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

Diagnóstico de Influência em Modelos de Volatilidade Estocástica

Simoni Fernanda Martim

Orientador: Prof. Dr. Mauricio Enrique Zevallos Herencia

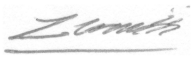
Dissertação apresentada junto ao Departamento de Estatística do Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas, para obtenção do Título de MESTRE em Estatística.

Campinas
2009

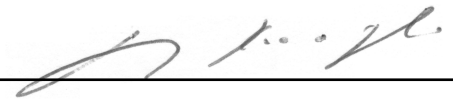
DIAGNÓSTICO DE INFLUÊNCIA EM MODELOS DE VOLATILIDADE ESTOCÁSTICA

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Simoni Fernanda Martim e aprovada pela comissão julgadora.

Campinas, 31 de agosto de 2009.



Prof. Dr. Mauricio Enrique Zevallos Herencia
Orientador



Prof. Dr. Luiz Koodi Hotta
Co-orientador

Banca Examinadora:
1. Prof. Dr. Mauricio Enrique Zevallos Herencia (IMECC - UNICAMP)
2. Profa. Dra. Clélia Maria de Castro Toloi (IME - USP)
3. Prof. Dr. Ricardo Sandes Ehlers (ICMC - USP)

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

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Crisllene Queiroz Custódio – CRB8 / 7966

Martim, Simoni Fernanda
M362d Diagnóstico de influência em modelos de volatilidade estocástica / Simoni Fernanda Martim -- Campinas, [S.P. : s.n.], 2009.

Orientadores : Mauricio Enrique Zevallos Herencia ; Luiz Koodi Hotta
Dissertação (Mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Observações influentes. 2. Valores estranhos (Estatística). 3. Diagnóstico. 4. Séries temporais. 5. Finanças - Estatística. I. Zevallos Herencia, Mauricio Enrique. II. Hotta, Luiz Koodi. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica.

Título em inglês: Influence diagnostics in stochastic volatility models

Palavras-chave em inglês (Keywords): 1. Influential observations. 2. Outliers (Statistics). 3. Diagnostics. 4. Time series. 5. Finance - Statistics.

Área de concentração: Séries temporais financeiras

Titulação: Mestre em Estatística

Banca examinadora: Prof. Dr. Mauricio Enrique Zevallos Herencia (IMECC-UNICAMP)
Profa. Dra. Clélia Maria de Castro Toloi (IME – USP)
Prof. Dr. Ricardo Sandes Ehlers (ICMC - USP)

Data da defesa: 31/08/2009

Dissertação de Mestrado defendida em 31 de agosto de 2009 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

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

Prof(a). Dr(a). CLÉLIA MARIA DE CASTRO TOLOI

Prof(a). Dr(a). RICARDO SANDES EHLERS

Agradecimentos

Agradeço primeiramente a Deus pois sem Ele eu não chegaria até aqui.

Aos meus pais, Teresa Alves Martim e Armando Antonio Martim, por sempre me darem apoio e incentivo durante essa jornada.

A Cristiano Amâncio Vieira Borges, pela ajuda, companheirismo, apoio e incentivo dados até o último momento.

Aos amigos que fiz durante todo o percurso do mestrado. Em especial: Alejandro, Lucia, Jonas, José Clelto, Rignaldo, Rodrigo (piadinha) e Rodrigo (xester).

Ao professor Maurício Zevallos, pela orientação, paciência, dedicação, apoio e atenção durante a elaboração deste nosso trabalho.

A Capes, pelo apoio financeiro a este projeto.

Aos professores do Departamento de Estatística do IMECC - UNICAMP, pelos ensinamentos concedidos.

Aos participantes da banca examinadora, pelas sugestões.

Resumo

O diagnóstico de modelos é uma etapa fundamental para avaliar a qualidade do ajuste dos modelos. Nesse sentido, uma das ferramentas de diagnóstico mais importantes é a análise de influência. Peña (2005) introduziu uma forma de analisar a influência em modelos de regressão, a qual avalia como cada ponto é influenciado pelos outros na amostra. Essa estratégia de diagnóstico foi adaptada por Hotta e Motta (2007) na análise de influência dos modelos de volatilidade estocástica univariados. Nesta dissertação, é realizado um estudo de diagnóstico de influência para modelos de volatilidade estocástica univariados assimétricos, assim como para modelos de volatilidade estocástica multivariados. As metodologias propostas são ilustradas através da análise de dados simulados e séries reais de retornos financeiros.

Abstract

Model diagnostics is a key step to assess the quality of fitted models. In this sense, one of the most important tools is the analysis of influence. Peña (2005) introduced a way of assessing influence in linear regression models, which evaluates how each point is influenced by the others in the sample. This diagnostic strategy was adapted by Hotta and Motta (2007) on the influence analysis of univariate stochastic volatility models. In this dissertation, it is performed a study of influence diagnostics of asymmetric univariate stochastic volatility models as well as multivariate stochastic volatility models. The proposed methodologies are illustrated through the analysis of simulated data and financial time series returns.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Os modelos estatísticos são ferramentas úteis para extrair e compreender as características essenciais de um conjunto de dados. Durante a construção de um modelo, é importante examinar os pontos amostrais que estão sendo usados, com o objetivo de determinar se existe alguns deles que estejam controlando propriedades importantes no modelo. Os elementos do conjunto de dados que efetivamente controlam aspectos da análise são considerados como influentes. Em particular, uma observação é dita *influente* se ela produzir alterações relevantes no resultado da análise quando for excluída ou submetida a uma pequena perturbação.

O diagnóstico de influência é feito assumindo o modelo como correto e, então, investigando a robustez das conclusões a pequenas modificações, a fim de avaliar a qualidade do ajuste do modelo para um conjunto de dados. Como formas de modificações, podemos ter desde a simples exclusão de observações ou perturbações em direções específicas.

Na literatura, existem várias medidas de influência no contexto de regressão. Uma das ferramentas de diagnóstico mais conhecidas, chamada de estatística da distância de Cook, foi proposta por Cook (1977), e consiste em comparar o valor ajustado para a variável resposta quando uma determinada observação é excluída da análise com o valor ajustado sem excluir nenhuma observação. A medida da distância de Cook é uma medida para avaliar qual é o nível de influência isolada de uma observação suspeita sobre as estimativas dos parâmetros. Outra ferramenta bastante utilizada, conhecida como Influência Local, foi proposta por Cook (1986), e é baseada na medida afastamento da verossimilhança e, consequentemente, na avaliação da curvatura de Cook. Este método permite avaliar a

influência que pequenas perturbações podem exercer sobre a verossimilhança.

Peña (2005) introduziu uma nova forma de analisar influência em modelos de regressão linear. Ao invés de verificar como a exclusão de uma observação influencia as estimativas dos parâmetros e as predições como no caso da medida da distância de Cook, o objetivo é verificar como uma observação é influenciada pelas outras observações na amostra. Ou seja, avalia-se a mudança na predição de uma determinada observação quando cada observação da amostra é excluída. Peña (2005) mostrou que, através dessa análise de influência, pode-se encontrar também grupos de observações influentes similares, que são mais difíceis de serem detectados pela medida da distância de Cook.

A generalização da medida de influência proposta por Peña (2005) para o contexto de séries temporais não é direta, uma vez que as observações de séries temporais são dependentes. Sendo assim, não se deve excluir uma observação, mas sim considerá-la uma observação faltante. Recentemente, Motta e Hotta (2007) estudaram a influência proposta por Peña (2005) em modelos de volatilidade estocástica univariados sem assimetria, onde apresentaram uma estatística baseada na estimativa da volatilidade, assim como uma metodologia baseada em marcas de referência para determinar se, estatisticamente, as observações são influentes.

Por outro lado, o primeiro modelo de volatilidade estocástica multivariado foi proposto por Harvey et al (1994), o qual considera que os retornos podem estar correlacionados entre eles, assim como as volatilidades. Aqui, a estimação é realizada baseada no método de quase-máxima verossimilhança, utilizando a representação na forma de espaço de estados linear. Outros modelos de volatilidade estocástica multivariados foram sugeridos na literatura. Assim, Quintana e West (1987) sugeriram um modelo fatorial multiplicativo, e Danielsson (1998) propôs um modelo assimétrico. Posteriormente, Jacquier et al (1999) consideraram um modelo de volatilidade estocástica fatorial aditivo, onde a volatilidade estocástica está presente em um fator comum, e o coeficiente de correlação entre os retornos observados é variante no tempo. Outros modelos fatoriais também foram considerados na literatura, como em Shephard (1996), Pitt e Shephard (1999) e Aguilar e West (2000).

O objetivo desta dissertação é apresentar um estudo da análise de influência no modelo de volatilidade estocástica univariado assimétrico, proposto por Harvey e Shephard (1996), e também nos seguintes modelos de volatilidade estocástica multivariados: modelo com correlações constantes generalizado, proposto por Harvey et al (1994), e modelo

fatorial aditivo, proposto por Jacquier et al (1999). Nesta dissertação, vamos propor uma medida de influência para a análise dos modelos de volatilidade estocástica multivariados, baseada no método proposto por Peña (2005) e que foi adaptado por Motta e Hotta (2007) para modelos de volatilidade univariados. Também, esta dissertação ilustra o uso da metodologia de limiares baseada em simulações de Monte Carlo no diagnóstico de influência para os modelos de volatilidade estocástica e também para modelos de regressão linear. O uso da técnica de diagnóstico de influência será ilustrado através da análise de séries simuladas e séries reais de retornos financeiros, e a estatística de influência será obtida tanto pela estimativa filtrada da volatilidade quanto pela estimativa suavizada da mesma. As análises de influência serão ilustradas através do estudo de séries simuladas e de séries financeiras reais.

1.1 Breve descrição da Dissertação

No capítulo 2 serão abordados os conceitos de modelos de espaço de estados, assim como a estimação dos parâmetros por máxima verossimilhança e o diagnóstico dos resíduos do modelo ajustado. Estes conceitos serão abordados com o propósito de fornecer uma estrutura teórica para desenvolvimentos nos capítulos posteriores.

O capítulo 3 apresenta os modelos de volatilidade estocástica univariados e também dois modelos de volatilidade estocástica multivariados, assim como as respectivas representações aproximadas na forma de espaço de estados, a fim de tratarmos da estimação dos modelos pelo método de quase-máxima verossimilhança.

No capítulo 4 serão apresentadas a estatística de influência proposta por Peña (2005) e sua adaptação para diagnosticar influência em modelos de volatilidade estocástica univariados proposta por Motta e Hotta (2007). Também, será descrita detalhadamente a metodologia de limiares proposta por Motta e Hotta (2007) para identificar observações influentes baseada em simulações de Monte Carlo. Por fim, essa metodologia é ilustrada através de dados simulados e da análise de dados reais, através do ajuste do modelo de volatilidade estocástica univariado assimétrico.

No capítulo 5 será apresentada uma generalização da medida de influência proposta por Motta e Hotta (2007) para modelos de volatilidade estocástica multivariados, e será descrita detalhadamente a metodologia de limiares para identificar observações influentes

baseada em simulações de Monte Carlo. Por fim, serão apresentadas algumas simulações e aplicações na análise de séries reais, utilizando o modelo com correlações constantes generalizado e o modelo fatorial aditivo.

O Capítulo 6 apresenta as conclusões e futuras linhas de pesquisas para o estudo realizado nesta dissertação.

Capítulo 2

Modelos de espaço de estados

Uma classe bastante geral de modelos, denominados Modelos de Espaço de Estados, foi introduzida por Kalman (1960) e Kalman e Bucy (1961). Esses modelos têm sido extensivamente utilizados para modelar dados provenientes da economia, da área médica e de ciências do solo, dentre outras áreas.

A representação de um modelo na forma de espaço de estados fornece uma abordagem flexível para a análise de séries temporais, possibilitando o uso do Filtro de Kalman como ferramenta básica para a estimação dos estados e a manipulação de valores faltantes. Os recursos de Kalman representam uma das maiores contribuições na teoria moderna de controle. Um excelente tratamento na análise de séries temporais baseada na forma de espaço de estados pode ser encontrado em Harvey (1989) e também em Durbin e Koopman (2001).

Neste capítulo serão apresentados os conceitos de modelos de espaço de estados lineares e gaussianos, os modelos não-lineares e não-gaussianos, os algoritmos do Preditor, Filtro e Suavizador de Kalman e a estimação por máxima verossimilhança. Esses conceitos serão apresentados aqui com o propósito de fornecer uma estrutura teórica para desenvolvimentos abordados nos capítulos posteriores.

2.1 Modelos de espaço de estados lineares e gaussianos

Nesta seção, será apresentado o conceito de modelos de espaço de estados lineares e gaussianos e as técnicas para análise desses modelos, como a estimação pelo método de máxima-verossimilhança e diagnóstico do modelo.

Um modelo de espaço de estados linear e gaussiano para uma série temporal $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$ consiste de duas equações. A primeira, conhecida como *equação de observação*, expressa a observação $\boldsymbol{y}_t$ como uma função linear de uma variável de estado $\boldsymbol{\alpha}_t$, e a segunda equação, conhecida como *equação de estado*, determina o estado $\boldsymbol{\alpha}_t$ em termos do estado anterior $\boldsymbol{\alpha}_{t-1}$. Assim, o modelo de espaço de estados é

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{H}_t), \\ \boldsymbol{\alpha}_t &= \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t), \end{aligned} \tag{2.1}$$

$t = 1, \ldots, T$ e $\boldsymbol{\alpha}_1 \overset{iid}{\sim} N(\boldsymbol{a}, \boldsymbol{P})$, onde

$$\begin{aligned} &\boldsymbol{y}_t, \boldsymbol{c}_t, \boldsymbol{\varepsilon}_t : N \times 1, \qquad \boldsymbol{Z}_t : N \times m, \qquad \boldsymbol{H}_t : N \times N, \\ &\boldsymbol{\alpha}_t, \boldsymbol{d}_t, \boldsymbol{\eta}_t : m \times 1, \qquad \boldsymbol{T}_t : m \times m, \qquad \boldsymbol{Q}_t : m \times m, \end{aligned}$$

as perturbações $\boldsymbol{\varepsilon}_t$ e $\boldsymbol{\eta}_t$ são não-correlacionadas entre si e $\boldsymbol{\alpha}_t$, o vetor de estados, é não-observável. Tanto a equação de observação quanto a equação de estado são lineares nos parâmetros $\boldsymbol{c}_t$ e $\boldsymbol{Z}_t$, $\boldsymbol{d}_t$ e $\boldsymbol{T}_t$ respectivamente, ou seja, são equações que envolvem um parâmetro multiplicado por um coeficiente, somado a outro parâmetro e a um erro. Também, tanto a observação $\boldsymbol{y}_t$ quanto o estado $\boldsymbol{\alpha}_t$ possuem distribuição normal.

A primeira equação em (2.1) segue a estrutura de um modelo de regressão linear onde $\boldsymbol{\alpha}_t$ varia com o tempo, enquanto que a segunda equação representa um modelo autoregressivo de primeira ordem.

Quando os parâmetros $\Theta_t = (\boldsymbol{c}_t, \boldsymbol{Z}_t, \boldsymbol{H}_t, \boldsymbol{d}_t, \boldsymbol{T}_t, \boldsymbol{Q}_t)$ forem constantes no tempo, o sistema será dito *invariante no tempo*, e denotaremos por $\Theta = (\boldsymbol{c}, \boldsymbol{Z}, \boldsymbol{H}, \boldsymbol{d}, \boldsymbol{T}, \boldsymbol{Q})$.

2.2 Os recursos de Kalman

Nesta seção, vamos considerar três problemas fundamentais associados à estimação do vetor de estados em (2.1). Aqui, o interesse é encontrar estimadores ótimos (estimadores com erro quadrático médio mínimo) para o vetor de estados $\boldsymbol{\alpha}_t$ em termos das observações $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$. A estimação de $\boldsymbol{\alpha}_t$, $t = 1, \ldots, T$, em termos de:

- $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$ está relacionada ao problema de predição,
- $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t$ está relacionada ao problema de filtragem,
- $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$ está relacionada ao problema de suavização.

Cada um desses problemas pode ser resolvido recursivamente usando um algoritmo apropriado. Nesta seção, veremos o Preditor, o Filtro e o Suavizador de Kalman, os quais serão utilizados nos capítulos posteriores.

2.2.1 O Preditor de Kalman

O Preditor de Kalman, proposto por Kalman (1960), é um algoritmo de estimação recursivo. No modelo de espaço de estados gaussiano, ele fornece estimadores ótimos para o vetor de estados $\boldsymbol{\alpha}_t$ e para sua matriz de covariâncias baseado na informação disponível até o instante $t-1$: $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}\}$. O Preditor de Kalman fornece, então,

$$\boldsymbol{a}_{t|t-1} = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}) \quad \text{e} \quad \boldsymbol{P}_{t|t-1} = \mathrm{Var}(\boldsymbol{\alpha}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}),$$

para $t = 1, \ldots, T$, onde $\boldsymbol{a}_{t|t-1}$ é a estimativa do vetor de estados $\boldsymbol{\alpha}_t$ e $\boldsymbol{P}_{t|t-1}$ é a estimativa da matriz de covariâncias do vetor de estados.

Para o modelo (2.1), com as condições iniciais $\boldsymbol{a}_1 = \boldsymbol{a}$ e $\boldsymbol{P}_1 = \boldsymbol{P}$, o Preditor de Kalman é dado por

$$\begin{aligned}
\boldsymbol{v}_t &= \boldsymbol{y}_t - \boldsymbol{c}_t - \boldsymbol{Z}_t\boldsymbol{a}_{t|t-1} \\
\boldsymbol{F}_t &= \boldsymbol{Z}_t\boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{H}_t \\
\boldsymbol{K}_t &= \boldsymbol{T}_t\boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{F}_t^{-1} \\
\boldsymbol{a}_{t+1|t} &= \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{a}_{t|t-1} + \boldsymbol{K}_t\boldsymbol{v}_t \\
\boldsymbol{P}_{t+1|t} &= \boldsymbol{T}_t\boldsymbol{P}_{t|t-1}\boldsymbol{T}_t' + \boldsymbol{Q}_t - \boldsymbol{K}_t\boldsymbol{F}_t\boldsymbol{K}_t',
\end{aligned} \tag{2.2}$$

para $t = 1, \ldots, T$, onde

$$\boldsymbol{v}_t : N \times 1, \qquad \boldsymbol{F}_t : N \times N, \qquad \boldsymbol{K}_t : m \times N,$$
$$\boldsymbol{a}_{t+1|t} : m \times 1, \qquad \boldsymbol{P}_{t+1|t} : m \times m,$$

$\boldsymbol{v}_t$ representa o erro da predição de $\boldsymbol{y}_t$ dado $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}\}$ e $\boldsymbol{F}_t = \mathrm{Var}(\boldsymbol{v}_t)$ representa a variância da predição.

Quando a distribuição dos estados é desconhecida, Koopman et al (1999) sugerem inicializar o Preditor de Kalman com $\boldsymbol{a}_1 = 0$ e $\boldsymbol{P}_1 = kI$, com $k = 10^6$, sendo I uma matriz identidade $m \times m$.

2.2.2 O Filtro de Kalman

Uma vez que a observação $\boldsymbol{y}_t$ está disponível, a estimação de $\boldsymbol{\alpha}_t$ pode ser atualizada. As *equações de atualização* são dadas por

$$\begin{aligned} \boldsymbol{a}_t &= \boldsymbol{a}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{v}_t \\ \boldsymbol{P}_t &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{Z}_t\boldsymbol{P}_{t|t-1}, \end{aligned} \tag{2.3}$$

para $t = 1, \ldots, T$. Através de (2.2) e (2.3), obtemos o Filtro de Kalman, que fornece

$$\boldsymbol{a}_t = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t) \quad \text{e} \quad \boldsymbol{P}_t = \mathrm{Var}(\boldsymbol{\alpha}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t),$$

para $t = 1, \ldots, T$. Estas estimativas estão baseadas na informação $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$, e contém toda a informação necessária para fazer previsões de valores futuros tanto dos estados quanto das observações.

2.2.3 O Suavizador de Kalman

Vimos que o Filtro de Kalman fornece o valor estimado para o vetor de estado $\boldsymbol{\alpha}_t$ baseado na informação disponível até o instante t. O Suavizador de Kalman considera também toda a informação disponível após o instante t. Sendo assim, ele fornece as estimativas do vetor de estados $\boldsymbol{\alpha}_t$ e da sua matriz de covariâncias utilizando todo o conjunto de observações $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T\}$. Os valores de $\hat{\boldsymbol{\alpha}}_t = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$ e da matriz de covariâncias estimada $\boldsymbol{V}_t = \mathrm{Var}(\boldsymbol{\alpha}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$ são dados por

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_t &= \boldsymbol{a}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{r}_{t-1} \\ \boldsymbol{V}_t &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{N}_{t-1}\boldsymbol{P}_{t|t-1}, \end{aligned} \tag{2.4}$$

$t = 1, \ldots, T$, onde $\boldsymbol{r}_{t-1}$ e $\boldsymbol{N}_{t-1}$ são obtidos pelas equações do Suavizador de Kalman, dadas por

$$\begin{aligned} \boldsymbol{L}_t &= \boldsymbol{T}_t - \boldsymbol{K}_t \boldsymbol{Z}_t \\ \boldsymbol{r}_{t-1} &= \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \boldsymbol{v}_t + \boldsymbol{L}_t' \boldsymbol{r}_t \\ \boldsymbol{N}_{t-1} &= \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \boldsymbol{Z}_t + \boldsymbol{L}_t' \boldsymbol{N}_t \boldsymbol{L}_t, \end{aligned} \tag{2.5}$$

$t = T, \ldots, 1$, onde $\boldsymbol{r}_N = \boldsymbol{N}_N = \boldsymbol{0}$ e $\boldsymbol{v}_t, \boldsymbol{F}_t$ e $\boldsymbol{K}_t$ são obtidos através das equações do Preditor de Kalman (2.2).

2.2.4 Algoritmo com dados faltantes

Quando a observação $\boldsymbol{y}_t$ num determinado instante t está ausente, o vetor $\boldsymbol{v}_t$ e a matriz $\boldsymbol{K}_t$ do Preditor de Kalman são iguais a zero e, então, o Preditor de Kalman se reduz a

$$\begin{aligned} \boldsymbol{a}_{t+1|t} &= \boldsymbol{d}_t + \boldsymbol{T}_t \boldsymbol{a}_{t|t-1} \\ \boldsymbol{P}_{t+1|t} &= \boldsymbol{T}_t \boldsymbol{P}_{t|t-1} \boldsymbol{T}_t' + \boldsymbol{Q}_t, \end{aligned} \tag{2.6}$$

uma vez que $\boldsymbol{v}_t = 0$, $\boldsymbol{F}_t^{-1} = 0$ e $\boldsymbol{K}_t = 0$. A estimativa suavizada (2.4) - (2.5) no ponto t faltante se reduz a

$$\begin{aligned} \boldsymbol{r}_{t-1} &= \boldsymbol{T}_t' \boldsymbol{r}_t \\ \boldsymbol{N}_{t-1} &= \boldsymbol{T}_t' \boldsymbol{N}_t \boldsymbol{T}_t \\ \hat{\boldsymbol{\alpha}}_t &= \boldsymbol{a}_{t|t-1} + \boldsymbol{P}_{t|t-1} \boldsymbol{r}_{t-1} \\ \boldsymbol{V}_t &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1} \boldsymbol{N}_{t-1} \boldsymbol{P}_{t|t-1} \end{aligned} \tag{2.7}$$

(ver Durbin e Koopman, 2001).

2.3 Estimadores de máxima verossimilhança

Os parâmetros do modelo de espaço de estados gaussiano incluídos em (2.1), $\Theta_t = (\boldsymbol{c}_t, \boldsymbol{Z}_t, \boldsymbol{H}_t, \boldsymbol{d}_t, \boldsymbol{T}_t, \boldsymbol{Q}_t)$, podem ser estimados através do método de máxima verossimilhança, supondo que o estado inicial $\boldsymbol{\alpha}_0$ tem distribuição normal e que as perturbações $\boldsymbol{\varepsilon}_t$ e $\boldsymbol{\eta}_t$ são não-correlacionadas e conjuntamente normais. O cálculo da função de log-verossimilhança para quaisquer valores fixos dos parâmetros é bastante simples; ela pode

ser expressa por

$$l = -\frac{TN}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}(\ln|\boldsymbol{F}_t| + \boldsymbol{v}_t'\boldsymbol{F}_t^{-1}\boldsymbol{v}_t), \tag{2.8}$$

onde $\boldsymbol{v}_t$ é o erro de predição de $\boldsymbol{y}_t$ e $\boldsymbol{F}_t$ é a matriz de covariância de $\boldsymbol{v}_t$, ambos obtidos pelas equações do Preditor de Kalman (2.2). Para obter a estimativa de máxima verossimilhança dos parâmetros, um algoritmo de otimização apropriado deve ser usado para encontrar os valores desses parâmetros que maximizam o valor da função (2.8).

2.4 Diagnóstico do modelo

Nesta seção, apresentaremos os procedimentos para diagnóstico dos resíduos dos modelos ajustados, definidos em Harvey (1989) e em Harvey e Koopman (1992), através dos erros de predição e dos resíduos auxiliares.

2.4.1 Resíduos de predição

Em um modelo de espaço de estados linear e gaussiano (2.1), os erros de predição de $\boldsymbol{y}_t$ possuem distribuição normal e são independentes. Assim,

$$\boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{F}_t), \tag{2.9}$$

$t = 1, \ldots, T$, onde $\boldsymbol{v}_t$ e $\boldsymbol{F}_t$ são obtidos através do Preditor de Kalman (2.2). Os resíduos padronizados das observações são úteis no diagnóstico dos modelos, através de procedimentos gráficos e testes estatísticos, e são dados por

$$\tilde{\boldsymbol{v}}_t = \boldsymbol{F}_t^{-1/2}\boldsymbol{v}_t, \tag{2.10}$$

$\tilde{\boldsymbol{v}}_t = (\tilde{v}_{1t}, \ldots, \tilde{v}_{Nt})'$, $t = d+1, \ldots, T$, onde d representa o número de elementos não-estacionários do vetor de estados $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T)'$. Apenas os resíduos $\tilde{\boldsymbol{v}}_{d+1}, \ldots, \tilde{\boldsymbol{v}}_T$ são utilizados no diagnóstico, uma vez que os outros resíduos estão relacionados a elementos não-estacionários do vetor de estados. No caso usual, quando os parâmetros Θ_t do modelo de espaço de estados são valores estimados, os erros de predição (2.10) são valores aproximados.

Os resíduos padronizados da i-ésima série de dados são obtidos através de

$$\frac{\tilde{v}_{it} - \bar{\tilde{v}}_i}{\hat{\sigma}_i} \overset{iid}{\sim} N(0,1), \tag{2.11}$$

$t = d+1, \ldots, T$, onde $\bar{\tilde{v}}_i$ é a média dos resíduos da i-ésima série e $\hat{\sigma}_i$, a estimativa do desvio padrão, é obtida através de

$$\hat{\sigma}_i^2 = (T-d-1)^{-1} \sum_{t=d+1}^{T} (\tilde{v}_t - \bar{\tilde{v}}_i)^2. \tag{2.12}$$

Gráficos dos resíduos de predição

A maneira mais comum de se analisar o comportamento dos resíduos é através dos gráficos dos resíduos padronizados (2.11) *versus* o tempo. Através desse gráfico, pode-se verificar a homogeneidade da variância do erro, ou seja, se a variância dos resíduos da i-ésima série é constante ao longo do tempo. Se os resíduos estiverem dispersos em torno de uma linha horizontal, então o modelo parece ser adequado pois, para cada acréscimo no tempo, a amplitude dos resíduos se mantém constante. Dessa forma, se os dados atendem às premissas, os resíduos devem ficar alocados numa faixa horizontal, sem mostrar uma tendência positiva ou negativa.

Função de autocorrelação

Outra ferramenta de análise é a função de autocorrelação dos resíduos. A presença de correlação representa uma séria violação da não-dependência entre os resíduos. Se o modelo é adequado, as autocorrelações devem estar praticamente todas dentro dos limites de ± 2 desvios padrões.

Análise da soma cumulativa (CUSUM)

A soma cumulativa (CUSUM) é útil na análise de instabilidade dos parâmetros e mudanças estruturais, e é definida por

$$W_{it} = \hat{\sigma}_i^{-1} \sum_{j=d+1}^{t} v_{it}, \tag{2.13}$$

$i = 1, \ldots, N$ e $t = d+1, \ldots, T$, onde $\hat{\sigma}_i$, a estimativa do desvio padrão, é dada por (2.12).

A análise é feita através do gráfico de W_{it} *versus* t para $t = d+1, \ldots, T$. Quando o modelo não está bem especificado, pode haver um número desproporcional de resíduos com o mesmo sinal. Neste caso, o gráfico da soma cumulativa mostrará W_{it} se afastando do eixo horizontal em $W_{it} = 0$. Para testar a significância do afastamento de W_t do eixo horizontal, utiliza-se duas linhas simétricas acima e abaixo de $W_{it} = 0$. Essas linhas são construídas de tal forma que a probabilidade de W_{it} cruzar alguma delas é igual a um dado nível de significância. Harvey (1989) definiu as linhas de significância através da equação

$$W = \pm \left(a\sqrt{T-d} + 2a \frac{t-d}{\sqrt{T-d}} \right), \tag{2.14}$$

onde $a = 0,948$ para um nível de significância de 5%.

A função das linhas de significância dadas por (2.14) é avaliar o comportamento da soma cumulativa (2.13) ao longo do tempo, embora elas possam ser usadas para fornecer um teste de significância, onde rejeita-se a hipótese de especificação correta do modelo quando a soma cumulativa ultrapassar as linhas de significância em qualquer ponto.

Análise da soma dos quadrados cumulativa (CUSUMSQ)

A soma dos quadrados cumulativa (CUSUMSQ) é útil para detectar mudanças estruturais, e pode ser usada para complementar a análise da soma cumulativa. Essa medida é dada por

$$WW_{it} = \sum_{j=d+1}^{t} v_{ij}^2 \Big/ \sum_{j=d+1}^{T} v_{ij}^2, \tag{2.15}$$

$t = d+1, \ldots, T$ e $i = 1, \ldots, N$. A análise é feita através do gráfico de WW_{it} *versus* t para $t = d+1, \ldots, T$. Quando o modelo está corretamente especificado, WW_{it} tem uma distribuição Beta com média $(t-d)/(T-d)$. Por esse motivo, define-se as linhas de significância através da equação

$$WW = \frac{t-d}{T-d} \pm c_0, \tag{2.16}$$

onde o valor de c_0 é escolhido de tal forma que a probabilidade de WW_t cruzar alguma das linhas (2.16) é igual ao nível de significância desejado. A Tabela com os valores significativos de c_0 pode ser encontrada em Harvey (1989).

2.4.2 Resíduos auxiliares

Os resíduos auxiliares são definidos como os estimadores das perturbações associados aos componentes não-observados. Harvey e Koopman (1992) mostraram que os resíduos auxiliares dos estados podem ser obtidos através dos estimadores suavizados de $\boldsymbol{\eta}_t$ padronizados:

$$\tilde{\boldsymbol{\eta}}_t = \mathrm{Var}(\hat{\boldsymbol{\eta}}_t)^{-1/2}\hat{\boldsymbol{\eta}}_t, \tag{2.17}$$

$\tilde{\boldsymbol{\eta}}_t = (\tilde{\eta}_{1t}, \ldots, \tilde{\eta}_{Nt})'$, $t = 1, \ldots, T$, onde

$$\begin{aligned} \hat{\boldsymbol{\eta}}_t &= \mathrm{E}(\boldsymbol{\eta}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T) = \boldsymbol{Q}_t\boldsymbol{r}_t \\ \mathrm{Var}(\hat{\boldsymbol{\eta}}_t) &= \boldsymbol{Q}_t - \boldsymbol{Q}_t\boldsymbol{N}_t\boldsymbol{Q}_t', \end{aligned} \tag{2.18}$$

e $\boldsymbol{r}_t$ e $\boldsymbol{N}_t$ são obtidos através do Suavizador de Kalman (2.5). Esses resíduos são úteis no diagnóstico dos modelos através de procedimentos gráficos e testes estatísticos, embora sejam serialmente correlacionados. Levando em consideração a correlação em série, Harvey e Koopman (1992) propuseram um teste para excesso de curtose corrigido e modificaram o teste de normalidade de Bowman-Shenton. O uso destes testes para modelos de volatilidade estocástica é limitado, pois a distribuição das observações após a transformação logarítmica não é gaussiana, como veremos no Capítulo 3.

