uma frequência relativamente mais alta, i.e. possuem distribuição com caudas mais pesadas do que a da distribuição normal.

### 2. Conglomerados de Volatilidade.

Apesar de as séries dos retornos em geral não apresentarem auto-correlação, as séries dos retornos absolutos e dos quadrados dos retornos apresentam uma auto-correlação significativa com decaimento lento para defasagens superiores. Este fato nos fornece evidências de que dos retornos não são serialmente independentes. Esta dependência na distribuição dos retornos, relacionada com suas variâncias, no qual existem períodos de alta volatilidade onde a variabilidade tende a permanecer alta e períodos de baixa volatilidade, onde a variabilidade tende a permanecer baixo, é chamada de conglomerados de volatilidade.

### 3. Efeito Alavancagem (Leverage Effect).

Fenômeno típico de séries financeiras, o efeito alavancagem faz que as variações nos preços estejam correlacionadas com a volatilidade, ou seja, quando os preços de um determinado ativo estão em queda, sua volatilidade tende a ser maior do que num período de alta.

### 4. Memória Longa e Persistência.

Em séries com alta frequência (por exemplo: dados diários), observa-se uma alta persistência, da volatilidade, tendendo a permanecer próxima à observada no instante anterior. A essa alta persistência, damos o nome de memória longa. Este comportamento é típico de modelos com raízes quase unitárias. Isto faz que o efeito de IO's (Innovation Outlier) a serem definidos sejam propagados por um longo período.

### 5. Efeito Calendário.

A volatilidade pode ser influenciada pelo dia da semana, dia do mês e ainda pelo mês do ano. Isto se deve ao fato de, por exemplo, segunda-feira ter passado por três noites de acúmulo de

informações sem que os mercados tenham aberto, espera-se que a variabilidade dos ativos sejam maiores. Efeito similar ao da segunda-feira pode ocorrer também depois de feriados prolongados. A influência do dia do mês na volatilidade está associada principalmente à divulgação de dados macroeconômicos com frequência mensal ou, ainda, podemos relacionar efeitos mensais ou bimestrais. Também existe um fator associado aos meses do ano e, neste caso, janeiro tende a ter um comportamento diferenciado. Geralmente, no início do ano, os grandes investidores fazem uma avaliação de seus investimentos com possíveis realocações dos seus portfólios, ocasionando um aumento da volatilidade nos mercados financeiros.

Devido às características empíricas das séries financeiras, modelos lineares Gaussianos são incapazes de descrever e reproduzir os principais fatos estilizados observados acima. O modelo ARCH (Autoregressive Conditional Heteroscedasticity) proposto por Engle (1982), foi o precursor de uma série de modelos que litou a reproduziram algumas dessas características empíricas. Posteriormente, diversos desdobramentos deste modelo foram propostos (GARCH, I-GARCH, E-GARCH, M-GARCH), com o intuito de aperfeiçoar a caracterização dos fatos estilizados, como sugerido por Taylor (1986).

Taylor (1980) propôs o modelo  $AR(1)_{SV}$  como alternativa aos modelos da família  $xARCH$  na modelagem de séries financeiras de retornos. Harvey et alli (1994) apresentou as propriedades assintóticas dos estimadores de QMV, explorando o fato do modelo poder ser colocado na forma de espaço de estados onde a volatilidade é considerada uma componente não observável, diferentemente do que ocorre nos modelos da família  $xARCH$ , onde a volatilidade é função somente dos parâmetros e dos retornos passados. O modelo  $AR(1)_{SV}$  reproduz os principais fatos estilizados, com a vantagem de ser parcimonioso, além de ser generalizado para o caso multivariado e incorporar facilmente o efeito de covariáveis.

## 2.2. Volatilidade.

Uma característica de grande interesse na série de retornos do tipo descrito na seção anterior é a volatilidade ou a variabilidade dos retornos.

A volatilidade está diretamente associada à variabilidade dos preços de um determinado ativo. Se os preços variam muito, dizemos que este ativo é muito volátil. Usualmente, utilizamos a variância ou o desvio-padrão como uma medida da volatilidade e, por se tratar da variabilidade nos preços de ativos, a volatilidade é uma medida do risco envolvido. Portanto, uma das principais importâncias de estimarmos e prevermos a volatilidade, pois estaremos estimando possíveis ganhos ou perdas.

Outra aplicação imediata da estimação da volatilidade é a precificação de *opções*, utilizando a fórmula de *Black & Scholes*. Nesta fórmula, existem seis variáveis (preço no presente, preço no vencimento, tempo de maturação, taxa de juros no período, volatilidade e preço da opção), sendo que duas a volatilidade e o preço não são observáveis ou conhecidas. Ao descobrirmos o preço justo da opção, poderemos saber se o preço de mercado está caro ou não. O preço pode ser encontrado a partir da volatilidade estimada. Foram sugeridas na literatura diversas metodologias para estimarmos a volatilidade, sendo que cada uma tem peculiaridades e características próprias. Considerando-se que a fórmula de Black & Scholes é amplamente utilizada no mercado financeiro, a negociação de opções passa a ser equivalente à negociação de volatilidade, ou seja, podemos considerar a volatilidade como sendo uma mercadoria. E, conseqüentemente, sua estimação e previsão passam a ter um papel fundamental.



### 2.3. Modelos na forma de espaço de estado.

Os modelos na forma de *espaço de estado* são conhecidos há bastante tempo, mas somente com o desenvolvimento do filtro de Kalman (Kalman 1960; Kalman e Bucy 1961) eles passaram a ser largamente utilizados.

O modelo na forma de espaço de estados para uma série temporal ( $y_t$ ) multivariada com dimensão  $N$  é definido pelas seguintes equações:

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t \quad (2.2)$$

onde  $E(\varepsilon_t)=0$  e  $\text{Var}(\varepsilon_t)=H_t$

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t \quad (2.3)$$

onde  $E(\eta_t)=0$  e  $\text{Var}(\eta_t)=Q_t$

A *equação de observação* (2.2) é composta pelos seguintes elementos:  $\alpha_t$  o vetor de estados com dimensão  $m \times 1$ , que contém as informações relevantes do sistema;  $Z_t$  com dimensão  $N \times m$  estabelece a combinação linear dos estados que constituem a série  $Y_t$  observada; o vetor  $d_t$  com dimensão  $N \times 1$  incorpora o efeito de covariáveis na série  $y_t$ ; finalmente  $\varepsilon_t$  é um vetor  $N \times 1$  de erros serialmente não correlacionados com média zero e matriz de covariância  $H_t$ . Como em geral não é possível observar o vetor de estados, a adequada estimação dos elementos deste vetor constitui um papel fundamental na análise de modelos estruturais em estado de espaço;

A *equação de transição* (2.3.) em geral é descrita por um processo markoviano de primeira ordem, sendo possível a generalização para ordens superiores. Esta equação é composta por:  $T_t$  matriz  $m \times m$  que estabelece a combinação linear dos estados  $\alpha_{t-1}$ ;  $c_t$  vetor com dimensão  $m \times 1$  no qual incorpora-se o efeito de covariáveis nos estados;  $R_t$  é uma matriz  $m \times g$  que junto com  $\eta_t$  compõem o vetor de distúrbios serialmente não correlacionados.

Os elementos que compõem a matriz de covariância  $H_t$  dos erros da *equação de observação* ( $\varepsilon_t$ ), a matriz de covariância  $Q_t$  dos distúrbios da *equação de transição* e a matriz  $T_t$  são os parâmetros do modelo. Sob a suposição de estes erros e distúrbios terem distribuição Normal, estes parâmetros podem

ser estimados pelo método da máxima verossimilhança, e a função de verossimilhança é calculada utilizando-se a decomposição dos erros de predição que são fornecidos pelo filtro de Kalman.

Como podemos verificar os modelos na forma de espaço de estado são parcimoniosos e naturalmente definidos para séries temporais multivariadas. Já nos modelos ARIMA a generalização para o caso multivariado é mais complexa pois envolve uma grande quantidade de parâmetros.

As equações (2.2) e (2.3) compõe o *sistema de equações*, as matrizes  $Z_t$ ,  $d_t$ ,  $H_t$ ,  $T_t$ ,  $c_t$ ,  $R_t$  e  $Q_t$  estão indexadas pelo coeficiente  $t$ , ou seja, temos um modelo dinâmico onde sua estrutura evolui com o passar do tempo. Caso esta estrutura permaneça estática, as matrizes do *sistema de equações* não precisam ser indexadas por  $t$ .

### 2.3.1. Modelo de Volatilidade Estocástica

Sejam os retornos definidos como em (2.1). Tanto nos modelos da família xARCH como nos modelos estruturais eles são usualmente modelados como:

$$Y_t = \sigma_t \varepsilon_t \quad \text{onde } \varepsilon_t \sim N(0,1) \text{ i.i.d.} \quad (2.4)$$

A diferença entre estes modelos está na maneira com que descrevemos o processo de evolução dos  $\sigma_t$ . Nos modelos da família xARCH estes são função dos parâmetros, dos retornos passados e das inovações. Portanto, dado que os parâmetros sejam conhecidos, a volatilidade é uma componente observável.

Já nos modelos estruturais de volatilidade estocástica os  $\sigma_t$  constituem uma componente não observável, mesmo quando os parâmetros do modelo são conhecidos. O modelo estrutural escrito na forma de espaço de estados é representado da seguinte maneira:

Elevando-se ao quadrado e tomando-se o logaritmo em (2.4) temos:

$$\ln(Y_t^2) = \ln(\sigma_t^2) + \ln(\varepsilon_t^2) \quad (2.5)$$

E  $\ln(\sigma_t^2)$  é modelado como um processo auto-regressivo de primeira ordem.

$$\ln(\sigma_t^2) = \alpha + \gamma \ln(\sigma_{t-1}^2) + \eta_t \quad (2.6)$$

Como  $\varepsilon_t$  tem distribuição normal padrão, então  $\ln(\varepsilon_t^2)$  tem uma distribuição que chamaremos de *log-qui-quadrado*. Como esta variável aleatória tem  $E(\ln(\varepsilon_t^2)) \cong -1,27$  e  $\text{Var}(\ln(\varepsilon_t^2)) = \pi^2/2 \cong 4,93$ , então utilizaremos a transformação  $\xi_t = \ln(\varepsilon_t^2) + 1,27$ . A sequência  $\{\xi_t\}$  é iid com média zero e variância aproximadamente 4,93.

O modelo descrito pelas equações (2.5) e (2.6) será reescrito da seguinte forma:

$$\ln(Y_t^2) = \alpha^* + h_t + \xi_t \quad (2.7)$$

$$h_t = \gamma h_{t-1} + \eta_t \quad (2.8)$$

onde  $\alpha^* = \alpha/(1-\gamma) - 1,27$

$h_t = \ln(\sigma_t^2) - \alpha/(1-\gamma)$

As equações (2.7) e (2.8) definem um modelo estrutural representado na forma de espaço de estados, no qual a equação de medida é definida por um nível fixo mais uma componente auto-regressiva. As matrizes que compõem o sistema de equação são:

$$Z = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad T = \begin{bmatrix} 1 & 0 \\ 0 & \gamma \end{bmatrix} \quad H = 4.93 \quad \text{Var}(\eta_t) = \sigma_\eta^2$$

Portanto teremos a seguinte representação de espaço de estado para o modelo AR(1)\_SV:

$$Y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ h_t \end{bmatrix} + \xi_y \quad (2.9)$$

$$\begin{bmatrix} \alpha \\ h_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} \alpha \\ h_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix} \quad (2.10)$$

E pode-se mostrar que:

A variância não condicional é igual a:

$$\text{Var}(Y_t) = \exp\left(\frac{\sigma_\eta^2}{2(1-\gamma^2)} + \alpha\right) \quad (2.11)$$

e a autocorrelação serial de defasagem  $\tau$  p/ para  $Y_t^2$  é igual a:

$$\rho(\tau) = \frac{\exp(\sigma_\eta^2) - 1}{3 \exp(\sigma_\eta^2) - 1} \gamma^{\tau-1} \quad (2.12)$$

## 2.4 Filtro de Kalman.

Desenvolvido no início da década de 60 por Kalman, este procedimento tinha como principal objetivo prever e estimar a posição de um foguete a partir dos dados sobre a posição, velocidade e aceleração presentes fornecidos por telemetria. Esta técnica foi publicada originalmente em uma revista de engenharia, fora do contexto estatístico. Essa ferramenta passou a ser amplamente utilizada não só pelos engenheiros, com também pelos estatísticos.

O filtro de Kalman é um algoritmo muito útil em modelos que possam ser representados na forma de espaço de estados. Através do filtro de Kalman, é possível decompor o erro de predição e, sob a hipótese de estes terem distribuição Normal, é possível calcularmos a verossimilhança e então utilizarmos o método de máxima verossimilhança para estimarmos os parâmetros do modelo. Uma vez estimados os parâmetros, o filtro de Kalman possibilita a estimação e previsão dos estados  $\alpha_t$ . Estas estimativas possuem diversas propriedades, dentre elas a mais importante é que sob a hipótese de os componentes aleatórios  $\varepsilon_t$  e  $\eta_t$  terem distribuição normal, o filtro de Kalman fornece *estimativas de mínimo erro quadrático médio (MMSE) não viciadas*, para as componentes filtradas, suavizadas e inclusive para a predição um passo à frente mesmo quando a suposição de normalidade não é válida, o filtro de Kalman ainda nos fornece *estimativas lineares de mínimo erro quadrático médio (MMSLE)*, (Harvey 1989, cap. 3 pp. 111).

O estimador do estado  $\alpha_t$  considerando a informação até o instante  $t = 1 \dots T$ , é denotado por  $a_{t/t-1} = E(\alpha_t/Y_{t-1})$ , onde  $Y_{t-1} = (y_1, y_2, \dots, y_{t-1})$ , é chamado de estimador *filtrado* ou previsão um passo à frente. Como este estimador não utiliza a informação contida nos instantes presentes e futuros, ele é empregado na obtenção de previsões. Já o estimador do estado  $\alpha_t$  utilizando toda informação contida na amostra é denotado por  $a_{t/T} = E(\alpha_t/Y_T)$ , onde  $Y_T = (y_1, y_2, \dots, y_T)$ , é chamado de estimador *suavizado*. Como este estimador considera toda informação contida na amostra, este é empregado para análise do processo gerador dos dados.

## 2.5. Os diferentes tipos de outliers.

Existem diversos nomes para descrever observações espúrias ou inconsistentes com o modelo probabilístico: valores aberrantes, valores discrepantes ou, como utilizaremos nesta dissertação, *outliers*. Os outliers podem ser distinguidos em duas categorias: os que afetam somente as observações no tempo onde eles ocorrem e os que afetam as observações posteriores à sua ocorrência. Estes outliers são usualmente chamados de AO (*aditive outlier*) e IO (*innovation outlier*) respectivamente. Como o AO afeta apenas o nível da série e o IO a volatilidade, Hotta e Tsay (1999) chamam estes tipos de outliers de Level Outlier (LO) e Volatility outlier (VO - outlier de volatilidade), respectivamente, aos trabalhos em modelos da família ARCH.

Estas duas categorias de outliers podem ser incorporados ao modelo AR(1)\_SV através dos  $\Delta_t$  e  $\beta_t$  inseridos nas equações 2.7 e 2.8 da seguinte forma:

$$\ln(Y_t^2) = \alpha^* + h_t + \xi_t + \Delta_t \quad (2.13)$$

$$\text{onde } \Delta_t = 0 \text{ p/ } t \neq t_{AO} \text{ e } \Delta_t \neq 0 \text{ p/ } t = t_{AO}$$

$$h_t = \gamma h_{t-1} + \eta_t + \beta_t \quad (2.14)$$

$$\text{onde } \beta_t = 0 \text{ p/ } t \neq t_{IO} \text{ e } \beta_t \neq 0 \text{ p/ } t = t_{IO}$$

nas equações (2.15) e (2.16) temos a representação na forma de espaço de estado do modelo descrito pelas equações (2.13) e (2.14):

$$Y_t = \begin{bmatrix} 1 & 1 & I_{AO} & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ h_t \\ \Delta_t \\ \beta_t \end{bmatrix} + \xi_t \quad (2.15)$$

$$\begin{bmatrix} \alpha \\ h_t \\ \Delta_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \gamma & 0 & I_{IO} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ h_{t-1} \\ \Delta_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \\ 0 \\ 0 \end{bmatrix} \quad (2.16)$$

$$\text{onde: } I_{IO}=0 \text{ p/ } t \neq t_{IO} \text{ e } I_{IO}=1 \text{ p/ } t=t_{IO} \\ I_{AO}=0 \text{ p/ } t \neq t_{AO} \text{ e } I_{AO}=1 \text{ p/ } t=t_{AO}$$

No contexto dos modelos de séries temporais, os AOs são interpretados como um fenômeno local, que afeta somente a observação onde ocorreu e não propaga seus efeitos para observações futuras. Geralmente este tipo de outlier está associado a erros de digitação, erros no processo de armazenagem dos dados e principalmente a fatos extraordinários referentes ao processo gerador dos dados. O mais importante neste tipo de outlier é a não propagação de seu efeito para observações futuras - por isso a sua natureza estar relacionada com fatores exógenos ao processo gerador dos dados. Modelaremos este tipo de outlier através da equação de observação descrito pela equação 2.13. e consideraremos que  $\Delta_t \neq 0$  somente quando  $t = t_{AO}$ .

Já os outliers do tipo IO, que também são chamados de quebra estrutural, serão modelados através dos  $\beta_t$  descritos pela equação 2.14. As causas deste tipo de outlier podem ser as mais variadas. Apenas para citarmos um exemplo, a crise nos países asiáticos em julho de 1997 ocasionou um aumento significativo na volatilidade nas bolsas de valores do mundo inteiro.

## Capítulo III. Estimação dos parâmetros do modelo AR(1)\_SV.

Neste capítulo discutiremos a estimação dos parâmetros do modelo AR(1)\_SV dentre as diversas metodologias existentes, focaremos as análises no método da quase-verossimilhança. Apesar de ser aproximado, este método possui uma razoável eficiência computacional - característica essencial para o procedimento de varredura para a detecção de outliers, descrito no capítulo V. Também neste capítulo ,discutiremos os principais aspectos práticos para implementação deste método e realizaremos simulações para avaliação dos estimadores e dos intervalos de confiança.



### 3.1 Principais métodos para estimação dos parâmetros do modelo AR(1)\_SV.

- método dos momentos generalizados. Este método evita o problema associado à linearização do modelo assim como o cálculo da verossimilhança. Ele é simples de implementar e generalizar, mas a eficiência de seus estimadores é sub-ótima se comparada com os estimadores baseados na verossimilhança. Existem outros desdobramentos deste método tal como SMM (Simulation Method of Moments). (Andersen e Sørensen (1996)). Nenhum destes fornece estimativas instantâneas dos estados, isto é, da volatilidade.
- a abordagem bayesiana, proposta por Jacquier et alli (1994), tem apresentado resultados muito encorajadores mas tem a desvantagem da ineficiência computacional. Na abordagem Bayesiana, as distribuições a posteriori dos parâmetros e dos estados podem ser encontradas através dos métodos *MCMC* (*Monte Carlo Markov Chain*). Jacquier et alli propõe a utilização do amostrador de Gibbs com Metropolis-Hasttings. O método bayesiano se torna ainda mais ineficiente quando o tamanho da série cresce e quando os outliers são incorporados nos modelos.
- o método de máxima verossimilhança via Monte Carlo, proposto inicialmente por Koopman e Sadjmann (1997), consiste em decompor a verossimilhança em uma componente Gaussiana mais uma correção, que é obtida através de simulações. Este método, como demonstrado por Koopman, fornece estimativas com menor *erro quadrático médio* do que o método QML, mas sua eficiência computacional é muito inferior ao QML (porém melhor que a do MCMC). As correções utilizadas na verossimilhança também podem ser utilizadas para obtermos estimativas mais precisas dos estados do modelo.
- método da quase-verossimilhança (QML), proposto inicialmente por Harvey et alli (1994). O modelo AR(1)\_SV descrito na seção (1.3.1.) é caracterizado por três parâmetros  $\alpha$ ,  $\gamma$  e  $\sigma_\eta$ . Como vimos na seção (1.3.1.), o parâmetro  $\alpha$  é considerado um estado do modelo e portanto sua estimativa é obtida através do filtro de Kalman, restando-nos estimarmos  $\gamma$  e  $\sigma_\eta$ . Uma vez colocado este modelo na

forma de espaço de estado (1.3.), podemos utilizar o filtro de Kalman para encontrarmos os erros de predição um passo à frente (inovações) e seus respectivos erros quadráticos médios. Supondo que as inovações tenham distribuição normal e considerando os parâmetros conhecidos, poderíamos calcular a função densidade de probabilidade conjunta, encontrar a verossimilhança e, a partir dela, estimativas de máxima verossimilhança. Entretanto, no modelo AR(1)\_SV os distúrbios associados à equação de observação têm distribuição log-qui-quadrado. Conseqüentemente, o cálculo da verossimilhança não pode ser realizado diretamente como nas situações que temos a suposição de normalidade. O método da QML consiste em aproximar a distribuição log-qui-quadrado, que tem média aproximadamente igual  $-1,27$  e variância  $\pi^2/2 \approx 4,93$  para uma distribuição  $Normal(-1,27; \pi^2/2)$ . Na figura (3.1), verificamos as discrepâncias entre estas duas distribuições. Supondo que a distribuição das inovações obtidas pelo filtro de Kalman tem é gaussiana, a verossimilhança pode ser facilmente calculada. Este método tem a vantagem de ser computacionalmente eficiente.

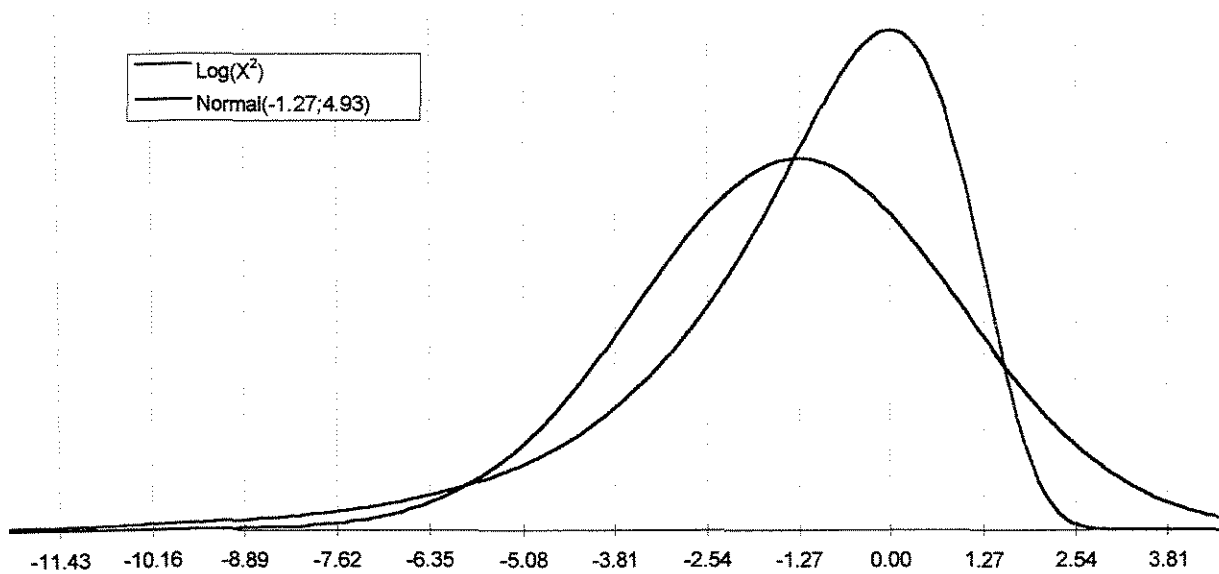


Figura 3.1. Distribuição log-qui-quadrado e normal(-1,27,4.93).

## 3.2. Aspectos práticos do método da máxima quase-verossimilhança.

Nesta seção trataremos dos aspectos práticos para implementação do método da QML para estimação dos parâmetros do modelo AR(1)\_SV. A verossimilhança não pode ser obtida como função explícita dos dados, portanto métodos numéricos serão necessários no cálculo e na maximização da quase-verossimilhança. Devido às restrições no espaço paramétrico, é aconselhável utilizar transformações nos parâmetros; porém, como veremos na seção 3.1.2, tais transformações podem acarretar problemas nos critérios de convergência e, para evitarmos estes problemas, sugeriremos algumas regras práticas.

### 3.2.1. Maximização numérica e reparametrização.

Apesar de a quase-verossimilhança ser obtida facilmente através das inovações e seus respectivos erros quadráticos médios, esta não pode ser escrita explicitamente como uma função das observações. Para maximizarmos esta função, utilizaremos o método numérico chamado **BFGS**, *Broyden-Fletcher-Goldfarb-Shanno* (Gill et alli, 1981). A restrição do espaço paramétrico é feita através de transformações. No modelo AR(1)\_SV,  $\gamma$  e  $\sigma_\eta$  estão restritos ao espaço definido por  $[-1;1] \times [0;\infty]$ . As transformações que utilizaremos são dadas pelas equações (3.1) e (3.2) e podem ser visualizadas nas figuras (3.2) e (3.3). Feitas as transformações, obteremos um novo conjunto de parâmetros ( $\omega$  e  $\psi$ ) que estão definidos em um novo espaço paramétrico definido em todo o  $\mathbb{R}^2$ .

As estimativas de máxima verossimilhança são invariantes em relação a transformações nos parâmetros. Esta propriedade pode ser estendida para estimativas de máxima quase-verossimilhança.

$$\gamma = \frac{\psi}{\sqrt{1+\psi^2}} \quad (3.1)$$

$$\sigma_\eta^2 = \exp(\omega) \quad (3.2)$$

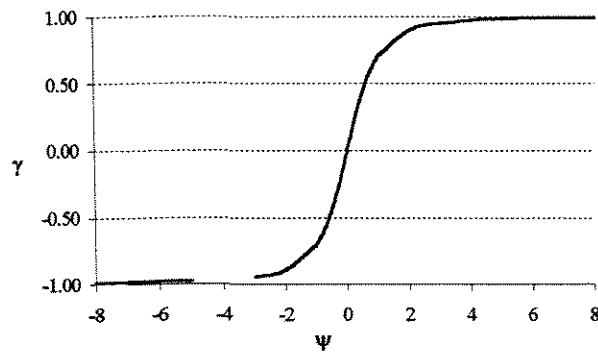


figura 3.2. Função de reparametrização de  $\gamma$ .