Os resíduos auxiliares podem ser úteis para explicar observações atípicas, como veremos no Capítulo 5.

2.5 Modelos de espaço de estados não-gaussianos

Os modelos de espaço de estados não-gaussianos serão de maior interesse devido ao enfoque dado aos modelos de volatilidade estocástica. Nesta seção, será apresentado o conceito de modelos de espaço de estados não-gaussianos, e também será apresentado um método utilizado para a estimação dos parâmetros dos mesmos.

Existem muitas formas de representar os modelos não-gaussianos na forma de espaço de estados. Durbin e Koopman (1997) e Shephard e Pitt (1997) consideraram a equação de estados da forma linear e gaussiana, como na equação (2.1), e a equação de observação

de uma forma mais geral. Assim, o modelo considerado é escrito da seguinte forma:

$$\begin{aligned} &\boldsymbol{y}_t \sim p(\boldsymbol{y}_t|\boldsymbol{\theta}_t), \qquad \boldsymbol{\theta}_t = \boldsymbol{c}_t + \boldsymbol{Z}_t\boldsymbol{\alpha}_t, \\ &\boldsymbol{\alpha}_t = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t), \end{aligned} \tag{2.19}$$

$t = 1, \ldots, T$, onde $p(\boldsymbol{y}_t|\boldsymbol{\theta}_t)$ é uma densidade conhecida e não-gaussiana, $\boldsymbol{\alpha}_1 \sim N(\boldsymbol{a}, \boldsymbol{P})$ e

$$\boldsymbol{y}_t, \boldsymbol{c}_t : N \times 1, \qquad \boldsymbol{\alpha}_t, \boldsymbol{d}_t, \boldsymbol{\eta}_t : m \times 1, \qquad \boldsymbol{T}_t : m \times m, \qquad \boldsymbol{Q}_t : m \times m.$$

Como no caso dos modelos de espaço de estados lineares e gaussianos, o objetivo é estimar os parâmetros do modelo e os vetores de estado. No caso dos modelos de espaço de estados lineares e gaussianos foi visto que, dados os parâmetros, o Filtro de Kalman fornece as estimativas dos estados. Para o caso não-gaussiano como em (2.19), não temos uma expressão analítica para as estimativas dos estados.

Jungbacker e Koopman (2005) utilizaram um modelo aproximado para os modelos de espaço de estados não-gaussianos que torna possível estimar os parâmetros e os vetores de estado dos mesmos. Essa mesma forma de aproximação para o modelo foi considerada nos trabalhos de Durbin e Koopman (1997) e Shephard e Pitt (1997). Dado um chute inicial $\boldsymbol{g}$ para a solução de $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}_1}, \ldots, \hat{\boldsymbol{\theta}_T})'$, a solução iterativa pode ser obtida aplicando-se o Suavizador de Kalman considerando o modelo de espaço de estados aproximado dado por

$$\begin{aligned} &\boldsymbol{x}_t = \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t = \boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{A}_t), \\ &\boldsymbol{\alpha}_t = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t), \end{aligned} \tag{2.20}$$

onde $\boldsymbol{A}_t = -\ddot{p}(\boldsymbol{y}_t|\boldsymbol{\theta}_t)\big|^{-1}_{\theta=g}$, $\boldsymbol{x}_t = \boldsymbol{g}_t + \boldsymbol{A}_t\, \dot{p}(\boldsymbol{y}_t|\boldsymbol{\theta}_t)\big|_{\theta=g}$,

$$\dot{p}(\boldsymbol{y}_t|\boldsymbol{\theta}_t) = \frac{\partial \ln p(\boldsymbol{y}_t|\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t} \quad \text{e} \quad \ddot{p}(\boldsymbol{y}_t|\boldsymbol{\theta}_t) = \frac{\partial^2 \ln p(\boldsymbol{y}_t|\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t \partial \boldsymbol{\theta}_t'}.$$

Dessa forma, através de um chute inicial $\boldsymbol{g}$, o algoritmo de suavização aplicado em (2.20) nos produz o novo valor $\boldsymbol{g}^+$. O processo se repete até a convergência e, assim, teremos $\hat{\boldsymbol{\theta}}$.

Os parâmetros do modelo de espaço de estados aproximado (2.20) podem ser estimados através do método de máxima verossimilhança descrito na Seção 2.3, e o diagnóstico do modelo ajustado pode ser realizado através dos procedimento descritos na Seção 2.4.

Capítulo 3

Modelos de volatilidade estocástica

Muitas séries temporais, em especial as séries financeiras, apresentam variâncias condicionais que mudam ao longo do tempo. A variância condicional dos retornos, mais conhecida como volatilidade, é de vital importância no mercado financeiro, pois a partir dessa pode-se, por exemplo, precificar opções e medir o risco de um ativo. Diante disso, uma grande variedade de modelos tem sido proposta para a modelagem da volatilidade.

Um dos modelos univariados propostos foi o modelo de volatilidade estocástica de Taylor (1986). Esse modelo tem como premissa o fato que a volatilidade presente depende dos valores passados da mesma, e é não observável. Uma importante característica deste modelo é que ele pode ser colocado na forma de espaço de estados. No caso multivariado, existem várias formulações do modelo de volatilidade estocástica, dentre as quais podemos destacar as formulações propostas por Harvey et al (1994) e Jacquier et al (1999).

Neste capítulo serão apresentados os modelos de volatilidade estocástica univariados e também dois modelos multivariados no contexto da modelagem de retornos. Por fim, trataremos da estimação dos modelos pelo método de quase-máxima verossimilhança.

3.1 Modelo univariado

Nos modelos de volatilidade estocástica, a variável dependente é o retorno. Quando o preço de um ativo de interesse é definido como p_t, o retorno pode ser definido como

$$y_t = \ln(p_t) - \ln(p_{t-1}). \tag{3.1}$$

Dizemos que uma série de retornos $y_1, \ldots, y_T$ segue um modelo de volatilidade estocástica univariado simétrico (MVE-U) se

$$y_t = \beta e^{h_t/2}\varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} N(0,1), \tag{3.2}$$

$t = 1, \ldots, T$, onde o termo h_t é a volatilidade (ou log-volatilidade) de y_t, é não-observado e independe dos valores passados de y_t. Na formulação mais simples proposta por Taylor (1986), h_t é modelado como um processo autoregressivo de primeira ordem

$$h_t = \phi h_{t-1} + \eta_t, \qquad \eta_t \overset{iid}{\sim} N(0, \sigma_\eta^2), \tag{3.3}$$

no qual η_t e ε_t são serialmente não-correlacionados e $|\phi| < 1$ para garantir que o processo h_t seja estacionário.

Modelo univariado assimétrico

O modelo de volatilidade estocástica univariado assimétrico (MVE-A) proposto por Harvey e Shephard (1996) assume assimetria através da correlação negativa entre as perturbações dos retornos $y_1, \ldots, y_T$ e da volatilidade, ou seja,

$$\begin{aligned} y_t &= \beta e^{h_t/2}\varepsilon_t, \\ h_{t+1} &= \phi h_t + \eta_t, \\ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} &\overset{iid}{\sim} N\left(\mathbf{0}, \begin{bmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{bmatrix}\right) \end{aligned} \tag{3.4}$$

$t = 1, \ldots, T$, onde $|\rho| < 1$. Pode-se observar que $corr(\varepsilon_t, \eta_t) = \rho$ para $t = 1, \ldots, T$.

Este modelo, diferentemente do modelo (3.2) - (3.3), consegue reproduzir o efeito de alavancagem usualmente encontrado nas séries de retornos, uma vez que, na volatilidade, o efeito dos retornos positivos é diferente do efeito dos retornos negativos com magnitude similar.

A estimação dos modelos de volatilidade estocástica torna-se mais simples quando este é escrito na forma de espaço de estados, como veremos a seguir.

3.1.1 Estimação

Os modelos de volatilidade estocástica não são tão simples de serem estimados, uma vez que não há uma forma analítica para a função de verossimilhança. Para contornar

este problema, Jacquier et al (1994) propuseram a análise desses modelos com enfoque bayesiano, onde utilizaram o método MCMC (*Markov Chain Monte Carlo*) na estimação dos parâmetros. Outra maneira de contornar o problema de estimação consiste em reescrever o modelo na forma de espaço de estados linear, dada por (2.1), e usar o procedimento de quase-máxima verossimilhança. Neste trabalho, vamos considerar dois métodos diferentes de se reescrever um modelo de volatilidade estocástica na forma de espaço de estados linear. O primeiro método é baseado na transformação logarítmica da equação de observação. Porém, após essa transformação, as perturbações $\boldsymbol{\varepsilon}_t$ obtidas não possuem distribuição gaussiana. Apesar disso, Harvey et al (1994) propõem assumir normalidade e trabalhar com a verossimilhança aproximada, chamada de quase-verossimilhança. Embora a estimação do modelo baseada nessa transformação não seja eficiente por se tratar de um método aproximado, ele apresenta estimadores consistentes para um tamanho amostral razoável e exige menor tempo computacional quando comparado à maximização da verossimilhança exata. O segundo método é baseado nas primeiras derivadas da log-densidade das observações, e também se trata de um método aproximado, com custo computacional mais elevado que o método anterior, mas que produz melhores estimativas da volatilidade $\boldsymbol{h}_t$ quando comparado ao primeiro método. A seguir, serão discutidos os modelos de volatilidade simétrico e assimétrico.

3.1.1.1 Modelo univariado simétrico

O modelo de volatilidade estocástica univariado (3.2) - (3.3) pode ser reescrito na forma de espaço de estados linear através da transformação logarítmica ou através do método baseado nas primeiras derivadas, visto na Seção 2.5. A seguir, serão mostrados esses dois métodos.

Transformação logarítmica

Para reescrever o MVE-U na forma de espaço de estados linear, podemos aplicar uma transformação logarítmica na equação (3.2), obtendo assim

$$\ln(y_t^2) = \ln(\beta^2) + h_t + \ln(\varepsilon_t^2). \tag{3.5}$$

Como ε_t tem distribuição normal com média zero e variância um, então $\ln(\varepsilon_t^2)$ tem

uma distribuição log-quiquadrado com média -1,27 e variância $\pi^2/2$. Considerando $\xi_t = \ln(\varepsilon_t^2) + 1,27$ e $y_t^* = \ln(y_t^2)$, e também considerando a equação (3.3), temos um modelo de espaço de estados linear

$$\begin{aligned} y_t^* &= -1,27 + \ln(\beta^2) + h_t + \xi_t, \qquad \xi_t \overset{iid}{\sim} (0, \pi^2/2) \\ h_t &= \phi h_{t-1} + \eta_t, \qquad \eta_t \overset{iid}{\sim} N(0, \sigma_\eta^2), \end{aligned} \tag{3.6}$$

mas não gaussiano, uma vez que ξ_t não segue uma distribuição normal. A função de log-verossimilhança é calculada através do Preditor de Kalman, como vimos na Seção 2.3.

Quando há observações iguais a zero, não podemos fazer a transformação logarítmica. Uma solução, sugerida por Fuller (1996, p. 495), é fazer a seguinte transformação baseada numa expansão de Taylor:

$$\ln(y_t^2) = \ln(y_t^2 + 0.02\ S_y^2) - \frac{0.02\ S_y^2}{y_t^2 + 0.02\ S_y^2}, \tag{3.7}$$

em que S_y^2 é a variância amostral da série $y_1, \ldots, y_N$.

Modelo aproximado baseado nas primeiras derivadas

O modelo gaussiano aproximado segue como na Seção 2.5, pelo método proposto por Jungbacker e Koopman (2005). No MVE-U (3.2) e (3.3) a densidade das observações é dada por

$$\ln p(y_t|h_t) = -\frac{1}{2}\ln(2\pi\beta^2) - \frac{1}{2}h_t - \frac{y_t^2}{2\beta^2}\exp(-h_t), \tag{3.8}$$

Assim,

$$\dot{p}(y_t|h_t) = -\frac{1}{2} + \frac{y_t^2}{2\beta^2}\exp(-h_t) \quad \text{e} \quad \ddot{p}(y_t|h_t) = -\frac{y_t^2}{2\beta^2}\exp(-h_t). \tag{3.9}$$

Dessa forma, através dos resultados obtidos na Seção 2.5, segue que

$$A_t = \frac{2\beta^2}{y_t^2}\exp(g_t) \tag{3.10}$$

e

$$x_t = g_t - \frac{1}{2}A_t + 1, \tag{3.11}$$

para o qual A_t é sempre positiva.

O modelo gaussiano aproximado para o MVE-U (3.2) - (3.3) é obtido através do chute inicial $g_t = 0$ e aplicando-se o Preditor e Suavizador de Kalman considerando

$$\begin{aligned} x_t &= h_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, A_t), \\ h_t &= \phi_t h_{t-1} + \eta_t, \qquad \eta_t \overset{iid}{\sim} N(0, \sigma_\eta^2), \end{aligned} \tag{3.12}$$

onde teremos novos valores para g_t. Esses novos valores serão usados para obter novos A_t e x_t. Esse processo recursivo é repetido até a convergência de g_t, onde teremos finalmente $\hat{h}_t$, $t = 1, \ldots, T$. Os parâmetros do modelo aproximado podem ser estimados através do método de máxima verossimilhança descrito na Seção 2.3.

3.1.1.2 Modelo univariado assimétrico

O modelo de volatilidade estocástica univariado assimétrico (3.4) pode ser reescrito na forma de espaço de estados linear através da transformação logarítmica ou através do método baseado nas primeiras derivadas, visto na Seção 2.5. A seguir, serão mostrados esses métodos.

Transformação logarítmica

Harvey e Shephard (1996) mostraram que o MVE-A pode ser escrito na forma de espaço de estados linear através de

$$\begin{aligned} y_t^* &= -1,27 + \ln(\beta^2) + h_t + \xi_t, \\ h_{t+1} &= As_t + \phi h_t + w_t, \\ \begin{pmatrix} \xi_t \\ w_t \end{pmatrix} &\sim \left(\mathbf{0}, \begin{bmatrix} \pi^2/2 & Bs_t \\ Bs_t & \sigma_\eta^2 - A^2 \end{bmatrix} \right) \end{aligned} \tag{3.13}$$

ou então através da forma que considera os erros não-correlacionados:

$$\begin{aligned} y_t^* &= -1,27 + \ln(\beta^2) + h_t + \xi_t, \\ h_{t+1} &= s_t \left(A + \frac{B}{\pi^2/2}(y_t^* - \ln(\beta^2) + 1,27) \right) + \left(\phi - \frac{Bs_t}{\pi^2/2} \right) h_t + \eta_t^*, \\ \begin{pmatrix} \xi_t \\ \eta_t^* \end{pmatrix} &\overset{iid}{\sim} \left(\mathbf{0}, \begin{bmatrix} \pi^2/2 & 0 \\ 0 & \sigma_\eta^2 - A^2 - \frac{B^2}{\pi^2/2} \end{bmatrix} \right), \end{aligned} \tag{3.14}$$

onde s_t é igual a 1 (-1) quando y_t for positivo (negativo), $A = 0,7979\rho\sigma_\eta$, $B = 1,1061\rho\sigma_\eta$ e a equação de observação não segue uma distribuição gaussiana. A função de log-verossimilhança é calculada através do Preditor de Kalman, como vimos na Seção 2.3.

Modelo aproximado baseado nas primeiras derivadas

O modelo gaussiano aproximado para o MVE-A (3.4) é obtido através do chute inicial $g_t = 0$ e aplicando-se o Preditor e Suavizador de Kalman considerando

$$\begin{aligned} x_t &= h_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, A_t), \\ h_{t+1} &= s_t \left(A + \frac{B}{\pi^2/2}(y_t^* - \ln(\beta^2) + 1,27) \right) + \\ &\quad + \left(\phi - \frac{B s_t}{\pi^2/2} \right) h_t + \eta_t^*, \qquad \eta_t^* \overset{iid}{\sim} N\left(0, \sigma_\eta^2 - A^2 - \frac{B^2}{\pi^2/2} \right), \end{aligned} \tag{3.15}$$

x_t e A_t são dados como na Seção 3.1, s_t é igual a 1 (-1) quando y_t for positivo (negativo), $A = 0,7979\rho\sigma_\eta$ e $B = 1,1061\rho\sigma_\eta$, onde teremos novos valores para g_t. Esses novos valores serão usados para obter novos A_t e x_t. Esse processo recursivo é repetido até a convergência de g_t, onde teremos finalmente $\hat{h}_t$, $t = 1, \ldots, T$. Os parâmetros do modelo aproximado podem ser estimados através do método de máxima verossimilhança descrito na Seção 2.3.

3.2 Modelos multivariados

Na literatura, existem várias formulações do modelo de volatilidade estocástica multivariado, como a formulação proposta por Harvey et al (1994) e Jacquier et al (1999). Nesta seção, vamos apresentar duas formulações que serão utilizadas na análise de influência desta dissertação. Essas formulações foram apresentadas no trabalho de Yu e Meyer (2006) para o caso bivariado.

3.2.1 Modelo com correlações constantes generalizado

Considere o modelo de volatilidade estocástica com correlações constantes generalizado (MVE-CCG) proposto por Harvey et al (1994):

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{B}\,\boldsymbol{\Omega}_t\,\boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon) \\ \boldsymbol{h}_t &= \boldsymbol{\Phi}\boldsymbol{h}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \end{aligned} \tag{3.16}$$

sendo $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$, onde y_{it} é o retorno observado no tempo t da série i, $t = 1, \ldots, T$ e $i = 1, \ldots, N$; $\boldsymbol{\Omega}_t = \text{diag}(\exp(\boldsymbol{h}_t/2))$, $\boldsymbol{h}_t = (h_{1t}, \ldots, h_{Nt})'$, onde h_{it} é a volatilidade no tempo t da série i; $\boldsymbol{B} = \text{diag}(\beta_1, \ldots, \beta_N)$, $\boldsymbol{\Phi} = (\phi_{11}, \ldots, \phi_{NN})'$; $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ e $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{Nt})'$ são vetores normais multivariados com média zero e matriz de covariância $\boldsymbol{\Sigma}_\varepsilon$ e $\boldsymbol{\Sigma}_\eta$, respectivamente, onde

$$\boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} 1 & \rho_{\varepsilon 12} & \ldots & \rho_{\varepsilon 1N} \\ \rho_{\varepsilon 12} & 1 & \ldots & \rho_{\varepsilon 2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\varepsilon 1N} & \rho_{\varepsilon 2N} & \ldots & 1 \end{pmatrix} \text{ e } \boldsymbol{\Sigma}_\eta = \begin{pmatrix} \sigma_1^2 & \rho_{\eta 12}\,\sigma_1\sigma_2 & \ldots & \rho_{\eta 1N}\,\sigma_1\sigma_N \\ \rho_{\eta 12}\,\sigma_1\sigma_2 & \sigma_2^2 & \ldots & \rho_{\eta 2N}\,\sigma_2\sigma_N \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\eta 1N}\,\sigma_1\sigma_N & \rho_{\eta 2N}\,\sigma_2\sigma_N & \ldots & \sigma_N^2 \end{pmatrix}.$$

Este modelo considera que os retornos podem estar correlacionados entre eles, assim como as volatilidades, e que os coeficientes de correlação $\rho_{\varepsilon ij}$ e $\rho_{\eta ij}$ são constantes ao longo do tempo.

Para ilustrar, o modelo bivariado é dado por

$$\boldsymbol{y}_t = \begin{pmatrix} \beta_1\, e^{h_{1t}} & 0 \\ 0 & \beta_2\, e^{h_{2t}} \end{pmatrix} \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N\left(\mathbf{0}, \begin{bmatrix} 1 & \rho_\varepsilon \\ \rho_\varepsilon & 1 \end{bmatrix}\right)$$

$$\boldsymbol{h}_t = \begin{pmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{pmatrix} \boldsymbol{h}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \overset{iid}{\sim} N\left(\mathbf{0}, \begin{bmatrix} \sigma_1^2 & \rho_\eta\,\sigma_1\sigma_2 \\ \rho_\eta\,\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right).$$

Aqui, pode-se observar que $corr(\varepsilon_{1t}, \varepsilon_{2t} | \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots) = \rho_\varepsilon$ e $corr(\eta_{1t}, \eta_{2t}) = \rho_\eta$ para $t = 1, \ldots, T$. Quando os elementos fora da diagonal principal de $\boldsymbol{\Sigma}_\varepsilon$ e $\boldsymbol{\Sigma}_\eta$ forem todos iguais a zero, cada elemento do vetor $\boldsymbol{y}_t$ segue um modelo de volatilidade estocástica univariado (3.2) - (3.3).

3.2.2 Modelo fatorial aditivo

Para resolver o problema computacional referente à estimação de um grande número de parâmetros em alguns modelos multivariados e também para capturar as características comuns entre retornos e volatilidades, alguns modelos de volatilidade estocástica fatoriais multivariados foram propostos na literatura, como o modelo de volatilidade estocástica fatorial aditivo (MVE-FA) apresentado por Jacquier et al (1999):

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{D} f_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)) \\ f_t &= \beta \exp(h_t/2) u_t, \qquad u_t \overset{iid}{\sim} N(0,1) \\ h_t &= \phi h_{t-1} + \sigma_\eta \eta_t, \qquad \eta_t \overset{iid}{\sim} N(0,1), \end{aligned} \tag{3.17}$$

sendo $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$, onde y_{it} é o retorno observado no tempo t da série i, $\boldsymbol{D} = (1, d_1, \ldots, d_{N-1})'$ e $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$.

Pode-se notar que a volatilidade estocástica está presente no fator comum, f_t. Este modelo assume que o fator comum segue um modelo de volatilidade estocástica univariado.

No caso bivariado, o modelo é dado por

$$\begin{aligned} \boldsymbol{y}_t &= \begin{pmatrix} f_t \\ df_t \end{pmatrix} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N\left(\mathbf{0}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right) \\ f_t &= \beta \exp(h_t/2)\, u_t, \qquad u_t \overset{iid}{\sim} N(0,1) \\ h_t &= \phi h_{t-1} + \sigma_\eta \eta_t, \qquad \eta_t \overset{iid}{\sim} N(0,1), \end{aligned} \tag{3.18}$$

onde o coeficiente de correlação entre y_{1t} e y_{2t} é dado por

$$\frac{d}{\sqrt{(1 + \sigma_1^2/\beta^2 \exp(-h_t))(d^2 + \sigma_2^2/\beta^2 \exp(-h_t))}}, \tag{3.19}$$

ou seja, o coeficiente de correlação entre os retornos observados é variante no tempo e depende da volatilidade h_t do fator comum.

3.2.3 Estimação

Assim como no caso univariado visto na Seção 3.1, os modelos de volatilidade estocástica multivariados podem ser estimados quando reescritos na forma de espaço de

estados linear, dada por (2.1), e usando-se o procedimento de quase-máxima verossimilhança. A seguir, serão discutidos o modelo com correlações constantes generalizado e o modelo fatorial aditivo.

3.2.3.1 Modelo com correlações constantes generalizado

O modelo com correlação constante generalizado (3.16) pode ser reescrito na forma de espaço de estados linear através da transformação logarítmica ou através do método baseado nas primeiras derivadas. A seguir, serão mostrados esses dois métodos.

Transformação logarítmica

Aplicando a transformação logarítmica no modelo (3.16), obtemos

$$\begin{aligned} \boldsymbol{y}_t^* &= -1,27\ \mathbb{1} + \ln(\boldsymbol{B}^2) + \boldsymbol{h}_t + \boldsymbol{\xi}_t, \qquad \boldsymbol{\xi}_t \overset{iid}{\sim} (\boldsymbol{0}, \boldsymbol{\Sigma}_\xi) \\ \boldsymbol{h}_t &= \boldsymbol{\Phi}\boldsymbol{h}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta), \end{aligned} \tag{3.20}$$

onde $\boldsymbol{y}_t^*$ e $\boldsymbol{\xi}_t$ são vetores $N \times 1$ com elementos $y_{it}^* = \ln(y_{it}^2)$ e $\xi_{it} = \ln(\varepsilon_{it}^2) + 1,27$, respectivamente, $\mathbb{1} = (1, \ldots, 1)'$ e $\ln(\boldsymbol{B}^2) = \text{diag}(\ln(\beta_1^2), \ldots, \ln(\beta_N^2))$. Harvey et al (1994) mostraram que o (i,j)-ésimo elemento de $\boldsymbol{\Sigma}_\xi$ é dado por $(\pi^2/2)\rho_{ij}^*$, onde $\rho_{ii}^* = 1$ e

$$\rho_{ij}^* = \frac{2}{\pi^2}\sum_{n=1}^{\infty} \frac{(n-1)!}{\prod_{k=1}^{n}(1/2+k-1)n}\rho_{\varepsilon ij}^{2n}, \quad i \neq j, \tag{3.21}$$

onde o sinal de $\rho_{\varepsilon ij}$ (correlação entre ε_{it} e ε_{jt}), $i \neq j$, é estimado como sendo positivo quando mais da metade dos pares $y_{it}\, y_{jt}$ for positivo, uma vez que o sinal de $\varepsilon_{it}\, \varepsilon_{jt}$ é o mesmo que o sinal do par $y_{it}\, y_{jt}$ correspondente. O método utilizado para obter a estimativa de quase-máxima verossimilhança dos parâmetros $\boldsymbol{B}$, $\boldsymbol{\Phi}$ e $\boldsymbol{\Sigma}_\eta$ segue como na Seção 2.3.

Modelo aproximado baseado nas primeiras derivadas

O modelo gaussiano aproximado segue como na Seção 2.5, pelo método proposto por Jungbacker e Koopman (2005). Vamos mostrar esse método apenas para o caso bivariado, sendo que o caso multivariado segue de maneira análoga. No MVE-CCG (3.16),

considerando o caso bivariado, a log-densidade das observações é dada por

$$\begin{aligned}\ln p(\boldsymbol{y}_t|\boldsymbol{h}_t) &= -\ln(2\pi) - \frac{1}{2}\left|(\boldsymbol{B}\boldsymbol{\Omega}_t)'\boldsymbol{\Sigma}_\varepsilon(\boldsymbol{B}\boldsymbol{\Omega}_t)\right| - \frac{1}{2}\boldsymbol{y}_t'\left((\boldsymbol{B}\boldsymbol{\Omega}_t)'\boldsymbol{\Sigma}_\varepsilon(\boldsymbol{B}\boldsymbol{\Omega}_t)\right)^{-1}\boldsymbol{y}_t \\ &= -\ln(2\pi) - \frac{1}{2}(h_{1t}+h_{2t}) - \frac{1}{2}\ln((1-\rho^2)(\beta_1^2\,\beta_2^2)) - \frac{1}{2}\boldsymbol{y}_t'\left((\boldsymbol{B}\boldsymbol{\Omega}_t)'\boldsymbol{\Sigma}_\varepsilon(\boldsymbol{B}\boldsymbol{\Omega}_t)\right)^{-1}\boldsymbol{y}_t,\end{aligned} \tag{3.22}$$

onde $\rho = \rho_{12} = \rho_{21}$ e

$$\boldsymbol{y}_t'\left(\boldsymbol{\Omega}_t'\boldsymbol{\Sigma}_\varepsilon\boldsymbol{\Omega}_t\right)^{-1}\boldsymbol{y}_t = \left(y_{1t}^2\,\frac{e^{-h_{1t}}}{\beta_1^2} + y_{2t}^2\,\frac{e^{-h_{2t}}}{\beta_2^2} - 2\,\rho\,y_{1t}\,y_{2t}\,\frac{e^{-(h_{1t}+h_{2t})/2}}{\beta_1\beta_2}\right)\;(1-\rho^2)^{-1}.$$

Dessa forma, segue que

$$\dot{p}(\boldsymbol{y}_t|\boldsymbol{h}_t) = \frac{1}{2(1-\rho^2)}\begin{pmatrix} -(1-\rho^2) + y_{1t}^2\dfrac{e^{-h_{1t}}}{\beta_1^2} - \rho\,y_{1t}\,y_{2t}\dfrac{e^{-(h_{1t}+h_{2t})/2}}{\beta_1\beta_2} \\ -(1-\rho^2) + y_{2t}^2\dfrac{e^{-h_{2t}}}{\beta_2^2} - \rho\,y_{1t}\,y_{2t}\dfrac{e^{-(h_{1t}+h_{2t})/2}}{\beta_1\beta_2}\end{pmatrix} \tag{3.23}$$

e

$$\ddot{p}(\boldsymbol{y}_t|\boldsymbol{h}_t) = \frac{e^{-(h_{1t}+h_{2t})/2}}{4(1-\rho^2)}\begin{pmatrix} -2\,y_{1t}^2\dfrac{e^{(-h_{1t}+h_{2t})/2}}{\beta_1^2} + \dfrac{\rho\,y_{1t}\,y_{2t}}{\beta_1\beta_2} & \dfrac{\rho\,y_{1t}\,y_{2t}}{\beta_1\beta_2} \\ \dfrac{\rho\,y_{1t}\,y_{2t}}{\beta_1\beta_2} & -2\,y_{2t}^2\dfrac{e^{(h_{1t}-h_{2t})/2}}{\beta_2^2} + \dfrac{\rho\,y_{1t}\,y_{2t}}{\beta_1\beta_2}\end{pmatrix}. \tag{3.24}$$

Assim, obtemos $\boldsymbol{A}_t = -\ddot{p}(\boldsymbol{y}_t|\boldsymbol{h}_t)|_{h=g}^{-1}$ e $\boldsymbol{x}_t = \boldsymbol{g}_t + \boldsymbol{A}_t\,\dot{p}(\boldsymbol{y}|\boldsymbol{h})|_{h=g}$.

O modelo gaussiano aproximado é então obtido através do chute inicial $\boldsymbol{g}_t = \boldsymbol{0}$ e aplicando-se o Preditor e Suavizador de Kalman considerando

$$\begin{aligned}&\boldsymbol{x}_t = \boldsymbol{h}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{A}_t), \\ &\boldsymbol{h}_t = \boldsymbol{d}_t + \boldsymbol{T}_t\boldsymbol{h}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t),\end{aligned} \tag{3.25}$$

onde teremos novos valores para $\boldsymbol{g}_t$, que serão usados para obter novos $\boldsymbol{A}_t$ e $\boldsymbol{x}_t$. Esse processo recursivo é repetido até a convergência de $\boldsymbol{g}_t$, onde teremos $\hat{\boldsymbol{h}}_t$, $t = 1, \ldots, T$. Os parâmetros do modelo aproximado podem ser estimados através do método de máxima verossimilhança descrito na Seção 2.3.

3.2.3.2 Modelo fatorial aditivo

O modelo fatorial aditivo pode ser reescrito na forma de espaço de estados linear através do método baseado nas primeiras derivadas, como veremos a seguir.