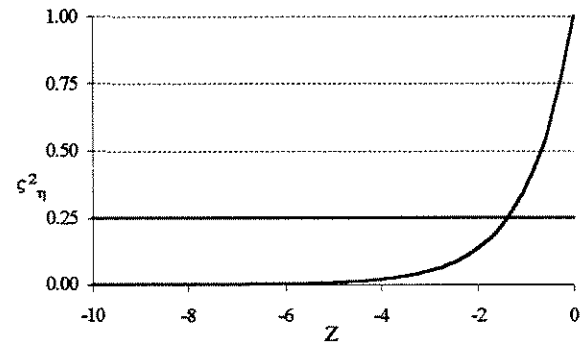


figura 3.3. Função de reparametrização  $\sigma_\eta^2$ .

### 3.2.2. Critérios de convergência.

Os seguintes critérios são utilizados para verificarmos a convergência do método numérico BFGS:

$$\text{critério i) - Verossimilhança: } |l(\theta_i) - l(\theta_{i+1})| / |l(\theta_i)| < \varepsilon_1$$

$$\text{critério ii) - Gradiente: } \frac{1}{m} \sum_{j=1}^m |q_j(\theta_i)| < \varepsilon_2$$

$$\text{critério iii) - Parâmetros: } \frac{1}{m} \sum_{j=1}^m |\theta_{i+1,j} - \theta_{i,j}| / |\theta_{i,j}| < \varepsilon_3$$

onde:  $\theta_{i,j}$  é o  $j$ -ésimo elemento do vetor de parâmetros na  $i$ -ésima iteração.

$l(\theta_i)$  é a log-verossimilhança (que no caso será a log-quase-verossimilhança).

$q_j(\theta_i)$  é o  $j$ -ésimo elemento do vetor de escores  $q(\theta_i)$ .

Quando os três critérios são atendidos, dizemos que temos uma *convergência forte*, ou seja, temos fortes evidências de que estamos próximos o suficiente do máximo local. Quando isto não ocorre, não podemos afirmar que o valor encontrado não está próximo do máximo local, pois existem casos em que temos um máximo local sem que todos os 3 critérios sejam atendidos. Por exemplo, em situações nas quais o ponto de máximo se encontra no limite do espaço paramétrico, o gradiente pode não ser próximo de zero, e portanto o *critério ii* não será atendido. Ainda, casos em que a função  $l(\theta_i)$  é achatada na região próxima ao ponto de máximo, o critério iii não será atendido nesta região. Nestes casos, ambos os critérios i e ii serão atendidos. A estes casos, quando somente o critério iii não é atendido, damos o nome de *convergência fraca*. Portanto, quando não temos uma convergência forte,

torna-se necessária uma análise mais detalhada dos resultados dos critérios e dos parâmetros encontrados, para concluirmos se o ponto de máximo está próximo o suficiente ou não.

Devido às transformações realizadas nos parâmetros, estimativas do parâmetro  $\gamma$  muito próximas ou iguais a um (1) e estimativas do parâmetro  $\sigma_{\eta}^2$  muito próximas ou iguais a zero corresponderão a estimativas dos parâmetros  $\psi$  e  $\omega$  tendendo a mais infinito e menos infinito, respectivamente. Conseqüentemente, a *condição iii* não será atendida. Para evitar problemas na aritmética do ponto flutuante, utilizaremos as seguintes regras:

$$\hat{\psi} > 20 \quad \equiv \quad \hat{\gamma} > 0.999 \quad \Rightarrow \quad \hat{\gamma} = 1$$

$$\hat{\omega} < -10 \quad \equiv \quad \hat{\sigma}_{\eta}^2 < 5 * 10^{-5} \quad \Rightarrow \quad \hat{\sigma}_{\eta}^2 = 0$$

### 3.3. Experimentos via simulação de Monte Carlo.

Para estudarmos as propriedades dos estimadores de máxima quase-verossimilhança, realizaremos experimentos com o uso do método de Monte Carlo. Estes experimentos serão conduzidos para que tenhamos 1000 estimativas dos parâmetros do modelo AR(1)\_SV para séries com 1000 e 2000 observações com os seguintes conjuntos de parâmetros:

$$\Theta_1 = \{\alpha=0; \gamma=0.98; \sigma_\eta=0.12\}$$

$$\Theta_2 = \{\alpha=0; \gamma=0.95; \sigma_\eta=0.20\}$$

Para que as séries simuladas não tenham influência de seus valores iniciais, geramos séries com um excedente de 100 observações em relação ao tamanho inicialmente desejado, e descartamos as 100 primeiras observações. Desta forma, anulamos a influência indesejada e continuamos a trabalhar com séries de tamanho adequado. Em todas as simulações foram utilizados  $h_t=0$  como valores iniciais.

Estes valores foram escolhidos com base nos parâmetros estimados para modelos de volatilidades descritos na literatura, com exceção do parâmetro  $\alpha$  que, por se tratar de um parâmetro de locação, não deve influenciar as estimativas dos outros dois parâmetros. Nestes experimentos utilizaremos como valor inicial do método iterativo BFGS os parâmetros verdadeiros. Como a convergência deste procedimento pode ser afetada pelos valores de inicialização, numa situação em que os parâmetros sejam desconhecidos, devemos modificar estes valores iniciais para averiguar o tipo de convergência. Para inicialização do filtro de Kalman, utilizaremos  $a_0 = 0$  e  $P_0 = 10^6$ . Como geralmente o parâmetro  $\alpha$  é desconhecido e como na nossa formulação ele é um estado, é necessário dar uma estimativa inicial. Podemos inicializá-lo com a média amostral das 20 observações, conforme sugerido por Harvey et alli (1992).

Os resultados dos experimentos serão analisados através de estatísticas descritivas das estimativas dos parâmetros, tais como: média, mediana, percentis e desvio padrão. Analisaremos a distribuição das estimativas dos parâmetros e dos parâmetros transformados através das densidades estimadas por um método não-paramétrico. Esta análise tem por objetivo comparar os resultados obtidos nas séries com 1000 e 2000 observações, além de verificar as propriedades das distribuições tais como assimetria e peso das caudas.

Finalmente analisaremos a confiança dos intervalos construídos, utilizando o método proposto por Ruiz (1994).

### 3.3.1 Estimativas dos parâmetros.

Os resultados dos experimentos estão descritos a seguir. Na tabela 3.1. podemos verificar a frequência com que as estimativas dos parâmetros convergiram para os limites do espaço paramétrico. No modelo com conjuntos de parâmetros  $\Theta_1$  e com a série de tamanho 1000, observamos uma grande incidência de estimativas do parâmetro  $\gamma$  iguais a 1 (7.1%). Entretanto esta frequência diminui significativamente quando aumentamos o tamanho da série (0.7%) ou quando temos o conjunto de parâmetros  $\Theta_2$ .

Tabela 3.1. Frequência observada nas estimativas dos parâmetros para os respectivos tamanho das séries e os conjuntos de parâmetros  $\Theta_1 = \{\alpha=0; \gamma=0.98; \sigma_\eta=0.12\}$ ;  $\Theta_2 = \{\alpha=0; \gamma=0.95; \sigma_\eta=0.20\}$ .

|            | n    | $\gamma = 1$ e $\sigma_\eta = 0$ | $\gamma = 1$ e $\sigma_\eta > 0$ | $\gamma < 1$ e $\sigma_\eta = 0$ | $\gamma < 1$ e $\sigma_\eta > 0$ |
|------------|------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| $\Theta_1$ | 1000 | 5                                | 66                               | 8                                | 921                              |
|            | 2000 | 0                                | 7                                | 0                                | 993                              |
| $\Theta_2$ | 1000 | 0                                | 6                                | 0                                | 994                              |
|            | 2000 | 0                                | 2                                | 0                                | 998                              |

Nos gráficos (3.3) e (3.4) temos as densidades empíricas marginais das estimativas dos parâmetros. Através destas densidades verificamos que a dispersão das estimativas dos parâmetros diminui consideravelmente quando aumentamos o tamanho da série para 2000 observações. Este fato também pode ser verificado nas tabelas (3.2), (3.3), (3.4) e (3.5) com as estatísticas descritivas, nas quais a variância das estimativas diminui significativamente. Nestas tabelas também verificamos que as estimativas dos parâmetros  $\gamma$  e  $\sigma_\eta$  são viciadas, pois os intervalos de confiança 95% da média não cobrem o valor verdadeiro. Analisando-se também a média, observamos que o estimador de  $\gamma$  tende a subestimar o parâmetro verdadeiro e o estimador de  $\sigma_\eta$  tende a superestimar-lo. Nota-se que nas séries com 2000 observações o vício diminui consideravelmente, contudo o intervalo de confiança 95% permanece não cobrindo o valor verdadeiro.



Quando analisamos as estimativas dos parâmetro transformados, verificamos que as estimativas dos parâmetros  $\psi$  não são viciadas, pois os I.C. 95% cobrem o valor verdadeiro. Não obstante, devemos analisar este resultado com cautela, pois nota-se que a densidade das estimativas, apresentadas no gráfico (3.4), que problemas de convergência para os limites do espaço paramétrico ocasionaram grandes valores nas estimativas do parâmetro  $\psi$  aumentando a variância e puxando a média destas estimativas para cima. Eliminando as estimativas que convergiram para o limite do espaço paramétrico, verificamos que o parâmetro  $\omega$  superestima na média, os valores verdadeiros, pois os I.C. 95% não cobrem o valor verdadeiro.

Finalmente, analisando as estimativas dos estados  $\alpha$ 's, verificamos que o filtro de Kalman fornece estimativas não viciadas para todos os conjuntos de parâmetros e para todos os tamanhos amostrais, pois todos os intervalos de confiança cobriram o valor verdadeiro. Através do gráfico (3.5), que mostra a estimativa da distribuição das estimativas e através da tabela (3.6) que apresenta as estatísticas descritivas das 1000 estimativas, verificamos que a variância da estimativa deste estado depende do conjunto de parâmetros referentes ao processo gerador dos dados e, obviamente, depende do tamanho da série analisada.

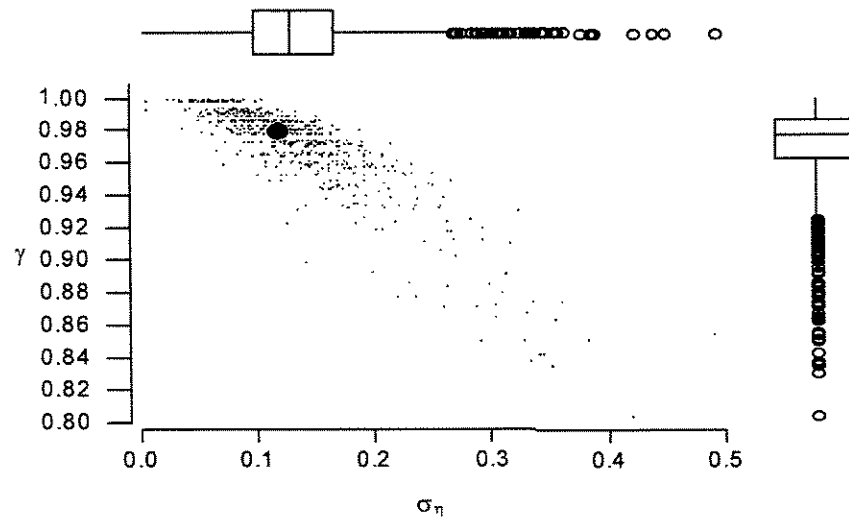


Gráfico 3.1. Dispersão das estimativas de QML do modelo AR(1)\_SV com conjunto de parâmetros  $\Theta_1$  e séries com 1000 observações. O ponto vermelho representa os parâmetros verdadeiros

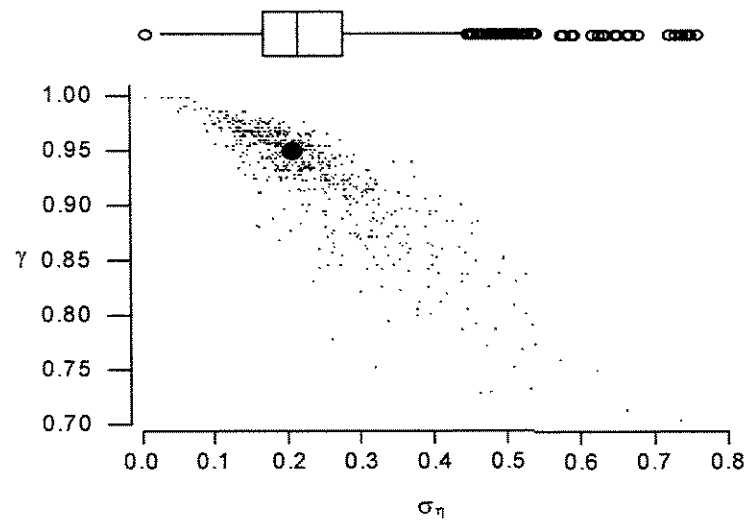


Gráfico 3.2. Dispersão das estimativas de QML do modelo AR(1)\_SV com conjunto de parâmetros  $\Theta_2$  e séries com 1000 observações. O ponto vermelho representa os parâmetros verdadeiros.

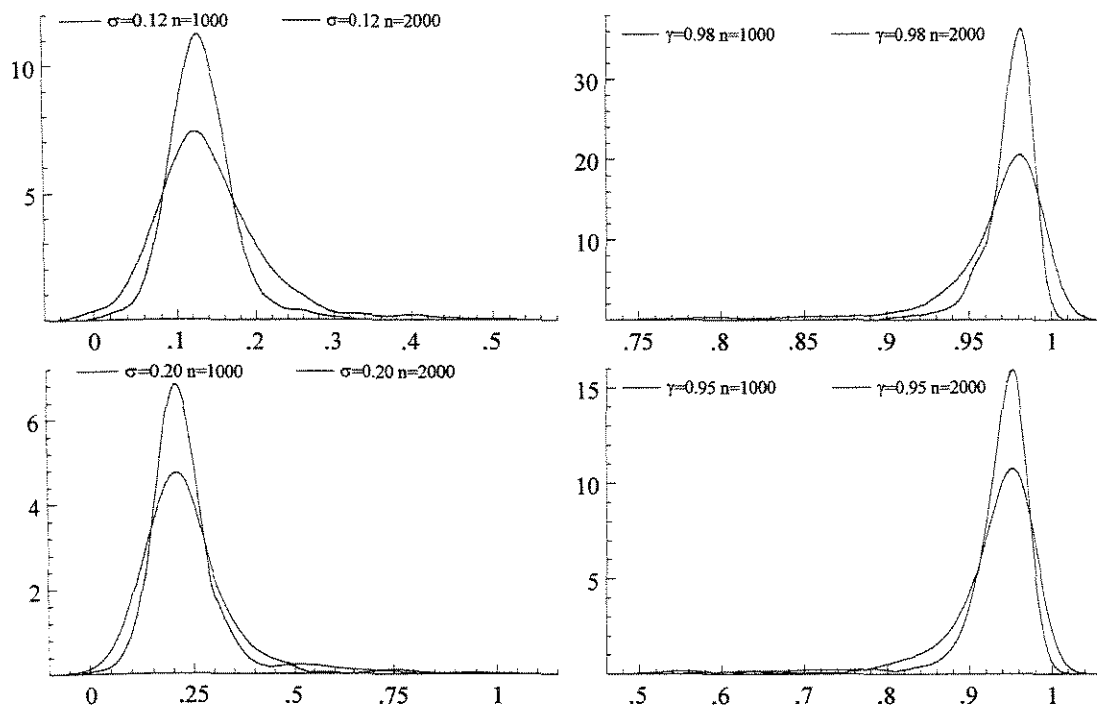


Gráfico 3.3. Densidade de probabilidade das estimativas de QML dos parâmetros  $\gamma$  e  $\sigma_\eta$  estimada através de 1000 replicações de séries com 1000 e 2000 observações.  
(legenda: parâmetros verdadeiros e tamanhos amostrais).

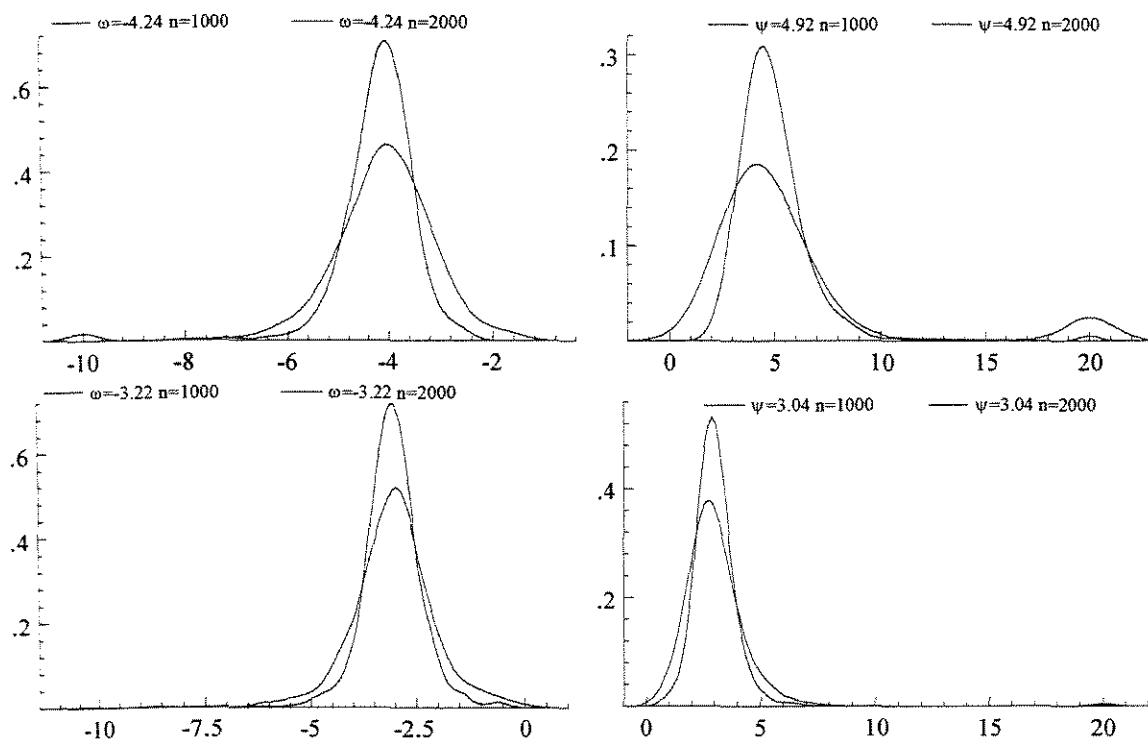


Gráfico 3.4. Densidade de probabilidade das estimativas de QML dos parâmetros  $\psi$  e  $\omega$  estimada através de 1000 replicações de séries com 1000 e 2000 observações.  
(legenda: parâmetros verdadeiros e tamanhos amostrais).

Tabela 3.2. Estatísticas descritivas de 1000 estimativas de máxima quase-verossimilhança de séries de tamanho 1000 e 2000. Entre parênteses os parâmetros verdadeiros.

| tamanho da série | 1000            | 2000            | 1000                   | 2000            |
|------------------|-----------------|-----------------|------------------------|-----------------|
| Estatística      | $\gamma$ (0.98) |                 | $\sigma_{\eta}$ (0.12) |                 |
| média            | 0.947           | 0.973           | 0.158                  | 0.132           |
| I.C. 95%         | [0.939 ; 0.955] | [0.971 ; 0.975] | [0.150 ; 0.165]        | [0.129 ; 0.135] |
| desvio padrão    | 0.127           | 0.030           | 0.120                  | 0.053           |
| percentil 95%    | 0.999           | 0.998           | 0.330                  | 0.200           |
| Mediana          | 0.975           | 0.978           | 0.133                  | 0.127           |
| percentil 5%     | 0.841           | 0.947           | 0.055                  | 0.073           |

Tabela 3.3. Estatísticas descritivas de 1000 estimativas de máxima quase-verossimilhança de séries de tamanho 1000 e 2000. Entre parênteses os parâmetros verdadeiros.

| tamanho da série | 1000            | 2000            | 1000              | 2000              |
|------------------|-----------------|-----------------|-------------------|-------------------|
| Estatística      | $\psi$ (4.92)   |                 | $\omega$ (-4.24)  |                   |
| média            | 5.612           | 4.982           | -4.146            | -4.165            |
| I.C. 95%         | [5.341 ; 5.883] | [4.854 ; 5.110] | [-4.225 ; -4.067] | [-4.208 ; -4.122] |
| desvio padrão    | 4.370           | 2.062           | 1.276             | 0.697             |
| percentil 95%    | 20.000          | 7.723           | -2.309            | -3.219            |
| Mediana          | 4.582           | 4.666           | -4.139            | -4.121            |
| percentil 5%     | 1.626           | 2.950           | -5.964            | -5.233            |

Tabela 3.4. Estatísticas descritivas de 1000 estimativas de máxima quase-verossimilhança de séries de tamanho 1000 e 2000. Entre parênteses os parâmetros verdadeiros.

| Tamanho da série | 1000            | 2000            | 1000                   | 2000            |
|------------------|-----------------|-----------------|------------------------|-----------------|
| Estatística      | $\gamma$ (0.95) |                 | $\sigma_{\eta}$ (0.20) |                 |
| média            | 0.903           | 0.935           | 0.253                  | 0.225           |
| I.C. 95%         | [0.894 ; 0.912] | [0.931 ; 0.938] | [0.243 ; 0.263]        | [0.220 ; 0.230] |
| Desvio padrão    | 0.145           | 0.050           | 0.158                  | 0.086           |
| Percentil 95%    | 0.982           | 0.975           | 0.546                  | 0.353           |
| Mediana          | 0.942           | 0.946           | 0.219                  | 0.211           |
| percentil 5%     | 0.691           | 0.871           | 0.102                  | 0.128           |

Tabela 3.5. Estatísticas descritivas de 1000 estimativas de máxima quase-verossimilhança de séries de tamanho 1000 e 2000. Entre parênteses os parâmetros verdadeiros.

| tamanho da série | 1000            | 2000            | 1000              | 2000             |
|------------------|-----------------|-----------------|-------------------|------------------|
| Estatística      | $\psi$ (3.04)   |                 | $\omega$ (-3.22)  |                  |
| média            | 3.001           | 2.998           | -3.052            | -3.109           |
| I.C. 95%         | [2.887 ; 3.115] | [2.927 ; 3.069] | [-3.120 ; -2.984] | [-3.154 ; 3.064] |
| desvio padrão    | 1.844           | 1.148           | 1.099             | 0.729            |
| percentil 95%    | 4.986           | 4.369           | -1.286            | -2.081           |
| Mediana          | 2.847           | 2.922           | -3.103            | -3.112           |
| percentil 5%     | 0.991           | 1.772           | -4.490            | -4.128           |

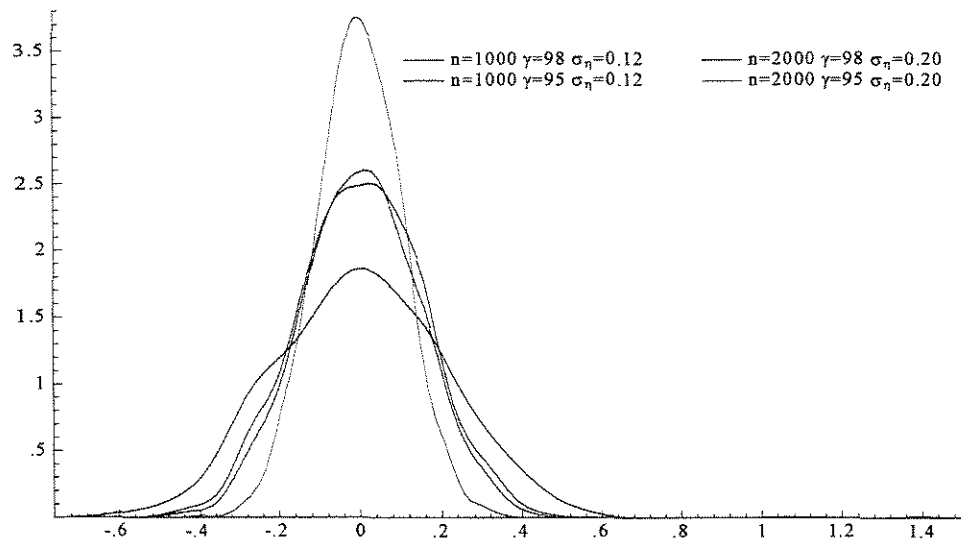


Gráfico 3.5. Densidade de probabilidade das estimativas de QML dos parâmetros  $\alpha$  estimada através de 1000 replicações de séries com 1000 e 2000 observações. (legenda: parâmetros verdadeiros e tamanhos amostrais).

Tabela 3.6. Estatísticas descritivas de 1000 estimativas de máxima quase-verossimilhança de séries de tamanho 1000 e 2000. Entre parênteses os parâmetros verdadeiros.

| tamanho da série | 1000                       | 2000              | 1000                       | 2000             |
|------------------|----------------------------|-------------------|----------------------------|------------------|
| Estatística      | $\alpha$ (0.00) $\Theta_1$ |                   | $\alpha$ (0.00) $\Theta_2$ |                  |
| média            | 0.001                      | 0.009             | -0.007                     | -0.003           |
| I.C. 95%         | [-0.012 ; .014]            | [-0.0001 ; 0.018] | [-0.016 ; 0.002]           | [-0.009 ; 0.003] |
| desvio padrão    | 0.208                      | 0.147             | 0.147                      | 0.104            |
| percentil 95%    | 0.346                      | 0.253             | 0.233                      | 0.168            |
| Mediana          | 0.002                      | 0.001             | -0.003                     | -0.005           |
| percentil 5%     | -0.321                     | -0.242            | -0.268                     | -0.179           |

### 3.3.2 Intervalos de confiança.

Encontradas as estimativas de máxima quase-verossimilhança, podemos construir intervalos de confiança. Usualmente construímos os intervalos de confiança utilizando a propriedade de que a distribuição dos estimadores de máxima verossimilhança é assintoticamente normal com matriz de variância e covariância função da *matriz de informação*. Dunsmuir (1979) mostrou que a matriz de variância e covariância das estimativas de máxima verossimilhança são definidas por:

Seja,

$$\sqrt{T}(\hat{\Theta} - \Theta) \stackrel{d}{\approx} N(0, C(\Theta))$$

onde.

$$C(\Theta) = 2A^{-1} + A^{-1}BA^{-1} \quad (3.3.)$$

Com A e B envolvendo integrais de funções das derivadas de funções geradora espectral do processo em relação a  $(\gamma, \text{e } \sigma_{\eta}^2)$  e o coeficiente do excesso de curtose do ruído (da forma reduzida do modelo na forma ARIMA) em relação à distribuição normal. É necessário que os dois processos de ruídos tenham quarto momento finito e que os parâmetros não estejam na fronteira do espaço paramétrico.

E no caso das estimativas de máxima quase-verossimilhança para o modelo de volatilidade estocástica, Ruiz, E. (1994) mostrou que as matrizes A e B podem ser escritas da seguinte forma:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \quad (3.4)$$

$$a_{11} = 4 \left[ \frac{1}{2(1-\lambda^2)} + \frac{\lambda^2 a}{(a^2 - b^2)^{3/2}} + \frac{1}{b^2} \left( 1 + \frac{a^3}{(a^2 - b^2)^{3/2}} - \frac{2a}{(a^2 - b^2)^{1/2}} \right) + \frac{2\lambda}{(a^2 - b^2)^{3/2}} - \frac{1}{q} + \frac{2}{q} \left( \frac{\lambda^2}{(a^2 - b^2)^{1/2}} + \frac{a}{b^2} \left( \frac{a}{(a^2 - b^2)^{1/2}} - 1 \right) + \left( 1 - \frac{a}{(a^2 - b^2)^{1/2}} \right) \right) \right] \quad (3.5)$$

$$a_{12} = \frac{4}{\pi} \left[ \frac{1}{q} \left( \frac{\lambda}{(a^2 - b^2)^{1/2}} - \frac{1}{b} \left( 1 - \frac{a}{(a^2 - b^2)^{1/2}} \right) \right) + \frac{\lambda a}{(a^2 - b^2)^{3/2}} + \frac{b}{(a^2 - b^2)^{3/2}} \right] \quad (3.6)$$

$$a_{22} = \frac{4a}{\pi^4 (a^2 - b^2)^{3/2}} \quad (3.7)$$

$$B = \frac{16}{(a^2 - b^2)^3} \left[ \begin{array}{cc} q^2 (\lambda a + b)^2 & \frac{-q((1 + \lambda^2)a + 2\lambda b)(\lambda a + b)}{\pi^2} \\ \frac{-q((1 + \lambda^2)a + 2\lambda b)(\lambda a + b)}{\pi^2} & \frac{((1 + \lambda^2)a + 2\lambda b)^2}{\pi^4} \end{array} \right] \quad (3.8)$$

onde  $a = q + 1 + \lambda^2$ ,  $b = -2\lambda$  e  $q = 2\sigma_\eta^2/\pi^2$



Assumindo que a distribuição das estimativas de máxima quase-verossimilhança também seja aproximadamente normal, construiremos os intervalos de confiança utilizando os dados gerados nos experimentos de Monte Carlo realizados na seção anterior. Os intervalos de confiança foram construídos somente para as estimativas que foram diferentes dos limites do espaço paramétrico; as estimativas que convergiram para estes limites não foram consideradas neste estudo. Os intervalos foram calculados para os parâmetros originais e a confiança utilizada foi de 90% e 95%. Para analisarmos a qualidade dos intervalos construídos utilizando esta metodologia, verificamos a proporção de intervalos que cobriram os valores verdadeiros dos parâmetros. Os resultados estão descritos na tabela (3.7.).

tabela 3.7. Confiança dos intervalos estimada pelo método de Monte Carlo.

| Conjunto de Parâmetros | Confiança Nominal | Parâmetro |       |                   |       |
|------------------------|-------------------|-----------|-------|-------------------|-------|
|                        |                   | $\gamma$  |       | $\sigma^2_{\eta}$ |       |
|                        |                   |           |       | tamanho da série  |       |
|                        |                   | 1000      | 2000  | 1000              | 2000  |
| $\Theta_1$             | 90%               | 0.931     | 0.962 | 0.944             | 0.930 |
|                        | 95%               | 0.959     | 0.980 | 0.961             | 0.950 |
| $\Theta_2$             | 90%               | 0.921     | 0.972 | 0.893             | 0.895 |
|                        | 95%               | 0.948     | 0.983 | 0.931             | 0.931 |

Os resultados foram satisfatórios, pois obtivemos uma confiança próxima da confiança nominal. Apenas o parâmetro  $\gamma$  com séries de 2000 observações é que houve uma superestimação da variância e consequentemente a confiança foi maior que a confiança nominal desejada.

## Capítulo IV - Detecção de outliers através dos resíduos auxiliares.

Neste capítulo, trataremos da detecção de outliers nos modelos AR(1)\_SV através dos resíduos auxiliares, que são estimativas dos distúrbios aleatórios dos componentes não observáveis. Esta metodologia foi inicialmente proposta por Harvey e Koopman (1992), doravante denotado por HK, no contexto de modelos estruturais com perturbações Gaussianas. Na seção 4.1, apresentaremos as características dos resíduos auxiliares. Mostraremos que mesmo com a correta especificação do modelo e, condicionam-se aos parâmetros verdadeiros, estes resíduos terão a desvantagem de serem serialmente correlacionados e correlacionados entre si.

Descreveremos na seção 4.2 os testes propostos por HK para detecção de outliers em modelos de *tendência localmente linear* e em modelos de *passeio ao acaso mais ruído branco*. Por se tratarem de testes assintóticos, assume-se que os parâmetros do modelo são estimados sem erro, ou com influência do erro desprezível para aplicação dos mesmos. Estes testes são baseados nas estatísticas de excesso de curtose e de assimetria. Lomnicki (1961) incorporou a correlação serial na distribuição assintótica dos momentos amostrais e das estatísticas de excesso de curtose e assimetria.

Como os distúrbios aleatórios da equação de observação tem uma distribuição log-qui-quadrado, os resíduos auxiliares irregulares terão uma distribuição aproximadamente log-qui-quadrada, visto que estes são estimativas das componentes aleatórias da equação de observação. Conseqüentemente, a distribuição destes resíduos terá uma acentuada assimetria e uma elevada curtose, que dificulta a utilização das estatísticas propostas por HK. Por este motivo, realizaremos a transformação inversa descrita na subseção 4.1.1. para obtermos uma distribuição aproximadamente Normal.

Na seção 3.3.1 do capítulo III, mostramos que os parâmetros do modelo AR(1)\_SV são estimados com erro significativo pelo método QML, mesmo em séries com 1000 e 2000 observações. Na seção 4.2.1., veremos que a distribuição das estatísticas propostas por HK têm uma forte dependência com os parâmetros geradores do modelo. Portanto, a distribuição de referência para realização dos testes estatísticos utilizando as estatísticas de excesso de curtose e assimetria não podem ser encontradas através da metodologia proposta por HK, nos casos em que os parâmetros foram

estimados por QML. Outro agravante para utilização dos testes assintóticos nesses casos está relacionado ao fato de que a distribuição do componente irregular não possui distribuição normal, e por isso as distribuições assintóticas encontradas por Lomnicki não são válidas. Apesar de a distribuição destes resíduos ser visualmente semelhante à Normal, sua curtose é maior indicando caudas mais pesadas e, conseqüentemente, invalidando os resultados obtidos por Lomnick.

Além das estatísticas propostas por HK analisaremos também o máximo dos resíduos auxiliares: já que outliers geralmente estão associados com resíduos de grande magnitude, esta seria uma estatística útil na detecção dos mesmos. Na seção 4.3. mostraremos que a distribuição do máximo também tem uma forte dependência com os parâmetros geradores do modelo.

Porém, os resíduos auxiliares são úteis para como instrumento de análise exploratória, pois valores com grande magnitude geralmente estão associados a outliers. Nos casos discorridos, a utilização de estatísticas que podem resumir a informação contida na amostra fica muito prejudicada. Isso ocorre devido à grande variabilidade dos estimadores de QML. A conclusão seria a mesma no caso de se utilizar estimadores de máxima verossimilhança dado que estes estimadores também possuem grande variabilidade (ver estudos de Jaquier et alli, 1994).

Concluimos este capítulo com a apresentação de duas séries simuladas e seus respectivos resíduos auxiliares estimados pelo filtro de Kalman. Na primeira série não teremos outliers de nenhuma natureza e assim, poderemos comparar estes resultados com os resultados obtidos pela segunda série, nas quais teremos os mesmos parâmetros do modelo e introduziremos três outliers, sendo dois AOs na posição 250 e 750 e um IO na posição 500. As influências destes outliers poderão ser notadas nos resíduos auxiliares e em suas respectivas estatísticas K(excesso de curtose) e S(assimetria).

## 4.1. Resíduos Auxiliares.

Usualmente diagnósticos de especificação do modelo são feitos através das *inovações*, isto é, estimativas dos erros de previsão um passo à frente. No entanto, em modelos estruturais com componentes não observáveis, existem outros resíduos disponíveis. Estes *resíduos auxiliares* são estimativas dos distúrbios aleatórios do modelo de espaço de estados.

Nesta seção, discutiremos as características dos resíduos auxiliares. Estas serão muito importantes e mostraremos que, mesmo que o modelo tenha sido corretamente especificado, estes resíduos serão serialmente correlacionados e correlacionados entre si, dificultando, desta forma, a detecção de outliers e sobretudo a identificação de sua posição e seu tipo.

Utilizaremos a seguinte notação para descrevermos um processo ARMA(p,q)

$$\phi(L) X_t = \theta(L) \varepsilon_t \quad (4.1.)$$

$$\text{onde: } \phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 \dots - \phi_p L^p)$$

$$\theta(L) = (1 - \theta_1 L - \theta_2 L^2 - \theta_3 L^3 \dots - \theta_q L^q)$$

$$L \text{ é o operador diferença tal que: } L^j(X_t) = X_{t-j}$$

$$\varepsilon_t \text{ é um ruído branco } E(\varepsilon_t) = 0 \text{ e } \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$$

O modelo AR(1)\_SV, descrito no capítulo I pelas equações (2.7) e (2.8), também pode ser representado da seguinte forma:

$$X_t = \ln(Y_t^2) - \alpha^* = \ln(\sigma_t^2) + \xi_t = \eta_t / (1 - \gamma L) + \xi_t \quad (4.2)$$

A equação (4.2) mostra que  $X_t$  é o resultado da soma de um processo AR(1) com  $\phi_1 = \gamma$  e um ruído branco com distribuição log-quadrado, ambos independentes entre si. O resultado desta soma é um processo ARMA(1,1) na forma  $(1 - \gamma L)X_t = (1 - \theta_1 L)\varepsilon_t$ , no qual o parâmetro  $\theta_1$  correspondente à componente média móvel é facilmente encontrado através da autocovariância de primeira ordem e da variância do processo.

Portanto temos que:

$$X_t \sim \text{ARMA}(1,1) \quad \phi_1 = \gamma \quad (4.3.)$$

$$\theta_1 = \varphi = \frac{q + \gamma^2 + 1 - \sqrt{(q + \gamma^2 + 1)^2 - 4\gamma^2}}{2\gamma} \quad (4.4.)$$

onde  $q = \sigma_\eta^2 / \sigma_\varepsilon^2$

Modelos estruturais lineares univariados na forma de espaço de estado sempre podem ser representados através da soma de processos ARIMAs, que também é outro processo ARIMA. De modo geral temos:

$$y_t = \sum_{i=0}^m \mu_{it} = \sum \frac{\theta_i(L)}{\phi_i(L)} \varepsilon_{it} = \frac{\theta(L)}{\phi(L)} \varepsilon_t \quad (4.5.)$$

onde:  $\varepsilon_{it}$  e  $\varepsilon_t$  são ruídos branco independentes entre si com média zero e  $\text{Var}(\varepsilon_{it}) = \sigma_i^2$  e  $\text{Var}(\varepsilon_t) = \sigma^2$ .

Os *estimadores lineares de mínimo erro quadrático médio* (MMSLE - minimum mean square linear estimator) de  $\mu_{it}$  e  $\varepsilon_{it}$ , para observações longe dos extremos da série são dados por, Bell (1984):

$$\hat{\mu}_{it} = \frac{\frac{|\theta_i(L)|^2}{|\phi_i(L)|^2} \sigma_i^2}{\frac{|\theta(L)|^2}{|\phi(L)|^2} \sigma^2} y_t \quad (4.6)$$

onde  $|\theta(L)|^2 = \theta(L) \cdot \theta(F)$  e  $F = L^{-1}$

$$\hat{\varepsilon}_{it} = \frac{\phi(F) \theta_i(F) \sigma_i^2}{\phi_i(F) \theta(F) \sigma^2} \varepsilon_t \quad (4.7)$$

As estimativas dos resíduos auxiliares obtidas através do filtro de Kalman também são MMSLE. Consequentemente os resíduos auxiliares de um modelo AR(1)\_SV seguem:

*estimativa do distúrbio aleatório correspondente a componente auto-regressivas (resíduo auxiliar estrutural):*

$$\hat{\eta}_t = \frac{\sigma_\eta^2}{(1 - \theta F)\sigma^2} \varepsilon_t \quad \text{AR(1)} \quad (4.8)$$

*estimativa do distúrbio aleatório correspondente a componente irregular (resíduo auxiliar irregular):*

$$\hat{\zeta}_t = \frac{(1 - \gamma F)\sigma_\varepsilon^2}{(1 - \theta F)\sigma^2} \varepsilon_t \quad \text{ARMA(1,1)} \quad (4.9)$$

Na tabela 4.1. temos alguns valores típicos dos parâmetros do modelo AR(1)\_SV e o respectivo parâmetro  $\phi$  calculado através da equação (4.4). com os parâmetros  $\gamma$  e  $\sigma_\eta^2$  conhecidos. Nesta tabela, verificamos que os resíduos auxiliares referentes à componente irregular da equação de observação seguirão um processo ARMA(1,1), cujas raízes dos polinômios são muito próximas. Desta forma, ocorrerá um quase cancelamento destes polinômios na equação (4.9), e conseqüentemente, a estrutura de autocorrelação apresentará uma estrutura de autocorrelação próxima a de um ruído branco.

Já a estimativa dos distúrbios da componente auto-regressiva do modelo AR(1)\_SV seguem um processo AR(1) com uma forte dependência das inovações posteriores, pois o parâmetro  $\theta$  está relativamente próximo de 1. Este fato dificultará a detecção de IOs, uma vez que este tipo de outlier propagará seus efeitos nos resíduos próximos à sua ocorrência, influenciando as estimativas do estado  $h_t$  (log-volatilidade) e dos resíduos auxiliares.

tabela 4.1. Valores de  $\phi$  calculados pela equação 4.4 para alguns valores típicos dos parâmetros do modelo AR(1)\_SV.

| $\gamma$ | $\sigma_\eta^2$   | $\sigma_\varepsilon^2$ | $q$                                   | $\theta$ |
|----------|-------------------|------------------------|---------------------------------------|----------|
| 0,98     | 0,12 <sup>2</sup> | 4,93                   | 0,12 <sup>2</sup> /4,93 $\cong$ 0,003 | 0,94     |
| 0,95     | 0,20 <sup>2</sup> | 4,93                   | 0,20 <sup>2</sup> /4,93 $\cong$ 0,008 | 0,90     |
| 0,90     | 0,30 <sup>2</sup> | 4,93                   | 0,30 <sup>2</sup> /4,93 $\cong$ 0,018 | 0,84     |

#### 4.1.1. Transformação inversa do resíduo auxiliar irregular.

As estimativas dos *resíduos auxiliares irregulares* devem apresentar uma distribuição aproximadamente log-qui-quadrado. Portanto, utilizaremos a seguinte transformação para obtermos uma distribuição aproximadamente normal padrão:

$$\kappa_t = \text{senal}_t \sqrt{\exp\left(\frac{\pi}{\sqrt{2 * \text{Var}(\hat{\xi}_t)}} * \hat{\xi}_t - 1.27\right)} \quad (4.10)$$

onde:  $\text{senal}_t = -1$  se  $\text{retorno}_t < 0$

$\text{senal}_t = 1$  se  $\text{retorno}_t \geq 0$

Na formulação do modelo AR(1)\_SV na seção 1.x., verificamos que a distribuição log-qui-quadrado tem média aproximadamente -1.27 e variância igual a  $\pi^2/2 \cong 4.93$ . Por este motivo, realizamos uma transformação na equação (4.10) de tal forma que os resíduos auxiliares tenham a mesma média e a mesma variância da distribuição log-qui-quadrado. Consequentemente, o exponencial destes resíduos terá distribuição aproximadamente qui-quadrado com 1 grau de liberdade e a raiz quadrada terá distribuição aproximadamente Normal, exceto o sinal, que pode ser facilmente recuperado pelos retornos.

## 4.2. Testes assintóticos para detecção de outliers.

Na seção 1.5, mostramos a existência de diferentes tipos de outliers cada qual com características intrínsecas. Veremos na seção 4.3 como a presença de outliers afetam os resíduos auxiliares e suas respectivas distribuições. Um dos efeitos observados é a mudança na medida de assimetria e um aumento da curtose. Baseados nestes fatos, HK propuseram que a detecção dos outlier seja realizada através das estatísticas  $K$  (excesso de curtose) e  $N=S^2 + K^2$  (uma combinação entre o excesso de curtose e a medida de assimetria) dos resíduos auxiliares. Uma vez que se tratam de testes assintóticos, assume-se que os parâmetros do modelo são estimados sem erro, ou com influência do erro desprezível.

Na seção 4.1. mostramos a relação dos parâmetros do modelo com a autocorrelação e a correlação entre os resíduos auxiliares. Esta autocorrelação é incorporada na distribuição assintótica das estatísticas  $K$  e  $S$ , como demonstrado por Lomnicki (1961) da seguinte forma:

Seja  $m_\alpha$  um estimador para o  $\alpha$ -ésimo momento central

$$m_\alpha = T^{-1} \sum_{t=1}^T (x_t - \bar{x})^\alpha \quad (4.11)$$

e defina

$$k(\alpha) = \sum_{\tau=-\infty}^{\infty} \rho_\tau^\alpha \quad \alpha = 2, 3 \text{ e } 4. \quad (4.12)$$

onde:  $\rho_\tau$  é a autocorrelação de  $\tau$ -ésima ordem.

$$\sqrt{b_1} = m_3 / m_2^{3/2} \quad (\text{medida de assimetria}) \quad (4.13)$$

$$b_2 = m_4 / m_2^2 \quad (\text{medida de curtose}) \quad (4.14)$$



Lomnicki demonstrou que, sob a suposição de normalidade, as estatísticas K e S, definidas pelas equações 4.15 e 4.16, têm distribuição assintoticamente **normal padrão**.

$$K = (b_2 - 3) / \sqrt{24k(4)/T} \quad (4.15)$$

$$S = \sqrt{\frac{b_1 T}{6k(3)}} \quad (4.16)$$

Os testes para detecção de outliers propostos por HK são baseados nestas estatísticas que, por sua vez, utilizam o 3º e o 4º momentos amostrais. Estes momentos são sensíveis à valores extremos e conseqüentemente, os testes sugeridos possuem poder de detecção, mesmo com a autocorrelação dos resíduos auxiliares.

A suposição de normalidade é necessária para que as estatísticas K e S tenham distribuição assintoticamente **normal padrão**. Mostraremos na seção 4.3 que quando esta suposição não é satisfeita, ainda assim a distribuição destas estatísticas será assintoticamente Normal, mas a média e a variância apresentarão uma forte dependência com os parâmetros e as estatísticas K e S e conseqüentemente terão distribuição assintótica normal com média e variância diferentes de zero e um.

### 4.2.1. Comparação das distribuições das estatísticas de interesse para diferentes conjuntos de parâmetros.

Mostraremos nesta subsecção que a distribuição das estatísticas K e S referentes aos resíduos auxiliares do modelo AR(1)\_SV têm uma forte dependência com os parâmetros geradores do modelo, sendo que esta dependência não ocorre somente em relação à variância: para diferentes conjuntos de parâmetros, as médias das distribuições assintóticas também serão distintas.

Como os parâmetros do modelo AR(1)\_SV são estimados com incerteza significativa pelo método da QML e outros métodos, a distribuições das estatísticas K e S têm uma forte dependência nos parâmetros geradores, não será possível encontrar uma distribuição de referência tal qual como HK obtiveram. Por este motivo, não realizaremos testes com as estatísticas K e S. Utilizaremos os resíduos auxiliares apenas com uma ferramenta para análise exploratória.

Para compararmos as distribuições das estatísticas de interesse, foram simuladas 3000 séries para três conjuntos de parâmetros distintos:

$$\Theta_1=(\alpha=0; \gamma=0.98; \sigma_\eta=0.12)$$

$$\Theta_2=(\alpha=0; \gamma=0.95; \sigma_\eta=0.20)$$

$$\Theta_3=(\alpha=0; \gamma=0.90; \sigma_\eta=0.30)$$

Estes parâmetros foram escolhidos com base no diagrama de dispersão das estimativas de QML dos parâmetros  $\gamma$  e  $\sigma_\eta$ , descritos no capítulo II pelos gráficos (3.1) e (3.2). Através destes gráficos verificamos que os conjuntos  $\Theta_1$  e  $\Theta_2$  são relativamente centrais à dispersão das estimativas, e o conjunto  $\Theta_3$  pode ser considerado localizado na extremidade do diagrama de dispersão.

Cada série simulada foi composta por 1000 observações. Assumindo os parâmetros conhecidos, foram calculados os resíduos auxiliares e suas respectivas estatísticas de interesse. Desta forma pudemos construir a função distribuição de probabilidade empírica através destas 3000 amostras simuladas e conseqüentemente compararmos suas distribuições, no gráfico 4.4 e algumas estatísticas descritivas na tabela 4.2 para o resíduo auxiliar irregular transformado e no gráfico 4.5 e tabela 4.3 para o resíduo auxiliar estrutural. Os resultados mostram que:

1. Os resíduos auxiliares irregulares mostraram-se simétricos porém com curtose maior do que a da distribuição normal. Observamos que as médias da medida de assimetria (S) das 3000 amostras simuladas para cada conjunto de parâmetros não foram estatisticamente diferente de zero. Ainda, as médias do excesso de curtose para cada conjunto de parâmetros foram estatisticamente diferentes entre si e diferentes de zero.
2. Os resíduos auxiliares estruturais apresentaram, em média, coeficiente de assimetria negativo, e foram os que apresentaram a maior dependência com os parâmetros geradores do modelo. O excesso de curtose dos resíduos auxiliares estruturais não apresentou diferenças tão grandes como a assimetria, porém foram suficientes para que uma aproximação não fosse válida.
3. A distribuição do máximo dos resíduos auxiliares irregulares e estruturais também apresentaram uma forte dependência nos parâmetros geradores do modelo.
4. Dado que os parâmetros geradores do modelo são iguais a  $\Theta_1$ , vimos, no capítulo 3, que as estimativas de QML referentes a esta série poderiam estar, com uma probabilidade alta, próximos de  $\Theta_2$  e com menor probabilidade próximos de  $\Theta_3$ . Nestas situações, se assumíssemos que as estimativas de QML são muito próximas dos parâmetros geradores do modelo, ou seja, que os parâmetros foram estimados quase que sem erro ou com influência do erro desprezível. As distribuições de referência das estatísticas de interesse poderiam ser obtidas através de simulação, mas como veremos nos gráficos a seguir, para um tamanho nominal do teste de 5%, teríamos que os tamanhos reais do teste calculados através desta metodologia ficariam muito aquém do desejado.

Os resultados obtidos pelas simulações citadas acima estão descritas a seguir nos gráficos (4.1) e (4.2) e pelas tabelas (4.2) e (4.3).