Modelo aproximado baseado nas primeiras derivadas

Vamos mostrar esse método apenas para o caso bivariado, sendo que o caso multivariado segue de maneira análoga. No modelo de volatilidade estocástica multivariado (3.17), considerando o caso bivariado, a log-densidade das observações é dada por

$$\ln p(y_t|h_t) = -\ln(2\pi) - \frac{1}{2}\ln(\beta e^{h_t}\sigma_2^2 + d^2\beta e^{h_t}\sigma_1^2 + \sigma_1^2\sigma_2^2) - \frac{1}{2}\boldsymbol{y}_t' \begin{pmatrix} \beta e^{h_t} + \sigma_1^2 & d\beta e^{h_t} \\ d\beta e^{h_t} & d^2\beta e^{h_t} + \sigma_2^2 \end{pmatrix}^{-1} \boldsymbol{y}_t, \tag{3.26}$$

Assim,

$$\dot{p}(\boldsymbol{y}_t|h_t) = \frac{\sigma_1^2\sigma_2^2(-d^2\sigma_1^2 - 2d^2\beta e^{h_t} + 2y_1y_2d - \sigma_2^2) - d^4\beta e^{h_t}\sigma_1^2 + y_2^2d^2\sigma_1^4 - \beta e^{h_t}\sigma_2^4 + y_1^2\sigma_2^4}{\beta e^{-h_t}(\beta e^{h_t}\sigma_2^2 + d^2\beta e^{h_t}\sigma_1^2 + \sigma_1^2\sigma_2^2)^2} \tag{3.27}$$

e

$$\begin{aligned} \ddot{p}(\boldsymbol{y}_t|\theta_t) = & -\frac{2d^2\beta e^{h_t}\sigma_1^4\sigma_2^4 + d^4\beta e^{h_t}\sigma_1^6\sigma_2^2 + 2y_1y_2d^2\beta e^{h_t}\sigma_1^2\sigma_2^4 + 2y_1y_2d^3\beta e^{h_t}\sigma_1^4\sigma_2^2}{2e^{-h_t}(e^{h_t}\sigma_2^2 + d^2\beta e^{h_t}\sigma_1^2 + \sigma_1^2\sigma_2^2)^3} - \\ & -\frac{2y_1y_2d\sigma_1^4\sigma_2^4 + y_1^2d^2\beta e^{h_t}\sigma_1^2\sigma_2^4 + y_2^2d^2\beta e^{h_t}\sigma_1^4\sigma_2^2 + y_2^2d^4\beta e^{h_t}\sigma_1^6}{2e^{-h_t}(e^{h_t}\sigma_2^2 + d^2\beta e^{h_t}\sigma_1^2 + \sigma_1^2\sigma_2^2)^3} - \\ & -\frac{y_2^2d^2\sigma_1^6\sigma_2^2 + \beta e^{h_t}\sigma_1^2\sigma_2^6 + d^2\sigma_1^6\sigma_2^4 + y_1^2\beta e^{h_t}\sigma_2^6 - y_1^2\sigma_1^2\sigma_2^6 + \sigma_1^4\sigma_2^6}{2e^{-h_t}(e^{h_t}\sigma_2^2 + d^2\beta e^{h_t}\sigma_1^2 + \sigma_1^2\sigma_2^2)^3}. \end{aligned} \tag{3.28}$$

Assim, obtemos $A_t = -\ddot{p}(\boldsymbol{y}_t|h_t)|_{h=g}^{-1}$ e $x_t = g_t + A_t\,\dot{p}(\boldsymbol{y}|h)|_{h=g}$.

O modelo gaussiano aproximado é então obtido através do chute inicial $g_t = 0$ e aplicando-se o Preditor e Suavizador de Kalman considerando

$$\begin{aligned} x_t &= h_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, A_t), \\ h_t &= d_t + T_t h_{t-1} + \eta_t, \qquad \eta_t \sim N(0, Q_t), \end{aligned} \tag{3.29}$$

onde teremos novos valores para g_t, que serão usados para obter novos A_t e x_t. Esse processo recursivo é repetido até a convergência de g_t, onde teremos $\hat{h}_t$, $t = 1, \ldots, T$. Os parâmetros do modelo aproximado podem ser estimados através do método de máxima verossimilhança descrito na Seção 2.3.

Capítulo 4

Estatística de Influência

O diagnóstico de modelos é uma das etapas fundamentais para avaliar a qualidade do ajuste do modelo para um conjunto de dados. Uma das ferramentas de diagnóstico é a *análise de influência*. Em particular, uma observação é dita influente se ela produzir alterações relevantes no resultado da análise quando a mesma for excluída, como proposto por Cook (1977), ou submetida a uma pequena perturbação, como na análise de influência local proposta por Cook (1986).

Na literatura, a estatística da distância de Cook (Cook, 1977), uma estatística comumente usada na análise de influência, se baseia na exclusão de observações do conjunto de dados e na posterior investigação das mudanças ocorridas nas estimativas dos parâmetros. Também, Cook (1986) introduziu o primeiro trabalho com respeito a influência local, onde propõe um método de diagnóstico a partir de uma medida chamada afastamento da verossimilhança, aproveitando assim, ideias elementares da geometria diferencial.

Como alternativa para a medida da distância de Cook ou a Influência Local, Peña (2005) introduziu uma nova forma de analisar influência em modelos de regressão linear. Ao invés de verificar como a exclusão de uma observação afeta as estimativas dos parâmetros e as predições como no caso da estatística da distância de Cook, é verificado como uma observação é influenciada pelas outras observações na amostra. Ou seja, é avaliada a mudança na predição de uma determinada observação quando cada observação da amostra é excluída. Também, através dessa análise de influência, pode-se encontrar blocos de valores influentes similares, que são mais difíceis de serem detectados pela medida da distância de Cook. A forma de análise de influência proposta por Peña (2005) é

simples de ser implementada, além de ter uma interpretação intuitiva, e pode ser usada como uma alternativa para a medida da distância de Cook.

Em séries temporais, o trabalho pioneiro que utiliza a metodologia de influência proposta por Peña (2005) como ferramenta de diagnóstico para detectar observações influentes em modelos de volatilidade estocástica univariados simétricos foi desenvolvido por Motta e Hotta (2007). Neste capítulo, esse estudo será extendido para os modelos de volatilidade estocástica univariados assimétricos. Como contribuição adicional, será avaliado o uso de estimadores filtrados da volatilidade assim como a relação com os resíduos de predição e resíduos auxiliares dos estados.

Inicialmente, neste capítulo será apresentada a abordagem de Cook (1977) e a proposta de Peña (2005), e serão ilustrados dois exemplos com dados simulados e também dados reais. Em seguida, será apresentada a estatística para diagnosticar influência em modelos de volatilidade estocástica univariados proposta por Motta e Hotta (2007), e serão mostradas duas aplicações em modelos de volatilidade estocástica assimétricos através de estimadores filtrados e suavizados da volatilidade, utilizando dados financeiros. Também, será apresentada uma metodologia baseada em marcas de referência para determinar quais observações são estatisticamente influentes.

4.1 Influência em modelos de regressão linear

Nesta seção, vamos apresentar a estatística proposta por Peña (2005) para a análise de influência em modelos de regressão linear e compará-la à abordagem de Cook (1977), uma vez que a medida da distância de Cook é uma das medidas de influência mais simples e mais utilizadas em aplicações.

Considere o modelo de regressão linear múltipla

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i, \qquad u_i \overset{iid}{\sim} N(0, \sigma^2), \tag{4.1}$$

$i = 1, \ldots, N$, onde y_i é a variável resposta, $\boldsymbol{x}_i = (1, x_{2i}, \ldots, x_{pi})'$ é o vetor de variáveis explanatórias e $\boldsymbol{\beta}$ é o vetor $N \times 1$ de parâmetros desconhecidos.

Definimos por $\boldsymbol{X}$ a matriz $N \times p$ de posto p onde a i-ésima linha é $\boldsymbol{x}_i'$. Denotamos por $\hat{\boldsymbol{\beta}}$ o estimador de mínimos quadrados $\boldsymbol{\beta}$ dada por $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$; por $\hat{\boldsymbol{y}} =$

$(\hat{y}_1, \ldots, \hat{y}_N)' = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ o vetor dos preditores; por $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ o vetor dos resíduos; e por $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ a matriz de projeção.

Seja $\hat{\boldsymbol{\beta}}_{(i)}$ o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ quando a observação i é excluída, e $\hat{\boldsymbol{y}}_{(i)} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{(i)}$ o vetor dos preditores correspondente. Cook (1977) propôs a seguinte estatística, conhecida como estatística da distância de Cook, para medir a influência de uma determinada observação:

$$D_i = \frac{\|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{(i)}\|^2}{ps^2} = \frac{\sum_{j=1}^{N}(\hat{y}_j - \hat{y}_{j(i)})^2}{ps^2}, \tag{4.2}$$

onde $s^2 = \boldsymbol{e}'\boldsymbol{e}/(N-p)$.

A estatística da distância de Cook é uma medida do efeito da retirada de um determinada observação i. Esta medida compara os preditores obtidos com todas as observações e os preditores obtidos sem a i-ésima observação no ajuste do modelo de regressão. Valores altos de D_i indicam observações atípicas. Como a medida da distância de Cook tem distribuição aproximadamente $F_{(p;\,N-p)}$, um critério para a comparação da magnitude dos valores de D_i mencionado por Kutner et al. (2004) consiste em compará-los com $F_{(p;\,N-p;\,0{,}2)}$, o percentil 20% da distribuição F com p graus de liberdade no numerador e $N-p$ no denominador. Se $D_i > F_{(p;\,N-p;\,0{,}2)}$, então a observação i pode ser considerada influente no modelo.

Ao invés de analisar o efeito global no vetor dos preditores devido a exclusão de uma observação, Peña (2005) propôs analisar como a exclusão de cada observação afeta a predição de uma observação em particular. Desta maneira, é verificado como cada observação está sendo influenciada pelas outras. No modelo de regressão, isto pode ser feito considerando os vetores

$$\boldsymbol{s}_i = (\hat{y}_i - \hat{y}_{i(1)}, \ldots, \hat{y}_i - \hat{y}_{i(N)}), \tag{4.3}$$

onde o j-ésimo termo deste vetor, $j = 1, \ldots, N$, mostra a diferença entre a predição de y_i com todas as observações e a predição de y_i sem a observação j. Sendo assim, Peña (2005) analisa o quão sensível é a predição da i-ésima observação quando se exclui cada observação da amostra.

Essa nova estatística é definida como o quadrado da norma do vetor $\boldsymbol{s}_i$ padronizado, ou seja,

$$S_i = \frac{\boldsymbol{s}_i'\boldsymbol{s}_i}{p\,\widehat{\mathrm{var}}(\hat{y}_i)} = \frac{\sum_{j=1}^{N}(\hat{y}_i - \hat{y}_{i(j)})^2}{ps^2 h_{ii}}, \tag{4.4}$$

onde h_{ii} corresponde ao i-ésimo elemento da diagonal de $\boldsymbol{H}$. Peña (2005) mostrou que (4.4) pode ser escrita como

$$S_i = \frac{1}{ps^2 h_{ii}} \sum_{j=1}^{n} \frac{h_{ji}^2 e_j^2}{(1-h_{jj})^2}, \tag{4.5}$$

onde e_j corresponde ao j-ésimo elemento do vetor de resíduos $\boldsymbol{e}$. A equação (4.5) exige menor tempo computacional que (4.4), uma vez que, para o cálculo de (4.4), é necessário recalcular os preditores cada vez que se retira uma determinada observação (o modelo de regressão é reajustado N vezes), enquanto que a equação (4.5) exige apenas um único ajuste do modelo.

A estatística S_i indica a influência da i-ésima observação. Se o valor de S_i for suficientemente "grande" ou "pequeno", diz-se que a i-ésima observação do específico conjunto de dados é influente. Ou seja, a i-ésima observação será influente se a medida S_i for notadamente "maior" ou "menor" do que as demais S_j, $i \neq j$, ou do que alguma referência específica, como será discutido na Seção 4.2. Por esta razão, as estatísticas S_i devem estar todas numa mesma ordem de grandeza, o que justifica a padronização dos vetores $\boldsymbol{s}_i$ na equação (4.4). Essa padronização consiste em uma mudança na escala do vetor $\boldsymbol{s}_i$ baseada no desvio padrão de $\hat{y}_i$. A padronização evita que um valor predito de y_i com variância muito alta faça diferença no resultado de S_i.

Peña (2005) mostrou também que a estatística S_i é uma combinação linear da medida da distância de Cook, e apresentou três propriedades dessa estatística proposta. A primeira delas é que, quando não há observações atípicas no conjunto de dados, o valor esperado da estatística é aproximadamente igual a $1/p$ no modelo de regressão. A segunda propriedade é que a distribuição da estatística é aproximadamente normal, enquanto que a medida da distância de Cook possui uma distribuição assintótica mais complicada. Por último, ele provou que, quando a amostra está contaminada por um grupo de valores atípicos com alta ou média alavancagem, a estatística S_i discrimina essas observações das demais.

4.1.1 Metodologia para diagnóstico de influência

Após calculadas as medidas da estatística (4.4), deve-se avaliar se estas são "grandes" ou "pequenas" para caracterizá-las como influentes, através de alguma referência es-

pecífica. Aqui, serão discutidos os procedimentos para o cálculo de *marcas de referência*, construídas a partir de simulações de Monte Carlo, e o consequente diagnóstico de influência.

Com o objetivo de construir níveis de referência para identificar as observações significativamente influentes, Peña (2005) propôs considerar os valores

$$m(S) \pm 4,5\, m(|S - m(S)|) \tag{4.6}$$

como *marcas de referência*, onde $m(S)$ é a mediana dos valores de S_i. Sendo assim, se a estatística S_i não satisfaz a condição

$$m(S) - 4,5\, m(|S - m(S)|) \;\leq\; S_i \;\leq\; m(S) + 4,5\, m(|S - m(S)|), \tag{4.7}$$

então a i-ésima observação é dita influente. As marcas de referência podem ser chamadas também de *valores de corte* ou *limiares*. Os *limiares de Peña* (4.6) foram obtidos baseando-se na distribuição normal da estatística S_i. Porém, ao extendermos a metodologia para outros modelos, a distribuição da estatística torna-se desconhecida, e os resultados (4.6) podem não ser adequados.

O trabalho de Motta e Hotta (2007) propõe o uso dos limiares construídos através de simulações de Monte Carlo como uma alternativa para investigar a significância das estatísticas de influência. Nesta seção, essa ideia é adaptada para os modelos de regressão linear. Assim, através de parâmetros pré-determinados, para a k-ésima replicação ($k = 1, \ldots, K$) é gerada uma amostra $y_1^{(k)}, \ldots, y_N^{(k)}$, da qual calculam-se as estatísticas $S_{max}^{(k)} = \max\limits_i S_i$ e $S_{min}^{(k)} = \min\limits_i S_i$. Seja M_1 o percentil 95% de $\{S_{max}^{(1)}, \ldots, S_{max}^{(K)}\}$ e M_2 o percentil 5% de $\{S_{min}^{(1)}, \ldots, S_{min}^{(K)}\}$. M_1 e M_2 são marcas de referência ou limiares. Se uma estatística S_i for maior que M_1 ou menor que M_2, então a observação i é individualmente influente.

Em resumo, considerando K replicações, os *limiares obtidos por simulação* são calculados da seguinte maneira:

1. Estime $\boldsymbol{\beta}$ considerando as observações $y_1, \ldots, y_N$.

2. Para $k = 1, \ldots, K$:

   (i) Simule $y_1^{(k)}, \ldots, y_N^{(k)}$ considerando $\hat{\boldsymbol{\beta}}$ e $\boldsymbol{X}$.

   (ii) Calcule a estatística (4.4) para $i = 1, \ldots, N$ a partir dos dados obtidos em (i).

(iii) Calcule as estatísticas $S_{max}^{(k)} = \max_i \{S_1^{(k)}, \ldots, S_N^{(k)}\}$ e $S_{min}^{(k)} = \min_i \{S_1^{(k)}, \ldots, S_N^{(k)}\}$ a partir das estatísticas obtidas em (ii).

3. Calcule M_1, o percentil 95% de $\{S_{max}^{(1)}, \ldots, S_{max}^{(K)}\}$, e M_2, o percentil 5% de $\{S_{min}^{(1)}, \ldots, S_{min}^{(K)}\}$.

4.1.2 Aplicações

Agora, vamos ilustrar o uso da estatística de influência e a técnica de limiares descrita na seção anterior através de dois exemplos. O primeiro exemplo trata-se da análise de dados simulados utilizando um modelo de regressão simples, onde foi inserido um bloco de valores atípicos, a fim de comparar os resultados obtidos a partir da metodologia de influência proposta por Peña (2005) e da medida da distância de Cook. O segundo exemplo trata-se de um conjunto de dados reais analisados no trabalho de Peña (2005), e que foram reanalisados aqui a fim de comparar as metodologias de limiares discutidas nesta seção e também para ilustrar o uso da medida da distância de Cook. Este exemplo apresenta uma situação onde há um grupo moderado de valores atípicos, e foi ajustado um modelo de regressão simples. Para o cálculo dos limiares por simulação, foram realizadas $K = 1000$ replicações. Os resultados foram obtidos através de programas escritos no *software R*.

4.1.2.1 Exemplo 1: simulação

Neste exemplo, vamos analisar o comportamento da estatística (4.4) na presença de um bloco de observações com alta alavancagem no conjunto de dados, ajustando um modelo de regressão linear simples. Vamos analisar também o comportamento da medida da distância de Cook (4.2), a fim de verificarmos se ambas as metodologias detectam corretamente os valores atípicos.

Foram simuladas $N = 100$ observações x e y independentes tais que $x_i \sim N(0,1)$ e $y_i \sim N(0,1)$, $i = 1, \ldots, 100$. Porém, foram inseridos os valores atípicos $x_i = 10$ e $y_i = 5$ para $i = 50, \ldots, 55$. Para esses valores simulados, foi ajustado o modelo de regressão linear

$$y_i = \beta_0 + x_i'\beta_1 + u_i, \qquad u_i \overset{iid}{\sim} N(0, \sigma^2). \tag{4.8}$$

Na Tabela 4.1 encontram-se as estimativas de máxima verossimilhança dos parâmetros do modelo (4.8). Segundo os quocientes t, a estimativa de β_0 não é significativa ao nível de 5%, mas a estimativa de β_1 é altamente significativa.

Tabela 4.1: Valores dos parâmetros estimados e seus respectivos erros-padrão e razão-t.

Parâmetro	Estimativa	Erro Padrão	Razão-t
β_0	0,1650	0,1048	1,57
β_1	0,4359	0,0405	10,8
σ	1,0202		

Nas Figuras 4.1 (a) e (b), podemos observar o bloco de valores atípicos das observações x_i e y_i simuladas. Através da Figura 4.1 (c) percebe-se que as observações atípicas localizadas em $x = 10$ e $y = 5$ (que estão sobrepostas) são observações com alta alavancagem, uma vez que elas "controlam" a linha de regressão.

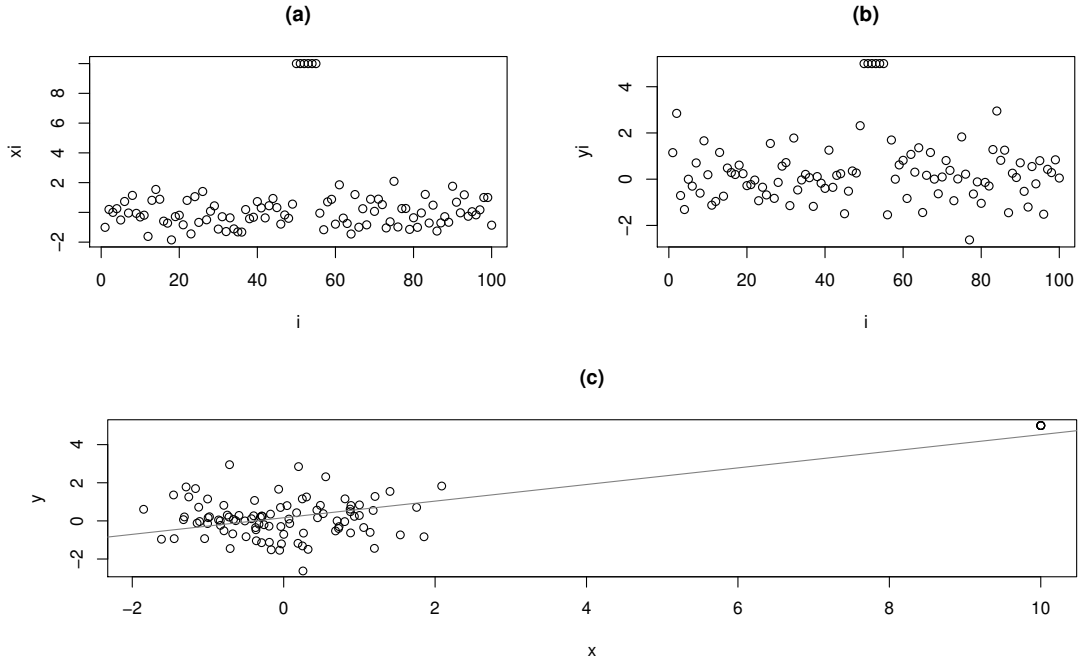


Figura 4.1: Gráficos das observações x_i e y_i, obtidas através de simulação, e a respectiva linha de regressão obtida através do modelo (4.8) ajustado.

A Figura 4.2 (a) apresenta os valores individuais da medida da distância de Cook

D_i, assim como o limiar $F_{(1;\,99;\,0{,}2)}$, e a Figura 4.2 (b) apresenta os valores individuais da estatística S_i, os limiares de Peña (linha contínua) e os limiares obtidos por simulações (linha tracejada). No gráfico da medida da distância de Cook, o bloco de valores atípicos não foi identificado, pois todos os valores de D_i obtidos são menores que $F_{(1;\,99;\,0{,}2)}$. Por outro lado, a estatística S_i indica claramente os dois grupos de dados. A comparação dos valores de S_i com os limiares obtidos por simulações (linha tracejada) confirma o grupo de valores atípicos, pois estes valores são menores que o limiar inferior. Porém, usando os limiares propostos por Peña (linha contínua) detectou-se incorretamente cinco observações (14, 26, 61, 75 e 90), já que estas são observações genuínas no conjunto de dados.

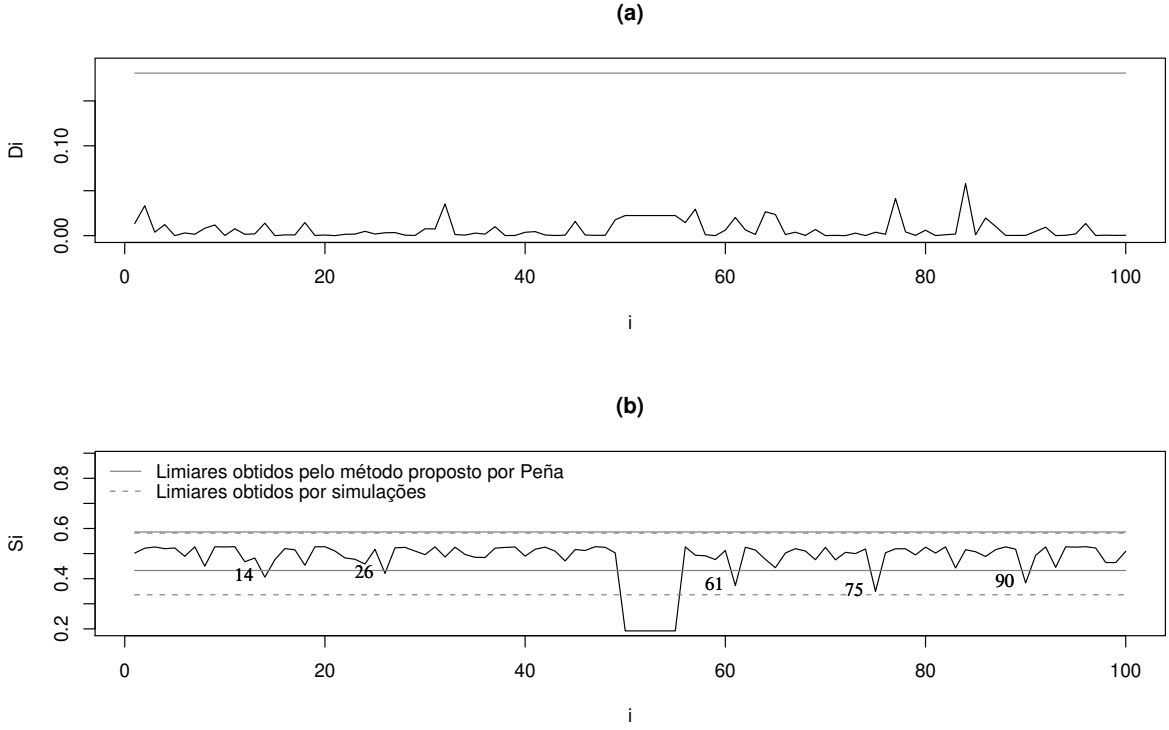


Figura 4.2: (a) Gráfico de D_i e o percentil 20% da distribuição F correspondente, e (b) gráfico de S_i com os limiares de Peña (linha contínua) e os limiares obtidos por simulações (linha tracejada).

Neste exemplo, os valores atípicos que foram inseridos nas variáveis são observações com alta alavancagem e, então, seus resíduos são próximos de zero. Dessa forma, os valores preditos para essas observações são pouco afetados quando se exclui qualquer observação da amostra, levando a valores de S_i mais baixos para essas observações do que para as

demais.

Através deste exemplo, notamos que os limiares obtidos por simulações foram mais eficazes para detectar as observações atípicas quando comparados aos limiares propostos por Peña (2005), pois eles detectaram apenas as observações atípicas, e não as genuínas, como as observações 14, 26, 61, 75 e 90. Também, a estatística proposta por Peña (2005) é eficaz para identificar grupos de observações atípicas similares, que não são detectados pela medida da distância de Cook.

4.1.2.2 Exemplo 2: dados reais

Neste exemplo, foram utilizados os dados do diagrama de Hertzsprung-Rusell, de Rousseeuw e Leroy (1987), apresentados na Figura 4.3. Esses dados correspondem ao grupo de estrelas CYG OB1, o qual consiste de 47 estrelas na direção de Cygnus. A variável x representa o logaritmo da temperatura na superfície da estrela e a variável y corresponde ao logaritmo da intensidade da luz.

Na Figura 4.3, podemos observar que as observações 11, 20, 30 e 34 estão bastante longe da configuração dos dados, os quais tem um comportamento linear entre x e y, e as observações 7 e 14 parecem estar um pouco longe da mesma configuração.

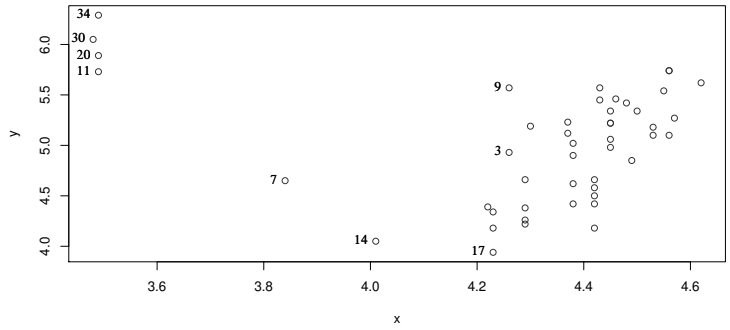


Figura 4.3: Logaritmo da temperatura versus logaritmo da intensidade da luz.

A Figura 4.4 (a) apresenta os valores individuais da medida da distância de Cook D_i, assim como o limiar $F_{(1;46;0,2)}$, e a Figura 4.4 (b) apresenta os valores individuais da estatística S_i, os limiares de Peña (linha contínua) e os limiares obtidos por simulações (linha tracejada). No gráfico da medida da distância de Cook, apenas as observações 30

e 34 foram detectadas como influentes quando comparamos os valores de D_i com o limiar $F_{(1;46;0,2)}$. Adicionalmente, a comparação dos valores de S_i com os limiares obtidos por simulações indica seis observações atípicas, as mesmas que aparecem longe da configuração linear dos dados na Figura 4.3. Porém, a comparação dos valores de S_i com os limiares de Peña indica mais três observações atípicas (3, 9 e 17) que, segundo a Figura 4.3, não parecem estar longe da configuração linear dos dados.

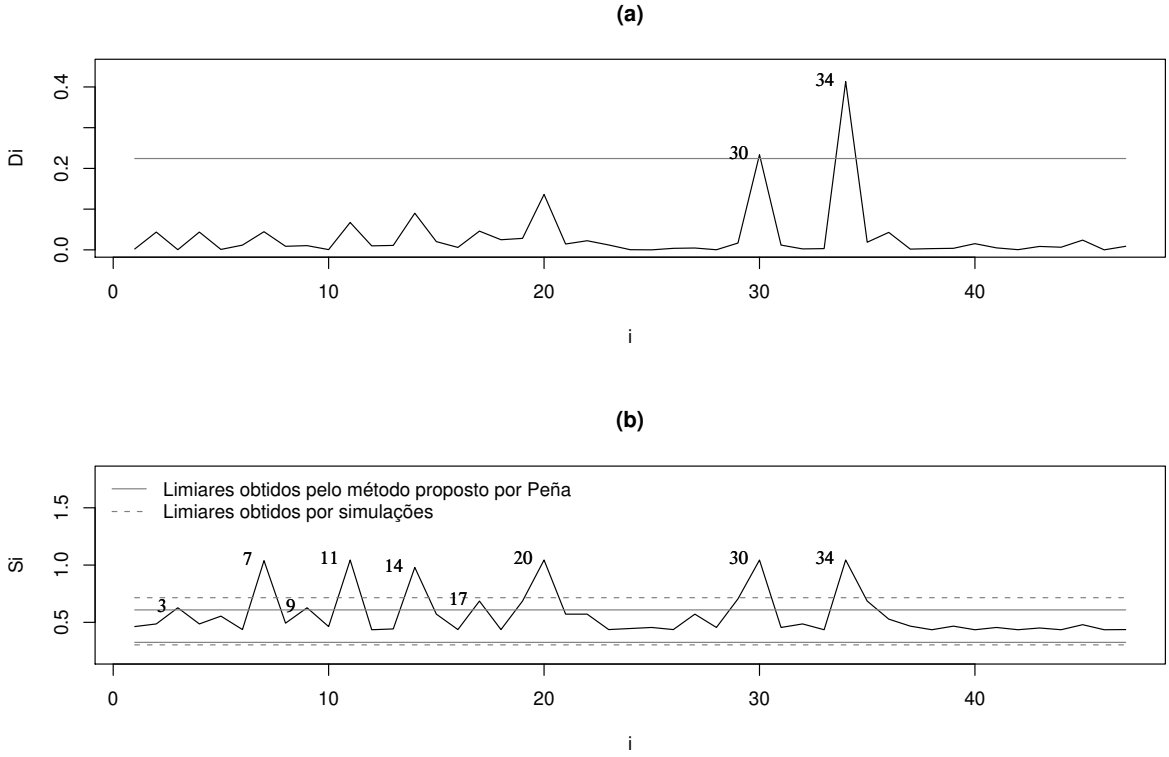


Figura 4.4: (a) Gráfico de D_i e o percentil 20% da distribuição F correspondente, e (b) gráfico de S_i com os limiares de Peña (linha contínua) e os limiares obtidos por simulações (linha tracejada).

Neste exemplo, notamos que os limiares obtidos por simulações foram mais eficazes para detectar as observações atípicas quando comparados aos limiares de Peña, pois detectaram as observações 7, 11, 14, 20, 30 e 37, que realmente estão longe da configuração linear dos dados, como mostra a Figura 4.3. Já os limiares de Peña detectaram também as observações 3, 9 e 17 que, segundo a Figura 4.3, não parecem estar longe da configuração dos dados. Também, a estatística proposta por Peña (2005) foi capaz de identificar observações atípicas que não foram detectadas pela medida da distância de Cook.

4.2 Influência em modelos de volatilidade estocástica univariados

O trabalho de Motta e Hotta (2007) generalizou a medida de influência proposta por Peña (2005) para o modelo de volatilidade estocástica univariado. Essa extensão não é direta, uma vez que as observações de séries temporais são dependentes e, além disso, os valores preditos das observações dos modelos de volatilidade são sempre iguais a zero. Nesta seção, utilizaremos a estatística para diagnosticar influência em modelos de volatilidade estocástica univariados simétricos proposta por Motta e Hotta (2007) nos modelos de volatilidade estocástica univariados assimétricos de Harvey e Shephard (1996), os quais não foram considerados no trabalho de Motta e Hotta (2007).

Para aplicar a estatística de influência definida por Peña (2005) nos modelos de volatilidade estocástica univariados, Motta e Hotta (2007) fizeram algumas modificações. Primeiro, como a sequência e o tempo das observações são importantes nos modelos de séries temporais, não se deve excluir uma observação, mas sim considerá-la uma observação faltante. Além disso, como o valor predito da observação y_t é sempre igual a zero, ele não pode ser usado para definir a estatística. Sendo assim, eles sugeriram usar as estimativas da volatilidade h_t.

Considere o modelo de volatilidade estocástica univariado assimétrico dado por

$$\begin{aligned} x_t &= \beta e^{h_t/2}\varepsilon_t, \\ h_{t+1} &= \phi h_t + \eta_t, \\ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} &\overset{iid}{\sim} N\left(\mathbf{0}, \begin{bmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{bmatrix}\right) \end{aligned} \tag{4.9}$$

$t = 1, \ldots, T$, onde o termo h_t é dito ser a volatilidade de x_t, é não-observado e independe dos valores passados de x_t; η_t e ε_t são serialmente não-correlacionados, $|\rho| < 1$ e $|\phi| < 1$ para garantir que o termo h_t seja estacionário.

Uma vez que o modelo (4.9) está escrito na forma de espaço de estados linear (2.1), denote por $\boldsymbol{y} = (y_1, \cdots, y_T) = (\ln x_1^2, \cdots, \ln x_T^2)$ o vetor de observações do modelo (2.1), por $\hat{\boldsymbol{\Psi}} = (\hat{\beta}, \hat{\phi}, \hat{\sigma}_\eta, \hat{\rho})$ o vetor das estimativas dos parâmetros baseadas em todas as observações e por $\hat{h}_t$ a estimativa de h_t obtida através do Filtro ou do Suavizador de Kalman aplicado no modelo aproximado, conforme foi visto na Seção 2.5. Considerando que $\boldsymbol{y}_{(t)}$

é o vetor de observações considerando a t-ésima observação como faltante, tem-se que a estimativa do conjunto de parâmetros baseada nesse vetor será representada por $\hat{\boldsymbol{\Psi}}_{(t)}$. Dessa forma, considerando o vetor $\boldsymbol{s}_t = (\hat{h}_t - \hat{h}_{t(1)}, \ldots, \hat{h}_t - \hat{h}_{t(T)})'$, a estatística para o modelo de volatilidade estocástica univariado é dada por

$$S_t = \frac{\boldsymbol{s}_t' \boldsymbol{s}_t}{\widehat{var}(\hat{h}_t)}, \tag{4.10}$$

$t = 1, \ldots, T$, onde $\widehat{var}(\hat{h}_t)$ é obtida através do Filtro ou do Suavizador de Kalman aplicado no modelo aproximado, conforme foi visto na Seção 2.5. O j-ésimo termo do vetor $\boldsymbol{s}_t$ mostra a diferença entre a estimação de h_t com o vetor de observações $\boldsymbol{y}$ completo e a estimação de h_t com o vetor $\boldsymbol{y}_{(j)}$, ou seja, considerando a j-ésima observação como faltante. Sendo assim, analisamos o quão sensível é a estimação de h_t quando se exclui cada observação y_j da amostra. A estatística S_t é, então, usada para verificar se a observação y_t é influente no modelo univariado.

O procedimento para encontrar as estatísticas é dado pelo seguinte algoritmo:

1. Estime os parâmetros $\boldsymbol{\Psi}$ do modelo de volatilidade baseado na amostra completa, e encontre $\hat{h}_t$ e $\widehat{var}(\hat{h}_t)$ para $t = 1, \ldots, T$.

2. Para $t = 1, \ldots, T$:

   (i) Considere a estimativa $\hat{h}_t$ encontrada no item (1).

   (ii) Para $j = 1, \ldots, T$:

   - Defina o vetor $\boldsymbol{y}_{(j)}$, considerando a observação y_j como faltante.
   - Reestime o modelo com base em $\boldsymbol{y}_{(j)}$, obtendo $\hat{\boldsymbol{\Psi}}_{(j)}$.
   - Encontre $\hat{h}_{t(j)}$, a estimativa de $\hat{h}_t$ baseada em $\hat{\boldsymbol{\Psi}}_{(j)}$ e em $\boldsymbol{y}_{(j)}$.

   (iii) Calcule a estatística S_t.

4.2.1 Metodologia para diagnóstico de influência

Após calculadas as medidas da estatística (4.10), deve-se avaliar se estas são "grandes" ou "pequenas" para caracterizá-las como influentes. Aqui, serão discutidos os procedimentos para o cálculo de *marcas de referência*, construídas a partir de simulações de Monte Carlo, e o consequente diagnóstico de influência.

O cálculo de níveis de referência é análogo ao método apresentado na Seção 4.1, embora demanda um tempo computacional mais alto quando comparado ao caso dos modelos de regressão, uma vez que é necessário recalcular a estimativa de h_t cada vez que se retira uma determinada observação. Ou seja, para calcular (4.10), é necessário reestimar h_t T vezes, o que exige T ajustes do modelo de volatilidade estocástica.

No contexto de modelos de regressão linear, dizemos que a i-ésima observação de um específico conjunto de dados é influente quando o valor de S_i for suficientemente "grande" ou "pequeno". Peña (2005) mostrou que valores "pequenos" da estatística S_i estão relacionados a observações atípicas com alta alavancagem. Porém, quando se trata do ajuste do modelo de volatilidade estocástica univariado a um conjunto de dados, Motta e Hotta (2007) mostraram que há muitas observações com valores "pequenos" de S_i (geralmente próximos de zero), e que esses valores não estão relacionados a casos especiais, enquanto que valores "grandes" de S_i estão relacionados a observações influentes. Dessa forma, Motta e Hotta (2007) utilizam apenas o percentil 95% de $\{S_{max}^{(1)}, \ldots, S_{max}^{(K)}\}$ como limiar, sendo $S_{max} = \max_t S_t$ e K o número de replicações.

Assim, através de parâmetros pré-determinados, para a k-ésima replicação ($k = 1, \ldots, K$) é gerada uma série de dados $y_1^{(k)}, \ldots, y_T^{(k)}$, da qual calcula-se a estatística $S_{max}^{(k)} = \max_i S_t$. Seja M_1 o percentil 95% de $\{S_{max}^{(1)}, \ldots, S_{max}^{(K)}\}$. M_1 representa o limiar. Se a estatística S_t for maior que M_1, então a observação t é considerada influente.

Em resumo, considerando K replicações, o *limiar obtido por simulação* é calculado da seguinte maneira:

1. Encontre $\hat{\boldsymbol{\Psi}}$ considerando a série $y_1, \ldots, y_T$.

2. Para $k = 1, \ldots, K$:

   (i) Simule $y_1^{(k)}, \ldots, y_T^{(k)}$ considerando $\hat{\boldsymbol{\Psi}}$.

   (ii) Calcule a estatística (4.10) para $t = 1, \ldots, T$ a partir dos dados obtidos em (i).

   (iii) Calcule a estatística $S_{max}^{(k)} = \max_i \{S_1^{(k)}, \ldots, S_N^{(k)}\}$ a partir das estatísticas obtidas em (ii).

3. Calcule M_1, o percentil 95% de $\{S_{max}^{(1)}, \ldots, S_{max}^{(K)}\}$.

4.2.2 Aplicações

Agora, vamos ilustrar o uso da estatística (4.10) e a técnica para o cálculo do limiar descrita nesta seção através de três exemplos. O primeiro exemplo trata-se da análise de dados simulados com um *outlier* incorporado, e os outros exemplos tratam-se das análises de dados reais. Nos três exemplos foi ajustado o modelo de volatilidade estocástica univariado assimétrico. Para o cálculo dos limiares, foram realizadas 500 e 1000 replicações, a fim de analisar se ambos os casos são eficazes para detectar as observações influentes e, assim, diminuir o tempo computacional realizando menos replicações. As estimativas dos parâmetros foram obtidas através da transformação logarítmica e do método de quase-verossimilhança, uma vez que esse é o método que exige menor tempo computacional quando comparado à maximização da verossimilhança exata, e o erro-padrão foi calculado através da matriz de informação de Fisher. As volatilidades foram obtidas através do Filtro e do Suavizador de Kalman aplicado no modelo aproximado, visto na Seção 2.5. Para a maximização da quase-verossimilhança, foi utilizado o método numérico BFGS implementado no *software Ox*.

4.2.2.1 Simulação

Neste exemplo, vamos analisar o comportamento da estatística (4.10) na presença de um *outlier* incorporado no conjunto de dados simulados, a fim de verificarmos se a estatística detecta corretamente o valor atípico. Foi simulada uma série univariada de tamanho $N = 500$, e uma observação atípica foi inserida na componente y_{200}, com tamanho igual a cinco vezes o desvio padrão da série gerada. Os valores dos parâmetros usados para simulação foram escolhidos baseados nos valores presentes na literatura do modelo de volatilidade univariado assimétrico (ver Harvey e Shephard, 1996).

Para o MVE-A (4.9), a série foi gerada de acordo com o conjunto de parâmetros descritos na Tabela 4.2. Nesta mesma tabela, encontram-se as estimativas de quase-máxima verossimilhança dos parâmetros do modelo considerando as séries simuladas com e sem a presença do *outlier*. As estimativas são significativas segundo os quocientes t. Pode-se verificar que a presença do *outlier* quase não influenciou as estimativas dos parâmetros. Cada parâmetro real está dentro do respectivo intervalo de confiança construído tanto pela série sem a presença de *outlier* quanto pela série com *outlier*, supondo distribuição normal.

Tabela 4.2: Valores dos parâmetros para o processo gerador da série, estimativas dos parâmetros para a série com e sem outlier, e seus respectivos erros-padrão e razão-t.

		Série sem outlier			Série com outlier		
Parâmetro	Valor real	Estim.	E.P.	Razão-t	Estim.	E.P.	Razão-t
β	0,900	0,9088	0,0931	9,76	0,9133	0,0903	10,1
ϕ	0,975	0,9628	0,0221	43,6	0,9595	0,0227	42,3
σ	0,200	0,1770	0,0682	2,60	0,1818	0,0675	2,69
ρ	-0,700	-0,7246	0,2121	-3,42	-0,7140	0,2044	-3,49

A Figura 4.5 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. O gráfico de autocorrelação mostra que os erros de predição são não-correlacionados, o que indica que o modelo está corretamente especificado. Também, não há mudança estrutural significativa nos modelos a 5%, uma vez que as somas cumulativas não ultrapassam as linhas de significância em nenhum ponto.

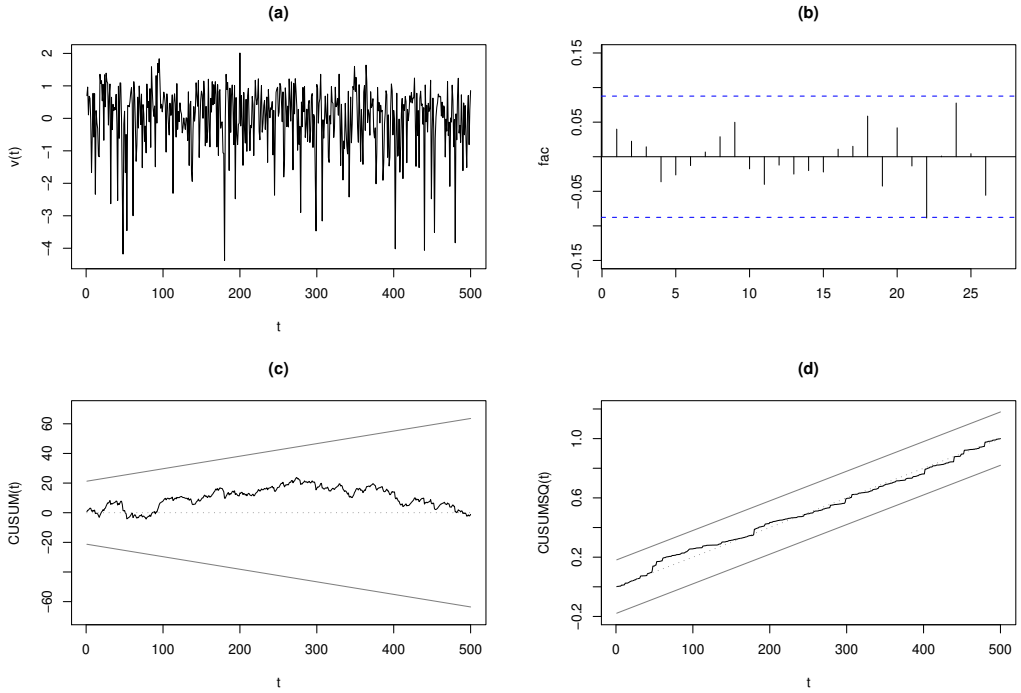


Figura 4.5: Análise residual do ajuste do MVE-A à série simulada: (a) resíduos das observações, (b) autocorrelação dos resíduos, (c) soma cumulativa e (d) soma dos quadrados cumulativa, e as respectivas linhas de significância a 5%.

A Figura 4.6 ilustra a série simulada, (a), e a estatística obtida utilizando o Filtro de Kalman, (b), e Suavizador de Kalman, (c). Na Figura 4.6 (a) pode-se observar que, além da observação 200 que foi inserida como *outlier*, as observações 93 e 95 são de "grande" magnitude em valor absoluto, sendo que ambas pertencem a um período de alta volatilidade. Na Figura 4.6 (b), onde a estatística foi obtida utilizando o Filtro de Kalman, a comparação dos valores de S_i com os limiares obtidos por simulações (6,5162 e 6,599, que correspondem a 500 e 1000 replicações, respectivamente) identifica corretamente a observação 200 como influente. Já a Figura 4.6 (c), onde utiliza-se o Suavizador de Kalman, a comparação dos valores de S_i com os limiares obtidos por simulações (6,1241 e 6,1078, que correspondem a 500 e 1000 replicações, respectivamente) identifica as observações 199 e 200 como influentes. Na Tabela 4.3 pode-se verificar a localização, valor da observação, valores das estatística, resíduos de predição e resíduos auxiliares dos estados das observações detectada como influentes e também das observações 93 e 95. Na tabela, pode-se observar que o resíduo de predição correspondente à observação 200 é significativo. Também, a observação 95, que é genuína no conjunto de dados, não foi detectada como influente, embora o resíduo auxiliar correspondente à esta observação seja significativo.

Tabela 4.3: Localização, valor da observação, valores das estatísticas e resíduos.

		Estatísticas		Resíduos	
Localiz.	Retorno	Filtro	Suavizador	Predição	Auxiliar
93	3,3411	5,0783	3,9967	1,6931	-0,7107
95	5,2711	5,2205	-5,0532	1,8366	2,1143
199	0,5379	6,5040	0,3283	-0,1497	0,4113
200	8,5831	8,6792	3,6217	2,0147	-1,8728

Nas Figuras 4.6 (b) e (c), vemos que a estatística apresentou um aumento significante do seu valor onde o valor atípico foi inserido, e esta conseguiu detectar este valor. Porém, na Figura 4.6 (c), a observação 199 foi detectada incorretamente. Isto pode ser explicado pelo fato da volatilidade ser uma estimativa suavizada, e observações vizinhas de valores influentes podem também ser detectadas como influentes. As observações genuínas 93 e 95 não foram identificadas em nenhum dos casos. Também, nas Figuras 4.6 (b) e (c), os limiares obtidos com 500 e 1000 replicações são valores bastante próximos e detectaram as mesmas observações.

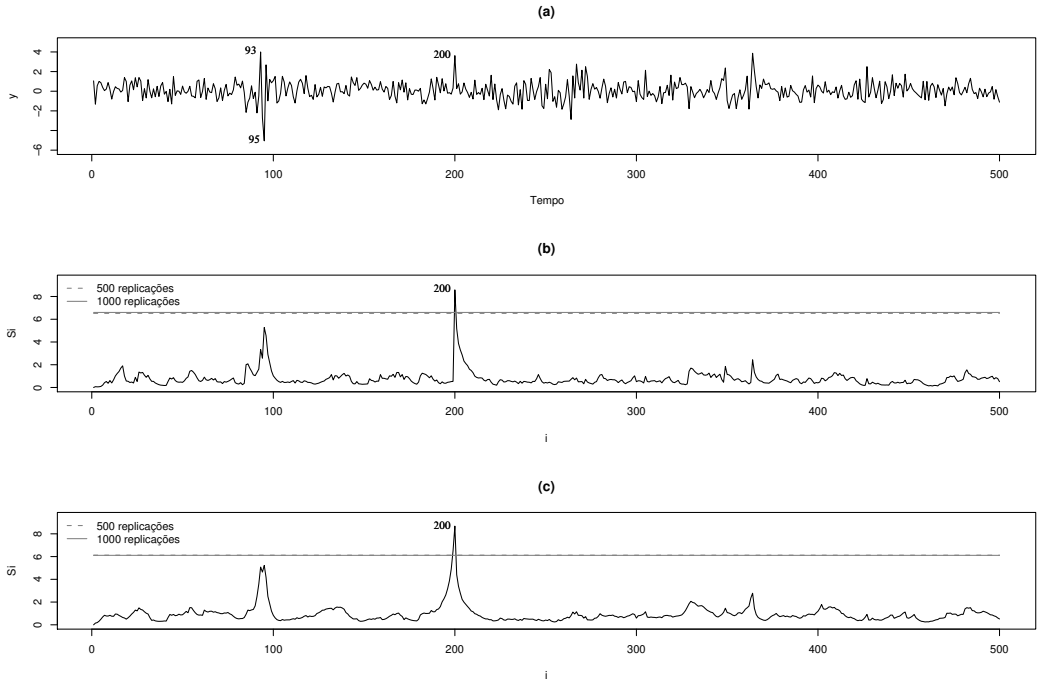


Figura 4.6: (a) Série simulada, (b) estatística de influência filtrada e (c) estatística de influência suavizada, e os limiares obtidos com 500 e 1000 replicações.

A partir deste exemplo notamos que, na análise de influência, o modelo de volatilidade estocástica univariado assimétrico conseguiu detectar corretamente o *outlier* incorporado à série, tanto no caso em que utilizamos o Filtro de Kalman quanto no caso do Suavizador de Kalman. No caso do Suavizador de Kalman, uma observação foi detectada incorretamente (observação 199), sendo que esta observação é vizinha da observação inserida como *outlier* (observação 200). As estimativas dos parâmetros do modelo foram significativas e próximas aos valores reais. Também, as observações genuínas de "grande" magnitude em valor absoluto não foram detectadas pela estatística de influência, mesmo no caso onde o resíduo auxiliar correspondente foi significativo. Por fim, os limiares obtidos com 500 replicações foram tão eficazes na análise de influência quanto os limiares obtidos com 1000 replicações, uma vez que são valores próximos e detectaram as mesmas observações.

4.2.2.2 Análise das séries dos retornos da bolsa da Alemanha

Neste exemplo, foi utilizada a série de retornos diários da bolsa de valores da Alemanha (DAX), no período de julho de 1999 a outubro de 2001, num total de 600 observações.

O resultado do ajuste do MVE-A (4.9) é apresentado na Tabela 4.4. Pode-se observar que as estimativas são significativas segundo os quocientes t.

Tabela 4.4: Estimativas dos parâmetros do MVE-A para a série da bolsa de valores da Alemanha e seus respectivos erros-padrão e razão-t.

Parâmetro	Estimativa	Erro-padrão	Razão-t
β	1,9318	0,1618	11,9
ϕ	0,9322	0,0325	28,7
σ	0,2595	0,0746	3,48
ρ	-0,4017	0,1640	-2,45

A Figura 4.7 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. Pela Figura 4.7 (b), temos correlação na defasagem 5, mas o teste de Ljung-Box reportou p-valor igual a 0,17. Também, não há mudança estrutural significativa nos modelos a 5%, uma vez que as somas cumulativas não ultrapassam as linhas de significância em nenhum ponto.

A Figura 4.8 ilustra a série de retornos, (a), e a estatística obtida utilizando o Filtro de Kalman, (b), e Suavizador de Kalman, (c). Na Figura 4.8 (a) pode-se observar que as observações 51, 467 e 523 parecem ser influentes. Na Figura 4.8 (b), onde a estatística foi obtida utilizando o Filtro de Kalman, a comparação dos valores de S_i com os limiares obtidos por simulações identifica as observações 51, 467, 468 e 523 como influentes. Na Figura 4.8 (c), onde utiliza-se o Suavizador de Kalman, apenas a observação 523 foi identificada como influente.

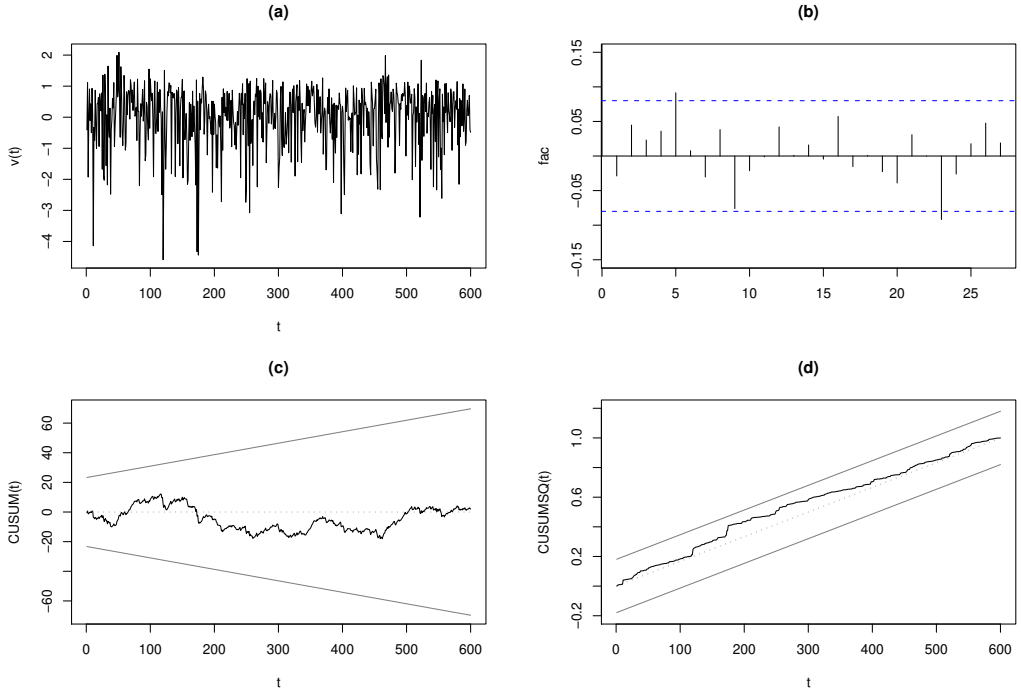


Figura 4.7: Análise residual do ajuste do MVE-A à série de retornos da DAX: (a) resíduos das observações, (b) autocorrelação dos resíduos, (c) soma cumulativa e (d) soma dos quadrados cumulativa, e as respectivas linhas de significância a 5%.

Na Tabela 4.5 pode-se verificar a data de ocorrência, localização, valor do retorno, valores das estatísticas, resíduo de predição e resíduo auxiliar do estado de cada observação detectada como influente. Os pontos detectados têm relação com fatores históricos. Assim, os pontos 467 e 468 correspondem aos dias 17 e 18 de setembro de 2001, os dois primeiros dias de transações logo após o fechamento do mercado devido ao atentado contra as torres do World Trade Center. Também, no início do ano 2000, algumas das principais bolsas despencaram devido a temores de elevação dos juros nos Estados Unidos e na Europa, e isto afetou o mercado da Alemanha. Na Tabela 4.5, pode-se observar que os retornos correspondentes às observações 51, 467 e 523 são valores de "grande" magnitude em valor absoluto. Também, apenas os resíduos auxiliares correspondentes às observações 51 e 467 são significativos. Adicionalmente, os limiares obtidos com 500 e 1000 replicações foram eficazes na análise de influência, pois ambos são valores extremamente próximos e detectaram as mesmas observações.

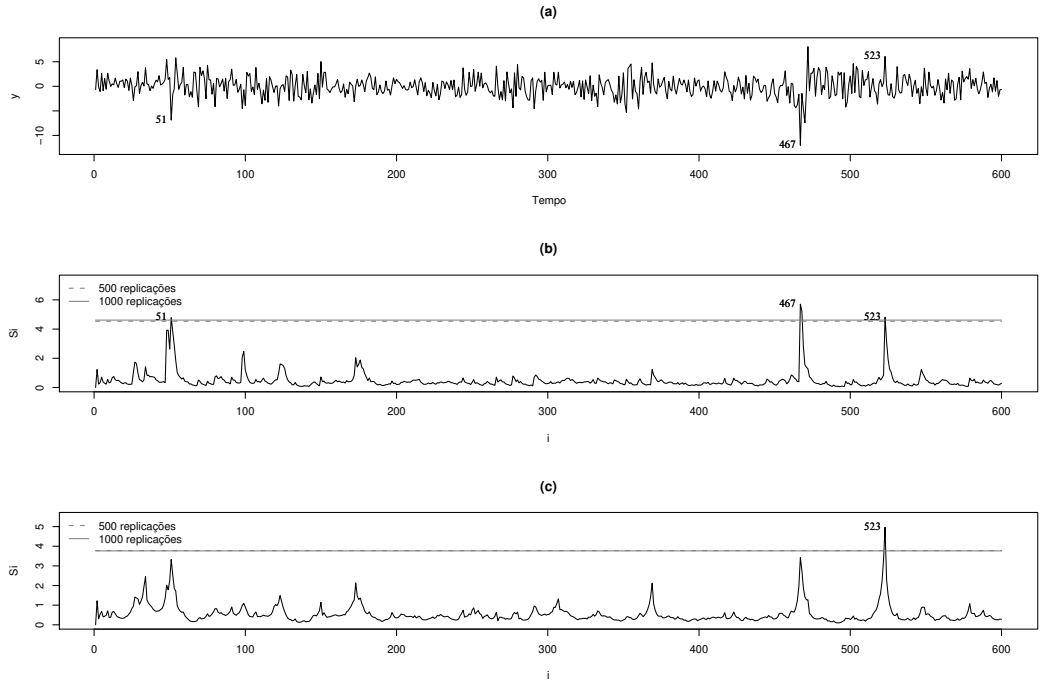


Figura 4.8: (a) Retornos da DAX, (b) estatística de influência filtrada e (c) estatística de influência suavizada, e os limiares obtidos com 500 e 1000 replicações.

Tabela 4.5: Valores influentes identificados: data de ocorrência, localização, valor da observação, valores das estatísticas e resíduos.

			Estatísticas		Resíduos	
Data	Localiz.	Retorno	Filtro	Suavizador	Predição	Auxiliar
04/01/2000	51	-6,8801	4,7814	3,3280	1,9456	2,0942
17/09/2001	467	-12,017	5,7083	3,4184	1,8451	2,2137
18/09/2001	468	-1,4524	5,2249	2,7654	-0,1547	0,9329
05/12/2001	523	6,0769	4,8124	4,9679	1,7081	-0,9924

4.2.2.3 Análise das séries dos retornos do Ibovespa

Neste exemplo, foi utilizada a série de retornos diários do índice da bolsa do Estado de São Paulo (Ibovespa), no período de janeiro de 1997 a janeiro de 2001, num total de 1000 observações.

O resultado do ajuste do MVE-A (4.9) é apresentado na Tabela 4.6. Pode-se observar que as estimativas são significativas segundo os quocientes t.

Tabela 4.6: Estimativas dos parâmetros do MVE-A para as séries de retornos do Ibovespa e seus respectivos erros-padrão e razão-t.

Parâmetro	Estimativa	Erro-padrão	Razão-t
β	2,4409	0,1838	13,3
ϕ	0,9428	0,0197	47,9
σ	0,2921	0,0627	4,66
ρ	-0,5942	0,1016	-5,85

A Figura 4.9 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. O gráfico de autocorrelação mostra que os erros de predição são não-correlacionados, o que indica que o modelo está corretamente especificado. Também, não há mudança estrutural significativa nos modelos a 5%, uma vez que as somas cumulativas não ultrapassam as linhas de significância em nenhum ponto.

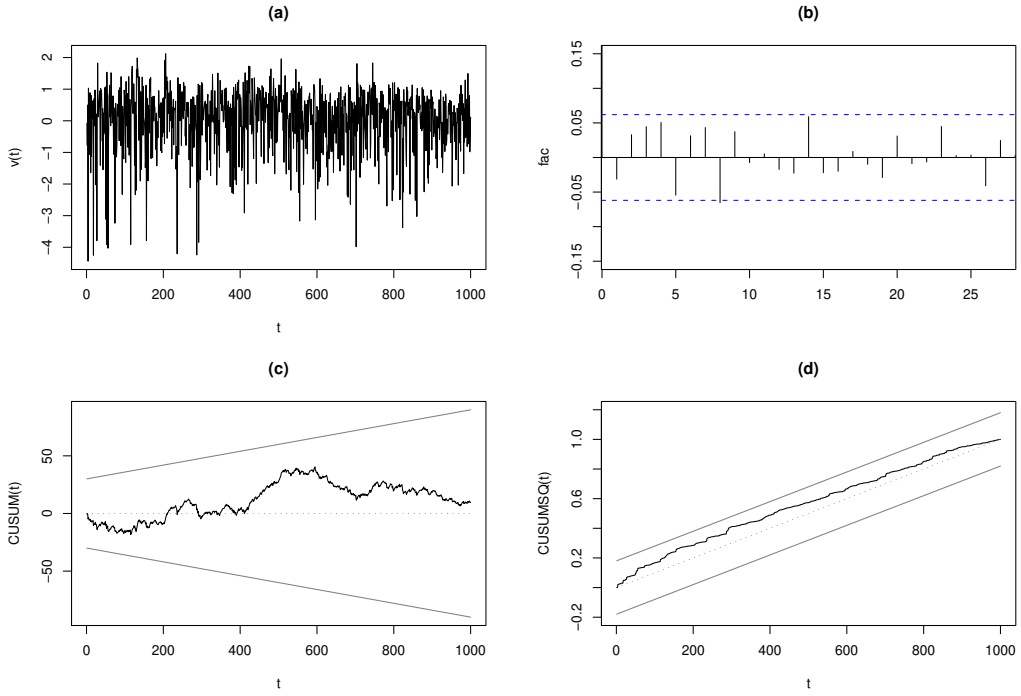


Figura 4.9: Análise residual do ajuste do MVE-A à série de retornos do Ibovespa: (a) resíduos das observações, (b) autocorrelação dos resíduos, (c) soma cumulativa e (d) soma dos quadrados cumulativa, e as respectivas linhas de significância a 5%.

A Figura 4.10 ilustra a série simulada e a estatística obtida utilizando o Filtro de Kalman (b) e Suavizador de Kalman (c). Na Figura 4.10 (a) pode-se observar que as

observações 206 e 507 parecem ser influentes. Na Figura 4.10 (b), onde a estatística foi obtida utilizando o Filtro de Kalman, a comparação dos valores de S_i com os limiares obtidos por simulações identifica as observações 204, 206 e 507 como influentes. Na Figura 4.10 (c), onde utiliza-se o Suavizador de Kalman, apenas a observação 507 foi identificada como influente.

Na Tabela 4.7 pode-se verificar a data de ocorrência, localização, valor do retorno, valores das estatísticas, resíduo de predição e resíduo auxiliar do estado de cada observação detectada como influente. Os pontos detectados têm relação com fatores históricos. Podemos citar o dia 23 de outubro de 1997, quando a bolsa de valores de Hong Kong caiu $10,4\%$ e afetou vários mercados, e a desvalorização do Real em janeiro de 1999. Na Tabela 4.7, pode-se observar que os retornos correspondentes às observações 204, 206 e 507 são valores de "grande" magnitude em valor absoluto. Também, o resíduo de predição correspondente à observação 206 e os resíduos auxiliares correspondentes às observações 204 e 206 são significativos. Adicionalmente, os limiares obtidos com 500 e 1000 replicações foram eficazes na análise de influência, pois ambos são valores extremamente próximos e detectaram as mesmas observações.

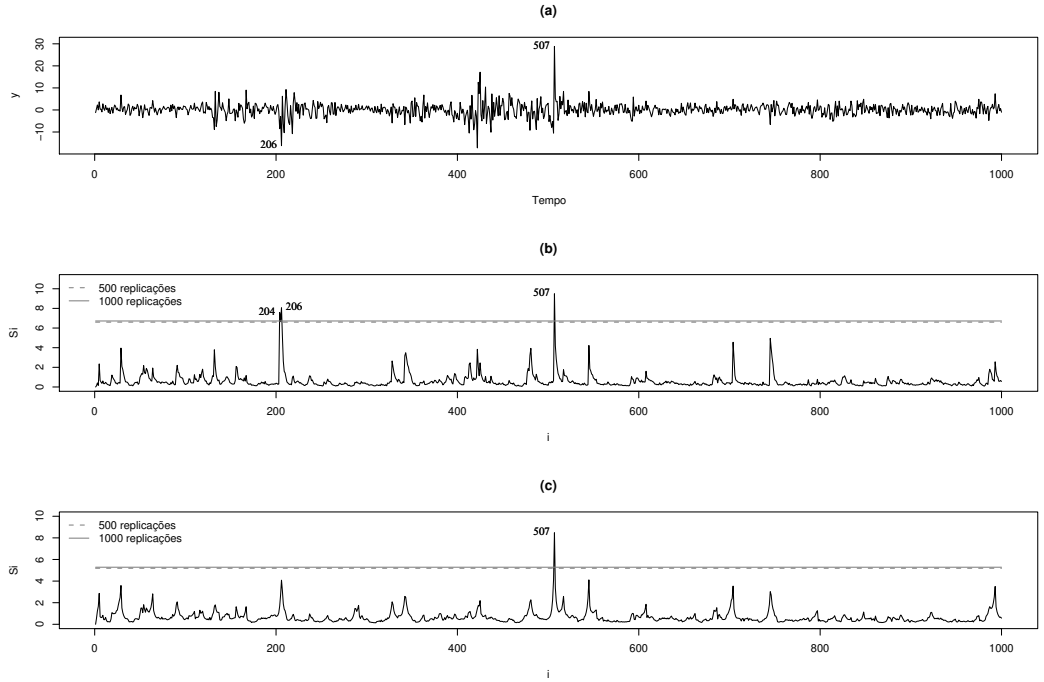


Figura 4.10: (a) Retornos do Ibovespa, (b) estatística de influência filtrada e (c) estatística de influência suavizada, e os limiares obtidos com 500 e 1000 replicações.