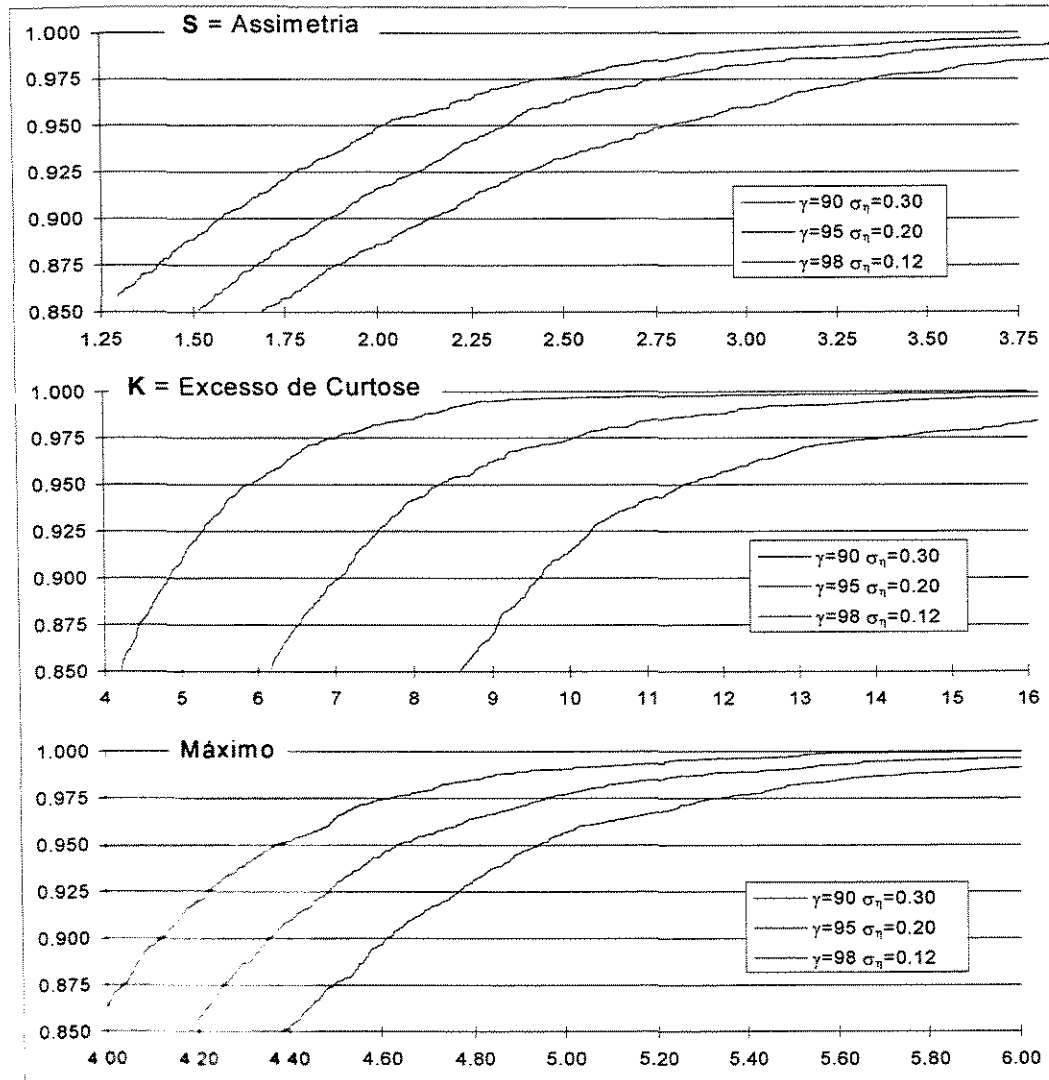


Gráfico 4.1. Comparação das distribuições empíricas das estatísticas de interesse do resíduo irregular

Tabela 4.2. Estatísticas descritivas das distribuições das estatísticas de interesse para cada conjunto parâmetro

|                    |           | $\gamma=90$ $\sigma_\eta=0.30$ | $\gamma=95$ $\sigma_\eta=0.20$ | $\gamma=98$ $\sigma_\eta=0.12$ |
|--------------------|-----------|--------------------------------|--------------------------------|--------------------------------|
| Assimetria         | Média     | -0,053                         | -0,046                         | -0,042                         |
|                    | Variância | 3,570                          | 2,271                          | 1,666                          |
| Excesso de Curtose | Média     | 5,766                          | 3,837                          | 2,301                          |
|                    | Variância | 31,205                         | 7,902                          | 4,294                          |
| Máximo             | Média     | 3.8143                         | 3.6665                         | 3.5215                         |
|                    | Variância | 0.4240                         | 0.2968                         | 0.2155                         |

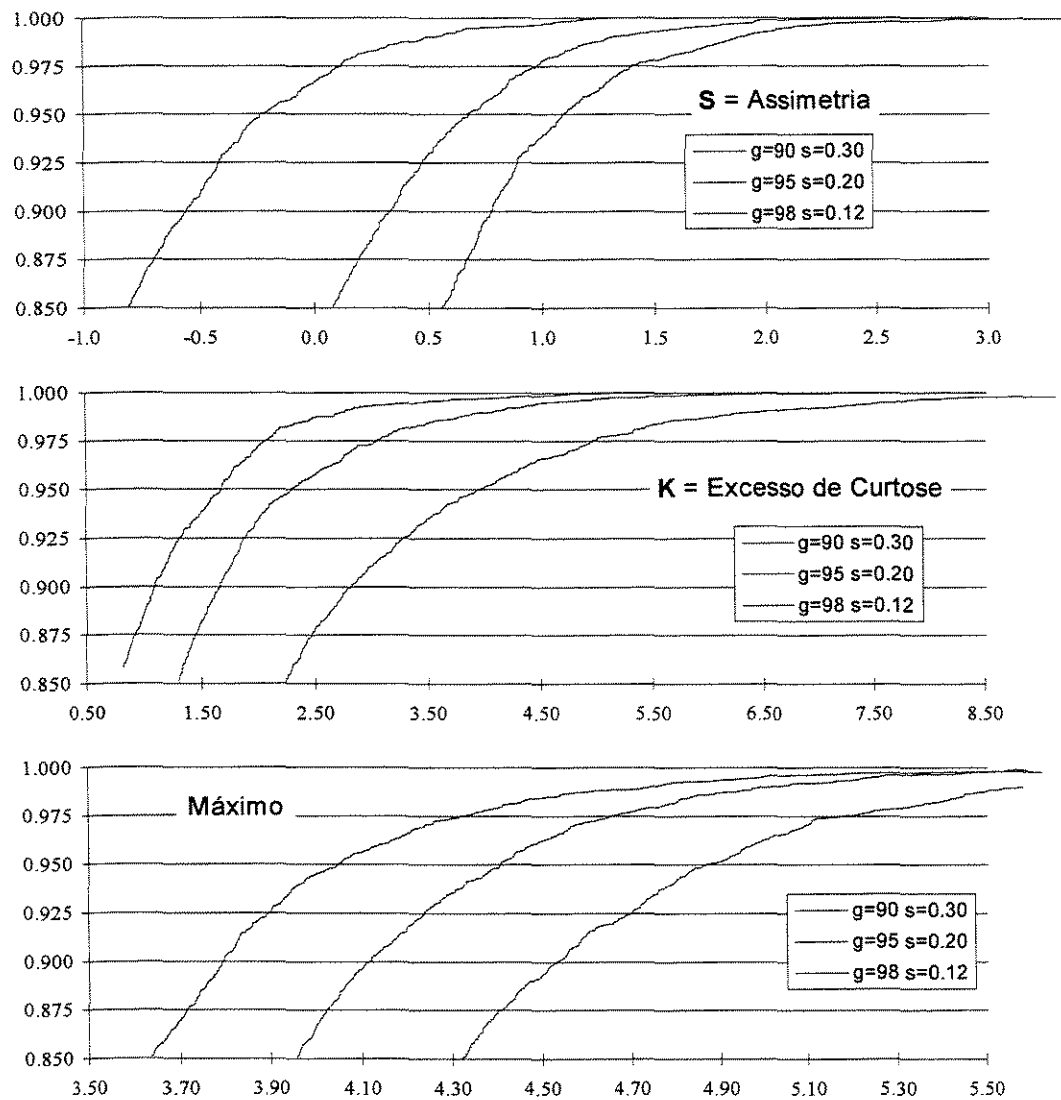


Gráfico 4.2. Comparação das distribuições empíricas das estatísticas de interesse do resíduo estrutural.

Tabela 4.3. Estatísticas descritivas das distribuições das estatísticas de interesse para cada conjunto parâmetro

|                               |           | $\gamma=0.90 \sigma_{\eta}=0.30$ | $\gamma=0.95 \sigma_{\eta}=0.20$ | $\gamma=0.98 \sigma_{\eta}=0.12$ |
|-------------------------------|-----------|----------------------------------|----------------------------------|----------------------------------|
| <b>S = Assimetria</b>         | Média     | -1,967                           | -0,938                           | -0,364                           |
|                               | Variância | 1,271                            | 1,035                            | 0,856                            |
| <b>K = Excesso de Curtose</b> | Média     | 0,949                            | 0,267                            | -0,029                           |
|                               | Variância | 2,436                            | 1,245                            | 0,789                            |
| <b>Máximo</b>                 | Média     | 3.735                            | 3.434                            | 3.195                            |
|                               | Variância | 0.361                            | 0.271                            | 0.240                            |

### 4.3. Análise dos resíduos auxiliares utilizando séries simuladas.

Nesta seção mostraremos, através de exemplos, as características dos resíduos auxiliares do modelo AR(1)\_SV. As duas séries-exemplo foram simuladas com 1000 observações utilizando os seguintes conjuntos de parâmetros:

**Série exemplo #1:**  $\alpha=-9,000$ ;  $\gamma=0.980$ ;  $\sigma_{\eta}=0.120$ ;

**Série exemplo #2:**  $\alpha=-9,000$ ;  $\gamma=0.980$ ;  $\sigma_{\eta}=0.120$ ;  $\beta_{500}=3,000$ ;  $\Delta_{250}=4,447$ ;  $\Delta_{750}=4,560$ ;

Temos duas séries com alta persistência, sendo que a primeira não possui outliers de nenhuma natureza e na segunda temos três outliers, dois AO's nas posições 250 e 750, e um IO na posição 500. O tamanho do IO foi escolhido de tal maneira que fosse aproximadamente 5 desvio-padrões da série auto-regressiva geradora do modelo, ou seja,  $5 * (0.12^2/(1-0.98^2))^{0.5} \cong 3$ . Já o tamanho dos AOs foram escolhidos de tal forma que fossem aproximadamente 5 desvio-padrões da série de retornos geradas pelo modelo AR(1)\_SV. Estes outliers foram escolhidos propositalmente com tamanhos relativamente grandes para que suas influências e características ficassem bem destacadas nos resíduos auxiliares.

Observamos claramente, através dos resíduos auxiliares obtidos nestas duas séries, as características apresentadas na seção anterior. A transformação inversa dos resíduos auxiliares irregulares conseguiu reproduzir uma distribuição aproximadamente normal. Entretanto, como veremos na seção seguinte, esta distribuição terá caudas mais pesadas que a da distribuição normal. Utilizando a série exemplo #1 que foi simulada sem outliers, podemos ilustrar o comportamento de tais resíduos sem observações espúrias para então através da série exemplo #2 com AO's e IO's mostrarmos a influência de tais outliers nos respectivos resíduos auxiliares.

**Série #1:**  $\alpha=-9,000$ ;  $\gamma=0.980$ ;  $\sigma_{\eta}=0.120$ ;

No gráfico (4.3) temos a série de retornos e logaritmos dos retornos ao quadrado da séries simulada #1. Nestes gráficos estão representados a volatilidades e a log-volatilidade verdadeiras.

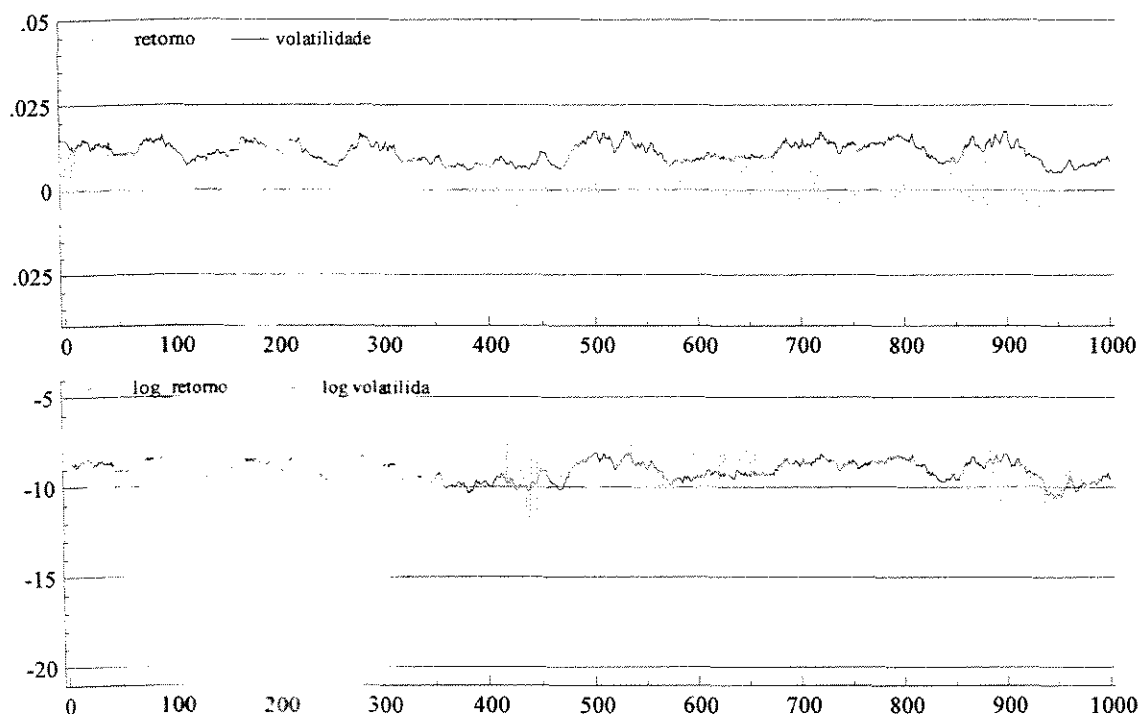


Gráfico 4.3. Série dos retornos e dos log-retornos<sup>2</sup> gerados por um modelo AR(1)\_SV.  
parâmetros reais:  $\alpha=-9,000$ ;  $\gamma=0.980$ ;  $\sigma_{\eta}=0.120$ ;

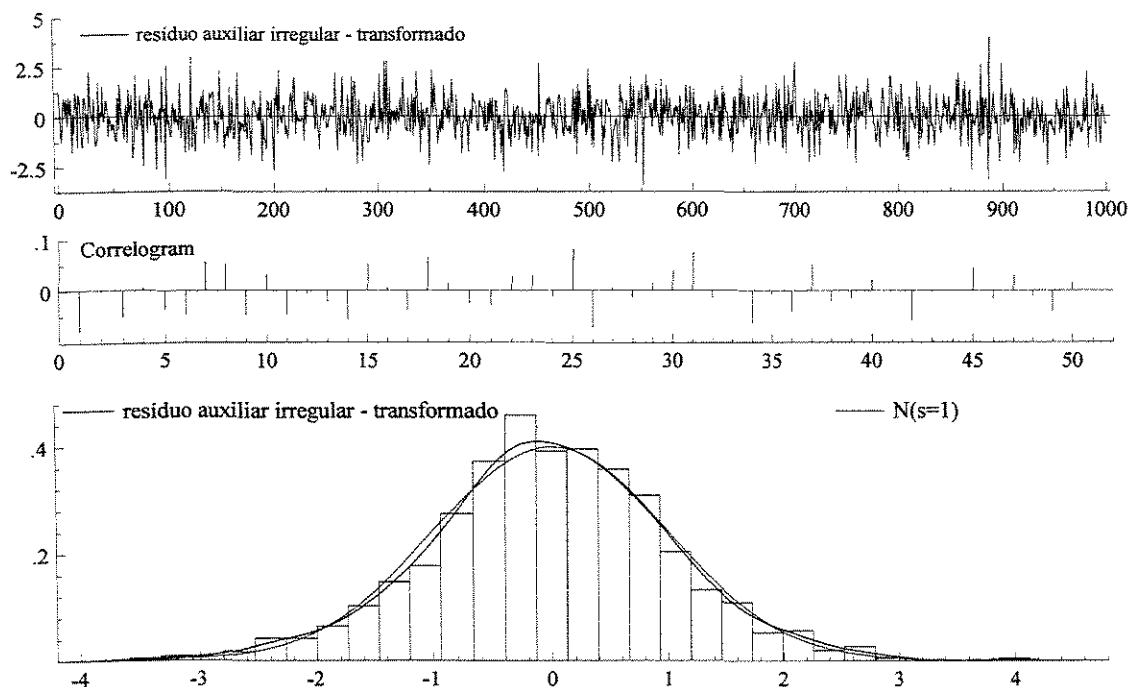


Gráfico 4.4. Resíduos Auxiliares Irregulares Transformados. Série dos resíduos, correlograma e histograma com estimativa da função densidade de probabilidade obtidos por um método não paramétrico e utilizando a aproximação normal. Assimetria  $S = -0.010$ ; Excesso de curtose  $K = 0.387$ ;

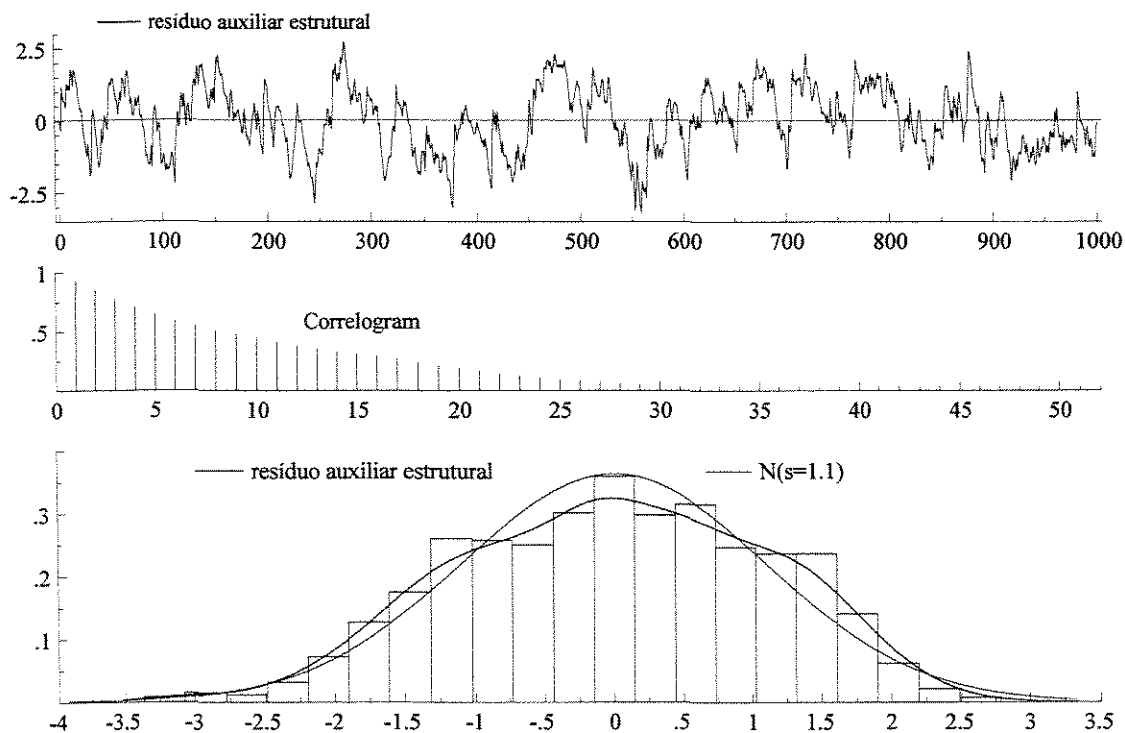


Gráfico 4.5. Resíduos auxiliares estruturais. Série dos resíduos, correlograma e histograma com estimativa da função densidade de probabilidade obtidos por um método não paramétrico e utilizando a aproximação normal. Assimetria  $S = -0.132$ ; Excesso de curtose  $K = -0.580$ ;



O gráfico (4.4) apresenta os resíduos auxiliares irregulares transformados pela equação (4.10). Verificamos que estes resíduos apresentam uma estrutura de autocorrelação semelhante a de um ruído branco e através do histograma com as densidades estimadas por um método não-paramétrico - pela aproximação normal - verificamos que sua distribuição é aproximadamente normal. Nota-se que o excesso de curtose é ligeiramente maior que zero (0,387), indicando caudas mais pesadas que a da distribuição normal. Na seção anterior foi mostrado através de simulações que além da curtose desta distribuição ser maior que a da distribuição normal, o peso das caudas destes resíduos têm uma forte dependência nos parâmetros geradores do modelo.

Já no gráfico (4.5) dos resíduos auxiliares estruturais, verificamos claramente a principal característica destes resíduos: uma forte autocorrelação de primeira ordem com decaimento exponencial para ordens superiores. Isto se deve ao fato destes resíduos seguirem um processo auto-regressivo de primeira ordem, no como o coeficiente deste processo é aproximadamente 0.94.

Nota-se que as estatísticas de assimetria e excesso de curtose foram próximas às médias das estatísticas encontradas nas simulações realizadas na seção anterior. No exemplo seguinte verificaremos a influência que estas estatísticas sofrem na presença de outliers.

**Série #2:**  $\alpha=-9,000$ ;  $\gamma=0.980$ ;  $\sigma_{\eta}=0.120$ ;  $\beta_{500}=3,000$ ;  $\Delta_{250}=4,447$ ;  $\Delta_{750}=4,560$ .

No gráfico (4.4) temos a série de retornos e logaritmos dos retornos ao quadrado da séries simulada #2. Nestes gráficos, estão representadas a volatilidade e a log-volatilidade verdadeiras.

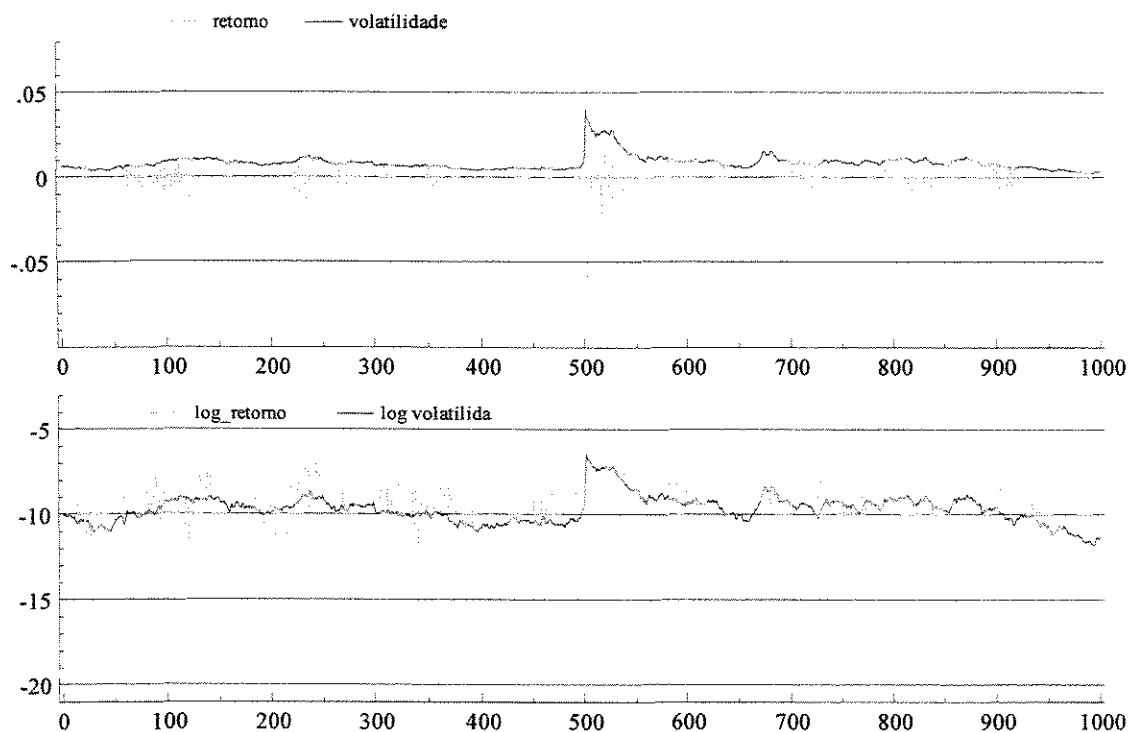


Gráfico 4.6. Série dos retornos e dos log-retornos<sup>2</sup> gerados por um modelo AR(1)\_SV.  
parâmetros reais:

$\alpha=-9,000$ ;  $\gamma=0.980$ ;  $\sigma_{\eta}=0.120$ ;  $\beta_{500}=3,000$ ; (AO)  $\Delta_{250}=4,447$ ; (AO)  $\Delta_{750}=4,560$

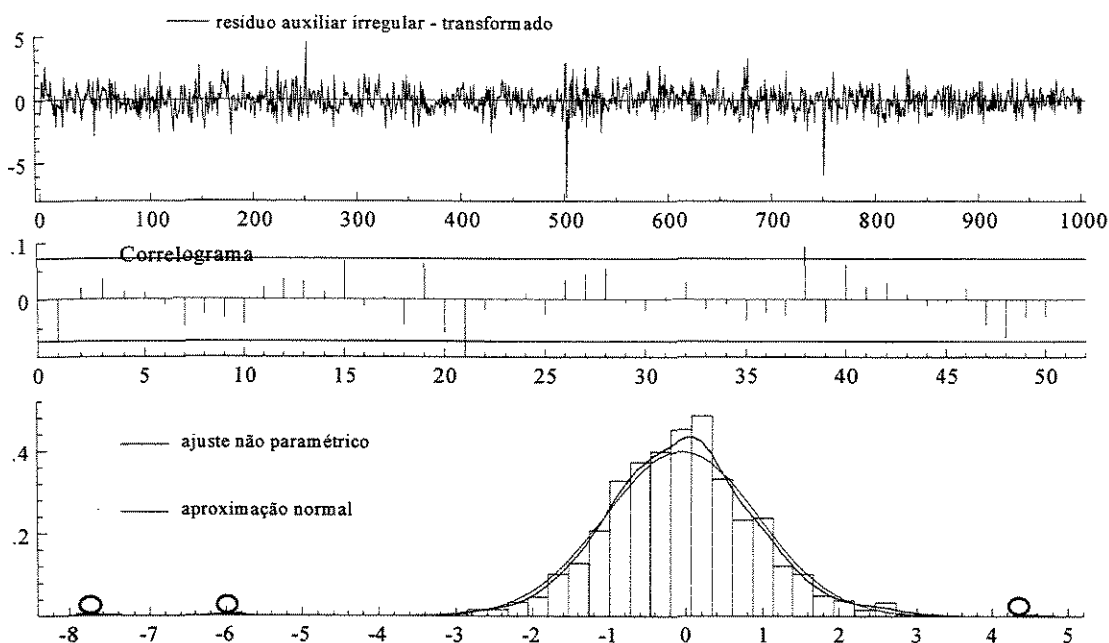


Gráfico 4.7. Resíduos Auxiliares Irregulares Transformados. Série dos resíduos, correlograma e histograma com estimativa da função densidade de probabilidade obtidos por um método não paramétrico e utilizando a aproximação normal. Os círculos vermelhos representam os outliers.

Assimetria  $S = -0.391$ ; Excesso de curtose  $K = 4.603$ ;

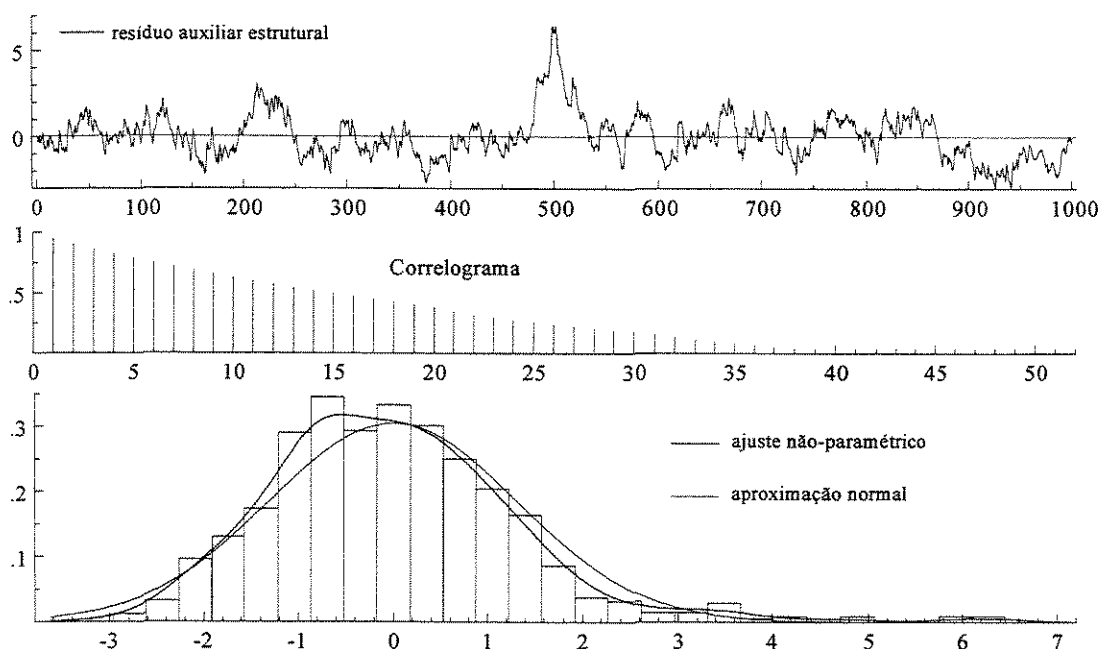


Gráfico 4.8. Descrição dos resíduos auxiliares estruturais. Série dos resíduos, correlograma e histograma com estimativa da função densidade de probabilidade obtidos por um método não paramétrico e utilizando a aproximação normal.