Tabela 4.7: Valores influentes identificados: data de ocorrência, localização, valor da observação, valores das estatísticas e resíduos.

			Estatísticas		Resíduos	
Data	Localiz.	Retorno	Filtro	Suavizador	Predição	Auxiliar
23/10/1997	204	-8,5105	7,5897	1,6737	1,8480	2,9648
27/10/1997	206	-16,214	8,0640	4,0834	2,0435	2,8678
15/01/1999	507	28,832	9,4900	8,4734	1,8878	-1,0632

Através destes exemplos, podemos ver que a estatística de influência S_t, em conjunto com a metodologia para o cálculo dos limiares, conseguiu detectar observações influentes que estão relacionadas à crises que impactaram diretamente o mercado financeiro. A estatística de influência obtida utilizando o Filtro de Kalman detectou mais observações influentes quando comparada à estatística obtida utilizando o Suavizador de Kalman. Também, os limiares obtidos com 500 replicações foram tão eficazes na análise de influência quanto o limiares obtidos com 1000 replicações, pois ambos são valores próximos e detectaram as mesmas observações. Por fim, vimos que algumas observações detectadas como influentes possuem resíduos não-significativos.

Capítulo 5

Influência em modelos de volatilidade estocástica multivariados

O trabalho de Motta e Hotta (2007) generalizou a medida de influência proposta por Peña (2005) para a análise de influência no modelo de volatilidade estocástica univariado. Neste capítulo, vamos propor uma generalização da medida de influência proposta por Motta e Hotta (2007) para o modelo de volatilidade estocástica com correlações constantes generalizado. Adicionalmente, será realizada a análise de influência do modelo de volatilidade estocástica fatorial aditivo. O desempenho das medidas de influência será avaliado através da análise de dados simulados e séries reais.

5.1 Estatística de influência

Nesta seção, apresentaremos uma generalização da medida de influência proposta por Motta e Hotta (2007), para diagnosticar influência em modelos de volatilidade estocástica multivariados.

Uma vez que o modelo está escrito na forma de espaço de estados linear (2.1), denote por $\boldsymbol{y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_i, \cdots, \boldsymbol{y}_N)$ a matriz de dados do modelo (2.1) formada pelas séries $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iT})'$, $i = 1, \cdots, N$, por $\hat{\boldsymbol{\Psi}}$ o vetor das estimativas dos parâmetros baseadas em todas as observações e por $\hat{h}_{it}$ a estimativa de h_{it} obtida através do Filtro ou do Suavizador de Kalman aplicado no modelo aproximado, conforme foi visto na Seção 2.5. Considerando que $\boldsymbol{y}_{(t)} = \left(\boldsymbol{y}_{1(t)}, \cdots, \boldsymbol{y}_{i(t)}, \cdots, \boldsymbol{y}_{N(t)}\right)$ é a matriz de dados onde cada série $\boldsymbol{y}_{i(t)}$ possui

a t-ésima observação como faltante, ou seja, o conjunto de observações $\{y_{1t}, \ldots, y_{Nt}\}$ é faltante, tem-se que a estimativa dos parâmetros será representada por $\hat{\boldsymbol{\Psi}}_{(t)}$. Desta forma, considerando os vetores $\boldsymbol{s}_{it} = (\hat{h}_{it} - \hat{h}_{it(1)}, \ldots, \hat{h}_{it} - \hat{h}_{it(T)})'$, a *estatística de influência global* para o modelo de volatilidade estocástica com correlações constantes generalizado é definida como

$$S_t = \sum_{i=1}^{N} \frac{\boldsymbol{s}_{it}'\boldsymbol{s}_{it}}{\widehat{var}(\hat{h}_{it})} + 2 \sum_{\substack{i,j=1 \\ i<j}}^{N} \frac{\boldsymbol{s}_{it}'\boldsymbol{s}_{jt}}{\widehat{cov}(\hat{h}_{it}, \hat{h}_{jt})}, \tag{5.1}$$

onde $\widehat{var}(\hat{h}_{it})$ e $\widehat{cov}(\hat{h}_{it}, \hat{h}_{jt})$ são obtidas através do Filtro ou do Suavizador de Kalman aplicado no modelo aproximado, conforme foi visto na Seção 2.5. O j-ésimo termo do vetor $\boldsymbol{s}_{it} = (\hat{h}_{it} - \hat{h}_{it(1)}, \ldots, \hat{h}_{it} - \hat{h}_{it(T)})'$ mostra a diferença entre a estimação de h_{it} com a matriz de dados $\boldsymbol{y}$ completa e a estimação de h_{it} com a matriz de dados $\boldsymbol{y}_{(j)}$. Sendo assim, analisamos o quão sensível é a estimação de $\boldsymbol{h}_t = (h_{1t}, \ldots, h_{Nt})'$ quando se exclui cada conjunto de observações $\{y_{1j}, \ldots, y_{Nj}\}$ da amostra.

Também, a *estatística de influência individual* para a i-ésima série é definida como

$$S_{it} = \frac{\boldsymbol{s}_{it}'\boldsymbol{s}_{it}}{\widehat{var}(\hat{h}_{it})}. \tag{5.2}$$

A estatística S_t pode se usada para verificar se a observação $\boldsymbol{y}_t$ é influente no modelo multivariado, enquanto que a estatística individual S_{it} pode se usada para verificar se a observação y_{it} de uma determinada série univariada i é influente no mesmo modelo.

O procedimento para calcular as estatísticas é dado pelo seguinte algoritmo:

1. Estime os parâmetros $\boldsymbol{\Psi}$ do modelo de volatilidade baseado na amostra completa, e encontre $\hat{h}_{it}$, $\widehat{var}(\hat{h}_{it})$ e $\widehat{cov}(\hat{h}_{it}, \hat{h}_{jt})$ para $i = 1, \ldots, N$ e $t = 1, \ldots, T$.

2. Para $t = 1, \ldots, T$:

   (i) Considere a estimativa $\hat{\boldsymbol{h}}_t = (\hat{h}_{1t}, \ldots, \hat{h}_{Nt})'$ encontrada no item (1).

   (ii) Para $j = 1, \ldots, T$:

   - Defina a matriz $\boldsymbol{y}_{(j)}$, considerando as observações $y_{1j}, \ldots, y_{Nj}$ como faltantes.
   - Reestime o modelo com base em $\boldsymbol{y}_{(j)}$, obtendo $\hat{\boldsymbol{\Psi}}_{(j)}$.
   - Encontre $\hat{\boldsymbol{h}}_{t(j)}$, a estimativa de $\hat{\boldsymbol{h}}_t$ baseada em $\hat{\boldsymbol{\Psi}}_{(j)}$ e em $\boldsymbol{y}_{(j)}$.

   (iii) Calcule as estatísticas S_t e S_{it}, $i = 1, \ldots, N$, utilizando (5.1) e (5.2).

5.2 Metodologia para diagnóstico de influência

Após calculadas as medidas da estatística (5.1) e (5.2), deve-se avaliar se estas são "grandes" ou "pequenas" para caracterizá-las como influentes, através de alguma referência específica. Nesta seção, serão discutidos os procedimentos para o cálculo de limiares, construídos a partir de simulações de Monte Carlo, e o consequente diagnóstico de influência.

O cálculo de níveis de referência através de simulações de Monte Carlo é análogo ao método apresentado na Seção 4.2. Considerando o modelo estimado para uma determinada série, para a k-ésima replicação ($k = 1, \ldots, K$) é simulada uma série de dados $y_1^{(k)}, \ldots, y_T^{(k)}$, da qual calcula-se a estatística $S_{max}^{(k)} = \max_i S_t$. O limiar M_1 para a estatística de influência global é calculado como sendo o percentil 95% de $\{S_{max}^{(1)}, \ldots, S_{max}^{(K)}\}$. Se uma estatística S_t for maior que M_1, então a observação t é considerada influente.

Em resumo, considerando K replicações, o limiar é calculado da seguinte maneira:

1. Encontre $\hat{\boldsymbol{\Psi}}$ considerando a série $y_1, \ldots, y_N$.

2. Para $k = 1, \ldots, K$:

   (i) Simule $y_1^{(k)}, \ldots, y_N^{(k)}$ considerando $\hat{\boldsymbol{\Psi}}$.

   (ii) Calcule a estatística (5.1) para $t = 1, \ldots, T$ a partir dos dados obtidos em (i).

   (iii) Calcule a estatística $S_{max}^{(k)} = \max_i \{S_1^{(k)}, \ldots, S_N^{(k)}\}$ a partir das estatísticas obtidas em (ii).

3. Calcule M_1, o percentil 95% de $\{S_{max}^{(1)}, \ldots, S_{max}^{(K)}\}$.

Os limiares para as estatísticas de influência individuais (5.2) são calculados de forma análoga.

5.3 Simulações

Nesta seção serão apresentadas as análises de séries bivariadas simuladas com *outliers* incorporados, a fim de estudarmos o comportamento da estatística de influência proposta

quando tratamos séries com valores atípicos. Para o cálculo dos limiares, foram realizadas 500 e 1000 replicações, a fim de verificar se ambos os casos são eficazes para detectar as observações influentes e, assim, diminuir o tempo computacional realizando menos replicações. As estimativas dos parâmetros do modelo com correlações constantes generalizado foram obtidas através da transformação logarítmica aplicada nos mesmos, enquanto que os parâmetros do modelo fatorial aditivo foram estimados através do modelo baseado nas primeiras derivadas, visto na Seção 2.5. As estimativas dos parâmetros dos modelos foram calculadas através do método de quase-verossimilhança, uma vez que esse é o método que exige menor tempo computacional quando comparado à maximização da verossimilhança exata, e o erro-padrão foi calculado através da matriz de informação de Fisher. As volatilidades foram obtidas através do Filtro e do Suavizador de Kalman aplicado no modelo aproximado, visto na Seção 2.5. Para a maximização da quase-verossimilhança, foi utilizado o método numérico BFGS implementado no *software Ox*.

Foram simuladas séries bivariadas de tamanho $N = 600$. Uma vez simuladas as séries, foram inseridas quatro observações atípicas nas componentes $y_{1,400}$, $y_{2,200}$, $y_{1,270}$ e $y_{2,270}$, com tamanhos iguais a cinco vezes o desvio padrão da respectiva série gerada. Ou seja, na primeira série foram inseridas observações nas posições 270 e 400, enquanto que na segunda série foram inseridas observações nas posições 200 e 270. Os valores dos parâmetros usados para simulação foram escolhidos baseados nos valores presentes na literatura dos modelos de volatilidade estocástica multivariados (ver Chan et al, 2006 e Jacquier et al, 1999).

5.3.1 Modelo com correlações constantes generalizado

Para o modelo (3.16), a série bivariada foi gerada de acordo com o conjunto de parâmetros descritos na Tabela 5.1. Nesta tabela, encontram-se também as estimativas de quase-máxima verossimilhança dos parâmetros do modelo considerando as séries simuladas com e sem a presença de *outliers*. As estimativas são significativas segundo os quocientes t. Pode-se verificar que a presença de *outliers* influenciou especialmente as estimativas dos parâmetros σ_1, σ_2, ρ_ε e ρ_η. Porém, os outros parâmetros são pouco influenciados pela presença dos mesmos. Cada parâmetro real está dentro do respectivo intervalo de confiança construído tanto pela série sem a presença de *outliers* quanto pela

série com *outliers*, supondo distribuição normal.

Tabela 5.1: Valores dos parâmetros para o processo gerador da série, estimativas dos parâmetros para as séries com e sem outliers, e seus respectivos erros-padrão e razão-t.

		Série sem outlier			Série com outlier		
Parâmetro	Valor real	Estim.	E.P.	Razão-t	Estim.	E.P.	Razão-t
β_1	1,00	0,9385	0,0994	9,44	0,9525	0,0973	9,79
β_2	1,50	1,4801	0,1742	8,50	1,5173	0,1852	8,19
ϕ_{11}	0,95	0,9428	0,0294	32,0	0,9351	0,0340	27,5
ϕ_{22}	0,97	0,9714	0,0193	50,3	0,9764	0,0179	54,6
σ_1	0,25	0,2805	0,0900	3,12	0,3015	0,0987	3,05
σ_2	0,20	0,1622	0,0646	2,51	0,1381	0,0666	2,07
ρ_ε	0,70	0,7085	0,0458	15,5	0,7213	0,0438	16,5
ρ_η	0,85	0,8992	0,1008	8,92	0,8801	0,1200	7,34

A Figura 5.1 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. Os gráficos de autocorrelação mostram que os erros de predição são não-correlacionados, o que indica que o modelo está corretamente especificado. Também, não há mudança estrutural significativa no modelo a 5%, uma vez que as somas cumulativas não ultrapassam as linhas de significância em nenhum ponto e são estáveis ao longo do tempo.

Para a série simulada é realizada a análise de influência utilizando os limiares simulados como descrito na Seção 5.2 e assim, espera-se identificar os *outliers*. Nas Figuras 5.2 (a) e (b) tem-se as séries simuladas, nas Figuras 5.2 (c), (d) e (e) tem-se a estatística de influência global e individuais obtidas utilizando o Filtro de Kalman, e nas Figuras 5.2 (f), (g) e (h) tem-se a estatística de influência global e individuais obtidas utilizando o Suavizador de Kalman. Nas Figuras 5.2 (a) e (b) pode-se observar que, além das observações atípicas inseridas nas séries simuladas, algumas outras observações poderiam ser consideradas como influentes devido a sua magnitude, como por exemplo as observações 33, 485, 578 e 580 da primeira série e as observações 479 e 485 da segunda série. Porém, as mesmas pertencem a um período de alta volatilidade.

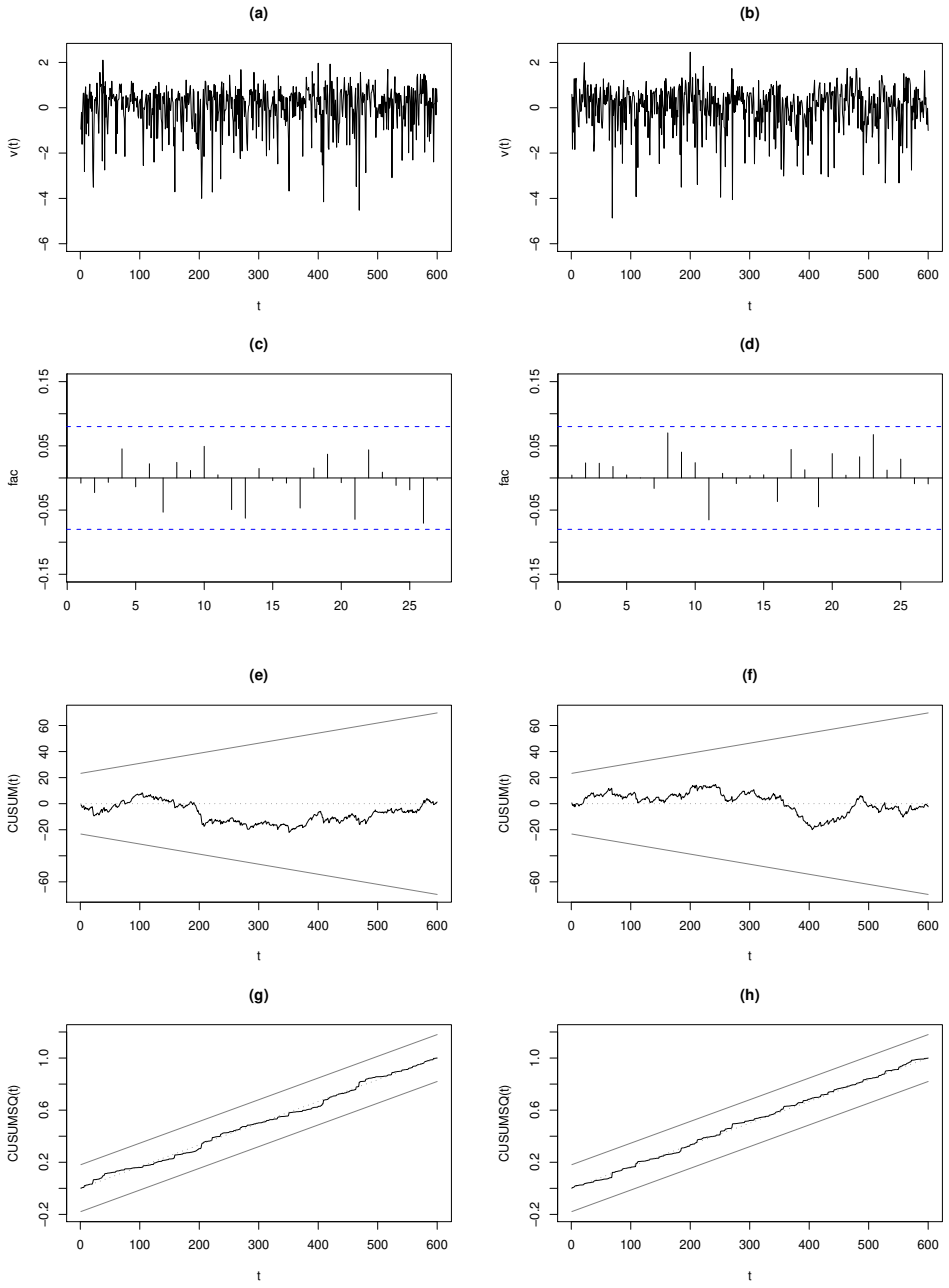


Figura 5.1: Análise residual do ajuste do MVE-CCG à série bivariada simulada: (a) e (b) resíduos das observações da 1ª e 2ª série, (c) e (d) autocorrelação dos resíduos da 1ª e 2ª série, (e) e (f) soma cumulativa da 1ª e 2ª série e as linhas de significância a 5% e (g) e (h) soma dos quadrados cumulativa da 1ª e 2ª série e as linhas de significância a 5%.

Podemos notar nas Figuras 5.2 (c) e (f) correspondentes a primeira série que as estatísticas apresentaram um aumento significante de seus valores nas posições 270 e 400,

justamente onde os *outliers* foram inseridos. O mesmo ocorreu com a segunda série, nas posições 200 e 270 (Figuras 5.2 (d) e (g)). Nas Figuras 5.2 (e) e (h), pode-se observar que as estatísticas globais obtidas respectivamente através do Filtro e Suavizador de Kalman conseguiram detectar os três valores atípicos inseridos na série bivariada, nas posições 200, 270 e 400, uma vez que, nestes pontos, elas são maiores que os limiares obtidos. Porém, tanto no caso do Filtro quanto do Suavizador de Kalman os limiares detectaram incorretamente algumas observações que são vizinhas de observações inseridas como *outliers*, especialmente no caso do Suavizador de Kalman (observações 201 e 271 no caso do Filtro de Kalman e observações 198, 199, 201, 202, 268, 269, 271, 272, 399 e 401 no caso do Suavizador de Kalman). Por fim, os limiares obtidos com 500 e 1000 replicações detectaram as mesmas observações e são valores bastante próximos.

Na Tabela 5.2 temos os valores dos resíduos de predição e resíduos auxiliares dos estados dos *outliers* e também das observações que são de "grande" magnitude em valor absoluto, mas que não foram inseridas como *outliers*. Nesta tabela, pode-se notar que apenas o resíduo auxiliar correspondente à observação 200 da série 2 é significativo ao nível de 5%. Já os resíduos de predição foram todos não-significativos. Ou seja, uma análise residual aplicada neste exemplo não seria suficiente para detectar os *outliers*.

Tabela 5.2: Valores dos resíduos para algumas observações.

Série	Localização	Retorno	Res. predição	Res. auxiliar
Série 1	33	-4,62	0,72	0,98
	270	-4,72	0,78	-1,10
	400	5,78	0,91	-1,31
	485	5,66	0,57	0,59
	578	-4,18	0,69	0,08
	580	6,24	0,65	0,21
Série 2	200	8,53	1,13	-2,01
	270	-6,96	0,80	-1,05
	479	-7,28	0,80	0,73
	485	9,46	0,61	0,19

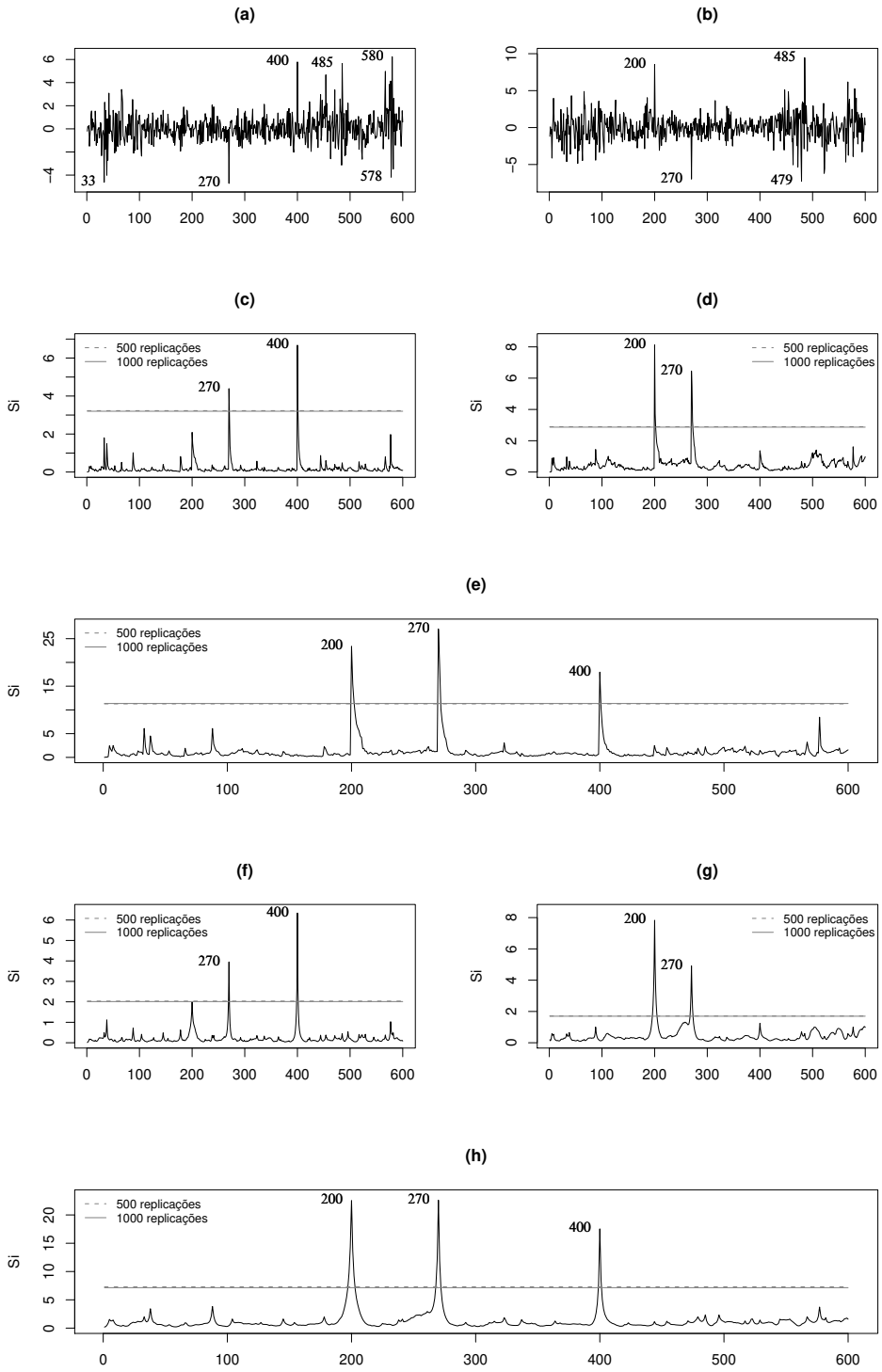


Figura 5.2: Figuras de influência para a série simulada: (a) e (b) séries, (c) e (d) estatísticas de influência individuais filtradas e (e) estatística de influência global filtrada, (f) e (g) estatísticas de influência individuais suavizadas e (h) estatística de influência global suavizada, e os limiares obtidos com 500 e 1000 replicações.

5.3.2 Modelo fatorial aditivo

Agora, a série bivariada foi gerada de acordo com o conjunto de parâmetros descritos na Tabela 5.3, onde encontram-se também as estimativas dos parâmetros do modelo considerando as séries simuladas com e sem os *outliers*. As estimativas são significativas segundo os quocientes t. Pode-se verificar que a presença de *outlier* influenciou em especial as estimativas dos parâmetros σ_{ε_1}, σ_{ε_2} e σ_η. Os outros parâmetros são pouco influenciados pela presença dos valores atípicos. Cada parâmetro real está dentro do respectivo intervalo de confiança construído tanto pela série sem a presença de *outliers* quanto pela série com *outliers*, supondo distribuição normal.

Tabela 5.3: Valores dos parâmetros para o processo gerador da série, estimativas dos parâmetros para as séries com e sem outliers, e seus respectivos erros-padrão e razão-t.

		Série sem outlier			Série com outlier		
Parâmetro	Valor real	Estim.	E.P.	Razão-t	Estim.	E.P.	Razão-t
d	1,20	1,2030	0,0551	21,8	1,2010	0,0681	17,6
β	1,00	0,9978	0,0612	14,7	0,9773	0,0716	12,2
ϕ	0,95	0,9695	0,0179	54,1	0,9615	0,0256	37,6
σ_{ε_1}	0,35	0,3599	0,0558	6,45	0,3712	0,0678	5,48
σ_{ε_2}	0,40	0,3698	0,0763	4,85	0,3921	0,0910	4,31
σ_η	0,20	0,2135	0,0308	3,69	0,2273	0,0379	3,36

A Figura 5.3 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. A Figura 5.3 (b) indica que os erros de predição são não-correlacionados, o que indica que o modelo está corretamente especificado. Também, as Figuras 5.3 (c) e (d) indicam que não há mudança estrutural significativa no modelo a 5%, uma vez que as somas cumulativas não ultrapassam as linhas de significância em nenhum ponto e são estáveis ao longo do tempo.

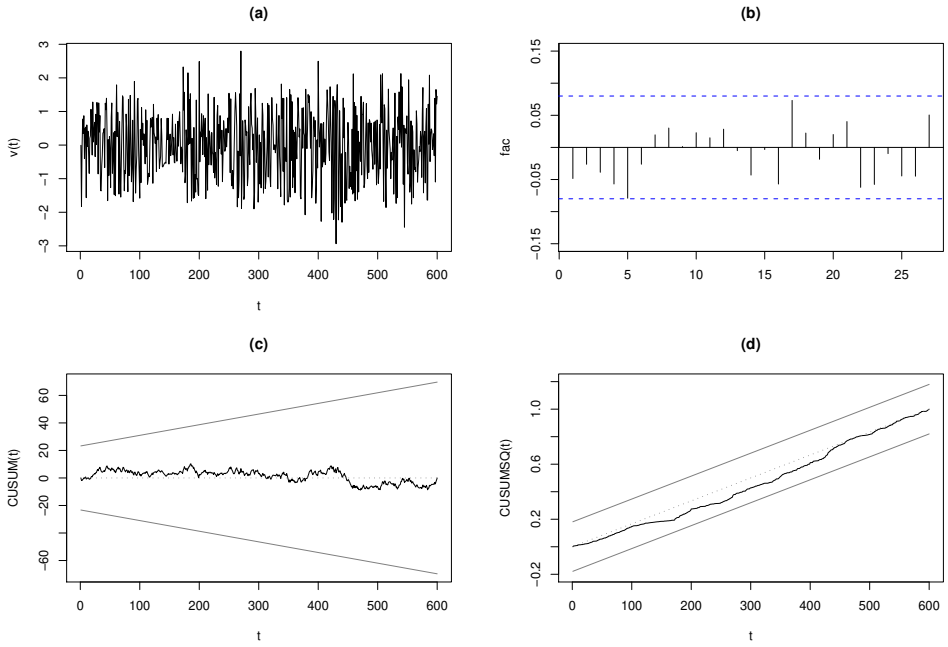


Figura 5.3: Análise residual do ajuste do MVE-FA à série simulada: (a) resíduos das observações, (b) autocorrelação dos resíduos, (c) soma cumulativa e (d) soma dos quadrados cumulativa, e as respectivas linhas de significância a 5%.

Para a série simulada é realizada a análise de influência utilizando os limiares descritos na Seção 5.2 e assim, espera-se identificar os *outliers* inseridos. Na Figura 5.4 temos as séries simuladas e as análises de influência para as duas séries. Para este modelo, temos apenas um gráfico de influência para cada caso (Figuras 5.4 (c) e (d)), pois a volatilidade estocástica h_t é comum para as duas séries. Nas Figuras 5.4 (a) e (b) pode-se observar que, além dos *outliers* inseridos nas séries simuladas, algumas outras observações parecem ser influentes, como por exemplo as observações 432 e 552. Porém, as mesmas pertencem a um período de alta volatilidade comum nas duas séries.

Nas Figuras 5.4 (c) e (d), pode-se observar que as estatísticas globais obtidas respectivamente através do Filtro e Suavizador de Kalman apresentaram um aumento significante de seus valores nas posições 200, 270 e 400, justamente onde os *outliers* foram inseridos. Assim como no modelo anterior, a estatística conseguiu detectar os *outliers* inseridos na série simulada, pois seu valor é maior que os limiares nos pontos onde há *outliers*. Porém, novamente os limiares detectaram incorretamente algumas observações que são vizinhas de observações inseridas como *outliers*, em especial no caso do Suavizador de Kalman (observação 271 no caso do Filtro de Kalman e observações 268, 269, 271 e 272 no caso

do Suavizador de Kalman). Também, os limiares obtidos com 500 e 1000 replicações detectaram as mesmas observações e são valores próximos.

Na Tabela 5.4 temos os valores dos resíduos de predição e resíduos auxiliares dos estados dos *outliers* e também das observações que são de "grande" magnitude em valor absoluto, mas que não foram inseridas como *outliers*. Nesta tabela, pode-se notar que os resíduos de predição são todos significativos ao nível de 5%, mesmo no caso das observações 432 e 552, que não foram inseridas como *outliers*. Já os resíduos auxiliares foram todos não-significativos. Ou seja, uma análise residual aplicada neste exemplo não seria suficiente para detectar os *outliers*.

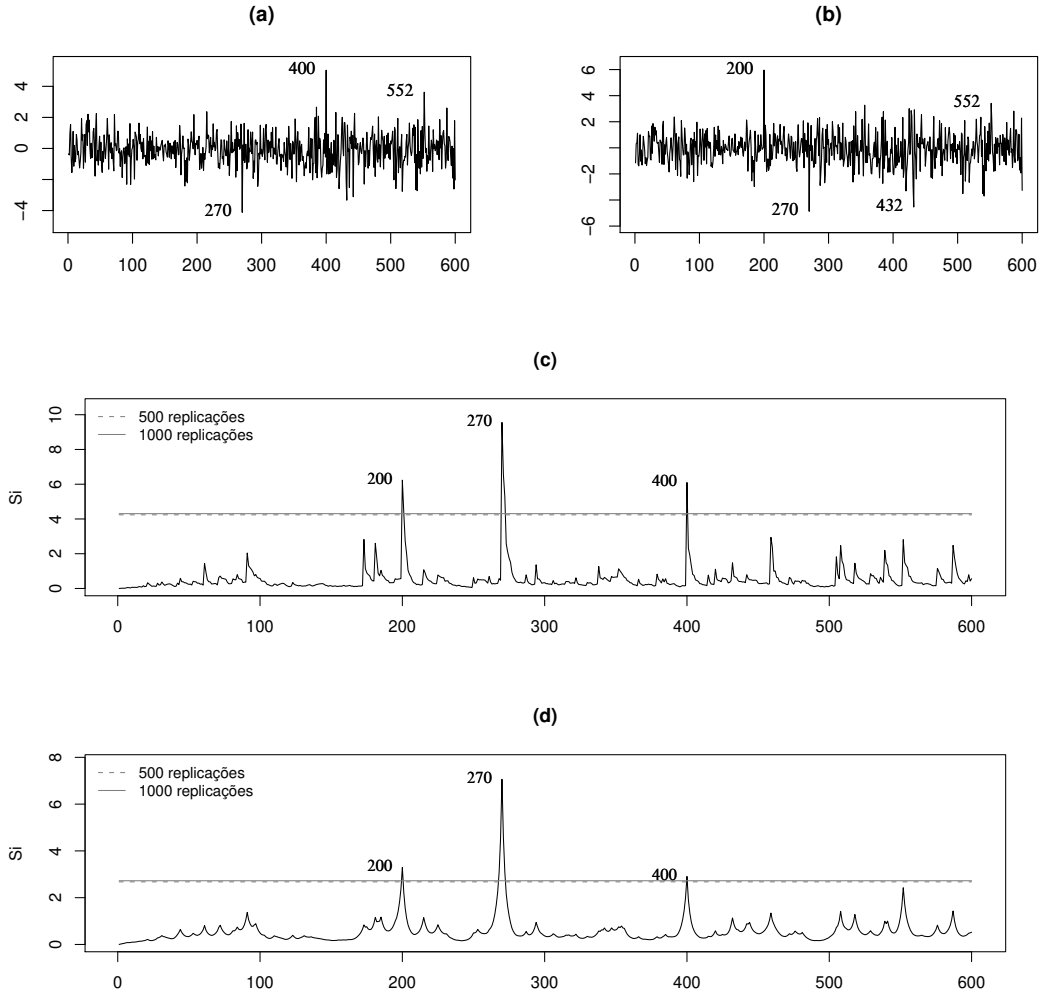


Figura 5.4: Figuras de influência para as séries simuladas: (a) e (b) séries, (c) estatística de influência filtrada e (d) estatística de influência suavizada, e os limiares obtidos com 500 e 1000 replicações.

Tabela 5.4: Valores dos resíduos para algumas observações.

Localização	Retorno série 1	Retorno série 2	Res. predição	Res. auxiliar
200	0,52	5,96	2,76	-0,58
270	-4,12	-4,87	3,11	-1,31
400	5,04	-1,95	2,77	-0,45
432	-3,33	-4,54	1,99	-0,71
552	3,62	3,41	2,16	-1,44

A partir dos experimentos de simulação pode-se ver que, na análise de influência, tanto o modelo com correlações constantes generalizado quando o modelo fatorial aditivo conseguiram detectar os *outliers* incorporados às séries, seja usando o Filtro de Kalman ou o Suavizador de Kalman. Em ambos os casos, algumas observações foram detectadas incorretamente, sendo que essas observações são vizinhas de observações inseridas como *outliers*. Além disso, as estimativas dos parâmetros de ambos os modelos foram significativas e próximas aos valores reais. Também, as observações de "grande" magnitude em valor absoluto e que não foram geradas como *outliers* não foram detectadas pela estatística de influência, mesmo nos casos onde os resíduos correspondentes são valores consideravelmente altos. Os limiares obtidos com 500 replicações foram tão eficazes na análise de influência quanto os limiares obtidos com 1000 replicações, uma vez que são valores próximos e detectaram as mesmas observações. Por fim, vemos que, quando há uma observação influente no conjunto de dados, a estatística de influência proposta pode detectar esta observação e também as observações vizinhas a esta, especialmente quando utilizamos o Suavizador de Kalman na estimação das volatilidades.

5.4 Aplicações

Nesta seção serão apresentadas as análises de séries reais de retornos (séries bivariadas), utilizando o método de influência proposto. Primeiramente, a análise de influência será realizada individualmente em cada série, ajustado o modelo de volatilidade estocástica univariado simétrico. A seguir, serão ajustados os modelos de volatilidade multivariados: modelo com correlações constantes generalizado e modelo fatorial aditivo, ambos apresentados no Capítulo 3. Para o cálculo dos limiares, foram realizadas 500

e 1000 replicações, a fim de analisar se ambos os casos são eficazes para detectar as observações influentes. As estimativas dos parâmetros do modelo univariado e do modelo com correlações constantes generalizado foram obtidas através da transformação logarítmica, enquanto que os parâmetros do modelo fatorial aditivo foram estimados através do modelo baseado nas primeiras derivadas. As estimativas dos parâmetros dos modelos foram calculadas através do método de quase-verossimilhança, e o erro-padrão foi calculado através da matriz de informação de Fisher. As volatilidades foram obtidas através do Filtro e do Suavizador de Kalman aplicado no modelo aproximado, visto na Seção 2.5. Para a maximização da quase-verossimilhança, foi utilizado o método numérico BFGS implementado no *software Ox*.

5.4.1 Análise das séries dos retornos do dólar da Austrália e da Nova Zelândia

A série estudada aqui contém 519 retornos semanais do dólar da Austrália e da Nova Zelândia, no período de janeiro de 1994 a dezembro de 2003. A Figura 5.5 ilustra as séries de retornos.

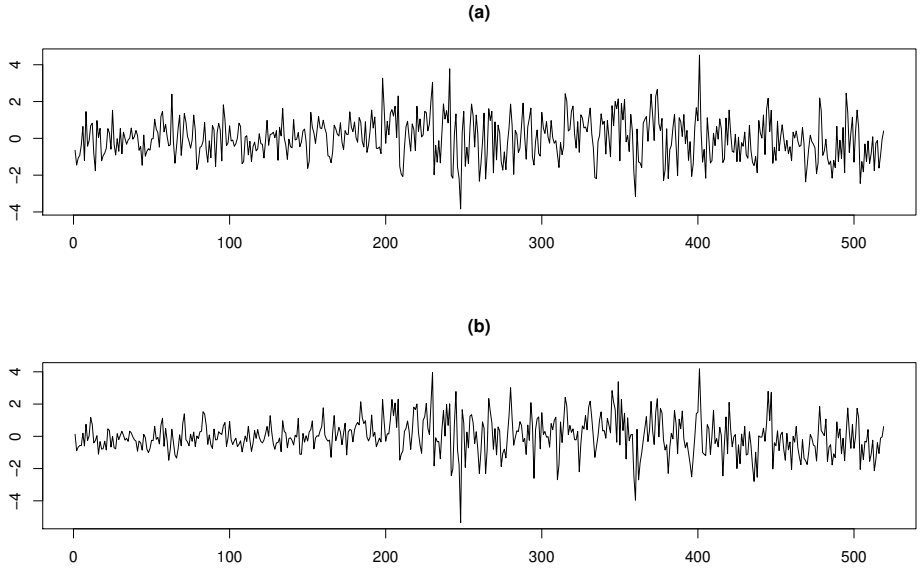


Figura 5.5: Retornos semanais do dólar (a) da Austrália e (b) da Nova Zelândia.

5.4.1.1 Modelos univariados

Os resultados do ajuste do MVE-U (3.2) - (3.3) para cada uma das séries são apresentados na Tabela 5.5. As estimativas são significativas segundo os quocientes t.

Tabela 5.5: Estimativas dos parâmetros do MVE-U para as séries de retornos do dólar da Austrália e da Nova Zelândia e seus respectivos erros-padrão e razão-t.

Série	Parâmetro	Estimativa	Erro-padrão	Razão-t
Austrália	β	1,0911	0,1401	7,79
	ϕ	0,9780	0,0175	56,0
	σ	0,1337	0,0556	2,40
Nova Zelândia	β	0,7031	0,2290	3,07
	ϕ	0,9952	0,0102	97,7
	σ	0,0888	0,0403	2,20

A Figura 5.6 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. Os gráficos de autocorrelação mostram que os erros de predição são não-correlacionados, o que indica que o modelo está corretamente especificado. Também, não há mudança estrutural significativa no modelo a 5%, uma vez que as somas cumulativas não ultrapassam as linhas de significância em nenhum ponto.

Para cada uma das séries é realizada a análise de influência utilizando os limiares descritos na Seção 5.2 e assim, espera-se identificar os valores atípicos. A Figura 5.7 ilustra as séries de retornos e as estatísticas de influência obtidas através do Filtro e Suavizador de Kalman. Nas Figuras 5.7 (a) e (b) pode-se observar que algumas observações parecem ser influentes, como por exemplo as observações 198, 241, 248 e 401 da primeira série e as observações 230, 248, 360 e 401 da segunda série. Na Tabela 5.6 pode-se verificar a data de ocorrência, localização, valor do retorno, valor da estatística, resíduo de predição e resíduo auxiliar do estado de cada observação detectada como influente.

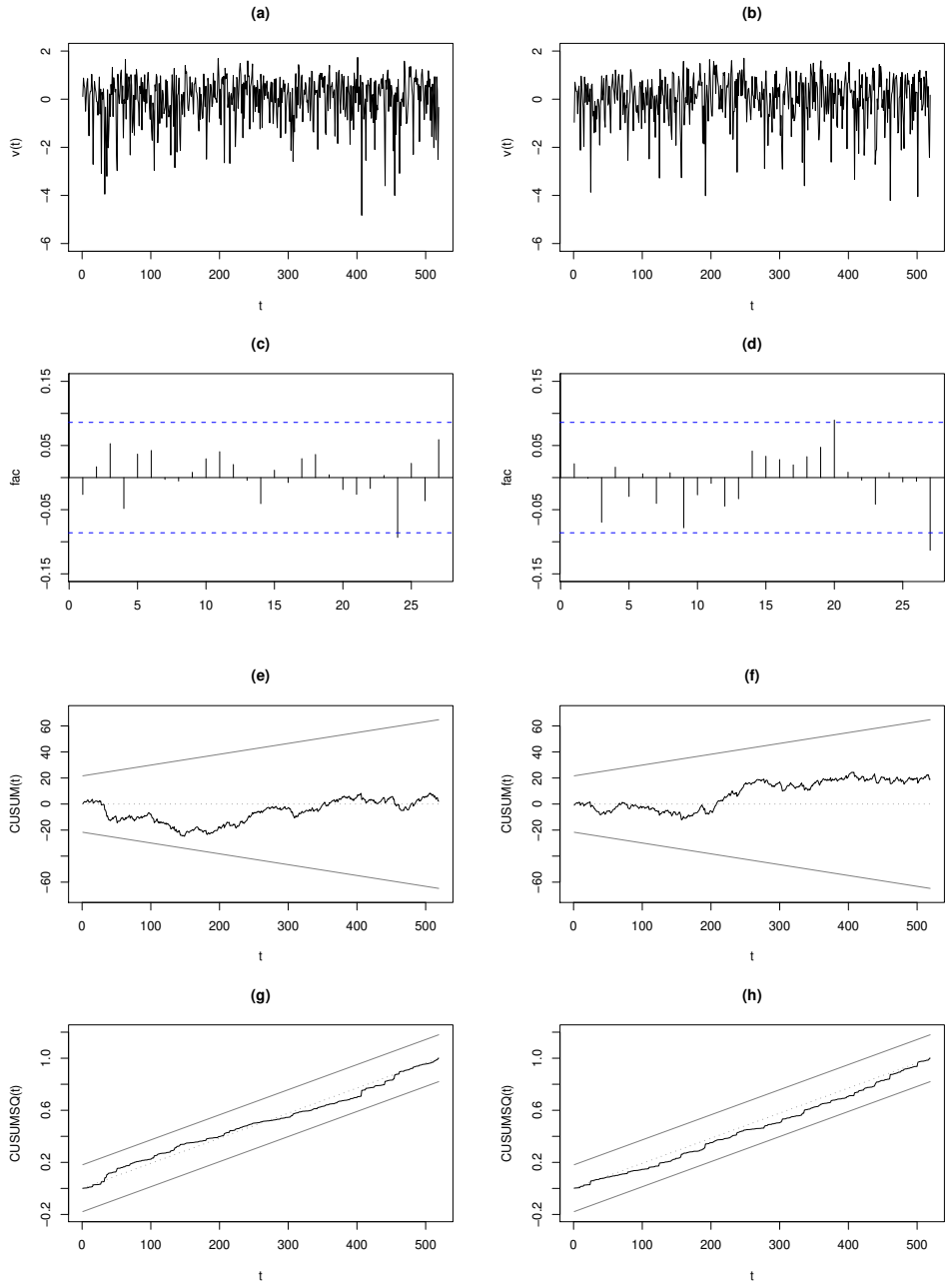


Figura 5.6: Análise residual do ajuste do MVE-U às séries de retornos do dólar da Austrália e da Nova Zelândia: (a) e (b) resíduos das observações da 1^a e 2^a série, (c) e (d) autocorrelação dos resíduos da 1^a e 2^a série, (e) e (f) soma cumulativa da 1^a e 2^a série e as linhas de significância a 5% e (g) e (h) soma dos quadrados cumulativa da 1^a e 2^a série e as linhas de significância a 5%.

Através da Tabela 5.6 e das Figuras 5.7 (c) e (e) correspondentes a primeira série, podemos ver que a estatística obtida pelo Filtro de Kalman detectou as observações 198,

401 e 402 como influentes, enquanto que a estatística obtida pelo Suavizador de Kalman detectou o bloco 400 a 402, e não detectou a observação 198. Através da Tabela 5.6 e das Figuras 5.7 (d) e (f) correspondentes a segunda série, vemos que a estatística obtida pelo Filtro de Kalman detectou a observação 248, enquanto que a estatística obtida pelo Suavizador de Kalman detectou o bloco 245 a 249, ou seja, a vizinhança da observação 248. Na mesma tabela, pode-se observar que os retornos 198, 248 e 401 detectados como influentes são valores de "grande" magnitude em valor absoluto, e que os resíduos de predição correspondentes a estes retornos também são relativamente grandes, mas não são significativos ao nível de 5%. Estas três observações têm relação com fatores históricos. Assim, em outubro de 1997, a crise da Ásia afetou o mercado financeiro mundial. A forte queda das bolsas em outubro de 1998 e o atentado contra as torres do World Trade Center em setembro de 2001 também foram detectados pela estatística proposta. Adicionalmente, os limiares obtidos com 500 e 1000 replicações detectaram as mesmas observações e são valores bastante próximos.

Tabela 5.6: Valores influentes identificados: data de ocorrência, localização, retorno, valor da estatística, resíduos de predição e resíduos auxiliares.

Série	Data	Local.	Retorno	Estatística	Res. pred.	Res. auxil.
	Filtro de Kalman					
Austrália	22/10/1997	198	3,2714	4,2489	1,5787	0,8137
	19/09/2001	401	4,5253	3,9081	1,6236	-1,0232
	26/09/2001	402	0,4474	3,5312	-0,4777	-0,9140
N. Zelândia	14/10/1998	248	-5,3533	3,3029	1,8229	0,4141
	Suavizador de Kalman					
Austrália	12/09/2001	400	0,9528	2,7647	0,2569	-0,3930
	19/09/2001	401	4,5253	3,6120	1,6236	-1,0232
	26/09/2001	402	0,4474	2,7627	-0,4777	-0,9140
N. Zelândia	23/09/1998	245	2,7738	2,3406	1,3176	1,1275
	30/09/1998	246	-0,8400	2,5428	0,2176	1,1109
	07/10/1998	247	-1,7606	2,8232	0,8670	0,9053
	14/10/1998	248	-5,3533	3,1649	1,8229	0,4141
	21/10/1998	249	1,6486	2,6059	0,7205	0,2223

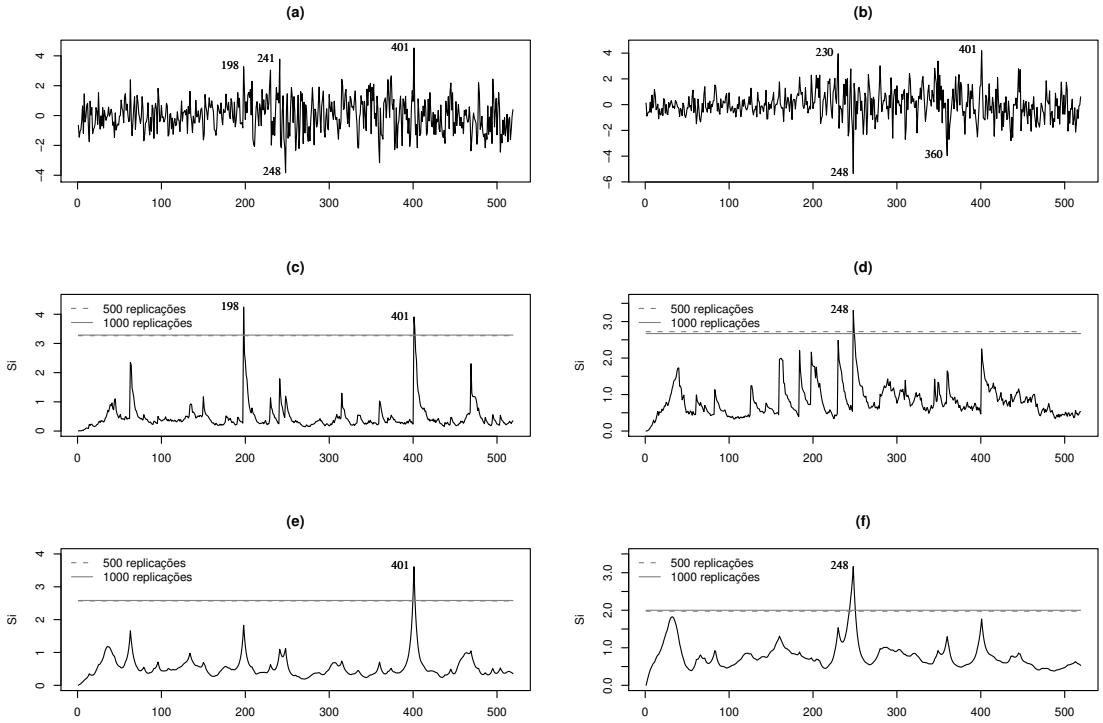


Figura 5.7: Figuras de influência para os retornos do dólar: (a) série da Austrália, (b) série da Nova Zelândia, (c) e (d) estatísticas de influência filtradas para a 1^a série e (e) e (f) estatísticas de influência suavizadas para a 2^a série, e os limiares obtidos com 500 e 1000 replicações.

5.4.1.2 Modelo com correlações constantes generalizado

Os resultados do ajuste do MVE-CCG (3.16) são apresentados na Tabela 5.7. As estimativas são significativas segundo os quocientes t. As estimativas das correlações ρ_ε e ρ_η são altamente significativas, o que justifica a adoção deste modelo no lugar dos modelos de volatilidade univariados.

A Figura 5.8 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. Os gráficos de autocorrelação mostram que os erros de predição são não-correlacionados, o que indica que o modelo está corretamente especificado. Também, não há mudança estrutural significativa no modelo a 5%, uma vez que as somas cumulativas não ultrapassam as linhas de significância em nenhum ponto e são estáveis ao longo do tempo.

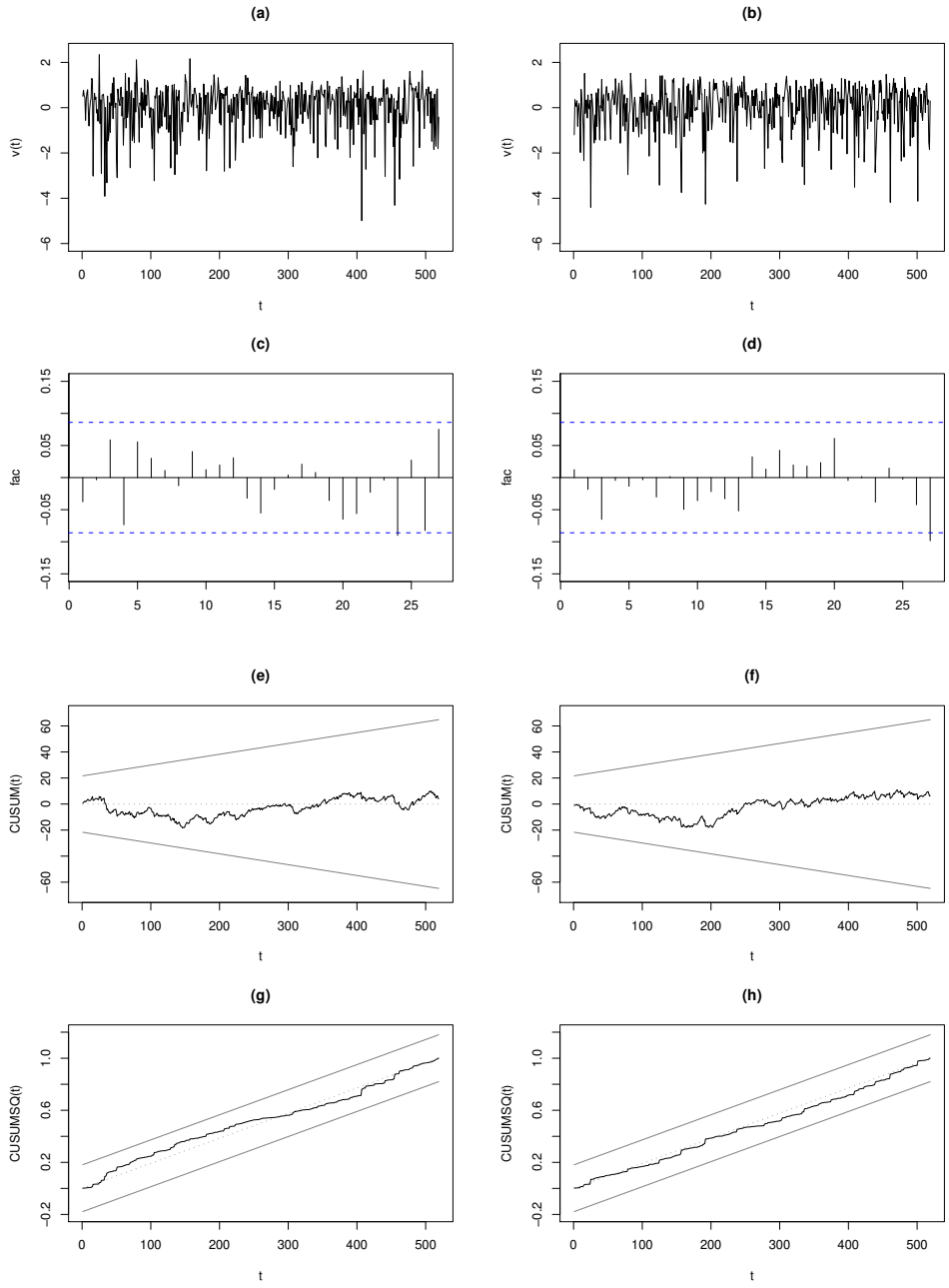


Figura 5.8: Análise residual do ajuste do MVE-CCG às séries de retornos do dólar da Austrália e da Nova Zelândia: (a) e (b) resíduos das observações da 1ª e 2ª série, (c) e (d) autocorrelação dos resíduos da 1ª e 2ª série, (e) e (f) soma cumulativa da 1ª e 2ª série e as linhas de significância a 5% e (g) e (h) soma dos quadrados cumulativa da 1ª e 2ª série e as linhas de significância a 5%.

Tabela 5.7: Estimativas dos parâmetros para as séries de retornos do dólar da Austrália e da Nova Zelândia e seus respectivos erros-padrão e razão-t.

Parâmetro	Estimativa	Erro-padrão	Razão-t
β_1	1,0565	0,1253	8,43
β_2	0,8638	0,1237	6,98
ϕ_{11}	0,9778	0,0158	62,0
ϕ_{22}	0,9841	0,0112	88,0
σ_1	0,1195	0,0498	2,40
σ_2	0,1079	0,0389	2,78
ρ_ε	0,7398	0,0440	16,8
ρ_η	0,8693	0,1307	6,65

Na Figura 5.9 temos os valores absolutos de cada retorno e as respectivas estimativas filtradas e suavizadas das volatilidades. Podemos verificar que estas conseguem captar os períodos de alta e baixa volatilidade nos retornos de cada série analisada.

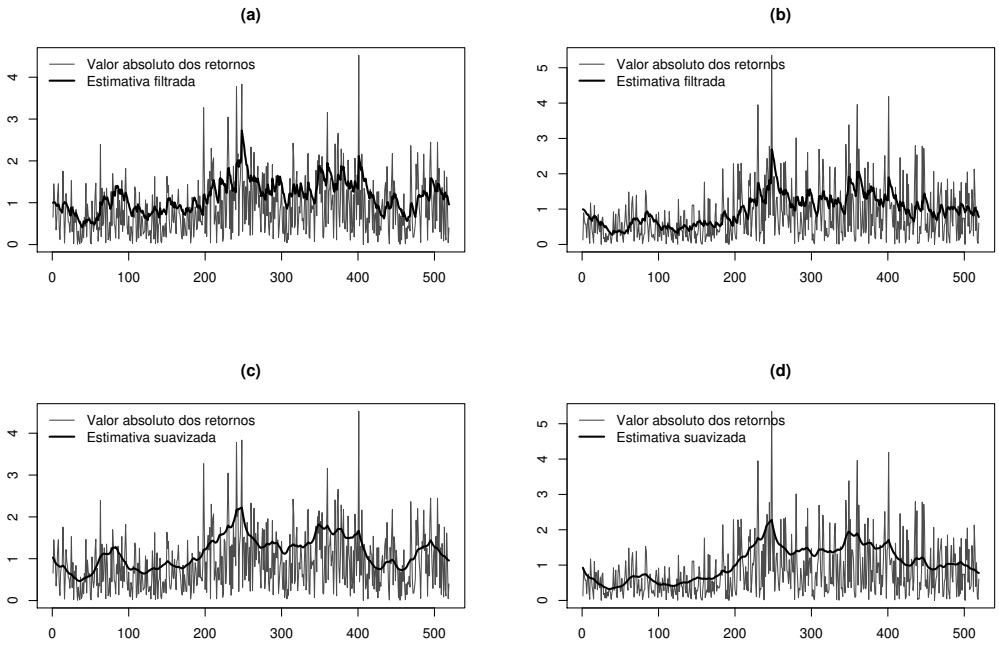


Figura 5.9: (a) e (c) Estimativas filtrada e suavizada da volatilidade e valores absolutos dos retornos do dólar da Austrália e (b) e (d) estimativas filtrada e suavizada da volatilidade e valores absolutos dos retornos do dólar da Nova Zelândia.

A Figura 5.10 ilustra as séries de retornos e a análise de influência para as duas séries. Na Tabela 5.8 pode-se verificar a data de ocorrência, localização, valor da estatística, valor do retorno, resíduo de predição e resíduo auxiliar do estado de cada observação detectada como influente. Através da tabela e das Figuras 5.10 (c) e (f) correspondentes a primeira série, pode-se ver que a estatística individual obtida pelo Filtro de Kalman detectou as observações 198 e 401 como influentes, enquanto que a estatística obtida pelo Suavizador de Kalman detectou o bloco 400 a 402, ou seja, detectou também as observações vizinhas da 401, e não detectou a observação 198. Na Tabela 5.8 e nas Figuras 5.10 (d) e (g) correspondentes a segunda série, pode-se notar que a estatística obtida pelo Filtro de Kalman detectou as observações 248, 249 e 401, enquanto que a estatística obtida pelo Suavizador de Kalman detectou o bloco 245 a 250 e a observação 401, ou seja, detectou a vizinhança da observação 248. Na Tabela 5.8 e nas Figuras 5.10 (e) e (h), pode-se observar que a estatística global obtida pelo Filtro de Kalman detectou as observações 198, 248, 401 e 402, e a estatística global obtida pelo Suavizador de Kalman detectou as observações 247, 248, 400, 401 e 402. Na Tabela 5.8, pode-se observar que os retornos 198, 248 e 401 detectados como influentes são valores de "grande" magnitude em valor absoluto, e que apenas o resíduo auxiliar do estado correspondente às observações 198 e 245 podem ser considerados valores de "grande" magnitude, mas não significativos ao nível de 5%. Novamente, os limiares obtidos com 500 e 1000 replicações detectaram as mesmas observações e são valores próximos.

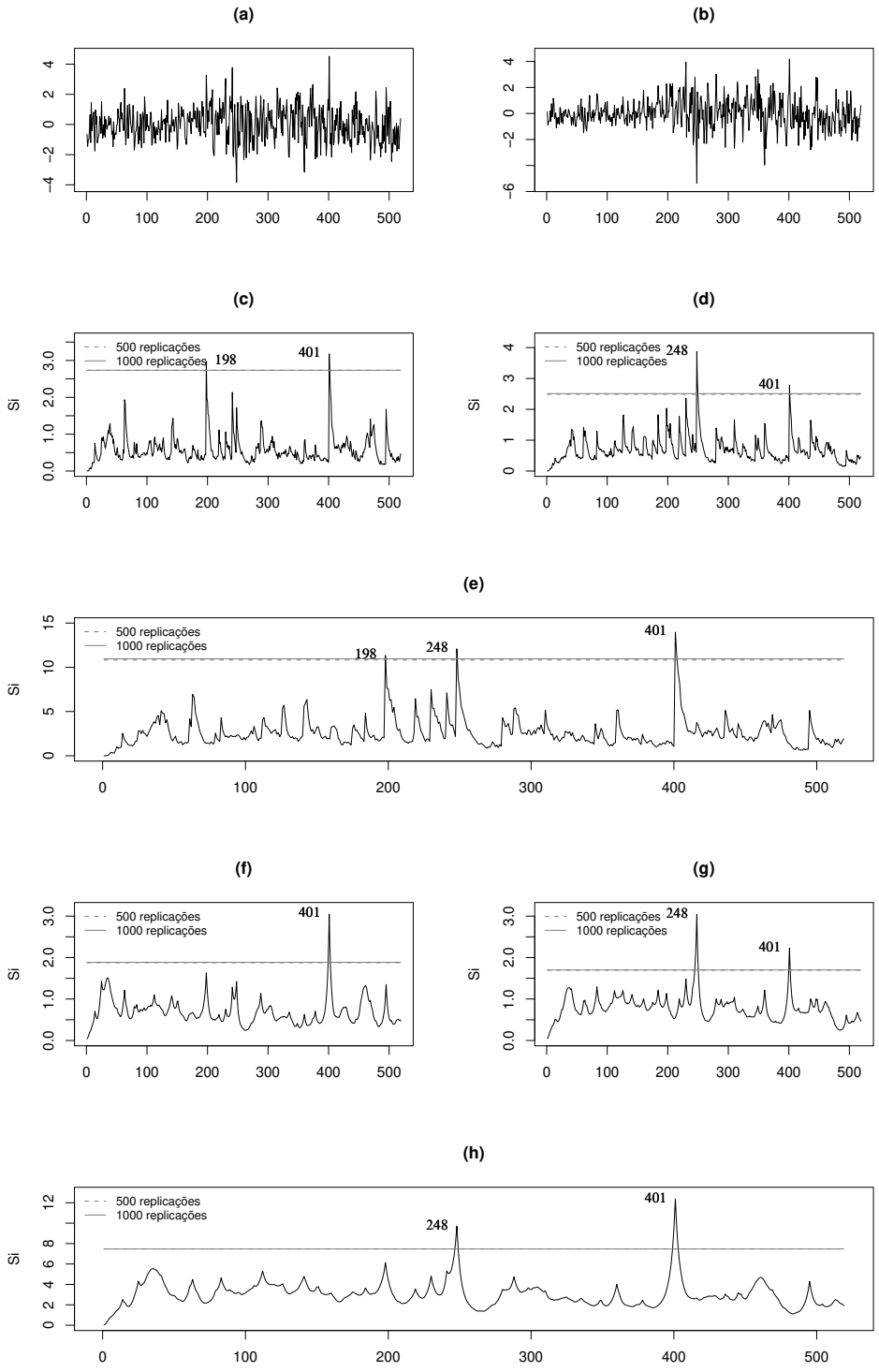


Figura 5.10: Figuras de influência para os retornos do dólar: (a) série da Austrália, (b) série da Nova Zelândia, (c) e (d) estatísticas de influência individuais filtradas e (e) estatística de influência global filtrada, (f) e (g) estatísticas de influência individuais suavizadas e (h) estatística de influência global suavizada, e os limiares obtidos com 500 e 1000 replicações.

Tabela 5.8: Valores influentes identificados: data de ocorrência, localização, valor da estatística, retorno, resíduos de predição e resíduos auxiliares.