Assimetria  $S = 1.050$ ; Excesso de curtose  $K = 2.871$ ;

Os gráficos (4.7) e (4.8) representam os resíduos auxiliares obtidos da série #2, sendo os efeitos dos outliers evidentes. No gráfico (4.7), observamos que nas posições 250 e 750 ocorreram duas observações discrepantes com o restante dos resíduos auxiliares irregulares, indicando-nos a presença de AOs. Nota-se também que próximo à posição 500 também existe um resíduo discrepante em relação aos demais. Inicialmente, associaríamos tal observação a um AO, porém, neste caso, sabemos que trata-se da existência de um IO. Ao analisarmos os resíduos auxiliares estruturais, verificamos que na região próxima a 500 existe um conjunto de observações que destoam do restante do conjunto de dados, indicando-nos a presença de um IO. Nas situações que não soubéssemos a natureza destes outliers, ficaríamos na dúvida se poderia ter ocorrido um AO na sequência de um IO, ou se poderiam ter ocorrido os dois ao mesmo tempo. No capítulo V, mostraremos que através do critério BIC poderemos quantificar a importância de cada tipo de outlier e assim poderemos melhor avaliar o tipo de outlier que poderia ter ocorrido.

O excesso de curtose de ambos os resíduos auxiliares também mostrou-se muito alto, indicando-nos a presença de observações espúrias. O excesso de curtose dos resíduos auxiliares irregulares transformados foi de 4.603 e dos resíduos auxiliares estruturais foi de 2.871. Lembrando que, sob a hipótese nula de que não existem outliers, estas estatísticas devem ser próximas de zero.

## **Capítulo V - Procedimento de varredura para detecção e estimação de outliers.**

Neste capítulo, será proposta uma metodologia para detecção e estimação dos efeitos dos diferentes tipos de outliers.

Na seção 5.1 apresentaremos o critério BIC, que será utilizado para detecção dos outliers. Este método é baseado na log-verossimilhança, no número de parâmetros do modelo e na quantidade de observações. Utilizaremos a aproximação para o log-quase-verossimilhança, que aproxima a distribuição das componentes aleatórias da equação de observação, que é log-qui-quadrado, para uma distribuição normal com mesma média e variância. O critério BIC se mostrou razoável para detecção dos IO's, que estão relacionados com os componentes aleatórios da equação de transição que no modelo AR(1)\_SV possuem distribuição gaussiana. Para a detecção dos AO's, o critério BIC se mostrou muito conservador e, mesmo em situações extremas, não detectou estes tipos de outliers. Isto se deve ao fato de a aproximação da distribuição log-qui-quadrado para normal apresentar distorções importantes na cauda superior, região onde ocorrem os AO's. O critério AIC detectou muitos outliers nos exemplos estudados. Isto já era esperado porque o critério AIC tende a ser não parcimonioso, especialmente para séries longas. Por estes motivos, optou-se apenas pela utilização do critério BIC.

Na seção 5.2, descreveremos o procedimento de varredura para detecção de outliers. Este procedimento consiste em, a cada instante  $t$  da série de interesse, avaliar o efeito de cada tipo de outlier separadamente, ou seja, a cada instante  $t$ , serão estimados três modelos, que contém respectivamente IO, ou AO e IO e AO. Para cada modelo e para cada instante da série, seria necessário obtermos os estimadores ou QML para cada um dos três modelos, que tornaria o procedimento computacionalmente ineficiente. Para melhorarmos a eficiência computacional, utilizaremos a informação contida nos resíduos auxiliares e só avaliaremos os instantes em que os resíduos auxiliares indicarem a possível presença de uma observação aberrante, ou seja, somente serão avaliados os instantes com resíduos auxiliares de grande magnitude.

Na seção 5.3 aplicaremos o procedimento nas séries exemplo #1(sem outliers), série exemplo #2 (com outliers) e em dados reais (IBOVESPA de 04/Jul/1994 a 07/Abr/2000 com a 1424 observações). Os outliers na série exemplo #2 foram escolhidos para ressaltar os efeitos dos outliers em um modelo AR(1)\_SV de modo que os efeitos dos outliers nas estatísticas de K e S e nos resíduos auxiliares se tornaram evidentes. Além disso, o procedimento de varredura se mostrou preciso na identificação e localização dos outliers.

## 5.1. Critério BIC para detecção de outliers.

Amplamente utilizado como critério para seleção de modelos, o critério BIC (*Bayes Information Criterion - Schwartz*) é baseado na log-verossimilhança com penalizações pelo número de parâmetros ( $p$ ) e pelo tamanho da amostra ( $T$ ). (Schwarz, 1978).

$$BIC = l(\hat{\theta}) - p \cdot \ln(T)/2 \quad (5.1.)$$

onde:  $l(\hat{\theta})$  representa a log-verossimilhança  
 $\hat{\theta}$  o estimador de máxima verossimilhança  
 $p$  o número de parâmetros no modelo  
 $T$  é o número de observações.

No contexto desta dissertação,  $l(\theta)$  representa a log-quase-verossimilhança e  $\hat{\theta}$ , o estimador de máxima quase-verossimilhança, sendo cada outlier representado por um novo parâmetro, como descrito pelas equações 2.13 e 2.14.

O efeito de um determinado tipo de outlier na posição  $t$  será considerado importante se o critério BIC do modelo com este outlier for maior que o critério BIC do modelo sem este outlier.

Na seção 5.3 veremos que este critério detecta com precisão os IO's que estão relacionados aos distúrbios aleatórios da equação de transição que possuem distribuição normal. Entretanto, conforme mencionamos anteriormente, o critério BIC mostrou-se muito conservador para detecção dos AO's e mesmo em situações extremas (Série exemplo #2), não detectou qualquer outlier desta natureza. Este extremo conservadorismo se deve à aproximação da distribuição log-qui-quadrado para distribuição normal. Na figura. 2.1 comparamos as duas distribuições e observamos que a aproximação não é adequada na cauda superior, justamente a região onde estariam localizados os AO's. A distribuição normal atribui um peso consideravelmente maior para observações nesta região do que se considerarmos verdadeira a distribuição a log-qui-quadrado. Conseqüentemente, o método da quase-verossimilhança atribuirá menos importância que a verdadeira distribuição log-qui-quadrado e, desta forma, estará subestimando os efeitos destes AO's. Portanto, a inclusão de um parâmetro que caracterize este tipo de outlier será menos penalizado, e ao invés da penalização original, substituiremos por  $p \cdot \ln(T)/4$ . Na seção 5.3 mostraremos que após feita esta modificação, o critério BIC modificado conseguirá detectar outliers desta natureza.

## 5.2. Procedimento de varredura para detecção de outliers.

Nesta seção, descreveremos a metodologia para detecção dos diferentes tipos de outliers. Utilizaremos a representação do modelo descrita na seção (2.3.1) e cada outlier será representado por um novo parâmetro e, no caso de uma ocorrência de um IO e AO no mesmo instante  $t$ , teremos dois parâmetros, um para cada tipo de outlier. O procedimento de varredura consiste em cada instante  $t$ , avaliarmos a importância do IO, do AO e dos dois simultaneamente. Portanto, para cada instante  $t$ , teremos três modelos a serem estimados e, conseqüentemente três BIC's. O problema de estimarmos os três modelos para todos os instantes da série de interesse é com a eficiência computacional deste procedimento. Em uma série com 1000 observações teríamos que estimar 3000 modelos utilizando o método de BFGS, que apesar de relativamente eficiente para uma única maximização, tomaria muito tempo na maximização de 3000 modelos. Portanto não serão avaliados todos as observações da série de interesse - serão avaliados somente os instantes que apresentarem resíduos auxiliares maiores que um determinado ponto de corte  $L$ . No caso dos resíduos auxiliares transformados, avaliaremos os instantes em que  $|k_t| > 2$  e  $\xi_t > 2$ . Este corte se deve ao fato de que estamos tratando dos resíduos auxiliares padronizados, que têm uma distribuição semelhante à normal, como apresentado no capítulo IV. Desta forma, avaliaremos a presença de AO's e IO's simultaneamente quando pelo menos um dos critérios for satisfeito. Utilizando esta regra, para filtrarmos somente os instantes com alguma evidência de presença de outliers, conseguiremos melhorar significativamente a eficiência computacional do procedimento de varredura.

### 5.2.1. O procedimento de varredura passo-a-passo.

|        |            |                                    |
|--------|------------|------------------------------------|
| Sejam: | M1         | - modelo com IO na posição t       |
|        | M2         | - modelo com AO na posição t       |
|        | M3         | - modelo com IO e AO na posição t. |
|        | $\eta_t$   | - resíduos auxiliares estruturais. |
|        | $\kappa_t$ | - resíduos auxiliares irregulares. |

i) Estimar o modelo AR(1)\_SV sem a presença de outliers e obter os resíduos auxiliares irregulares e estruturais e o critério BIC.

ii) Varrer a série obtendo as estimativas de QML e os BIC's dos três modelos apresentados, utilizando as seguintes regras:

Se  $\eta_t > 2$ , obter as estimativas de QML e o BIC do M1 no instante t.

Se  $|\kappa_t| > 2$ , obter as estimativas de QML e o BIC do M2 no instante t.

Se  $|\kappa_t| > 2$  ou  $\eta_t > 2$ , obter as estimativas de QML e o BIC do M3 no instante t.

iii) Utilizar o critério BIC e o BIC modificado para AO e selecionar os outliers detectados em ii).

iv) Obter as estimativas de QML, BIC e os resíduos auxiliares do modelo com os outliers selecionados.

v) Repetir os passos ii, iii, iv até que nenhum outlier seja detectado.



### 5.3. Aplicação do procedimento.

Utilizaremos o procedimento proposto na seção 5.2 nas três séries exemplos: duas simuladas (série exemplo #1 e #2 utilizadas no capítulo IV) e uma série de dados reais (IBOVESPA - Índice da Bolsa de Valores de São Paulo) no período de 04/julho1994 à 07/abril/2000 correspondendo a 1424 observações. A série exemplo #1 não possui outliers de nenhum tipo, já a série exemplo #2 possui dois AO's nas posições 250 e 750 e um IO na posição 500. Na série de retornos do IBOVESPA deve existir vários outliers das duas naturezas já que no período analisado, ocorreram diversas crises nos mercados financeiros mundiais (México, Ásia, Rússia e Brasil).

Avaliaremos o desempenho do procedimento proposto, utilizando estas três séries exemplos sendo que, nas séries simuladas, poderemos comparar com a volatilidade verdadeira e, na série exemplo #2, poderemos analisar os efeitos de considerarmos ou não um outlier na estimativa da volatilidade.

### 5.3.1. Série simulada sem outliers (série exemplo #1).

A série exemplo #1 possui 1000 observações e foi simulada sem outliers com o conjunto de parâmetros ( $\alpha=-9,000$ ;  $\gamma=0.980$ ;  $\sigma_\eta=0.120$ ). Na tabela 5.1 apresentamos os parâmetros estimados, a log-quase-verossimilhança e as estatísticas de K e S para o modelo AR(1)\_SV sem considerar a presença de outliers. O gráfico dos retornos ( $Y_t$ ), do log do quadrado dos retornos ( $\log(Y_t^2)$ ), e dos resíduos auxiliares do modelo sem outliers estão no capítulo IV e representados respectivamente nos graficos 4.3, 4.4 e 4.5.

Tabela 5.1. Estimativas do modelo AR1\_SV sem considerar a presença de outliers. Série exemplo #1 (sem outliers). Entre parenteses os parâmetros verdadeiros.

| Parâmetro                | Estimativa | IC 95%         |
|--------------------------|------------|----------------|
| $\alpha$ (9.000)         | -9.050     | -9.189; -8.912 |
| $\gamma$ (0.980)         | 0.956      | 0.834; 0.981   |
| $\sigma_\eta^2$ (0.0144) | 0.024      | 0.007; 0.086   |
| Log-verossimilhança      | -2231.790  |                |
| BIC                      | -2238.698  |                |
| S estrutural             | -0.132     |                |
| K estrutural             | -0.580     |                |
| S Irregular              | -0.010     |                |
| K Irregular              | 0.387      |                |

Tabela 5.2. Resultados da varredura para série #1.  
As colunas Dif representam a diferença entre as log-quase-verossimilhança entre os modelos com outliers e sem outliers.

| T   | Dif L<br>IO | Dif L<br>AO | Dif L<br>IO & AO | Retorno | Log ( $X^2$ ) |
|-----|-------------|-------------|------------------|---------|---------------|
| 73  |             | 1.08        | 4.31             | 0.47%   | -10.74        |
| 125 |             | 1.05        | 1.06             | 1.85%   | -7.98         |
| 169 |             | 0.96        | 0.96             | 0.99%   | -9.23         |
| 202 |             | 1.03        | 1.92             | 0.61%   | -10.22        |
| 212 | 2.94        |             | 3.73             | 0.10%   | -13.91        |
| 224 |             | 1.39        | 1.41             | 0.61%   | -10.21        |
| 283 |             | 1.19        | 1.40             | 1.11%   | -9.00         |
| 288 | 3.21        |             | 3.78             | 1.25%   | -8.76         |
| 360 |             | 1.36        | 1.95             | 1.09%   | -9.05         |
| 524 |             | 1.07        | 1.59             | 3.15%   | -6.91         |
| 711 |             | 1.22        | 1.33             | 0.94%   | -9.33         |
| 732 |             | 1.21        | 4.50             | 1.49%   | -8.41         |
| 744 |             | 1.00        | 4.11             | 0.57%   | -10.35        |
| 792 |             | 1.13        | 1.22             | 0.62%   | -10.16        |
| 810 |             | 1.11        | 1.59             | 0.65%   | -10.09        |
| 993 |             | 1.00        | 1.85             | 0.24%   | -12.06        |

Diferença entre as log-quase-verossimilhanças para considerar um outlier importante, segundo o critério BIC e BIC modificado para AO:

$$IO > \ln(1000)/2 \cong 3.45;$$

$$AO > \ln(1000)/4 \cong 1.73.;$$

$$IO \& AO > \ln(1000)/2 + \ln(1000)/4 \cong 5.18$$

Com os resultados apresentados na tabela 5.1 e 5.2, verificamos que não existem evidências da presença de outliers, pois a curtose (K) estrutural e irregular estão próximas de zero, os resíduos auxiliares não apresentam observações discrepantes e os resultados obtidos do procedimento de varredura não indicam a presença de nenhum outlier, uma vez que em nenhum momento o critério BIC foi atingido para considerarmos um outlier importante.

No gráfico 5.3 temos os retornos com as volatilidades verdadeiras e estimadas pelo modelo AR1\_SV.

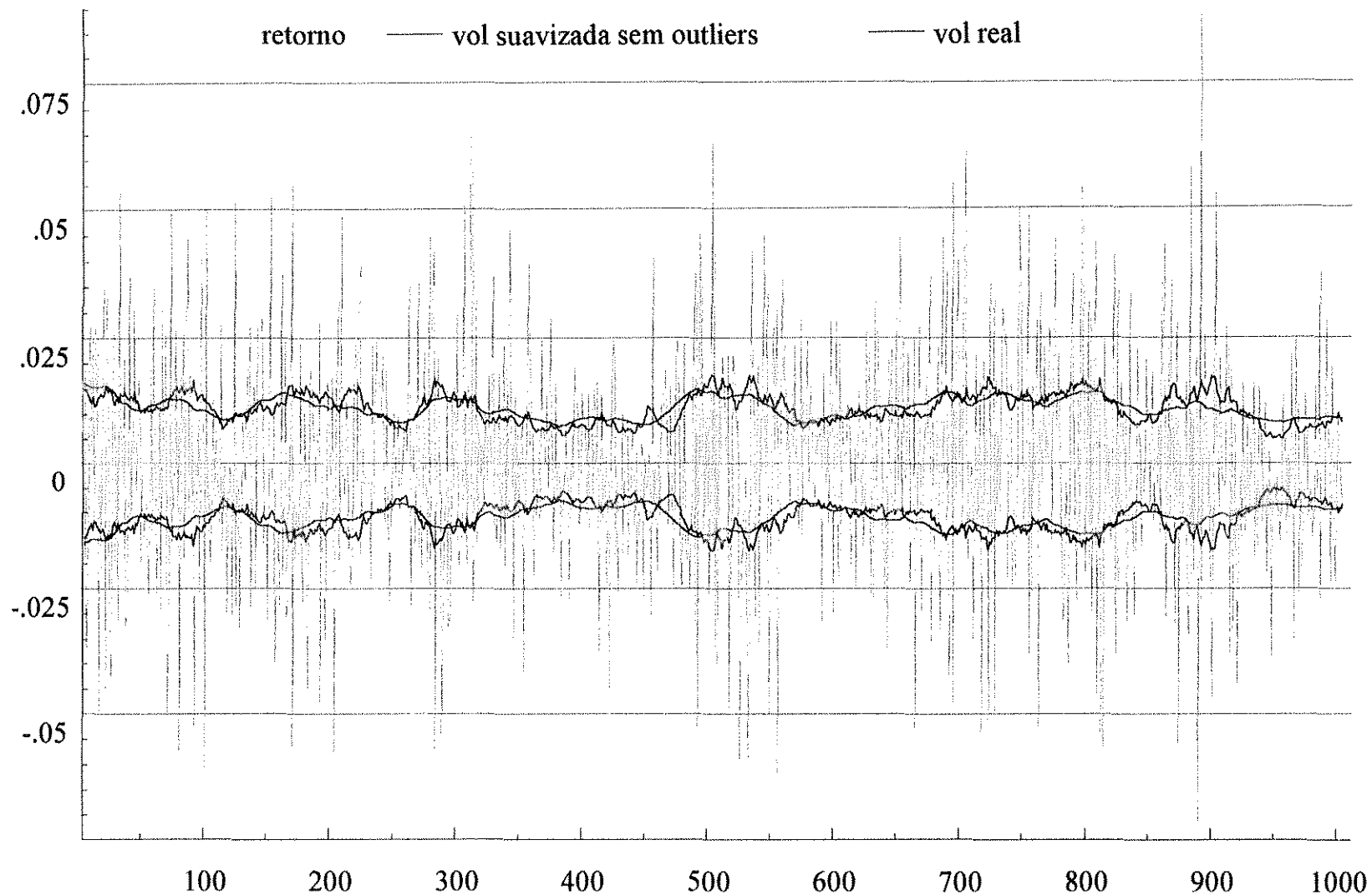


Gráfico 5.1. Retornos com volatilidade verdadeira e estimada pelo modelo AR1\_SV. Série exemplo #1 (sem outliers).

### 4.3.2. Série simulada com outliers.

A série exemplo #2 possui 1000 observações e foi simulada com outliers. Foi utilizado o conjunto de parâmetros ( $\alpha=-9,000$ ;  $\gamma=0.980$ ;  $\sigma_\eta=0.120$ ;  $\Delta_{250}=4,447$ ;  $\Delta_{750}=4,560$ ).

Na tabela 5.3 apresentamos os parâmetros estimados, a log-quase-verossimilhança, o critério BIC e as estatísticas de K e S para o modelo AR(1)\_SV sem considerar a presença de outliers. O gráfico dos retornos ( $Y_t$ ), do log do quadrado dos retornos ( $\log(Y_t^2)$ ) e dos resíduos auxiliares do modelo sem outliers estão no capítulo IV, nos gráficos 4.6, 4.7 e 4.8.

Tabela 5.3. Estimativas do modelo AR1\_SV sem considerar a presença de outliers. Série exemplo #2 (com outliers). Entre parênteses os parâmetros verdadeiros.

| Parâmetro              | Estimativa | IC 95%         |
|------------------------|------------|----------------|
| $\alpha$ (-9.000)      | -8.573     | -8.711; -8.435 |
| $\gamma$ (0.980)       | 0.983      | 0.923; 0.993   |
| $\sigma_\eta^2$ (0.12) | 0.014      | 0.006; 0.035   |
| Log-verossimilhança    | -2217.220  |                |
| BIC                    | -2224.128  |                |
| S estrutural           | 1.050      |                |
| K estrutural           | 2.870      |                |
| S irregular            | -0.390     |                |
| K irregular            | 4.602      |                |

Os resultados apresentados mostram claramente a existência de outliers. Na tabela 5.3 verificamos que as curtoses (K) dos resíduos auxiliares irregulares (4.602) e estruturais (2.870) indicam a presença de observações discrepantes. No Gráfico 4.7 e 4.8 nos quais estão representados os resíduos auxiliares irregulares e estruturais, respectivamente, verificamos que os AO's estão nas posições 250 e 750 e o IO próximo da posição 500. É importante lembrarmos que, devido à forte correlação serial dos resíduos auxiliares estruturais, os resíduos próximos à posição de sua ocorrência também são afetados.

Nas tabelas 5.4 e 5.5, apresentamos os resultados do procedimento de varredura até que nenhum outro outlier seja detectado.

Tabela 5.4. Resultados da primeira varredura para série #2.  
As colunas difs representam a diferença entre as log-quase-verossimilhança entre os modelos com outliers e sem outliers.

| t             | Dif<br>IO     | Dif<br>AO    | Dif<br>IO&AO  | Retorno       | Log (X <sup>2</sup> ) |
|---------------|---------------|--------------|---------------|---------------|-----------------------|
| 47            |               | 1.205        | 1.432         | -3.24%        | -5.590                |
| 107           |               | 1.105        | 1.277         | -3.71%        | -5.316                |
| 121           |               | 1.012        | 1.601         | -3.74%        | -5.302                |
| 146           |               | 1.070        | 1.554         | 3.86%         | -5.240                |
| 152           |               | 0.936        | 1.240         | -3.38%        | -5.505                |
| 169           |               | 0.936        | 0.972         | 2.97%         | -5.762                |
| 177           |               | 1.145        | 1.616         | -3.33%        | -5.534                |
| 212           |               | 1.039        | 3.157         | 3.48%         | -5.446                |
| <b>*</b> 250  |               | <b>1.879</b> | <b>1.976</b>  | <b>7.92%</b>  | <b>-3.802</b>         |
| 428           |               | 1.067        | 1.067         | -2.99%        | -5.752                |
| 487           | 9.778         |              | 9.848         | -1.99%        | -6.569                |
| 489           | 10.505        |              | 10.517        | -1.65%        | -6.938                |
| 490           | 9.992         |              | 10.067        | -0.91%        | -8.139                |
| 491           | 9.960         |              | 10.036        | -0.91%        | -8.120                |
| 492           | 9.931         |              | 9.936         | 1.58%         | -7.023                |
| 493           | 9.502         |              | 11.387        | 0.17%         | -11.498               |
| 494           | 10.803        |              | 12.234        | -0.24%        | -10.820               |
| 495           | 11.954        |              | 11.665        | 0.03%         | -15.004               |
| 496           | 13.340        |              | 13.497        | 1.00%         | -7.939                |
| 497           | 13.296        |              | 13.438        | -3.37%        | -5.512                |
| 498           | 14.010        |              | 14.081        | -1.22%        | -7.543                |
| 499           | 13.810        |              | 14.341        | 0.61%         | -8.935                |
| <b>**</b> 500 | <b>14.336</b> | <b>1.208</b> | <b>14.909</b> | <b>6.26%</b>  | <b>-4.272</b>         |
| 501           | 12.573        | 2.869        | 14.550        | -16.25%       | -2.365                |
| 502           | 10.220        |              | 10.224        | -1.92%        | -6.640                |
| 503           | 9.768         |              | 10.288        | -5.24%        | -4.629                |
| 504           | 8.633         |              | 8.652         | 2.06%         | -6.492                |
| 505           | 8.216         |              | 8.533         | 3.96%         | -5.187                |
| 506           | 7.436         |              | 7.696         | -3.56%        | -5.399                |
| 507           | 6.802         | 1.001        | 7.595         | 6.27%         | -4.270                |
| 516           | 5.993         | 0.000        | 6.421         | -4.52%        | -4.923                |
| 519           |               | 1.044        | 6.182         | 7.01%         | -4.044                |
| 672           |               | 1.055        | 5.704         | 4.26%         | -5.043                |
| 675           |               | 1.354        | 5.968         | 5.27%         | -4.617                |
| 680           |               | 1.082        | 5.996         | -4.18%        | -5.081                |
| 713           |               | 0.908        | 5.614         | 3.73%         | -5.308                |
| 734           |               | 0.962        | 5.674         | -3.37%        | -5.513                |
| <b>*</b> 750  |               | <b>2.356</b> | <b>6.896</b>  | <b>-7.41%</b> | <b>-3.935</b>         |
| 829           |               | 1.019        | 5.774         | 3.98%         | -5.176                |

\* AO detectados ; \*\* IO detectados;

Diferença entre as log-quase-verossimilhanças para considerar um outlier importante:

$$IO > \ln(1000)/2 \cong 3.45;$$

$$AO > \ln(1000)/4 \cong 1.73.;$$

$$IO \& AO > \ln(1000)/2 + \ln(1000)/4 \cong 5.18$$

Tabela 5.5. Resultados da segunda varredura para série #2.  
As colunas difs representam a diferença entre as log-quase-verossimilhança entre os modelos com outliers e sem outliers.

| T   | Dif<br>IO | Dif<br>AO | Dif<br>IO & AO | Retorno | log(x <sup>2</sup> ) |
|-----|-----------|-----------|----------------|---------|----------------------|
| 108 |           | 0.85      | 1.09           | 2.94%   | -5.79                |
| 121 |           | 1.11      | 2.07           | -3.74%  | -5.30                |
| 139 |           | 0.73      | 0.74           | -2.77%  | -5.90                |
| 146 |           | 1.15      | 1.38           | 3.86%   | -5.24                |
| 151 |           | 0.75      | 0.77           | -2.77%  | -5.91                |
| 152 |           | 0.99      | 1.17           | -3.38%  | -5.50                |
| 169 |           | 0.91      | 0.93           | 2.97%   | -5.76                |
| 174 |           | 0.80      | 1.20           | 2.67%   | -5.97                |
| 177 |           | 1.08      | 1.35           | -3.33%  | -5.53                |
| 210 | 2.51      |           | 2.98           | 0.39%   | -9.83                |
| 211 | 2.92      |           | 3.84           | 0.27%   | -10.57               |
| 212 | 3.61      | 1.04      | 4.06           | 3.48%   | -5.45                |
| 213 | 2.75      |           | 2.78           | -1.49%  | -7.14                |
| 214 | 2.43      |           | 3.57           | -0.22%  | -10.93               |
| 215 | 3.18      |           | 3.31           | -0.71%  | -8.62                |
| 216 | 3.29      |           | 3.30           | 1.41%   | -7.26                |
| 217 | 3.00      |           | 3.08           | -1.92%  | -6.64                |
| 218 | 2.55      |           | 2.57           | 0.96%   | -8.03                |
| 219 | 2.49      |           | 2.70           | -2.42%  | -6.17                |
| 222 | 2.25      | 0.85      | 2.67           | -3.19%  | -5.62                |
| 223 |           | 0.97      | 2.20           | 3.53%   | -5.42                |
| 226 | 2.29      |           | 2.46           | 2.30%   | -6.27                |
| 227 |           | 0.76      | 2.21           | -3.03%  | -5.72                |
| 229 | 2.41      |           | 2.41           | 1.26%   | -7.49                |
| 230 | 2.24      |           | 2.33           | 1.99%   | -6.57                |
| 233 | 2.11      |           | 2.16           | -1.82%  | -6.75                |
| 241 |           | 0.81      | 1.56           | 3.30%   | -5.56                |
| 243 |           | 0.82      | 1.13           | -3.31%  | -5.55                |
| 276 |           | 0.90      | 1.13           | -3.18%  | -5.63                |
| 359 |           | 0.89      | 0.93           | -2.81%  | -5.87                |
| 428 |           | 1.08      | 1.11           | -2.99%  | -5.75                |
| 579 |           | 0.92      | 2.06           | 4.40%   | -4.98                |
| 581 |           | 0.73      | 1.36           | 3.75%   | -5.30                |
| 591 |           | 1.19      | 1.45           | 5.27%   | -4.62                |
| 596 |           | 0.87      | 0.94           | 3.99%   | -5.17                |
| 637 |           | 0.68      | 0.75           | -2.79%  | -5.89                |
| 647 |           | 0.94      | 1.07           | 3.47%   | -5.45                |
| 667 | 2.04      | 0.69      | 2.35           | -2.88%  | -5.82                |
| 672 |           | 1.15      | 1.79           | 4.26%   | -5.04                |
| 675 |           | 1.46      | 1.54           | 5.27%   | -4.62                |
| 680 |           | 1.15      | 1.17           | -4.18%  | -5.08                |
| 695 |           | 0.76      | 0.90           | -3.04%  | -5.72                |
| 713 |           | 1.01      | 1.01           | 3.73%   | -5.31                |
| 734 |           | 0.97      | 1.13           | -3.37%  | -5.51                |
| 759 |           | 0.85      | 1.14           | 2.97%   | -5.76                |
| 763 |           | 0.69      | 1.90           | -2.61%  | -6.02                |
| 802 |           | 0.86      | 1.07           | -3.30%  | -5.55                |
| 821 |           | 0.70      | 1.29           | -2.89%  | -5.82                |
| 829 |           | 1.07      | 2.11           | 3.98%   | -5.18                |

Diferença entre as log-quase-verossimilhanças para considerar um outlier importante:

$$IO > \ln(1000)/2 \cong 3.45;$$

$$AO > \ln(1000)/4 \cong 1.73.;$$

$$IO \& AO > \ln(1000)/2 + \ln(1000)/4 \cong 5.18$$

Os resultados do procedimento de varredura apresentados nas tabelas 5.3 e 5.4 mostraram que os outliers foram detectados corretamente, ou seja, na posição verdadeira e do tipo correto. Após considerarmos os outliers detectados na primeira varredura, a segunda passagem do procedimento não apresentou evidências da existência de novos outliers. As estimativas do modelo final estão apresentadas na tabela 5.6.