Série	Data	Local.	Estatística	Retorno	Res. pred.	Res. auxil.
	Filtro de Kalman					
Austrália	22/10/1997	198	2,9762	3,2714	0,5743	1,5904
	19/09/2001	401	3,1786	4,5253	0,5497	-0,9474
	14/10/1998	248	3,8747	-5,3533	0,6802	0,5502
N. Zelândia	21/10/1998	249	2,8575	1,6486	0,4631	0,6278
	19/09/2001	401	2,7838	4,1859	0,5820	-0,7802
Austrália	22/10/1997	198	11,344			
e	14/10/1998	248	12,089			
N. Zelândia	19/09/2001	401	13,974			
	26/09/2001	402	11,296			
	Suavizador de Kalman					
	12/09/2001	400	2,3194	0,9528	0,0154	-0,3187
Austrália	19/09/2001	401	3,0494	4,5253	0,5497	-0,9474
	26/09/2001	402	2,3210	0,4474	-0,1789	-0,9765
	23/09/1998	245	1,8283	2,7738	0,5592	1,5736
	30/09/1998	246	2,0941	-0,8400	0,0134	1,4975
	07/10/1998	247	2,4860	-1,7606	0,2860	1,1998
N. Zelândia	14/10/1998	248	3,0359	-5,3533	0,6802	0,5502
	21/10/1998	249	2,3451	1,6486	0,4631	0,6278
	28/10/1998	250	1,8513	0,8176	-0,0290	0,5509
	19/09/2001	401	2,2285	4,1859	0,5820	-0,7802
	07/10/1998	247	7,9464			
Austrália	14/10/1998	248	9,7011			
e	12/09/2001	400	9,4288			
N. Zelândia	19/09/2001	401	12,354			
	26/09/2001	402	9,4725			

5.4.1.3 Modelo fatorial aditivo

Os resultados do ajuste do MVE-FA (3.17) são apresentados na Tabela 5.9. As estimativas são significativas segundo os quocientes t.

Tabela 5.9: Estimativas dos parâmetros para as séries de retornos do dólar da Austrália e da Nova Zelândia e seus respectivos erros-padrão.

Parâmetro	Estimativa	Erro-padrão	Razão-t
d	1,1730	0,0621	18,89
β	0,6560	0,0570	11,50
ϕ	0,9917	0,0057	173,0
σ_{ε_1}	0,6546	0,0328	19,95
σ_{ε_2}	0,3342	0,0679	4,921
σ_{η}	0,1091	0,0243	4,490

A Figura 5.11 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. Pela Figura 5.11 (b), temos correlação nas defasagens 3 e 9, mas o teste de Ljung-Box reportou p-valores iguais a 0,12, 0,30, 0,18 e 0,24 correspondentes às defasagens 3, 6, 9 e 12, respectivamente. Também, as Figuras 5.11 (c) e (d) indicam que não há mudança estrutural significativa no modelo a 5%, uma vez que as somas cumulativas não ultrapassam as linhas de significância em nenhum ponto e são estáveis ao longo do tempo.

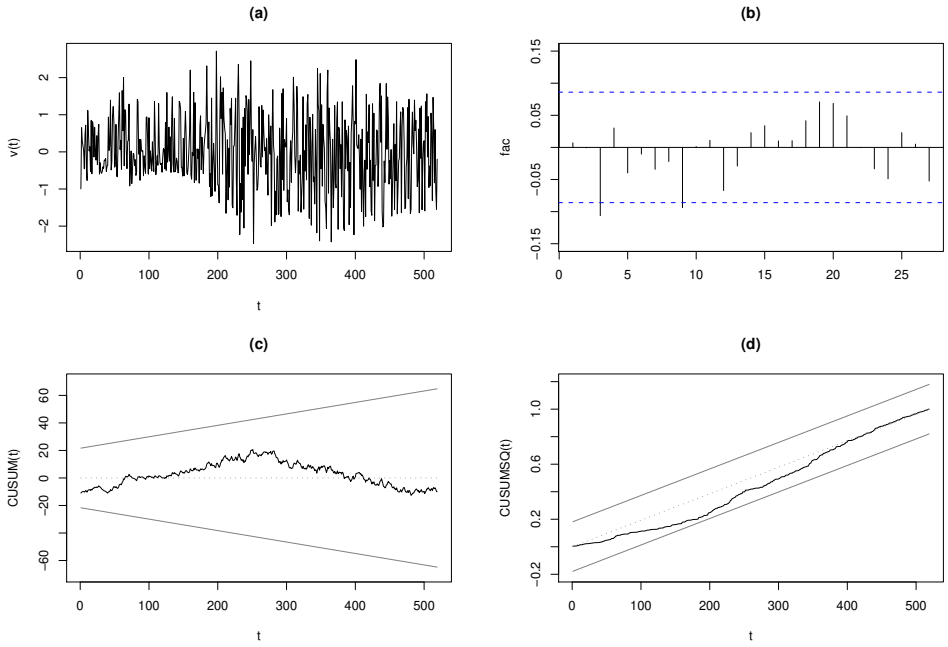


Figura 5.11: Análise residual do ajuste do MVE-FA às séries de retornos do dólar da Austrália e da Nova Zelândia: (a) resíduos das observações, (b) autocorrelação dos resíduos, (c) soma cumulativa e (d) soma dos quadrados cumulativa, e as respectivas linhas de significância a 5%.

Na Figura 5.12 temos a estimativa filtrada e suavizada da volatilidade. Podemos verificar que estas conseguem captar os períodos de alta e baixa volatilidade nos retornos das duas séries.

Neste modelo, o coeficiente de correlação entre os retornos observados é variante no tempo, o que é confirmado pelo gráfico da correlação *versus* o tempo ilustrado na Figura 5.13. Pode-se observar que a correlação é uma função crescente de h_t, ou seja, a medida que a volatilidade observada na Figura 5.12 (c) aumenta, a correlação também aumenta. Este gráfico indica que as duas séries de retornos estão bastante relacionadas, uma vez que a correlação é alta em vários períodos.

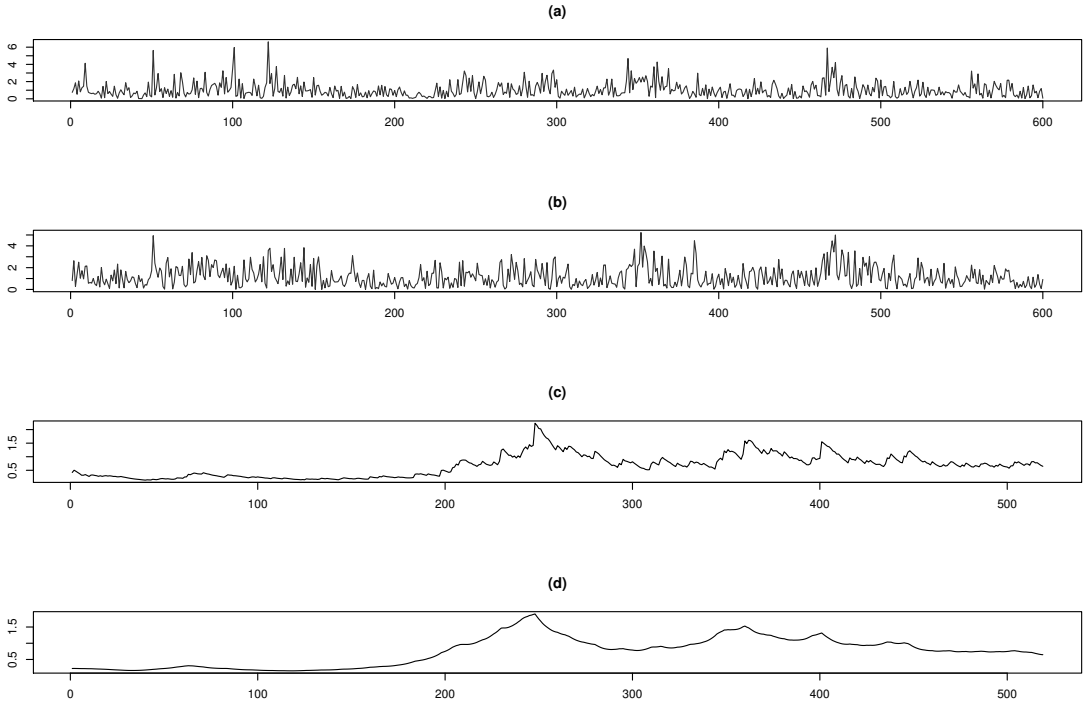


Figura 5.12: (a) e (b) Valores absolutos dos retornos, (c) estimativa filtrada e (d) estimativa suavizada da volatilidade comum dos retornos do dólar da Austrália e da Nova Zelândia.

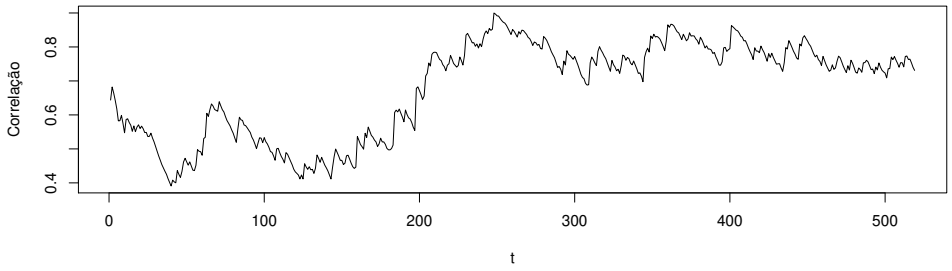


Figura 5.13: Correlação entre os retornos observados.

A Figura 5.14 ilustra as séries de retornos e a estatística de influência proposta. Na Tabela 5.10 pode-se verificar a data de ocorrência, localização, valor da estatística, resíduo de predição e resíduo auxiliar do estado de cada observação detectada como influente. Através dessa tabela e das Figuras 5.14 (c) e (d) pode-se ver que a estatística global obtida através do Filtro de Kalman detectou as observações 198, 248, 401 e 402, e a

estatística global obtida pelo Suavizador de Kalman detectou os blocos 246 a 250 e 399 a 403, ou seja, a vizinhança das observações 248, 401 e 402, e não detectou a observação 198. Na mesma tabela, pode-se observar que os resíduos de predição correspondentes às observações 198, 248 e 401 são valores significativos ao nível de 5%. Novamente, os limiares obtidos com 500 e 1000 replicações detectaram as mesmas observações e são valores próximos.

Tabela 5.10: Valores influentes identificados: data de ocorrência, localização, valor da estatística e resíduos.

Data	Localização	Estatística	Res. predição	Res. auxiliar
Filtro de Kalman				
22/10/1997	198	3,0130	2,9469	1,6436
14/10/1998	248	3,2069	2,6611	-1,1204
19/09/2001	401	3,4867	2,6970	-1,0855
26/09/2001	402	2,5230	0,2081	-1,0453
Suavizador de Kalman				
30/09/1998	246	1,5456	-0,0727	1,2667
07/10/1998	247	1,8331	0,9467	1,1893
14/10/1998	248	2,1894	2,6611	-1,1204
21/10/1998	249	1,8595	0,2464	-1,0967
28/10/1998	250	1,5731	-0,2752	-0,9348
05/09/2001	399	1,6720	0,8932	1,2770
12/09/2001	400	1,9501	0,8161	1,2492
19/09/2001	401	2,3055	2,6970	-1,0855
26/09/2001	402	1,9666	0,2081	-1,0453
03/10/2001	403	1,6779	0,1254	-0,9740

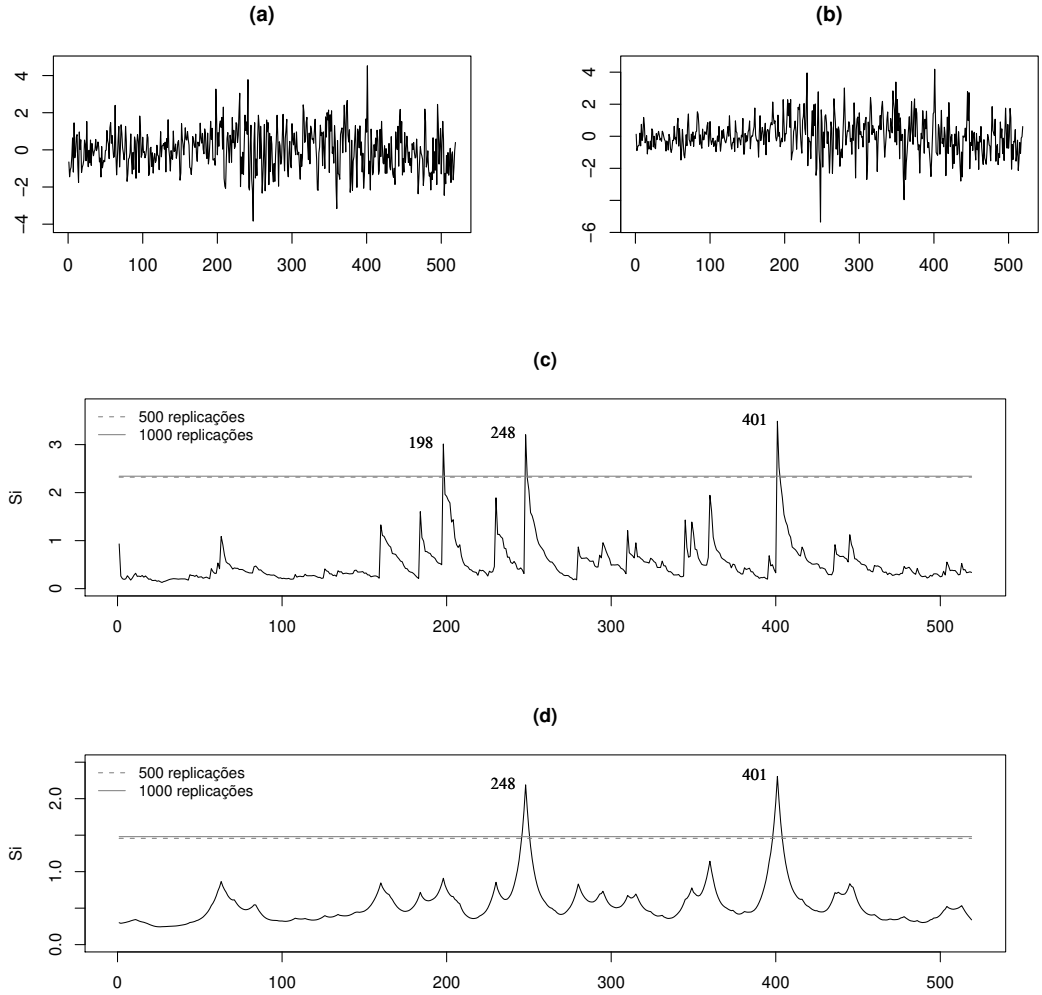


Figura 5.14: Figuras de influência para os retornos do dólar: (a) série da Austrália, (b) série da Nova Zelândia, (c) estatística de influência filtrada e (d) estatística de influência suavizada, e os limiares obtidos com 500 e 1000 replicações.

Através na análise das séries dos retornos semanais do dólar da Austrália e da Nova Zelândia, podemos ver que, tanto no ajuste do modelo de volatilidade univariado quanto no ajuste dos modelos multivariados, as observações com "grande" magnitude em valor absoluto e que foram detectadas como influentes estão relacionadas à crises que impactaram diretamente o mercado financeiro, e são as observações 198, 248 e 401. Tanto no caso do Filtro quanto do Suavizador de Kalman, algumas observações vizinhas destas observações foram também detectadas, em especial no caso do Suavizador de Kalman. Já os resíduos de predição correspondentes a estas três observações são valores significativos

ao nível de 5% apenas no MVE-FA.

5.4.2 Análise das séries dos retornos de Nova Iorque e da Inglaterra

Agora, será apresentada a análise das séries diárias dos retornos da bolsa de valores de Nova Iorque (NYSE) e da Inglaterra (FTSE), no período de julho de 1999 a outubro de 2001, num total de 600 observações. A Figura 5.15 ilustra as séries de retornos.

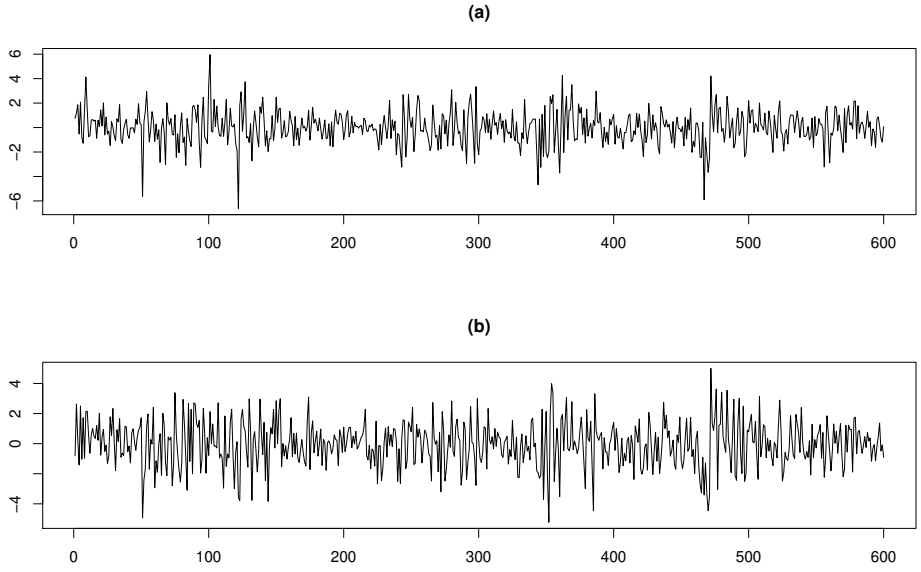


Figura 5.15: Retornos diários (a) da NYSE e (b) da FTSE.

5.4.2.1 Modelos univariados

Os resultados do ajuste do MVE-U (3.2) - (3.3) são apresentados na Tabela 5.11. As estimativas são significativas segundo os quocientes t.

A Figura 5.16 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. Pela Figura 5.16 (c), temos correlação nas defasagens 2 e 10, mas o teste de Ljung-Box reportou p-valores iguais a 0,06, 0,41, 0,54, 0,47 e 0,09 correspondentes às defasagens 2, 4, 6, 8 e 10, respectivamente. Também, pela Figura 5.16 (d), temos correlação na defasagem 4, mas o teste de Ljung-Box reportou um p-valor igual a 0,06. Também, não há mudança estrutural significativa no modelo a 5%, uma vez que as somas

cumulativas não ultrapassam as linhas de significância em nenhum ponto e são estáveis ao longo do tempo.

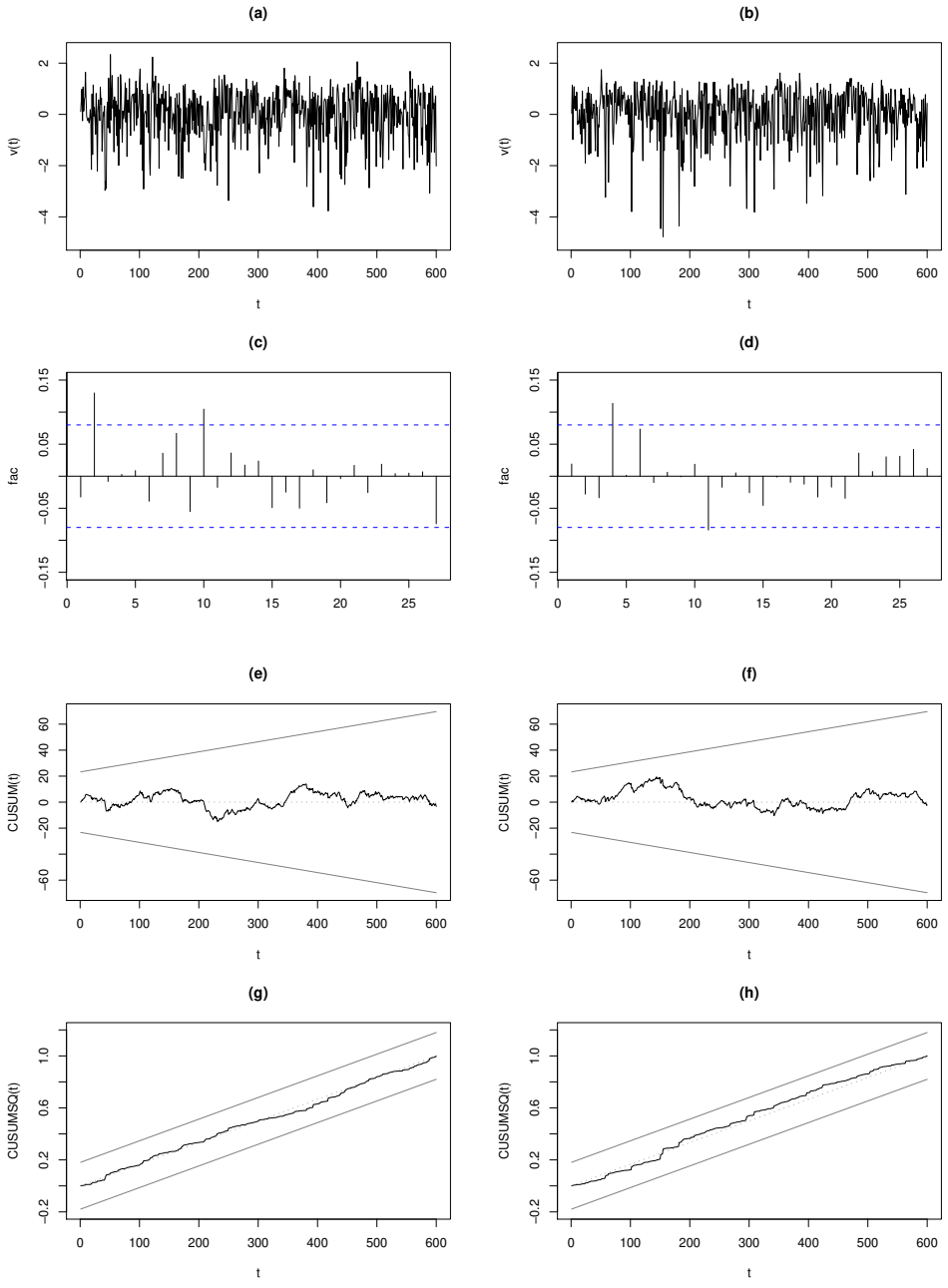


Figura 5.16: Análise residual do ajuste do MVE-U às séries de retornos da NYSE e da FTSE: (a) e (b) resíduos das observações da 1ª e 2ª série, (c) e (d) autocorrelação dos resíduos da 1ª e 2ª série, (e) e (f) soma cumulativa da 1ª e 2ª série e as linhas de significância a 5% e (g) e (h) soma dos quadrados cumulativa da 1ª e 2ª série e as linhas de significância a 5%.

Tabela 5.11: Estimativas dos parâmetros do MVE-U para as séries de retornos da NYSE e da FTSE e seus respectivos erros-padrão e razão-t.

Série	Parâmetro	Estimativa	Erro-padrão	Razão-t
NYSE	β	1,2664	0,1093	11,6
	ϕ	0,9362	0,0332	28,2
	σ	0,2365	0,0751	3,15
FTSE	β	1,4594	0,1191	12,3
	ϕ	0,9403	0,0436	21,6
	σ	0,2053	0,0948	2,16

Para cada uma das série é realizada a análise de influência utilizando os limiares descritos na Seção 5.2 e assim, espera-se identificar os valores atípicos. A Figura 5.17 ilustra as séries de retornos e a estatística de influência obtidas através do Filtro e Suavizador de Kalman. Nas Figuras 5.17 (a) e (b) pode-se observar que algumas observações parecem ser influentes, como por exemplo as observações 51, 101, 122 e 467 da primeira série e as observações 51, 352, 385 e 472 da segunda série. Na Tabela 5.12 pode-se verificar a data de ocorrência, localização, valor do retorno, valor da estatística, resíduo de predição e resíduo auxiliar do estado de cada observação detectada como influente. Através da tabela e das Figuras 5.17 (c) e (e) correspondentes a primeira série, pode-se ver que a estatística obtida pelo Filtro de Kalman detectou as observações 51, 52, 122 e 467 como influentes, enquanto que a estatística obtida pelo Suavizador de Kalman detectou o bloco 50 a 52 e a observação 122, e não detectou a observação 467. Nas Figuras 5.17 (d) e (f), pode-se notar que a estatística obtida pelo Filtro de Kalman detectou apenas a observação 51, enquanto que a estatística obtida pelo Suavizador de Kalman não detectou nenhuma observação. Na mesma tabela, pode-se observar que apenas os retornos 51, 122 e 467 detectados como influentes são valores de "grande" magnitude em valor absoluto, e que apenas os resíduos de predição das observações 51 e 122 são significativos ao nível de 5%. Dentre estas três observações, podemos citar a queda da bolsa Nasdaq em abril de 2000 e o atentado contra as torres do World Trade Center, quando o mercado reabriu apenas no dia 17 de setembro de 2001. Também, no início do ano 2000, algumas das principais bolsas despencaram devido a temores de elevação dos juros nos Estados Unidos e na Europa, e isto afetou tanto o mercado de Nova Iorque quanto o mercado da Inglaterra.

Adicionalmente, os limiares obtidos com 500 e 1000 replicações detectaram as mesmas observações e são valores bastante próximos.

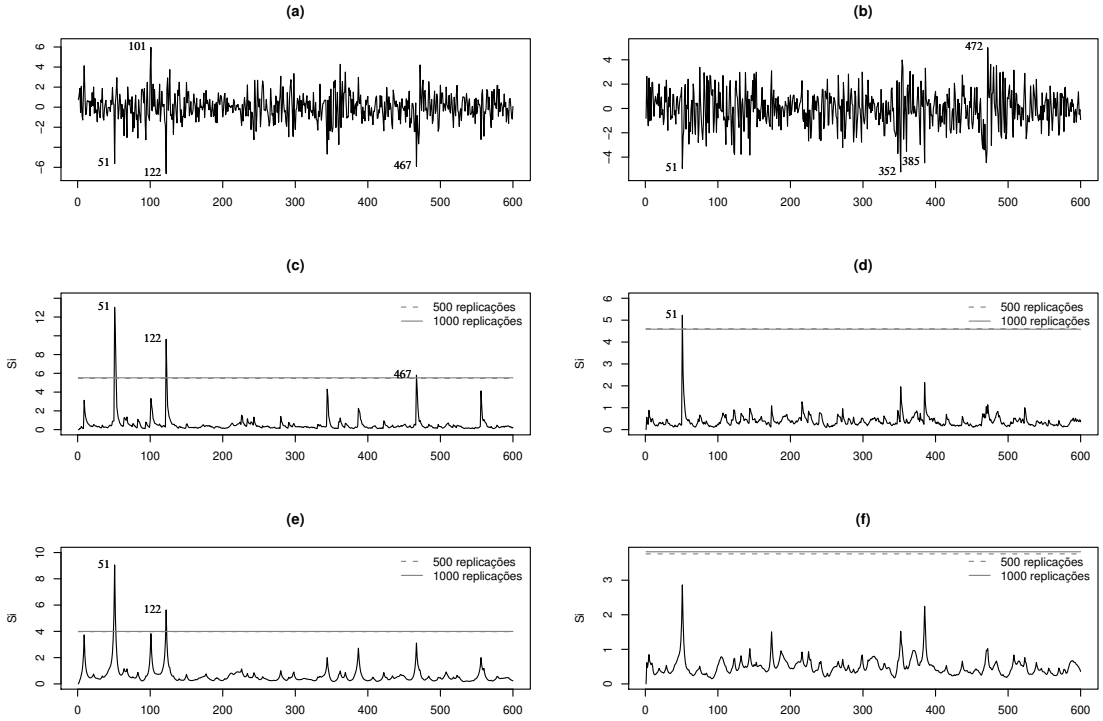


Figura 5.17: Figuras de influência para os retornos: (a) série da NYSE, (b) série da FTSE, (c) e (d) estatísticas de influência filtradas para a 1ª série e (e) e (f) estatísticas de influência suavizadas para a 2ª série, e os limiares obtidos com 500 e 1000 replicações.

Tabela 5.12: Valores influentes identificados: data de ocorrência, localização, retorno, valor da estatística e resíduos.

Série	Data	Local.	Retorno	Estatística	Res. predição	Res. auxiliar
	Filtro de Kalman					
NYSE	04/01/2000	51	-5,6193	13,037	2,1425	0,6926
	05/01/2000	52	0,4455	10,121	-0,2091	0,9003
	14/04/2000	122	-6,6197	9,6312	2,0507	1,0897
	17/09/2001	467	-5,9002	5,8050	1,8854	1,4142
FTSE	04/01/2000	51	-4,9364	5,2265	1,6677	0,7544
	Suavizador de Kalman					
NYSE	30/12/1999	50	0,2289	5,2899	-0,6717	1,6281
	04/01/2000	51	-5,6193	9,0572	2,1425	0,6926
	05/01/2000	52	0,4455	5,2481	-0,2091	0,9003
	14/04/2000	122	-6,6197	5,6303	2,0507	1,0897

5.4.2.2 Modelo com correlações constantes generalizado

Os resultados do ajuste do MVE-CCG (3.16) são apresentados na Tabela 5.13. As estimativas são significativas segundo os quocientes t. As estimativas das correlações ρ_ε e ρ_η são altamente significativas, o que justifica a adoção deste modelo no lugar do modelo de volatilidade univariado.

A Figura 5.18 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. Pela Figura 5.18 (c), temos correlação nas defasagens 2 e 10, mas o teste de Ljung-Box reportou p-valores iguais a 0,06, 0,25, 0,35, 0,39 e 0,08 correspondentes às defasagens 2, 4, 6, 8 e 10, respectivamente. Também, pela Figura 5.18 (d), temos correlação na defasagem 4, mas o teste de Ljung-Box reportou um p-valor igual a 0,11. Também, não há mudança estrutural significativa no modelo a 5%, uma vez que as somas cumulativas não ultrapassam as linhas de significância em nenhum ponto e são estáveis ao longo do tempo.

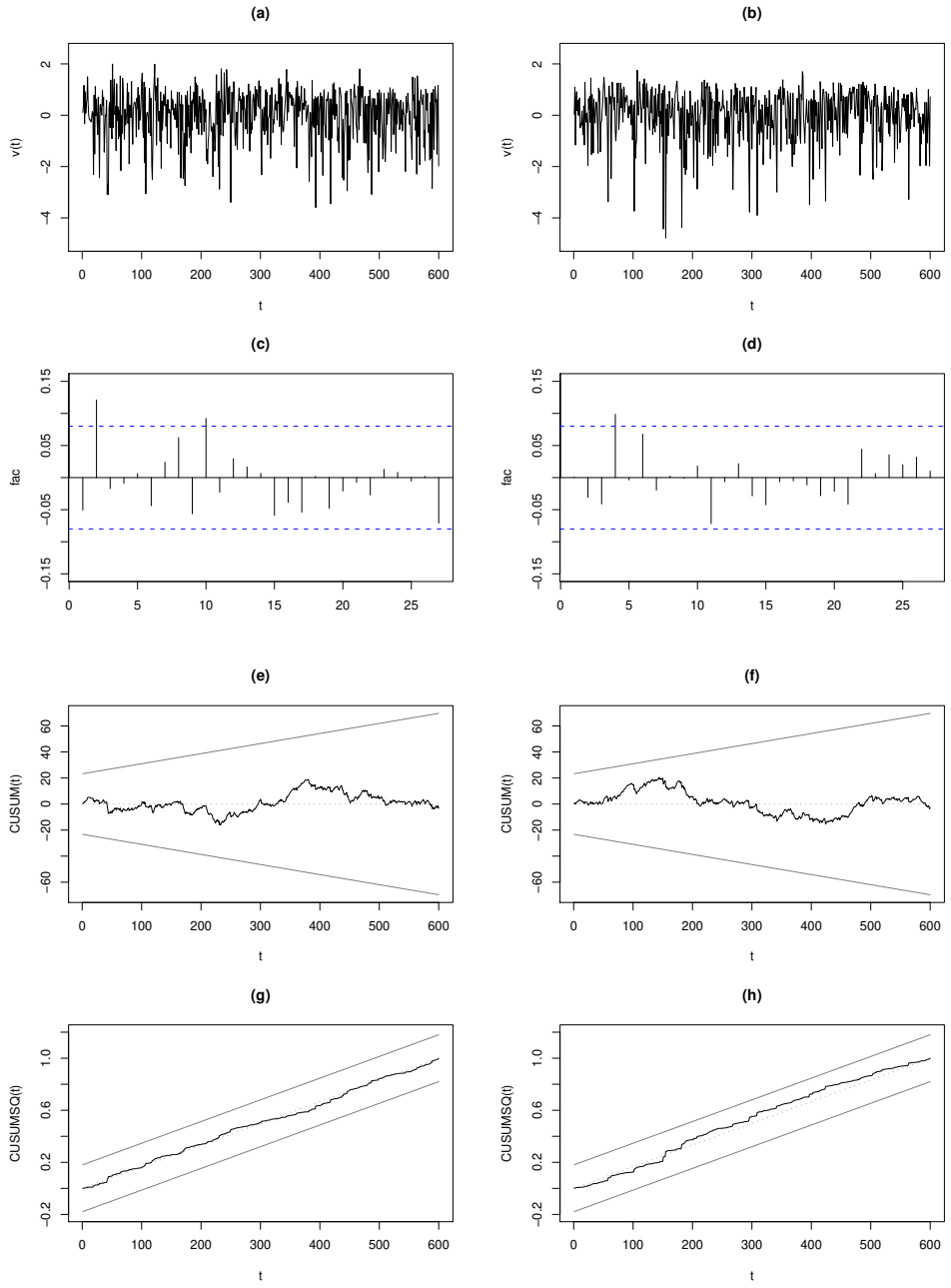


Figura 5.18: Análise residual do ajuste do MVE-CCG às séries de retornos da NYSE e da FTSE: (a) e (b) resíduos das observações da 1^a e 2^a série, (c) e (d) autocorrelação dos resíduos da 1^a e 2^a série, (e) e (f) soma cumulativa da 1^a e 2^a série e as linhas de significância a 5% e (g) e (h) soma dos quadrados cumulativa da 1^a e 2^a série e as linhas de significância a 5%.

Tabela 5.13: Estimativas dos parâmetros para as séries de retornos da NYSE e da FTSE e seus respectivos erros-padrão e razão-t.

Parâmetro	Estimativa	Erro-padrão	Razão-t
β_1	1,2690	0,1172	10,8
β_2	1,4703	0,1127	13,0
ϕ_{11}	0,9467	0,0274	34,6
ϕ_{22}	0,9201	0,0452	20,4
σ_1	0,2184	0,0648	3,37
σ_2	0,2486	0,0890	2,79
ρ_ε	0,5567	0,0897	6,21
ρ_η	0,8942	0,1058	8,45

Na Figura 5.19 temos os valores absolutos de cada retorno e as respectivas estimativas filtradas e suavizadas das volatilidades. Podemos verificar que estas conseguem captar os períodos de alta e baixa volatilidade nos retornos de cada série analisada.

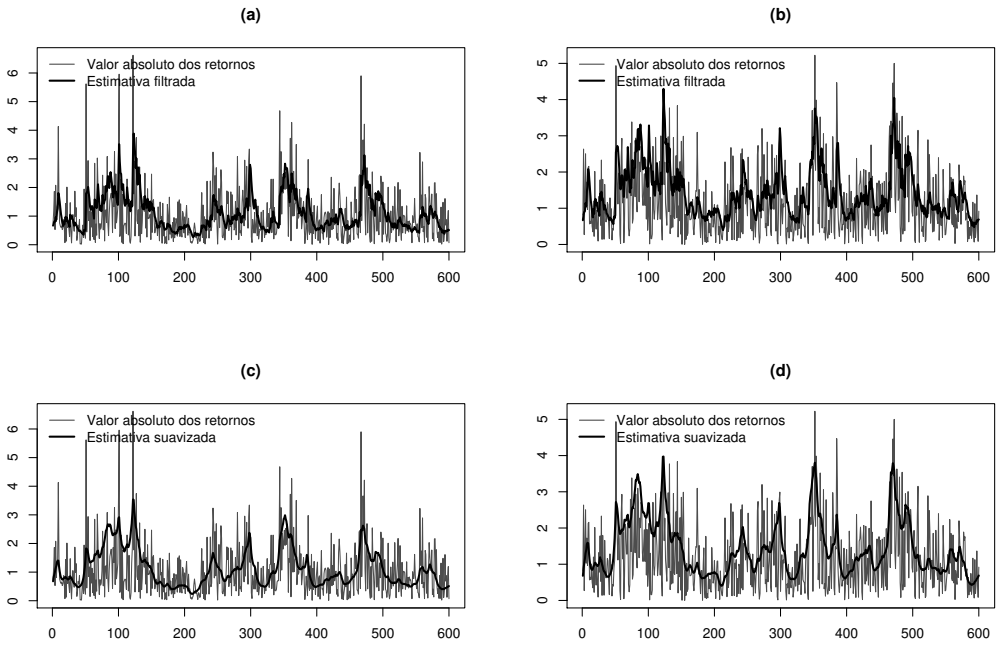


Figura 5.19: (a) e (c) Estimativas filtrada e suavizada da volatilidade e valores absolutos dos retornos do dólar da NYSE e (b) e (d) estimativas filtrada e suavizada da volatilidade e valores absolutos dos retornos do dólar da FTSE.

Para essa série é realizada a análise de influência utilizando os limiares descritos na Seção 5.2 e assim, espera-se identificar os valores atípicos. Na Tabela 5.14 podemos verificar a data de ocorrência, localização, valor da estatística, valor do retorno e resíduos de cada observação detectada como influente. Através dessa tabela e das Figuras 5.20 (c) e (f) correspondentes a primeira série, pode-se ver que a estatística individual obtida pelo Filtro de Kalman detectou as observações 51, 122 e 467 como influentes, enquanto que a estatística obtida pelo Suavizador de Kalman detectou apenas as observações 51 e 467. Nas Figuras 5.20 (d) e (g), pode-se notar que a estatística obtida pelo Filtro de Kalman detectou as observações 51 e 385, enquanto que a estatística obtida pelo Suavizador de Kalman detectou apenas a observação 385. Nas Figuras 5.20 (e) e (h), pode-se observar que a estatística global obtida pelo Filtro de Kalman conseguiu detectar as quatro observações, nas posições 51, 122, 385 e 467, e a estatística global obtida pelo Suavizador de Kalman detectou as observações 50, 51, 385 e 467, ou seja, não detectou a observação 122. Na mesma tabela, pode-se observar que os retornos 51, 122, 385 e 467 detectados como influentes são valores de "grande" magnitude em valor absoluto, e que apenas os resíduos auxiliares dos estados correspondentes às observações 122 e 467 são consideravelmente "grandes", mas não significativos ao nível de 5%. Estas observações são iguais as que foram detectadas na análise dos modelos univariados, com exceção da observação 385. Em maio de 2001, a queda da bolsa da Inglaterra ocorreu devido a possibilidade do banco central dos Estados Unidos não cortar a taxa básica de juros do país. Por fim, os limiares obtidos com 500 e 1000 replicações detectaram as mesmas observações e são valores próximos.

Tabela 5.14: Valores influentes identificados: data de ocorrência, localização, valor da estatística, retorno e resíduos.

Série	Data	Localiz.	Estatística	Retorno	Res. predição	Res. auxiliar
	Filtro de Kalman					
	04/01/2000	51	13,370	-5,6194	0,8141	0,7778
NYSE	14/04/2000	122	5,9216	-6,6197	0,8094	1,4290
	17/09/2001	467	6,4498	-5,9002	0,7353	1,8368
FTSE	04/01/2000	51	7,7073	-4,9364	0,6271	0,8290
	14/05/2001	385	6,5021	-4,4693	0,7195	-0,6071
NYSE	04/01/2000	51	49,471			
e	14/04/2000	122	19,303			
FTSE	14/05/2001	385	23,550			
	17/09/2001	467	20,857			
	Suavizador de Kalman					
NYSE	04/01/2000	51	6,0453	-5,6194	0,8141	0,7778
	17/09/2001	467	3,7746	-5,9002	0,7353	1,8368
FTSE	14/05/2001	385	3,7656	-4,4693	0,7195	-0,6071
NYSE	30/12/1999	50	11,647			
e	04/01/2000	51	21,670			
FTSE	14/05/2001	385	13,893			
	17/09/2001	467	11,942			

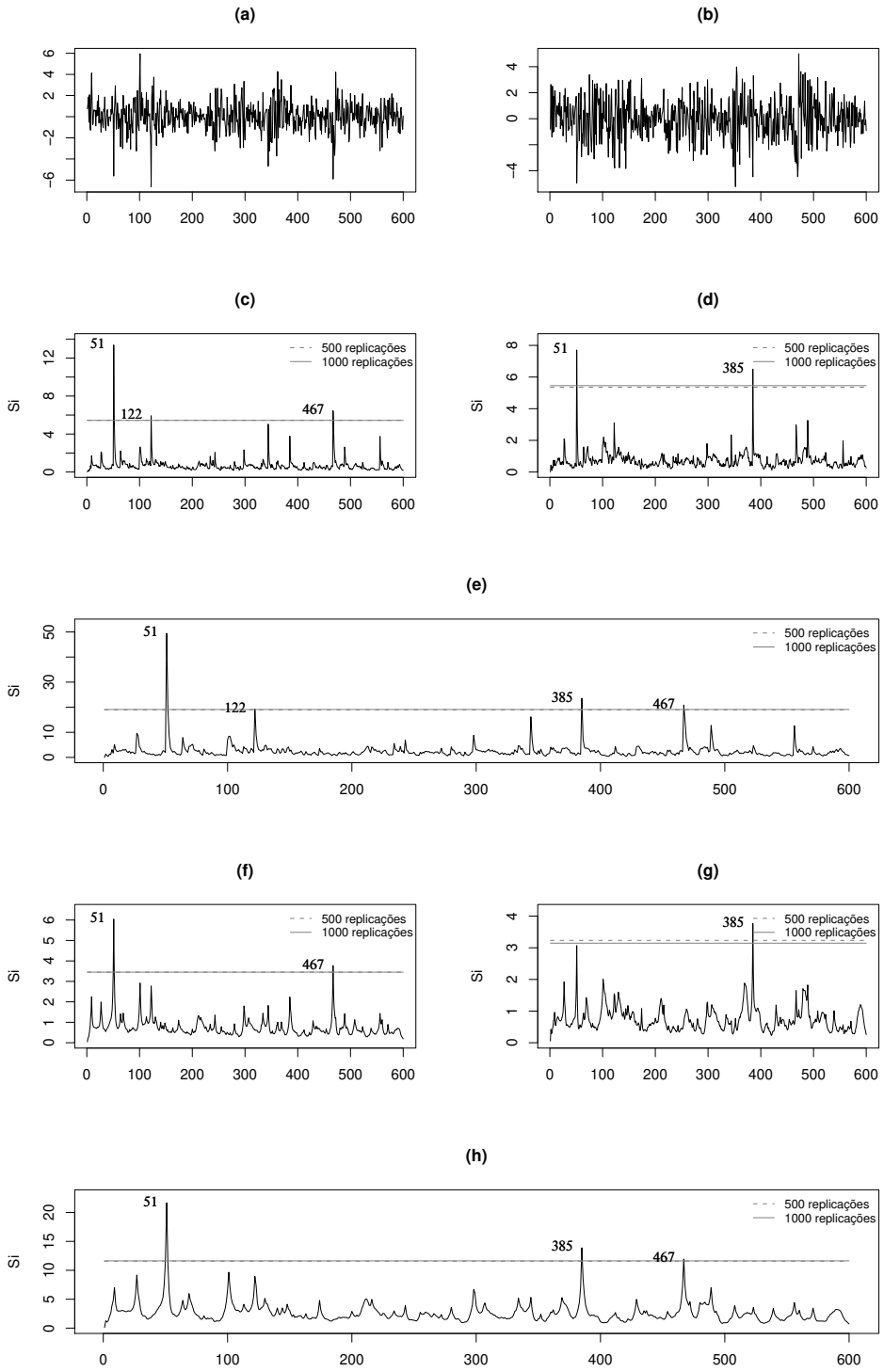


Figura 5.20: Figuras de influência para os retornos: (a) série da NYSE, (b) série da FTSE, (c) e (d) estatísticas de influência individuais filtradas e (e) estatística de influência global filtrada, (f) e (g) estatísticas de influência individuais suavizadas e (h) estatística de influência global suavizada, e os limiares obtidos com 500 e 1000 replicações.

5.4.2.3 Modelo fatorial aditivo

Os resultados do ajuste do MVE-FA são apresentados na Tabela 5.15. As estimativas são significativas segundo os quocientes t.

Tabela 5.15: Estimativas dos parâmetros para as séries de retornos da NYSE e da FTSE e seus respectivos erros-padrão.

Parâmetro	Estimativa	Erro-padrão	Razão-t
d	0,7801	0,1015	7,69
β	1,0380	0,2661	3,90
ϕ	0,9585	0,0216	44,5
σ_{ε_1}	0,7235	0,1171	6,18
σ_{ε_2}	1,2160	0,0559	21,7
σ_η	0,3233	0,0711	4,55

A Figura 5.21 fornece a análise residual através dos erros de predição, conforme vimos na Seção 2.4. Pela Figura 5.21 (b), temos correlação na defasagem 1, mas o teste de Ljung-Box reportou p-valores iguais a 0,06, 0,07, 0,13, 0,23, 0,27 e 0,38 correspondentes às defasagens 1, 2, 3, 4, 5 e 6, respectivamente. Também, as Figuras 5.21 (c) e (d) indicam que não há mudança estrutural significativa no modelo a 5%, uma vez que as somas cumulativas não ultrapassam as linhas de significância em nenhum ponto e são estáveis ao longo do tempo.

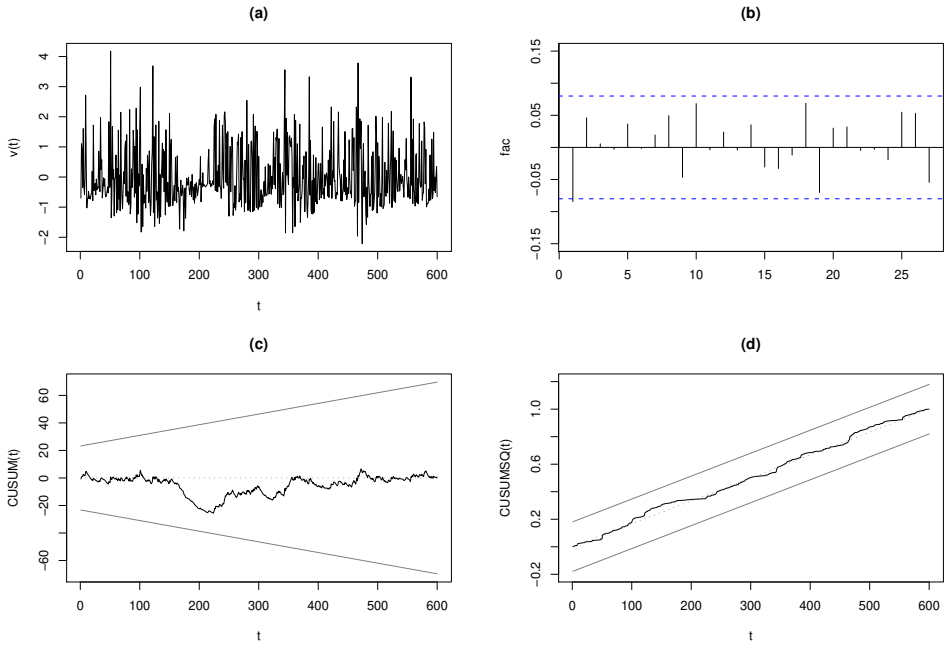


Figura 5.21: Análise residual do ajuste do MVE-FA às séries de retornos da NYSE e da FTSE: (a) resíduos das observações, (b) autocorrelação dos resíduos, (c) soma cumulativa e (d) soma dos quadrados cumulativa, e as respectivas linhas de significância a 5%.

Na Figura 5.22 temos a estimativa filtrada e suavizada da volatilidade. Podemos verificar que estas conseguem captar os períodos de alta e baixa volatilidade nos retornos das duas séries.

Neste modelo, o coeficiente de correlação entre os retornos observados é variante no tempo, o que é confirmado pelo gráfico da correlação *versus* o tempo ilustrado na Figura 5.23. Pode-se observar que a correlação é uma função crescente de h_t, e que as duas séries de retornos estão bastante relacionadas, uma vez que a correlação é alta em vários períodos.

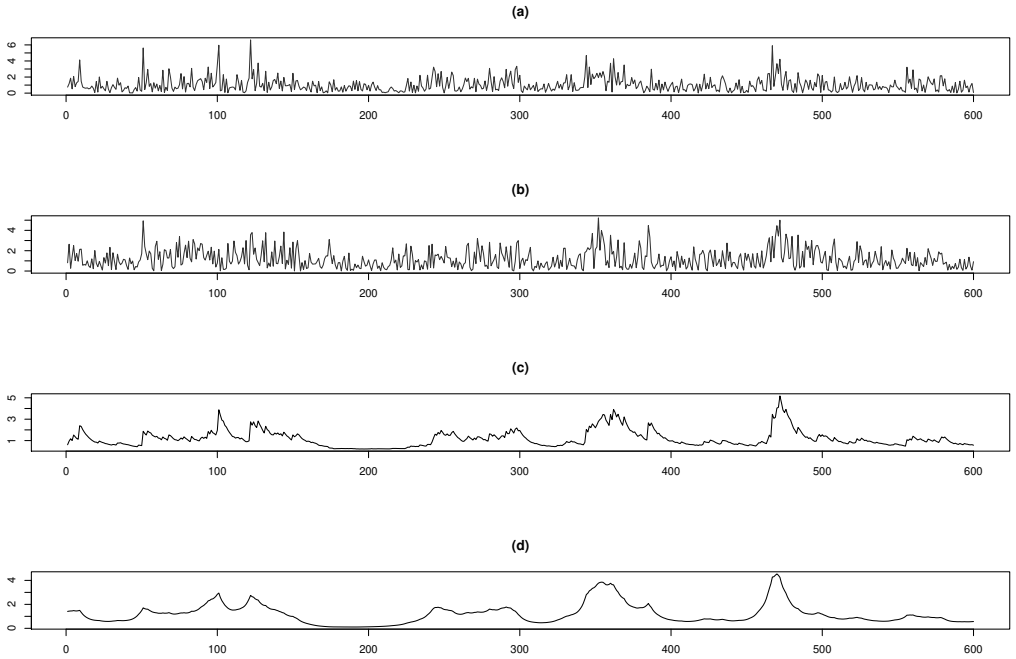


Figura 5.22: (a) e (b) Valores absolutos dos retornos e (c) estimativa filtrada e suavizada da volatilidade comum dos retornos do dólar da NYSE e da FTSE.

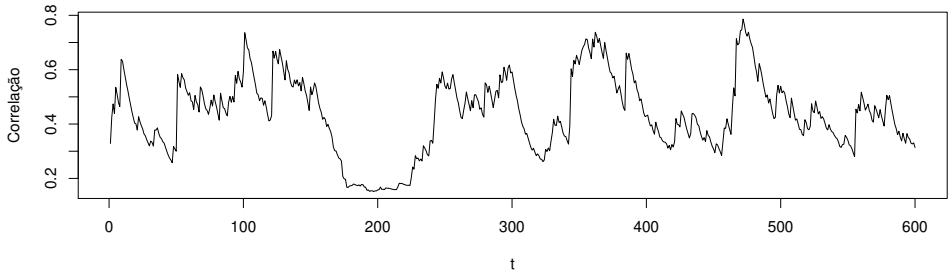


Figura 5.23: Correlação entre os retornos observados.

A Figura 5.24 ilustra as séries de retornos e a estatística de influência proposta. Na tabela 5.16 podemos verificar a data de ocorrência, localização, valor da estatística e resíduos de cada observação detectada como influente. Através da tabela e das Figuras 5.24 (c) e (d) pode-se ver que a estatística global obtida respectivamente através do Filtro de Kalman detectou as observações 51, 52, 122, 385 e 467, e a estatística global obtida pelo Suavizador de Kalman detectou as observações 50 a 52, 101 e 385, ou seja, detectou também as observações 50 e 101 e não detectou a observação 467. Na mesma tabela,

pode-se observar que os resíduos de predição correspondentes às observações 51, 101, 122, 385 e 467 são significativos ao nível de 5%. Também, os limiares obtidos com 500 e 1000 replicações detectaram as mesmas observações e são valores próximos.

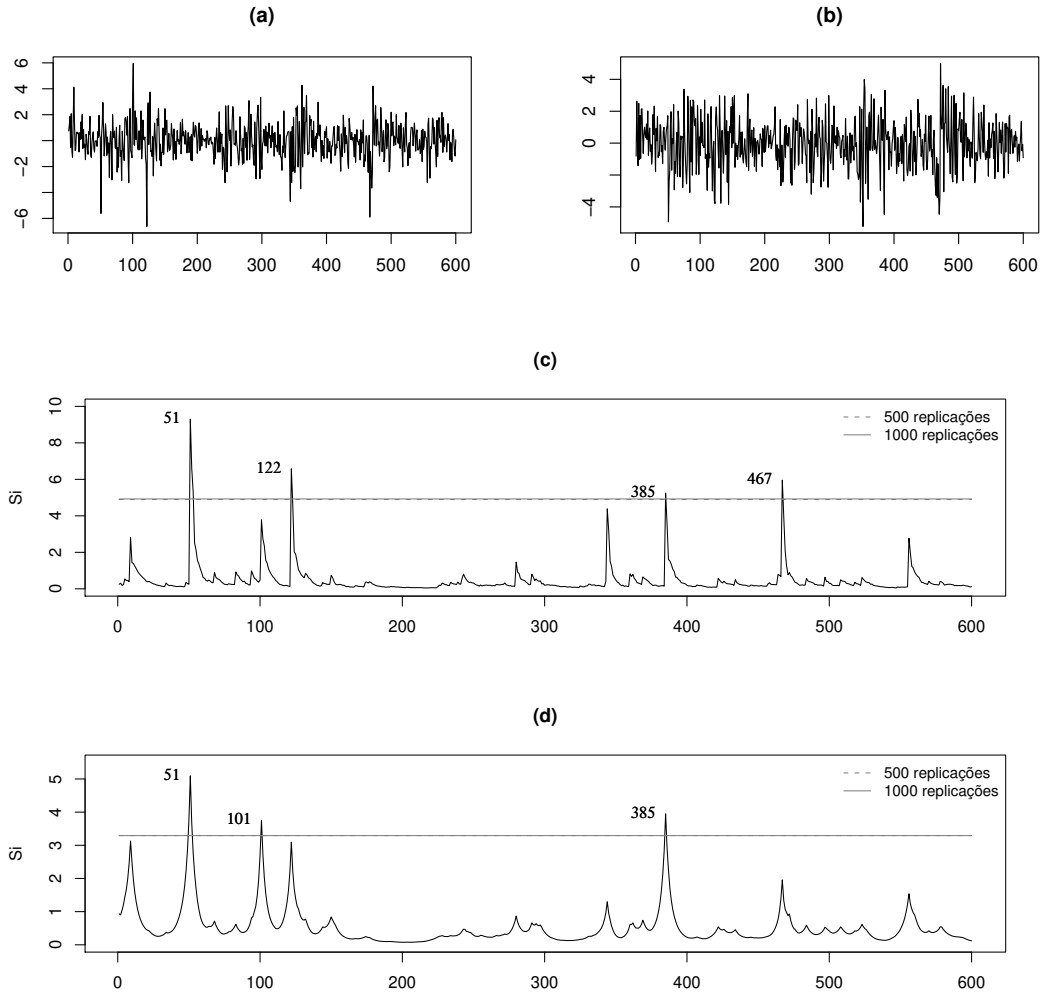


Figura 5.24: Figuras de influência para os retornos do dólar: (a) série da NYSE, (b) série da FTSE, (c) estatística de influência filtrada e (d) estatística de influência suavizada, e os limiares obtidos com 500 e 1000 replicações.

Tabela 5.16: Valores influentes identificados: data de ocorrência, localização, valor da estatística e resíduos.

Data	Localização	Estatística	Res. predição	Res. auxiliar
Filtro de Kalman				
04/01/2000	51	9,3029	3,2381	-0,4234
05/01/2000	52	6,6068	-1,0473	-0,0746
14/04/2000	122	6,5783	2,8579	-0,0103
14/05/2001	385	5,2508	2,5768	-1,2166
10/09/2001	467	5,9586	2,9306	0,9678
Suavizador de Kalman				
30/12/1999	50	3,7422	-0,2561	2,4451
04/01/2000	51	5,0936	3,2381	-0,4234
05/01/2000	52	3,6720	-1,0473	-0,0746
16/03/2000	101	3,7407	2,3113	-1,8170
14/05/2001	385	3,9480	2,5768	-1,2166

Através da análise das séries dos retornos da bolsa de valores de Nova Iorque (NYSE) e da Inglaterra (FTSE) podemos ver que, tanto no ajuste do modelo de volatilidade univariado quanto no ajuste dos modelos multivariados, as observações com "grande" magnitude em valor absoluto e que foram detectadas como influentes estão relacionadas à crises que impactaram diretamente o mercado financeiro. Tanto no caso do Filtro quanto do Suavizador de Kalman, algumas observações vizinhas dessas observações foram também detectadas, em especial no caso do Suavizador de Kalman. Já os resíduos de predição correspondentes a estas observações são valores significativos ao nível de 5% apenas no MVE-FA.

Capítulo 6

Conclusões e considerações finais

Esta dissertação tem seis contribuições. A primeira é ilustrar o uso da metodologia de influência proposta por Motta e Hotta (2007) para os modelos de volatilidade estocástica univariados assimétricos. A segunda contribuição é a proposta de uma medida de influência para a análise dos modelos de volatilidade estocástica multivariados com correlações constantes generalizados. A terceira consiste em ilustrar o uso da metodologia de influência para o modelo de volatilidade estocástica fatorial aditivo. Já a quarta contribuição consiste no uso das estatísticas de influência baseadas tanto na estimativa filtrada da volatilidade quanto na estimativa suavizada da mesma, sendo que a estimativa filtrada não foi considerada no trabalho de Motta e Hotta (2007). A quinta contribuição consiste na discussão sobre a associação entre observações influentes e resíduos de predição e/ou resíduos auxiliares dos estados. Por fim, a sexta contribuição consiste em ilustrar o uso da metodologia baseada em marcas de referência (limiares) simuladas proposta por Motta e Hotta (2007) no diagnóstico de influência para os modelos de volatilidade estocástica e também para modelos de regressão linear. A partir dessa metodologia, é possível determinar quais observações das amostras são significativamente influentes.

Foram simuladas séries bivariadas para o modelo com correlações constantes generalizado e para o modelo fatorial aditivo com *outliers* incorporados, e as estatísticas propostas foram aplicadas juntamente com a técnica de limiares obtidos via simulação. Em todos os exemplos analisados, as estatísticas de influência conseguiram detectar os *outliers* incorporados às séries. As estatísticas de influência baseadas na estimativa filtrada da volatilidade foram mais eficazes para identificar observações influentes quando

comparadas às estatísticas baseadas na estimativa suavizada, pois estas últimas tendem a identificar observações vizinhas de observações influentes, sendo que essas observações vizinhas não foram inseridas como *outliers*. Em geral, as estatísticas de influência baseadas na estimativa suavizada detectam mais observações influentes. Também, o método é eficaz no sentido de não identificar observações de "grande" magnitude em valor absoluto que não foram inseridas como *outliers* (observações genuínas).

Foram realizadas também as análises de influência de séries reais de retornos financeiros, através das estatísticas propostas e da técnica de limiares obtidos via simulação, considerando o modelo univariado assimétrico, modelo com correlações constantes generalizado e modelo fatorial aditivo. Verificou-se a identificação de observações atípicas (*outliers*) que estão relacionadas à crises que impactaram diretamente o mercado financeiro.

Os limiares foram construídos com base em 1000 replicações, mas, a fim de diminuir o custo computacional relacionado ao cálculo dos mesmos, foram construídos também limiares com 500 replicações. Desta forma, pode-se verificar que os limiares obtidos com 500 replicações foram tão eficazes na análise de influência quanto os limiares obtidos com 1000 replicações, uma vez que ambos conseguiram detectaram as mesmas observações.

Através das análises, observamos que nem sempre um resíduo de predição ou resíduo auxiliar significativo ao nível de 5% indica uma observação influente, e também que um resíduo não-significativo pode estar relacionado a uma observação influente. Sendo assim, a análise isolada dos valores dos resíduos pode não ser suficiente para identificar observações atípicas em um conjunto de dados.

No contexto de regressão linear, foram realizadas as análises de influência em dados simulados e também em dados reais. Foi utilizada a técnica de limiares obtidos por simulações de Monte Carlo, juntamente com os limiares propostos por Peña (2005). Verificou-se que os limiares obtidos por simulações foram mais eficazes para detectar as observações atípicas quando comparados aos limiares de Peña, pois detectaram apenas as observações que são influentes nos conjuntos de dados, ou seja, observações que foram inseridas como *outliers*, no caso simulado, ou observações que estão longe da configuração linear dos dados, no caso real.

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

1. Aplicar a metodologia de influência proposta a outros modelos de volatilidade es-

tocástica multivariados como, por exemplo, o modelo multivariado assimétrico;

2. Aplicar a metodologia de influência em modelos de dimensões maiores que três, que requer maior custo computacional.

Referências Bibliográficas

Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, 18(3), 338-357.

Chan, D., Kohn, R. and Kirby, C. (2006). Multivariate stochastic volatility models with correlated errors. *Econometric Reviews*, 25(2-3), 245-274.

Cook R. D. (1986). Assessment of local influence. *Journal of Royal Statistical Society*, Series B, 42(2), 133-169.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19(1), 15-18.

Danielsson, J. (1998). Multivariate stochastic volatility models: estimation and a comparison with VGARCH models. *Journal of Empirical Finance*, 5(2), 155-173.

Durbin, J. and Koopman, S. J. (2001). Time series analysis by state space methods. *Oxford University Press, New York.*

Fuller, W. A. (1996). Introduction to statistical time series. *New York: Wiley*, 2ed.

Harvey, A. C. (1989). Forecasting, structural time series models and the Kalman filter. *Cambridge: University Press.*

Harvey, A. C. and Koopman, S. J. (1992). Diagnostic checking of unobserved components time series models. *Journal of Business and Economic Statistics*, 10(4), 377-390.

Harvey, A. C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *The review of economic studies*, 61(2), 247-264.

Harvey, A. C. and Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, 14(1), 429-434.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12(4), 371-417.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1999). Stochastic volatility: univariate and multivariate extensions. Manuscrito não publicado, Universidade de Chicago.

Jungbacker, B. and Koopman, S. J. (2005). On importance sampling for state space models. *Tinbergen institute discussion papers*, 5, 117-124.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic engineering, Transactions ASMA, Series D*, 82(1), 35-45.

Kalman, R. E. and Bucy R. S. (1961). New Results in Linear Filtering and Prediction Theory. *Journal of basic engineering, Transactions ASMA, Series D*, 83(1), 95-108.

Koopman, S. J., Shephard, N. and Doornik, J. A. (1999). Statistical algorithms for models in state space using Ssfpack 2.2. *Econometrics Journal*, 2(1), 107-160.

Kutner, M. H., Nachtsheim, C. J., Neter, J. and Li, W. (2004). Applied Linear Statistical Models. *McGraw-Hill Irwin*, 5ed.

Motta, A. C. O. e Hotta, L. K. (2007). Influence in stochastic volatility models. In Third Brazilian Conference on Statistical Modelling in Insurance and Finance, São Sebastião. *Proceedings of the Third Brazilian Conference on Statistical Modelling in Insurance and Finance.* São Paulo, p. 320-323.

Peña, D. (2005). A new statistic for influence in linear regression. *Technometrics*, 47(1), 1-12.

Pitt, M. and Shephard, N. (1999). Time varying covariances: a factor stochastic volatility approach. In: Bernardo, J. M., Berger, J. O., David, A. P., Smith, A. F. M., eds. Bayesian Statistics 6. *Oxford University Press, London*, p. 547-570.

Quintana, J. M. and West, M. (1987). An analysis of international exchange rates using multivariate DLMs. *The Statistician*, 36(2-3), 275-281.

Rousseeuw, P. J. and Leroy, A. M. (1987). Robust regression and outlier detection. *New York: John Wiley.*

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In: Cox, D. R., Hinkley, D. V., Barndorff-Nielsen, O. E., eds. *Time Series Models in Econometrics, Finance and Other Fields*, London: Chapman and Hall, p. 1-67.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika*, 84(3), 653-667.

Taylor, S. J. (1986). Modeling financial time series. *New York: John Wiley.*

Yu, J. e Meyer, R. (2006). Multivariate stochastic volatility models: bayesian estimation and model comparison. *Econometrics Reviews*, 25(2-3), 361-384.