Tabela 5.6. Estimativas do modelo AR1\_SV considerando a presença dos outliers detectados pelo procedimento de varredura. Série exemplo #2 (com outliers). Entre parênteses os parâmetros originais.

| Parâmetros                 | Estimativa | IC 95%         |
|----------------------------|------------|----------------|
| $\alpha$ (-9.000)          | -8.773     | -8.911; -8.635 |
| $\gamma$ (0.980)           | 0.983      | 0.966; 0.990   |
| $\sigma^2_{\eta}$ (0.0144) | 0.005      | 0.002; 0.014   |
| $\beta_{500}$ (3.000)      | 3.393      | 2.975; 3.811   |
| $\Delta_{250}$ (4.447)     | 4.595      | 4.235; 4.955   |
| $\Delta_{750}$ (4.560)     | 4.744      | 4.358; 5.130   |
| Log-verossimilhança        | -2198.480  |                |
| BIC                        | -2212.296  |                |
| S estrutural               | -0.101     |                |
| K estrutural               | -0.285     |                |
| S irregular                | 0.074      |                |
| K irregular                | 0.381      |                |

Na tabela 5.6 verificamos que as estimativas dos parâmetros estão muito próximas dos parâmetros verdadeiros, exceto para  $\alpha$  e  $\sigma^2_{\eta}$ , cujos intervalos de confiança 95% não cobrem os parâmetros verdadeiros. Os resíduos auxiliares irregulares e estruturais do modelo considerando os outliers detectados pelo procedimento de varredura estão representados nos gráficos 5.2 e 5.3. Comparando-os com os resíduos auxiliares do modelo sem considerar os outliers, verificamos que a dispersão dos resíduos melhorou consideravelmente e não existem mais observações discrepantes. Consequentemente, as medidas de curtose (K) e as medidas de assimetria (S) diminuíram significativamente. Outro fato importante é que o critério BIC do modelo com outlier (-2212.3) é maior do que o critério BIC do modelo sem outlier (-2224.13). Todos estes fatos evidenciaram a importância de considerarmos os outliers no modelo.

No gráfico 5.4 temos os retornos, a volatilidade verdadeira e estimada, com e sem a presença de outliers. A inclusão do IO na posição 500 possibilitou a estimação da volatilidade de forma muito mais precisa do que se utilizássemos o modelo sem considerar a presença dos outliers.



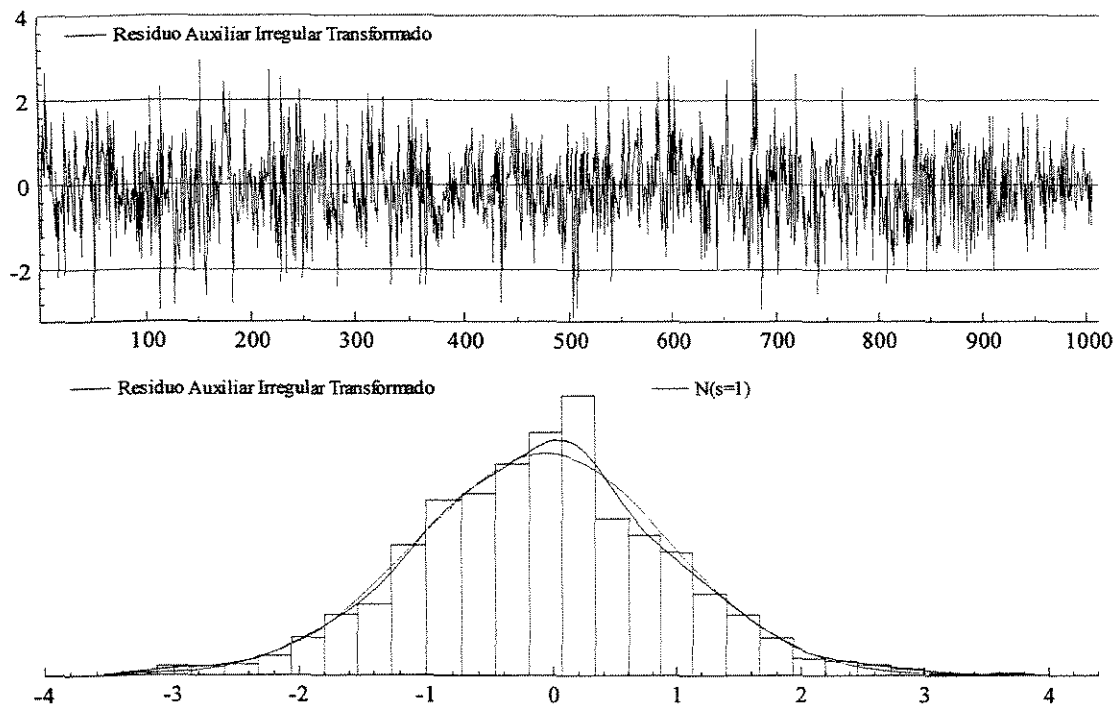


Gráfico 5.2. Resíduos auxiliares irregulares série exemplo #2 (com outliers) do modelo considerando os outliers detectados pelo procedimento de varredura.

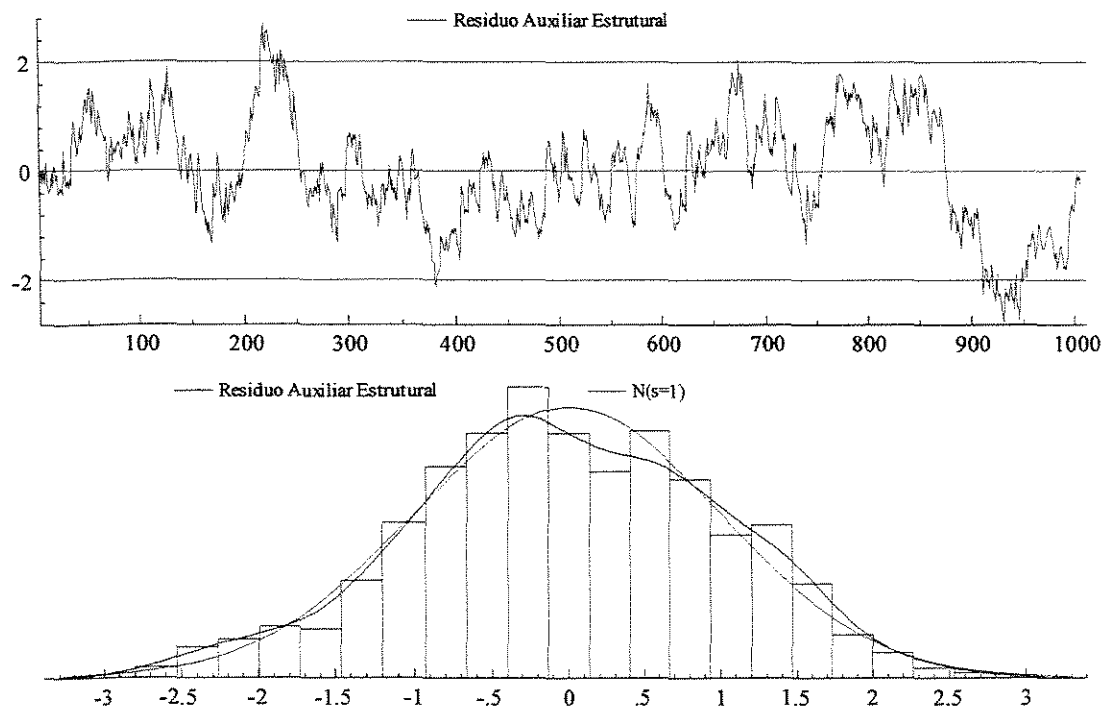


Gráfico 5.3. Resíduos auxiliares estruturais da série exemplo #2 (com outliers) do modelo considerando os outliers detectados pelo procedimento de varredura.

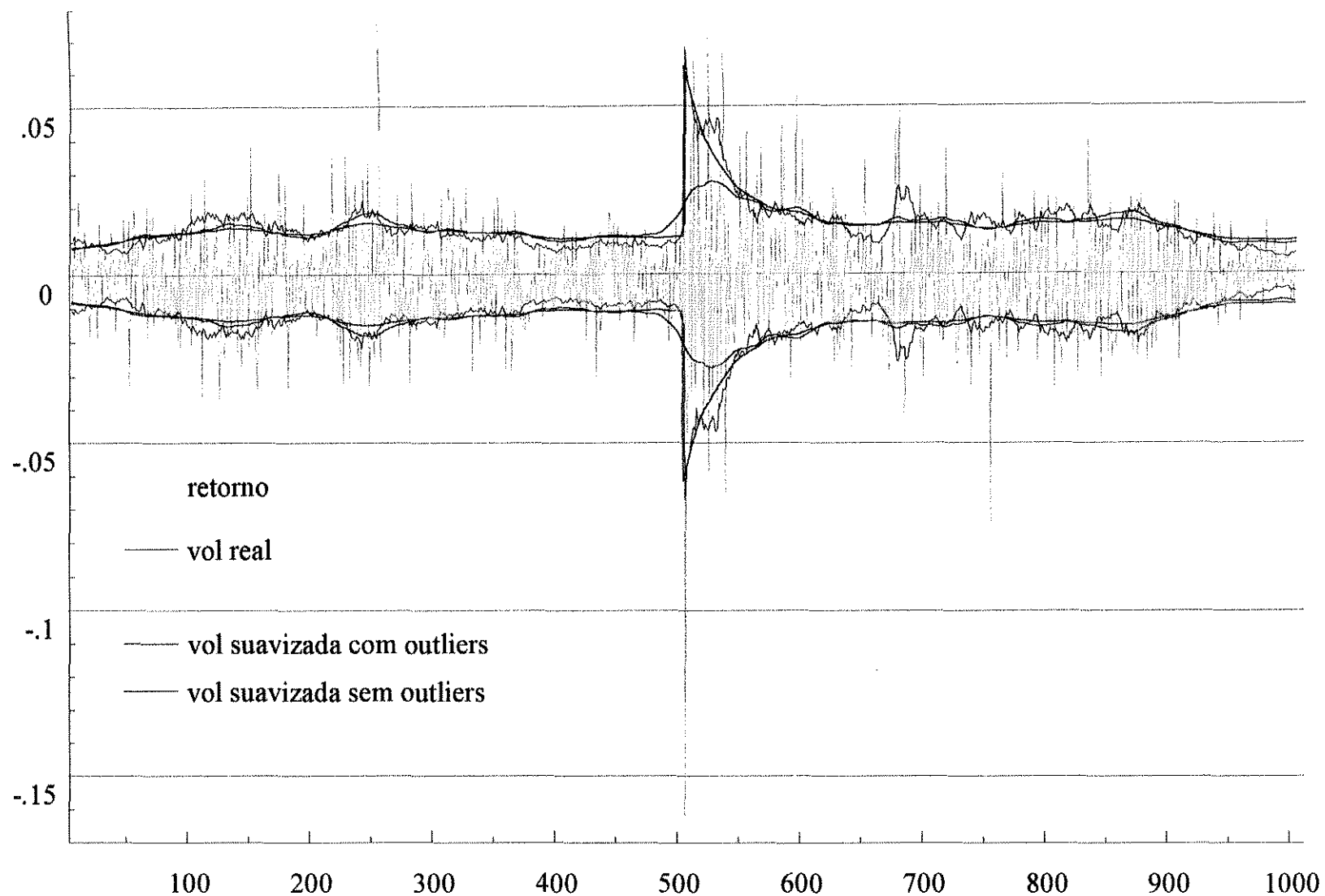


Gráfico 5.4. Retornos com volatilidade verdadeira e estimada pelo modelo AR1\_SV com os outliers detectados pelo procedimento de varredura. Série exemplo #2.

### 5.3.3. Dados reais Índice da Bolsa de Valores de São Paulo (BOVESPA).

Nesta seção aplicaremos o procedimento de varredura em uma série de dados reais. A série escolhida foi do IBOVESPA (Índice da Bolsa de Valores de São Paulo) no período de 04/Jul/1994 (início do plano real) à 07/Abr/2000. No período analisado, ocorreram diversas crises no mercado financeiro mundial sendo as principais: México (final de 1994 início de 1995), Ásia (segundo semestre de 1997), Rússia (agosto de 1998) e Brasil (início de 1999). Como veremos nas análises feitas nesta seção, estas crises tiveram efeitos importantes na volatilidade do IBOVESPA. Além das grandes crises, diversos fatos isolados contribuíram para fortes oscilações da Bolsa ocasionando vários AO's.

No gráfico 5.5, temos a distribuição de probabilidade dos retornos diários do IBOVESPA. Esta distribuição caracteriza o primeiro fato estilizado. Como podemos observar, trata-se de uma distribuição leptocúrtica (com caudas mais pesadas que a da distribuição normal). Os retornos maiores que 15% e menores que -15% provavelmente são os primeiros outliers detectados.

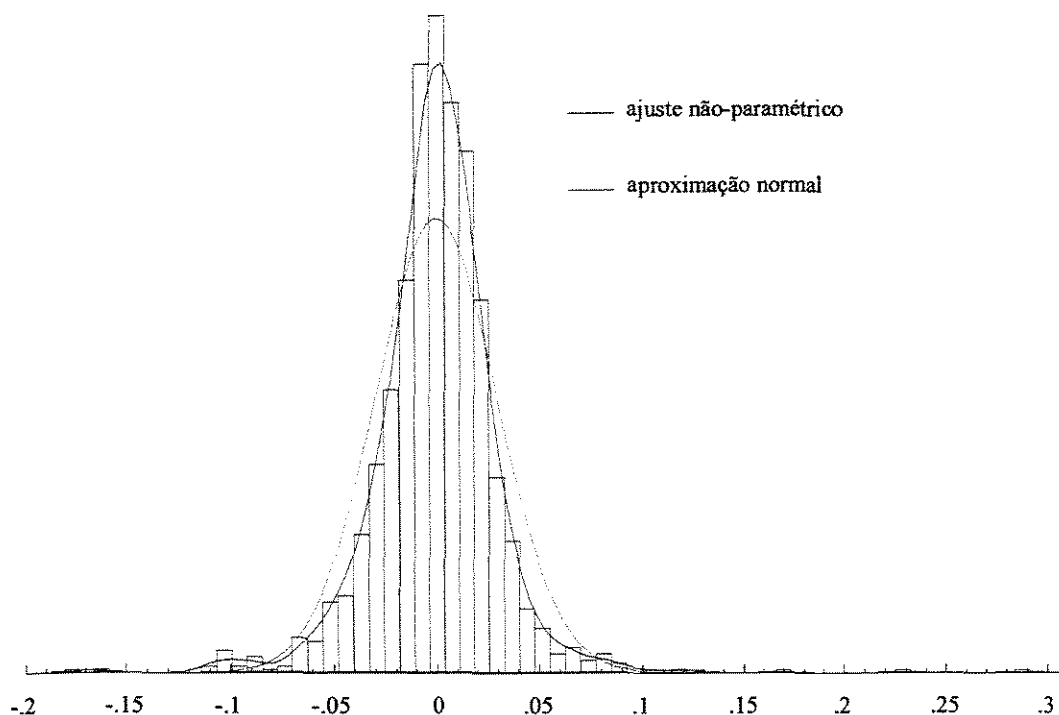


Gráfico 5.5. Histograma dos retornos do IBOVESPA.  
(de 04/Jul/1994 à 07/Abr/2000 - 1424 observações).

No gráfico 5.7, apresentamos o log do quadrado dos retornos ( $\log(Y_t^2)$ ) e os retornos( $Y_t$ ). Podemos observar outro fato estilizado, os conglomerados de volatilidade. Existem períodos de grande volatilidade e períodos de baixa volatilidade, sendo que os períodos de volatilidade alta podem ser associados às crises. Por exemplo: por volta da observação 150 que representa final de 1994 e início de 1995 ocorreu a crise do México; da observação 750 à 900 se situa a crise asiática; por volta da observação 1050 a crise da Rússia; e finalmente por volta da observação 1100 aconteceu a crise cambial brasileira. Os conglomerados estão melhor representados no gráfico dos retornos.

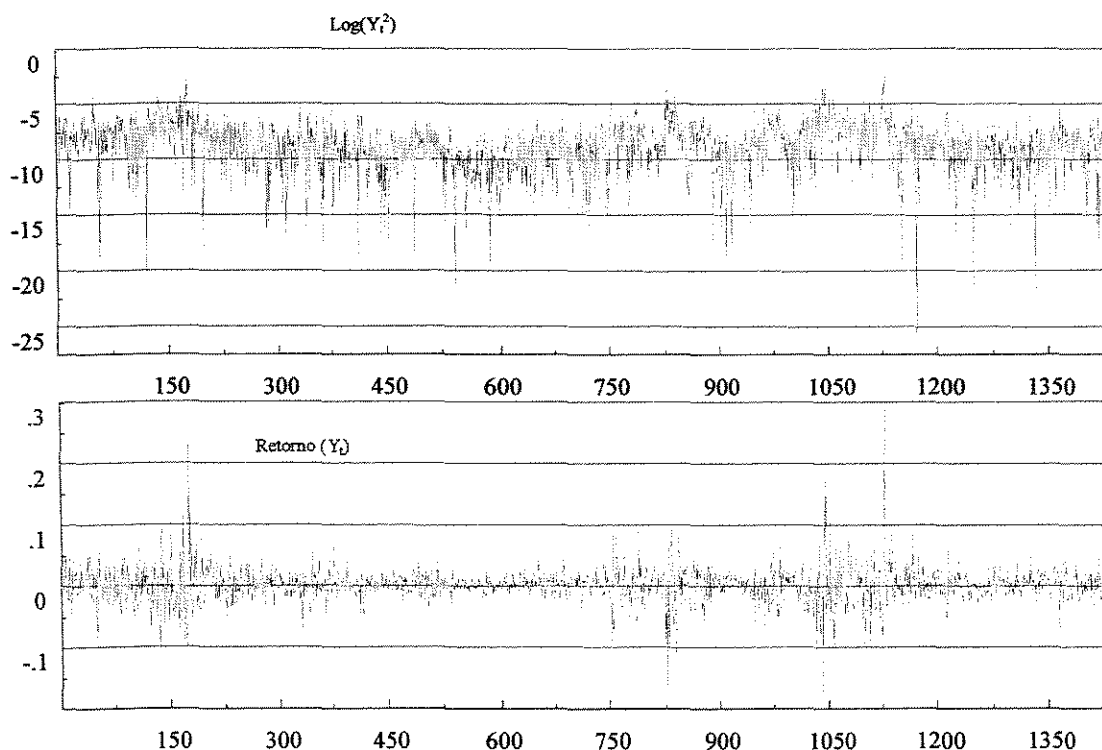


Gráfico 5.6.  $\log(Ret^2)$  e retornos da série do IBOVESPA  
( de 04/Jul/1994 à 07/Abr/2000) 1424 observações.

Na tabela 5.7 temos as estimativas do modelo AR(1)\_SV, a log-quase-verossimilhança, o critério BIC e as estatísticas de K e S sem considerar a presença de outliers. Apesar de não muito alto, o excesso de curtose do residuo auxiliar irregular indica a presença de AO's. Em relação às outras estatísticas, nada podemos concluir quanto à presença de outliers

Tabela 5.7. Estimativas do modelo AR(1)\_SV sem considerar a presença de outliers. Série IBOVESPA.

| Parâmetro           | Estimativa | IC 95%         |
|---------------------|------------|----------------|
| $\alpha$            | -8.939     | -9.069; -8.810 |
| $\gamma$            | 0.980      | 0.952; 0.989   |
| $\sigma^2_\eta$     | 0.048      | 0.026; 0.088   |
| Log-verossimilhança | -3251.910  |                |
| BIC                 | -3259.171  |                |
| S estrutural        | 0.076      |                |
| K estrutural        | 0.196      |                |
| S irregular         | 0.058      |                |
| K irregular         | 0.979      |                |

Nos gráficos 5.8 temos os resíduos auxiliares estruturais e suas respectivas densidades, de probabilidade. A presença de IO's também não pode ser detectada somente com base nesses resíduos. No gráfico 5.9 apresentamos os resíduos auxiliares irregulares transformados e sua respectiva densidade de probabilidade. Neste gráfico podemos selecionar vários candidatos ao AO's.

Nas tabelas 5.8, 5.9 e 5.10, apresentaremos os resultados do procedimento de varredura para a série do IBOVESPA. Na primeira varredura foram detectados diversos IO's e AO's, segundo o critério BIC proposto. Incorporando os outliers detectados e realizando novamente o procedimento de varredura, foi detectado mais um IO. E finalmente, na terceira varredura, não foram detectados mais nenhum outlier, dado que os detectados nas etapas anteriores já haviam sido incorporados ao modelo.

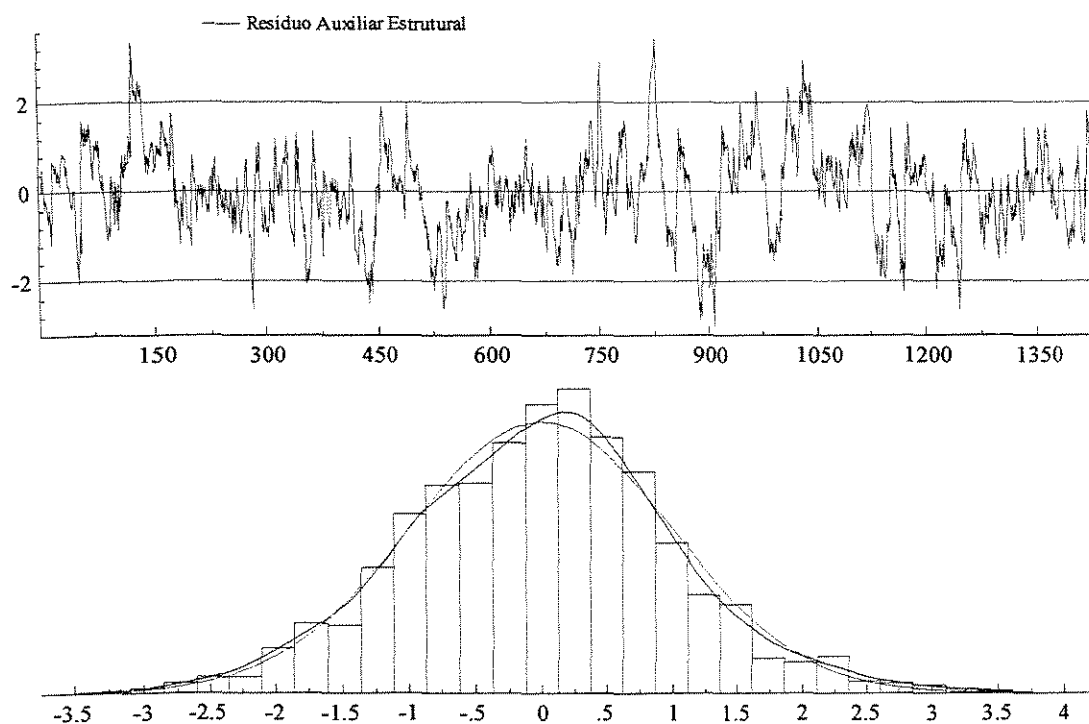


Gráfico 5.7. Resíduos auxiliares estruturais sem considerar a presença de outliers. Série IBOVESPA.

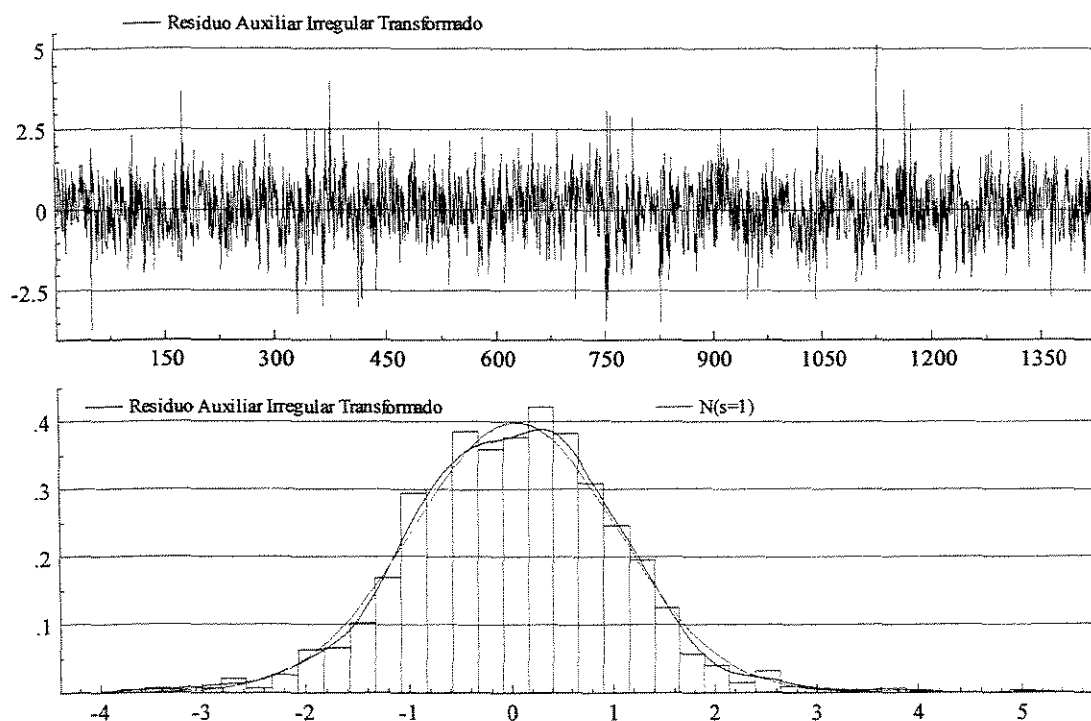


Gráfico 5.8. Resíduos auxiliares irregulares sem considerar a presença de outliers. Série IBOVESPA.

Tabela 5.8. Resultados da primeira varredura para série do IBOVESPA. As colunas dif's representam a diferença entre as log-quase-verossimilhança entre os modelos com outliers e sem outliers.

| Data         | DIF<br>IO | DIF<br>AO | DIF<br>IO & AO | Retorno | LogX^2 | Data         | DIF<br>IO | DIF<br>AO | DIF<br>IO & AO | Retorno | LogX^2 |
|--------------|-----------|-----------|----------------|---------|--------|--------------|-----------|-----------|----------------|---------|--------|
| * 05/Set/94  |           | 1.83      | 1.74           | -11.09% | -4.38  | 02/Mar/98    |           | 1.06      | 1.19           | 2.56%   | -7.42  |
| 29/Nov/94    |           | 0.92      | 0.94           | 6.03%   | -5.65  | 27/Abr/98    |           | 1.16      | 1.89           | -5.89%  | -5.63  |
| ** 20/Dez/94 | 5.60      |           | 5.71           | -2.55%  | -7.25  | 18/Mai/98    |           | 0.98      | 1.79           | -6.65%  | -5.39  |
| 21/Dez/94    | 5.23      |           | 5.30           | -6.54%  | -5.42  | 25/Mai/98    | 2.52      |           | 2.57           | -2.22%  | -7.52  |
| 22/Dez/94    | 3.85      |           | 3.94           | -6.29%  | -5.50  | 26/Mai/98    | 2.35      |           | 2.51           | -5.96%  | -5.60  |
| 23/Dez/94    | 2.74      |           | 2.77           | 3.04%   | -7.06  | 28/Jul/98    | 2.74      |           | 2.74           | -2.35%  | -7.41  |
| 26/Dez/94    | 2.49      |           | 2.75           | -1.70%  | -8.03  | 29/Jul/98    | 2.30      |           | 2.33           | 2.15%   | -7.79  |
| 27/Dez/94    | 2.71      |           | 2.71           | -3.92%  | -6.42  | 30/Jul/98    | 2.09      |           | 2.13           | 3.69%   | -6.66  |
| 28/Dez/94    | 2.24      |           | 2.52           | 1.96%   | -7.98  | 19/Ago/98    | 2.18      |           | 2.66           | -1.18%  | -8.70  |
| 29/Dez/94    | 2.49      |           | 2.50           | 3.71%   | -6.65  | 20/Ago/98    | 2.66      |           | 2.79           | -6.65%  | -5.39  |
| 02/Jan/95    | 2.19      |           | 3.29           | -0.80%  | -9.39  | ** 26/Ago/98 | 4.40      |           | 4.44           | -4.01%  | -6.38  |
| 03/Jan/95    | 3.19      |           | 3.19           | -5.25%  | -5.85  | 27/Ago/98    | 3.93      | 0.89      | 4.13           | -10.48% | -4.49  |
| 04/Jan/95    | 2.55      |           | 2.62           | -3.23%  | -6.80  | 28/Ago/98    | 2.57      |           | 3.02           | 1.94%   | -8.00  |
| 05/Jan/95    | 2.41      |           | 3.00           | 1.72%   | -8.26  | 31/Ago/98    | 3.02      |           | 3.04           | -4.15%  | -6.31  |
| 06/Jan/95    | 3.00      |           | 3.00           | -5.32%  | -5.83  | 01/Set/98    | 2.68      |           | 2.70           | 6.65%   | -5.46  |
| 09/Jan/95    | 2.47      |           | 2.47           | -5.36%  | -5.81  | 03/Set/98    | 2.54      |           | 2.64           | -9.01%  | -4.79  |
| * 10/Mar/95  |           | 2.03      | 1.58           | 22.81%  | -2.97  | 09/Set/98    | 2.87      |           | 3.17           | -2.83%  | -7.05  |
| * 14/Mar/95  |           | 1.95      | 0.76           | 12.18%  | -4.23  | * 10/Set/98  | 3.10      | 1.99      | 3.48           | -17.23% | -3.50  |
| 04/Ago/95    |           | 0.76      | 0.76           | 4.25%   | -6.37  | * 15/Set/98  |           | 1.97      | 1.31           | 17.12%  | -3.54  |
| 07/Ago/95    |           | 0.83      | 0.98           | 4.50%   | -6.25  | 03/Dez/98    |           | 1.84      | 1.41           | -9.20%  | -4.75  |
| 25/Out/95    |           | 0.76      | 0.89           | -4.16%  | -6.31  | 14/Dez/98    |           | 1.81      | 1.23           | -8.88%  | -4.82  |
| 26/Out/95    |           | 1.38      | 1.39           | -6.75%  | -5.36  | * 15/Jan/99  |           | 2.30      | 2.19           | 28.82%  | -2.50  |
| 10/Nov/95    |           | 0.77      | 1.28           | -3.84%  | -6.46  | * 01/Fev/99  |           | 1.87      | 2.34           | 8.44%   | -4.97  |
| 13/Nov/95    |           | 0.91      | 1.00           | -4.41%  | -6.19  | * 15/Mar/99  |           | 1.97      | 2.39           | 8.40%   | -4.98  |
| 16/Nov/95    |           | 1.05      | 1.22           | 5.29%   | -5.92  | 25/Mar/99    |           | 1.13      | 1.37           | 4.76%   | -6.14  |
| 30/Nov/95    |           | 0.99      | 2.64           | 4.58%   | -6.22  | 24/Mai/99    |           | 0.82      | 0.87           | -5.01%  | -5.94  |
| 14/Dez/95    |           | 1.27      | 1.74           | -4.36%  | -6.21  | 25/Mai/99    |           | 0.85      | 1.26           | -5.08%  | -5.92  |
| 20/Dez/95    |           | 1.02      | 1.02           | 4.14%   | -6.42  | 26/Mai/99    |           | 1.02      | 2.22           | 5.89%   | -5.70  |
| * 03/Jan/96  |           | 1.89      | 2.39           | 6.43%   | -5.52  | 16/Jun/99    |           | 0.98      | 1.01           | 4.03%   | -6.48  |
| 30/Jan/96    |           | 0.92      | 0.92           | 3.95%   | -6.52  | 24/Jun/99    |           | 0.74      | 0.93           | -3.11%  | -6.87  |
| 29/Fev/96    |           | 1.28      | 1.64           | -4.76%  | -6.04  | 26/Jul/99    |           | 0.82      | 1.01           | -2.90%  | -7.01  |
| 06/Mar/96    |           | 1.12      | 1.19           | -4.38%  | -6.21  | 30/Set/99    |           | 0.77      | 0.77           | -2.90%  | -7.00  |
| 07/Mar/96    |           | 1.17      | 1.74           | -4.51%  | -6.15  | 06/Out/99    |           | 1.04      | 1.74           | 3.79%   | -6.61  |
| 08/Abr/96    |           | 1.03      | 3.70           | -2.52%  | -7.27  | 03/Nov/99    |           | 1.41      | 1.41           | 4.85%   | -6.10  |
| 12/Abr/96    |           | 1.15      | 3.40           | 2.63%   | -7.36  | 04/Jan/00    |           | 1.14      | 1.37           | -6.59%  | -5.41  |
| 21/Jun/96    | 1.99      |           | 2.22           | 3.88%   | -6.56  |              |           |           |                |         |        |
| 26/Ago/96    |           | 0.91      | 1.97           | -2.86%  | -7.03  |              |           |           |                |         |        |
| 28/Ago/96    |           | 0.83      | 3.85           | 2.72%   | -7.30  |              |           |           |                |         |        |
| 22/Out/96    |           | 0.77      | 0.78           | -1.83%  | -7.88  |              |           |           |                |         |        |
| 29/Out/96    |           | 0.90      | 2.51           | 2.30%   | -7.65  |              |           |           |                |         |        |
| 11/Dez/96    |           | 0.90      | 0.91           | -2.63%  | -7.19  |              |           |           |                |         |        |
| 12/Fev/97    |           | 0.97      | 1.11           | 4.17%   | -6.41  |              |           |           |                |         |        |
| 04/Abr/97    |           | 1.03      | 1.18           | 4.24%   | -6.37  |              |           |           |                |         |        |
| 14/Mai/97    |           | 1.16      | 2.04           | -3.49%  | -6.65  |              |           |           |                |         |        |
| 12/Jun/97    |           | 0.82      | 1.02           | 3.10%   | -7.02  |              |           |           |                |         |        |
| 10/Jul/97    | 2.18      |           | 5.01           | -0.06%  | -12.70 |              |           |           |                |         |        |
| ** 11/Jul/97 | 4.21      |           | 4.23           | -3.40%  | -6.70  |              |           |           |                |         |        |
| 14/Jul/97    | 3.22      |           | 3.25           | -3.52%  | -6.63  |              |           |           |                |         |        |
| * 15/Jul/97  | 2.36      | 1.86      | 3.08           | -8.90%  | -4.81  |              |           |           |                |         |        |
| 16/Jul/97    |           | 1.82      | 1.89           | 8.45%   | -4.97  |              |           |           |                |         |        |
| 17/Jul/97    |           | 1.16      | 1.24           | -7.49%  | -5.15  |              |           |           |                |         |        |
| 22/Jul/97    |           | 1.26      | 1.47           | 8.02%   | -5.07  |              |           |           |                |         |        |
| 15/Ago/97    |           | 0.77      | 0.85           | -4.91%  | -5.98  |              |           |           |                |         |        |
| * 02/Set/97  |           | 1.90      | 1.21           | 9.02%   | -4.84  |              |           |           |                |         |        |
| 16/Out/97    | 2.93      |           | 3.03           | -1.99%  | -7.73  |              |           |           |                |         |        |
| 17/Out/97    | 2.84      |           | 3.77           | -0.72%  | -9.58  |              |           |           |                |         |        |
| 20/Out/97    | 3.74      |           | 4.13           | 1.72%   | -8.26  |              |           |           |                |         |        |
| ** 21/Out/97 | 4.07      |           | 4.26           | 2.43%   | -7.53  |              |           |           |                |         |        |
| 22/Out/97    | 4.05      |           | 6.07           | -0.46%  | -10.32 |              |           |           |                |         |        |
| 23/Out/97    | 5.89      | 0.77      | 5.95           | -8.50%  | -4.90  |              |           |           |                |         |        |
| 24/Out/97    | 4.35      |           | 4.50           | -3.02%  | -6.93  |              |           |           |                |         |        |
| * 27/Out/97  | 4.19      | 2.10      | 4.69           | -16.22% | -3.62  |              |           |           |                |         |        |
| 28/Out/97    | 2.36      |           | 2.39           | 6.22%   | -5.59  |              |           |           |                |         |        |
| * 30/Out/97  |           | 1.84      | 1.56           | -10.33% | -4.52  |              |           |           |                |         |        |
| * 12/Nov/97  |           | 1.92      | 0.88           | -10.76% | -4.44  |              |           |           |                |         |        |
| 25/Fev/98    |           | 0.77      | 4.93           | 2.18%   | -7.76  |              |           |           |                |         |        |

\* AO detectado; \*\* IO detectado;

diferença entre as log-quase-verossimilhanças para considerar um outlier importante.

$$IO > \ln(1424)/2 \cong 3.63;$$

$$AO > \ln(1424)/4 \cong 1.82;$$

$$IO \& AO > \ln(1424)/2 + \ln(1000)/4 \cong 5.45;$$

Tabela 5.9. Resultados da segunda varredura para série do IBOVESPA. As colunas dif's representam a diferença entre as log-quase-verossimilhanças entre os modelos com outliers e sem outliers.

| data                | Dif<br>IO   | Dif<br>AO | Dif<br>IO & AO | Retorno       | log(X^2)      | data      | Dif<br>IO | Dif<br>AO | Dif<br>IO & AO | Retorno | log(X^2) |
|---------------------|-------------|-----------|----------------|---------------|---------------|-----------|-----------|-----------|----------------|---------|----------|
| 06/Set/94           |             | 0.75      | 1.01           | 3.68%         | -6.60         | 23/Jun/99 |           | 0.73      | 1.02           | -1.64%  | -8.21    |
| 18/Nov/94           |             | 0.80      | 1.29           | -3.48%        | -6.72         | 23/Jul/99 |           | 0.76      | 1.07           | -2.36%  | -7.51    |
| 28/Nov/94           |             | 1.10      | 1.78           | -0.11%        | -13.60        | 29/Set/99 |           | 0.74      | 0.74           | -0.19%  | -12.51   |
| 22/Fev/95           |             | 0.80      | 1.16           | 2.69%         | -7.23         | 05/Out/99 |           | 0.99      | 1.44           | -1.31%  | -8.61    |
| 07/Mar/95           |             | 0.76      | 0.99           | -9.29%        | -4.75         | 01/Nov/99 |           | 1.34      | 1.38           | 0.42%   | -10.91   |
| 08/Mar/95           |             | 0.76      | 0.78           | -10.26%       | -4.55         | 03/Jan/00 |           | 1.28      | 2.01           | -1.06%  | -9.11    |
| 03/Ago/95           |             | 0.79      | 0.80           | 2.58%         | -7.31         | 10/Mar/00 |           | 0.74      | 0.79           | -2.12%  | -7.71    |
| 04/Ago/95           |             | 0.87      | 1.02           | 4.14%         | -6.37         | 20/Mar/00 |           | 0.94      | 0.95           | -0.03%  | -16.21   |
| 22/Ago/95           |             | 0.83      | 0.97           | 0.04%         | -15.56        |           |           |           |                |         |          |
| 24/Out/95           |             | 0.86      | 1.10           | -0.90%        | -9.43         |           |           |           |                |         |          |
| 25/Out/95           |             | 1.52      | 1.53           | -4.27%        | -6.31         |           |           |           |                |         |          |
| 09/Nov/95           |             | 0.82      | 1.21           | -0.21%        | -12.31        |           |           |           |                |         |          |
| 10/Nov/95           |             | 0.98      | 1.06           | -3.95%        | -6.46         |           |           |           |                |         |          |
| 14/Nov/95           |             | 1.15      | 1.27           | 0.16%         | -12.87        |           |           |           |                |         |          |
| 29/Nov/95           |             | 1.05      | 2.51           | 3.11%         | -6.94         |           |           |           |                |         |          |
| 13/Dez/95           |             | 1.21      | 1.36           | -0.02%        | -17.54        |           |           |           |                |         |          |
| 19/Dez/95           |             | 1.05      | 1.08           | -0.63%        | -10.13        |           |           |           |                |         |          |
| 29/Jan/96           |             | 0.97      | 0.97           | 0.44%         | -10.86        |           |           |           |                |         |          |
| 28/Fev/96           |             | 1.35      | 1.47           | 0.01%         | -18.59        |           |           |           |                |         |          |
| 05/Mar/96           |             | 1.23      | 1.40           | 0.76%         | -9.75         |           |           |           |                |         |          |
| 06/Mar/96           |             | 1.28      | 1.97           | -4.49%        | -6.21         |           |           |           |                |         |          |
| 03/Abr/96           |             | 0.89      | 3.47           | -0.27%        | -11.82        |           |           |           |                |         |          |
| 11/Abr/96           |             | 0.91      | 2.80           | -0.20%        | -12.42        |           |           |           |                |         |          |
| 30/Abr/96           | 2.11        |           | 2.39           | 0.18%         | -12.69        |           |           |           |                |         |          |
| 02/Mai/96           | 2.39        |           | 2.40           | -0.46%        | -10.75        |           |           |           |                |         |          |
| 03/Mai/96           | 2.17        |           | 2.19           | -0.98%        | -9.25         |           |           |           |                |         |          |
| 20/Jun/96           | 2.46        | 0.73      | 2.73           | -0.01%        | -18.41        |           |           |           |                |         |          |
| 23/Ago/96           |             | 0.88      | 3.22           | -0.85%        | -9.53         |           |           |           |                |         |          |
| 27/Ago/96           |             | 0.78      | 5.10           | 1.50%         | -8.40         |           |           |           |                |         |          |
| 21/Out/96           |             | 0.74      | 0.75           | -0.10%        | -13.81        |           |           |           |                |         |          |
| 28/Out/96           |             | 0.87      | 2.28           | -1.70%        | -8.15         |           |           |           |                |         |          |
| 10/Dez/96           |             | 0.91      | 0.95           | 0.17%         | -12.72        |           |           |           |                |         |          |
| 07/Fev/97           |             | 1.03      | 1.37           | 1.94%         | -7.89         |           |           |           |                |         |          |
| 03/Abr/97           |             | 1.09      | 1.32           | 0.59%         | -10.27        |           |           |           |                |         |          |
| 13/Mai/97           |             | 1.14      | 2.02           | -0.42%        | -10.94        |           |           |           |                |         |          |
| 11/Jun/97           |             | 0.90      | 0.94           | 0.42%         | -10.94        |           |           |           |                |         |          |
| 24/Jun/97           |             | 1.02      | 1.03           | 2.70%         | -7.23         |           |           |           |                |         |          |
| 30/Jun/97           |             | 0.96      | 2.48           | -1.61%        | -8.25         |           |           |           |                |         |          |
| 15/Jul/97           |             | 0.81      | 1.13           | -9.01%        | -4.81         |           |           |           |                |         |          |
| 21/Jul/97           |             | 0.86      | 2.78           | 0.87%         | -9.49         |           |           |           |                |         |          |
| 28/Ago/97           |             | 0.87      | 1.01           | -4.05%        | -6.41         |           |           |           |                |         |          |
| 08/Jan/98           |             | 0.83      | 1.58           | -2.38%        | -7.48         |           |           |           |                |         |          |
| 27/Fev/98           |             | 0.76      | 0.95           | 0.02%         | -16.70        |           |           |           |                |         |          |
| 17/Abr/98           | 3.78        |           | 4.12           | 0.04%         | -15.68        |           |           |           |                |         |          |
| 20/Abr/98           | 4.11        |           | 4.17           | -0.78%        | -9.71         |           |           |           |                |         |          |
| 22/Abr/98           | 3.37        |           | 3.45           | -2.85%        | -7.12         |           |           |           |                |         |          |
| 23/Abr/98           | 3.34        |           | 3.43           | -1.23%        | -8.79         |           |           |           |                |         |          |
| 24/Abr/98           | 3.32        | 1.21      | 3.86           | 1.23%         | -8.80         |           |           |           |                |         |          |
| 05/Mai/98           | 2.49        |           | 2.63           | 0.30%         | -11.63        |           |           |           |                |         |          |
| 13/Mai/98           | 2.52        |           | 3.35           | -0.28%        | -11.73        |           |           |           |                |         |          |
| 14/Mai/98           | 3.24        |           | 3.33           | -0.54%        | -10.46        |           |           |           |                |         |          |
| 15/Mai/98           | 3.23        | 1.08      | 3.71           | 1.48%         | -8.42         |           |           |           |                |         |          |
| 20/Mai/98           | 3.01        |           | 3.12           | 0.52%         | -10.53        |           |           |           |                |         |          |
| 21/Mai/98           | 3.06        |           | 4.54           | -1.46%        | -8.45         |           |           |           |                |         |          |
| <b>** 22/Mai/98</b> | <b>4.22</b> |           | <b>4.23</b>    | <b>-0.39%</b> | <b>-11.07</b> |           |           |           |                |         |          |
| 25/Mai/98           | 3.92        | 0.80      | 4.17           | -2.33%        | -7.52         |           |           |           |                |         |          |
| 26/Mai/98           | 2.87        |           | 2.88           | -6.07%        | -5.60         |           |           |           |                |         |          |
| 27/Mai/98           | 2.41        |           | 2.50           | 3.14%         | -6.92         |           |           |           |                |         |          |
| 16/Jun/98           |             | 0.88      | 1.00           | 1.80%         | -8.04         |           |           |           |                |         |          |
| 27/Jul/98           | 2.27        |           | 2.28           | -0.21%        | -12.31        |           |           |           |                |         |          |
| 03/Ago/98           |             | 0.88      | 1.21           | -2.90%        | -7.08         |           |           |           |                |         |          |
| 19/Ago/98           |             | 1.20      | 1.23           | -1.29%        | -8.70         |           |           |           |                |         |          |
| 11/Dez/98           |             | 0.88      | 2.08           | -4.66%        | -6.13         |           |           |           |                |         |          |
| 13/Jan/99           |             | 1.08      | 1.09           | -5.29%        | -5.88         |           |           |           |                |         |          |
| 24/Mar/99           |             | 0.99      | 1.40           | 0.00%         | -25.59        |           |           |           |                |         |          |
| 21/Mai/99           |             | 0.95      | 1.04           | -3.14%        | -6.92         |           |           |           |                |         |          |

\* AO detectado; \*\* IO detectado;

diferença entre as log-quase-verossimilhanças para considerar um outlier importante.

$$IO > \ln(1424)/2 \cong 3.63;$$

$$AO > \ln(1424)/4 \cong 1.82;$$

$$IO \& AO > \ln(1424)/2 + \ln(1000)/4 \cong 5.45;$$



Tabela 5.10. Resultados da segunda varredura para série do IBOVESPA.

Critério BIC para considerar um outlier importante:  $IO = \ln(1424)/2 \cong 3.63$ ;  $AO = \ln(1424)/4 \cong 1.82$ .

| data      | DIF<br>IO | DIF<br>AO | DIF<br>IO & AO | Retorno | log(X <sup>2</sup> ) |
|-----------|-----------|-----------|----------------|---------|----------------------|
| 06/Set/94 |           | 0.74      | 0.85           | 3.68%   | -6.60                |
| 19/Set/94 | 2.05      |           | 2.05           | -0.15%  | -13.04               |
| 30/Set/94 | 2.14      |           | 2.16           | 0.65%   | -10.08               |
| 18/Nov/94 |           | 0.81      | 1.29           | -3.48%  | -6.72                |
| 28/Nov/94 |           | 1.10      | 1.79           | -0.11%  | -13.60               |
| 22/Fev/95 |           | 0.83      | 1.30           | 2.69%   | -7.23                |
| 07/Mar/95 |           | 0.79      | 1.09           | -9.29%  | -4.75                |
| 08/Mar/95 |           | 0.79      | 0.84           | -10.26% | -4.55                |
| 03/Abr/95 |           | 0.73      | 0.76           | -3.02%  | -7.00                |
| 04/Ago/95 |           | 0.87      | 1.02           | 4.14%   | -6.37                |
| 22/Ago/95 |           | 0.80      | 0.93           | 0.04%   | -15.56               |
| 24/Out/95 |           | 0.90      | 1.18           | -0.90%  | -9.43                |
| 25/Out/95 |           | 1.57      | 1.59           | -4.27%  | -6.31                |
| 09/Nov/95 |           | 0.84      | 1.21           | -0.21%  | -12.31               |
| 10/Nov/95 |           | 1.01      | 1.09           | -3.95%  | -6.46                |
| 14/Nov/95 |           | 1.18      | 1.29           | 0.16%   | -12.87               |
| 29/Nov/95 |           | 1.06      | 2.36           | 3.11%   | -6.94                |
| 13/Dez/95 |           | 1.19      | 1.31           | -0.02%  | -17.54               |
| 19/Dez/95 |           | 1.04      | 1.06           | -0.63%  | -10.13               |
| 29/Jan/96 |           | 0.99      | 0.99           | 0.44%   | -10.86               |
| 28/Fev/96 |           | 1.37      | 1.43           | 0.01%   | -18.59               |
| 05/Mar/96 |           | 1.25      | 1.46           | 0.76%   | -9.75                |
| 06/Mar/96 |           | 1.30      | 2.01           | -4.49%  | -6.21                |
| 03/Abr/96 |           | 0.84      | 3.20           | -0.27%  | -11.82               |
| 11/Abr/96 |           | 0.84      | 2.48           | -0.20%  | -12.42               |
| 30/Abr/96 | 2.33      |           | 2.64           | 0.18%   | -12.69               |
| 02/Mai/96 | 2.63      |           | 2.64           | -0.46%  | -10.75               |
| 03/Mai/96 | 2.46      |           | 2.48           | -0.98%  | -9.25                |
| 06/Mai/96 | 2.36      |           | 2.60           | 0.89%   | -9.43                |
| 20/Jun/96 | 2.62      | 0.77      | 2.93           | -0.01%  | -18.41               |
| 29/Jul/96 |           | 0.78      | 2.01           | -1.11%  | -9.01                |
| 23/Ago/96 |           | 0.87      | 3.86           | -0.85%  | -9.53                |
| 27/Ago/96 |           | 0.77      | 5.68           | 1.50%   | -8.40                |
| 28/Out/96 |           | 0.84      | 2.15           | -1.70%  | -8.15                |
| 10/Dez/96 |           | 0.90      | 1.02           | 0.17%   | -12.72               |
| 07/Fev/97 |           | 1.06      | 1.52           | 1.94%   | -7.89                |
| 03/Abr/97 |           | 1.11      | 1.35           | 0.59%   | -10.27               |
| 13/Mai/97 |           | 1.12      | 1.95           | -0.42%  | -10.94               |
| 11/Jun/97 |           | 0.89      | 0.93           | 0.42%   | -10.94               |
| 24/Jun/97 |           | 1.04      | 1.05           | 2.70%   | -7.23                |
| 30/Jun/97 |           | 0.97      | 2.44           | -1.61%  | -8.25                |
| 15/Jul/97 |           | 0.83      | 1.17           | -9.01%  | -4.81                |
| 21/Jul/97 |           | 0.87      | 2.74           | 0.87%   | -9.49                |
| 28/Ago/97 |           | 0.88      | 1.02           | -4.05%  | -6.41                |
| 08/Jan/98 |           | 0.87      | 2.15           | -2.38%  | -7.48                |
| 17/Abr/98 | 2.35      |           | 2.55           | 0.04%   | -15.68               |
| 20/Abr/98 | 2.54      |           | 2.69           | -0.78%  | -9.71                |
| 24/Abr/98 |           | 1.43      | 2.54           | 1.23%   | -8.80                |
| 15/Mai/98 |           | 1.57      | 1.57           | 1.48%   | -8.42                |
| 03/Ago/98 |           | 0.86      | 1.13           | -2.90%  | -7.08                |
| 19/Ago/98 |           | 1.19      | 1.23           | -1.29%  | -8.70                |
| 10/Dez/98 | 2.38      |           | 2.42           | -0.67%  | -10.00               |
| 11/Dez/98 |           | 0.91      | 2.37           | -4.66%  | -6.13                |
| 11/Jan/99 |           | 0.79      | 1.22           | -5.85%  | -5.68                |
| 13/Jan/99 |           | 1.17      | 1.18           | -5.29%  | -5.88                |
| 24/Mar/99 |           | 0.94      | 1.36           | 0.00%   | -25.59               |
| 21/Mai/99 |           | 1.00      | 1.09           | -3.14%  | -6.92                |
| 24/Mai/99 |           | 1.03      | 1.43           | -5.12%  | -5.94                |

| data      | DIF<br>IO | DIF<br>AO | DIF<br>IO & AO | Retorno | log(X <sup>2</sup> ) |
|-----------|-----------|-----------|----------------|---------|----------------------|
| 25/Mai/99 |           | 1.19      | 2.17           | -5.19%  | -5.92                |
| 15/Jun/99 |           | 0.91      | 1.04           | 0.78%   | -9.71                |
| 23/Jul/99 |           | 0.73      | 1.08           | -2.36%  | -7.50                |
| 29/Set/99 |           | 0.73      | 0.73           | -0.19%  | -12.54               |
| 05/Out/99 |           | 0.98      | 1.30           | -1.31%  | -8.67                |
| 01/Nov/99 |           | 1.31      | 1.42           | 0.42%   | -10.94               |
| 13/Dez/99 | 2.50      |           | 2.52           | 0.08%   | -14.32               |
| 14/Dez/99 | 2.41      |           | 2.48           | -1.25%  | -8.77                |
| 27/Dez/99 | 2.48      |           | 2.51           | 0.25%   | -11.98               |
| 03/Jan/00 |           | 1.34      | 2.45           | -1.06%  | -9.10                |
| 10/Mar/00 |           | 0.75      | 0.77           | -2.12%  | -7.71                |
| 20/Mar/00 |           | 0.94      | 0.96           | -0.03%  | -16.25               |

\* AO detectados; \*\* IO detectado;

diferença entre as log-quase-verossimilhanças para considerar um outlier importante.

 $IO > \ln(1424)/2 \cong 3.63$ ; $AO > \ln(1424)/4 \cong 1.82$ ; $IO \& AO > \ln(1424)/2 + \ln(1000)/4 \cong 5.45$ ;

Na tabela 5.6, apresentamos os resultados do modelo incorporando os outliers detectados pelo procedimento de varredura. Observem que a curtose dos resíduos irregulares diminuiu de 0.9 para 0.3. Se analisarmos o gráfico 5.10 (principalmente o histograma) verificaremos que ainda existem pontos *soltos* que talvez pudessem ser considerados outlier, e, por este motivo, a curtose referente aos resíduos auxiliares irregulares transformados não foi muito afetada. Outro efeito observado é a mudança na estimativa do parâmetro  $\alpha$ . Este parâmetro representa a volatilidade média da série e com a detecção dos outliers a consequência natural é que a volatilidade média diminua. Comparando os modelos verificamos que a estimativa do parâmetro  $\alpha$  para o modelo com outliers (-9.605) foi menor que do modelo sem outliers (-8.939).

Tabela 5.11. Estimativas do modelo AR(1)\_SV considerando a presença dos outliers detectados pelo procedimento de varredura. Série IBOVESPA.

| Parâmetros  | Estimativa | IC 95%  |        |
|---|------------|---------|--------|
| $\alpha$  | -9.605     | -9.735; | -9.475 |
| $\gamma$  | 0.985      | 0.975;  | 0.990  |
| $\sigma^2_\eta$   | 0.014      | 0.005;  | 0.030  |
| IO 20/Dez/1994 (crise do México)  | 3.239      | 1.866;  | 4.611  |
| IO 11/Jul/1997 (crise da Ásia I)  | 2.280      | 1.031;  | 3.529  |
| IO 21/Out/1997 (crise da Ásia II)   | 2.314      | 0.868;  | 3.759  |
| IO 22/Mai/1998 (início da Crise)  | 1.877      | 0.536;  | 3.217  |
| IO 26/Ago/1998 (Rússia decreta moratória)   | 2.897      | 1.408;  | 4.385  |
| AO 05/Set/1994 (Queda do Ministro Rubens Ricupero da Economia)  | 4.084      | 2.447;  | 5.720  |
| AO 10/Mar/1995 (Introdução das intrabandas na cambiais)   | 4.397      | 3.092;  | 5.700  |
| AO 14/Mar/1995 (Pérsio Arida deixa o Banco Central)   | 3.188      | 2.079;  | 4.296  |
| AO 03/Jan/1996 (otimismo mercado financeiro mundial)  | 4.119      | 2.986;  | 5.251  |
| AO 15/Jul/1997 (Crise da Ásia I)  | 2.902      | 1.454;  | 4.350  |
| AO 02/Set/1997 (Crise da Ásia I)  | 3.684      | 2.141;  | 5.226  |
| AO 27/Out/1997 (Crise da Ásia II)   | 3.272      | 1.547;  | 4.996  |
| AO 30/Out/1997 (Crise da Ásia II)   | 2.487      | 1.424;  | 3.549  |
| AO 12/Nov/1997 (Crise da Ásia II)   | 2.942      | 1.32;   | 4.563  |
| AO 10/Set/1998 (crise economica mundial Bolsas do Mundo inteiro despencam, e aumento na taxa de juros doméstica). | 2.899      | 1.385;  | 4.412  |
| AO 15/Set/1998  | 2.975      | 1.926;  | 4.022  |
| AO 03/Dez/1998 (congresso não aprova cobrança dos inativos)   | 3.033      | 1.678;  | 4.386  |
| AO 15/Jan/1999 (desvalorização do Real)   | 5.361      | 3.863;  | 6.859  |
| AO 01/Fev/1999 (Chico Lopez deixa o Banco Central)  | 3.273      | 1.802;  | 4.744  |
| AO 15/Mar/1999 (principais bancos japoneses e europeus anunciam que linhas de crédito para o Brasil será mantido) | 3.987      | 2.362;  | 5.612  |
| log-verossimilhança   | -3210.070  |         |        |
| BIC (considerando outliers no modelo)   | -3262.714  |         |        |
| BIC (sem considerar outliers)   | -3259.171  |         |        |
| S estrutural  | -0.382     |         |        |
| K estrutural  | -0.037     |         |        |
| S irregular   | -0.150     |         |        |
| K irregular   | 0.302      |         |        |

No gráfico 5.9, apresentamos o resíduo auxiliar estrutural do modelo AR(1)\_SV incorporando os efeitos dos outliers e sua respectiva estimativa da densidade de probabilidade. Comparando com os do modelo sem outliers, verificamos que os picos superiores foram eliminados, sobrando somente o pico inferior para indicar a presença de INLIERS.

No gráfico 5.11 apresentamos o resíduo auxiliar irregular transformado. Comparando-o com os resíduos do modelo sem os outliers, verificamos que as principais observações discrepantes foram eliminadas e a distribuição ficou muito próxima da distribuição Normal. Entretanto, ainda sobraram algumas observações na cauda inferior, que contribuíram para a curtose ficar em 0.302.

Na tabela 5.12 temos uma lista com os possíveis outliers que não foram detectados pelo procedimento de varredura.

Tabela 5.12 Resultado do primeiro procedimento de varredura com os possíveis outliers não detectados

| Data      | DIF  | DIF  | DIF     | Retorno | LogK^2 |
|-----------|------|------|---------|---------|--------|
|           | IO   | AO   | IO & AO |         |        |
| 04/Jan/00 |      | 1.14 | 1.37    | -6.59%  | -5.41  |
| 26/Mai/99 |      | 1.02 | 2.22    | 5.89%   | -5.70  |
| 14/Dez/98 |      | 1.81 | 1.23    | -8.88%  | -4.82  |
| 03/Dez/98 |      | 1.84 | 1.41    | -9.20%  | -4.75  |
| 03/Set/98 | 2.54 |      | 2.64    | -9.01%  | -4.79  |
| 01/Set/98 | 2.68 |      | 2.70    | 6.65%   | -5.46  |
| 27/Ago/98 | 3.93 | 0.89 | 4.13    | -10.48% | -4.49  |
| 20/Ago/98 | 2.66 |      | 2.79    | -6.65%  | -5.39  |
| 26/Mai/98 | 2.35 |      | 2.51    | -5.96%  | -5.60  |
| 18/Mai/98 |      | 0.98 | 1.79    | -6.65%  | -5.39  |
| 27/Abr/98 |      | 1.16 | 1.89    | -5.89%  | -5.63  |
| 28/Out/97 | 2.36 |      | 2.39    | 6.22%   | -5.59  |
| 23/Out/97 | 5.89 | 0.77 | 5.95    | -8.50%  | -4.90  |
| 22/Jul/97 |      | 1.26 | 1.47    | 8.02%   | -5.07  |
| 17/Jul/97 |      | 1.16 | 1.24    | -7.49%  | -5.15  |
| 16/Jul/97 |      | 1.82 | 1.89    | 8.45%   | -4.97  |
| 16/Nov/95 |      | 1.05 | 1.22    | 5.29%   | -5.92  |
| 26/Out/95 |      | 1.38 | 1.39    | -6.75%  | -5.36  |
| 09/Jan/95 | 2.47 |      | 2.47    | -5.36%  | -5.81  |
| 06/Jan/95 | 3.00 |      | 3.00    | -5.32%  | -5.83  |
| 03/Jan/95 | 3.19 |      | 3.19    | -5.25%  | -5.85  |
| 22/Dez/94 | 3.85 |      | 3.94    | -6.29%  | -5.50  |
| 21/Dez/94 | 5.23 |      | 5.30    | -6.54%  | -5.42  |
| 29/Nov/94 |      | 0.92 | 0.94    | 6.03%   | -5.65  |

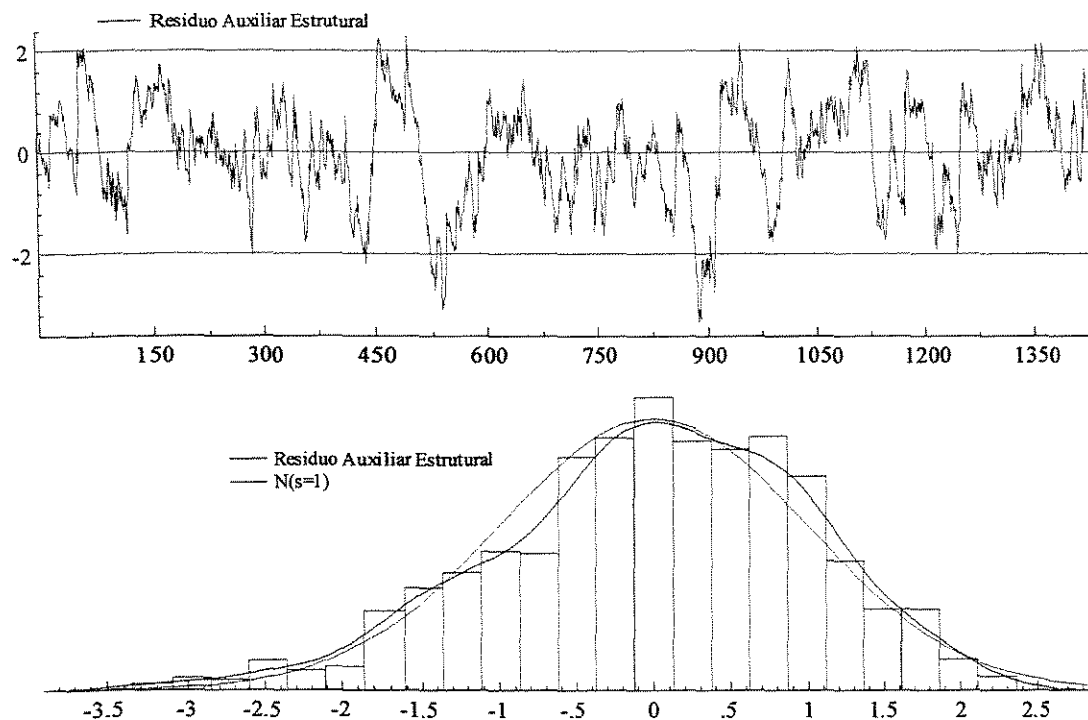


Gráfico 5.9 Resíduos auxiliares estruturais considerando os outliers detectados pelo procedimento de varredura. Série IBOVESPA

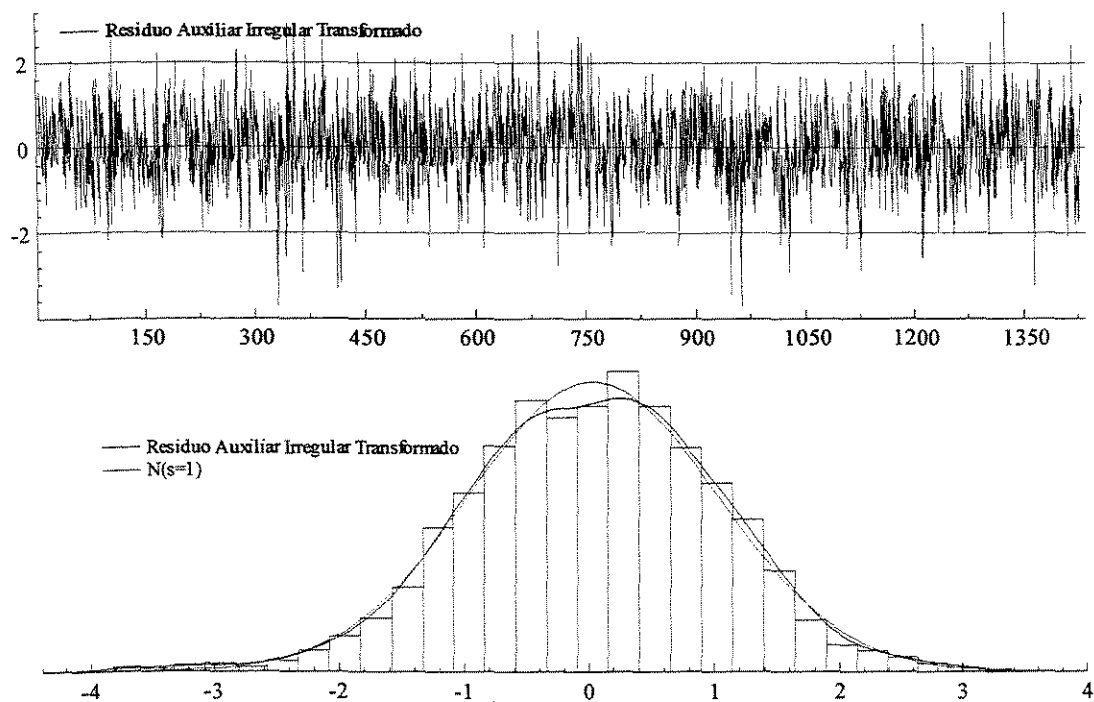


Gráfico 5.10 Resíduos auxiliares irregulares considerando os outliers detectados pelo procedimento de varredura. Série IBOVESPA

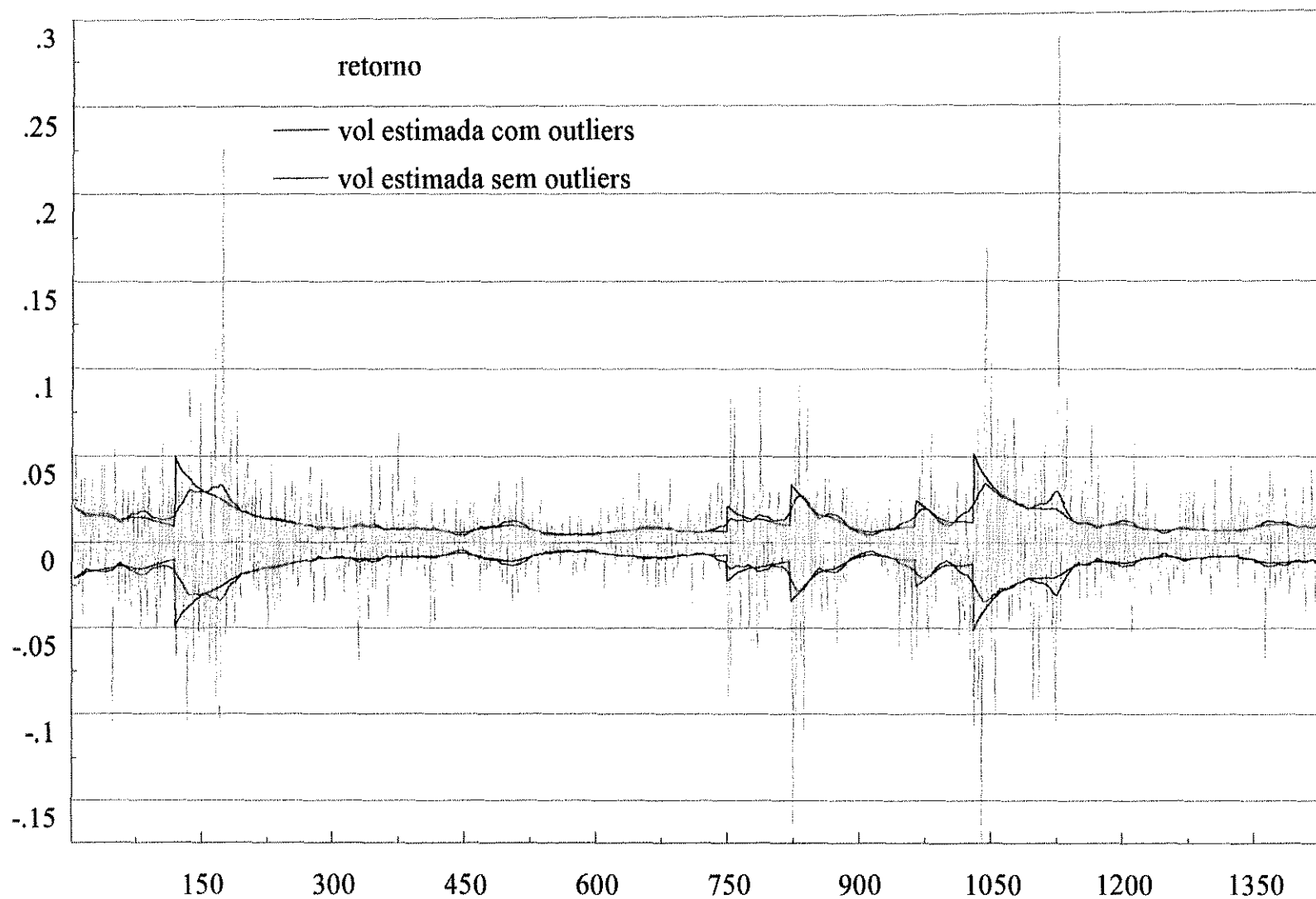


Gráfico 5.11. Retornos com volatilidade estimada com e sem os outliers detectados pelo procedimento de varredura. Série IBOVESPA.

## Conclusão.

A capacidade de detectar com eficiência e precisão outliers é extremamente útil, principalmente se levarmos em consideração os acontecimentos recentes do mercado financeiro mundial. Na última década, presenciamos um aumento da volatilidade dos mercados financeiros, representada por muitas crises financeiras em várias partes do mundo. Cada vez mais, uma turbulência num determinado país pode ter seus efeitos sentidos nos mercados do mundo inteiro. Dessa forma, são cada vez mais numerosos os fatos que influenciam um determinado mercado - índice de desemprego do país vizinho, quebra de uma determinada safra, resultado das eleições de um outro país etc.

O algoritmo proposto - que independe dos fatos - é um valioso instrumento analítico. Ele fornece informações quantitativas para podermos avaliar a magnitude de outliers. Cada outlier, entretanto, deve ter uma justificativa empírica, ou seja, ter uma causa real. Os resultados deste algoritmo, agregados a interpretações dos fatos empíricos, pode ajudar na tomada de decisões. Saber se um "dia nervoso" na bolsa de valores representa uma ruptura do padrão ou se foi apenas uma forte oscilação passageira pode mudar o rumo (ou resultado financeiro!) de muitas empresas e bancos!

O algoritmo proposto se mostrou preciso na detecção dos outliers, a um custo computacional não muito elevado. As melhorias que podem ser feitas, seguindo a mesma linha metodológica, implicam em aumento do custo computacional. Estas melhorias poderiam ser feitas calculando-se a verossimilhança exata. Entretanto, o método utilizado para o cálculo desta verossimilhança exata é baseado em métodos de estimação via simulação de Monte Carlo, que são caros computacionalmente. Considerando a surpreendente evolução da velocidade dos computadores, imaginamos que, em pouco tempo, este preço computacional pode não mais ser tão elevado. Portanto, acredito que tais métodos possam ser mais desenvolvidos pressupondo um elevado aumento da eficiência dos computadores. Um aspecto importante do algoritmo proposto é que ele pode ser inteiramente automatizado, restando ao usuário o acompanhamento do processo e, principalmente, a interpretação dos resultados gerados pelo algoritmo.

## Referências Bibliográficas.

- Bell, W. R. (1984). *Signal extration for nonstationary series*. The Annals os Statistics, 13, 646-664
- Black, F. (1976). *Studies in stock price volatility changes*. Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section, American Statistical Association, 177-181.
- Black, F. and M. Scholes (1973). *The pricing of options and corporate liabilities*, Journal of Political Economy 81, 637-654.
- Doornik, J.A. (1996), *Object-Oriented Matrix Programming using Ox*, London: International Thomson Business Press.
- Dunsmuir, W (1979), *A central limit theorem for parameter estimation in stationary time series and its applications to models and a random variance model*. Journal of Financial and Quantitative Analysis, 24, 267-284.
- Engle, R.F. (1982). *Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation*, Econometrika, 50, 987-1008.
- Gill, P. E., Murray, W. e Wright, M. H. (1981). *Practical Optimization*. New York. Acedemic Press.
- Harvey, A. C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. CAMBRIDGE University Press.
- Harvey, A. C. Koopman, S. J. (1992) *Diagnostic checking of unobserved-components time series models*. Journal of Business & Economic Statistics, Vol 10 No. 4 p. 377-389
- Harvey A. C., Ruiz, E., Shephard, N. (1994). *Multivariate stochastic variance models*. Review of Econometric Studies, 61:247-267.
- Hotta, L. K. e Tsay, R. S. (1999). *Outliers in GARCH models*. Documento de trabalho, Graduate School of Business, Universidade de Chicago, EUA.
- Jacquier E., Polson N. G e Rossi P. E. (1994) *Bayesian analysis of stochastic volatility models*. Journal of Business & Economic Statistic, Vol 12. No. 4 p. 371-417.
- Kalman, R. E. (1960). *A new approach to linear filtering and prediction problems*. Journal of Basic Ebginieering, transaction ASME. Series D82:35-45
- Kalman, R. E. e R. S. Bucy (1961). *New results in linear filtering and prediction theory*. Journal of Basic Engineering, Transactions AMSE, Series D 83:95-108.

- Koopman S. J. e Sandmann G. (1997) *Estimation of stochastic volatility models via monte carlo maximum likelihood*. Financial Markets Group and Department of Statistics, Lodon School of Economics
- Koopman, S. J. Harvey, A. C. Doornik J. A. e Shephard N. (1995) *STAMP 5.0 Structural Time Series Analyser, Modeller and Predictor*. Chapman & Hall
- Lomnicki, Z. A. (1961). *Tests for departure from normality in the base of linear stochastic processes*. *Metrika*, 4, 37-62.
- Ruiz, E. (1994) *Quasi-maximum likelihood etimation of stochastic volatility models*. *Journal of Econometrics*. 63 p. 289-306.
- Sakamoto, Y. Ishiguro, M. e Kitagawa, G. (1983) *Akaike Information Criterion Statistics*. D. Reidel Publishing Company.
- Schwarz, G. (1978). *Estimative of dimension of a model*. *Annals of Statistics*, v6, p. 461-464.
- Taylor, S. J. (1980). *Conjuctural models for trend in financial pricesm tests and forecasts*. *JRSS, A*, v. 143, p. 338-362.
- Taylor, S. J. (1986). *Modelling Finacial Time Series*. John Wiley & Sons.