
VALOR EM RISCO AUTO-REGRESSIVO CONDICIONAL:
O CASO DE ÍNDICES BRASILEIROS

TIAGO FELIPE DE MELO SALES SANTOS

Orientador: Prof. Dr. Sylvio José Pereira dos Santos

Co-orientador: Prof. Dr. Francisco Cribari Neto

Área de concentração: Estatística Aplicada

Dissertação submetida como requerimento parcial para obtenção do
grau de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, fevereiro de 2006

Santos, Tiago Felipe de Melo Sales
Valor em Risco Auto-regressivo Condicional : o
caso de índices brasileiros / Tiago Felipe de Melo
Sales Santos. – Recife : O Autor, 2006.

94 folhas : il., fig., tab

Dissertação (mestrado) – Universidade Federal de
Pernambuco. CCEN. Estatística, 2006.

Inclui bibliografia.

1. Estatística aplicada – Séries temporais. 2. Valor
em Risco (VaR) – Mercado financeiro – Risco de
investimento – Modelo CAViaR. I. Título.

311
310

CDU (2.ed.)
CDD (22.ed.)

UFPE
BC2006-148

Universidade Federal de Pernambuco

Mestrado em Estatística

22 de fevereiro de 2006

(data)

Nós recomendamos que a dissertação de mestrado de autoria de

Tiago Felipe de Melo Sales Santos

intitulada

Valor em Risco Auto-Regressivo Condicional; o caso de índices brasileiros

seja aceita como cumprimento parcial dos requerimentos para o grau de Mestre em Estatística.

Klaus Lito Pires Vasconcelos

Coordenador da Pós-Graduação em Estatística



Prof. Klaus L. P. Vasconcelos
Coordenador da
Pós-graduação em
Estatística, UFPE

Banca Examinadora:

Francisco Cribari Neto

Francisco Cribari Neto

orientador

Beatriz Vaz de Melo Mendes (UFRJ)

Beatriz Vaz de Melo Mendes (UFRJ)

Ana Katarina Campêlo

Ana Katarina Telles de Novaes Campêlo

Este documento será anexado à versão final da dissertação.

“Nothing in nature is random... A thing appears random
only through the incompleteness of our knowledge.”
— Benedict Spinoza

Agradecimentos

Aos meus pais, Francisco e Vera, pelo carinho, amor, compreensão e por terem me apoiado em todos os sentidos sem medir esforços.

À minha noiva, Karla, pela amizade e companheirismo e por ter me ajudado a retirar as pedras encontradas na difícil caminhada do mestrado, a quem devo sinceras desculpas pela ausência.

Ao meu orientador, professor Sylvio José Pereira dos Santos, por ter me acolhido desde nossa primeira conversa sobre a dissertação, pela cobrança constante, pelas excelentes sugestões e pela amizade surgida desde os tempos de graduação.

Ao meu co-orientador, professor Francisco Cribari Neto, pelas boas sugestões, pela seriedade e pela leitura cuidadosa.

À minha mestra e amiga, Jacira Guiro Marino, por acreditar em meu potencial e por oferecer sempre os conselhos mais centrados.

Aos meus amigos de graduação: Hugo Leonardo, Luciano Souza, Fábio Cavalcanti, Luís Dimas, Catarina Costa e Eduardo José, pelo incentivo, amizade e companheirismo.

Aos meus colegas de mestrado, em especial, a Renata Nunes e a Daniela Nava, grandes amizades reveladas nesses dois anos, por sempre me estimularem a estudar, mesmo nos momentos menos propícios, e a Francisco “Pacho”, pelo companheirismo.

Aos amigos Carlos Gadelha, Denis Santos, Andréa Rocha e Alexandre Simas, companheiros de jornada na graduação e também no mestrado, com quem pude travar as mais calorosas discussões sobre estatística.

À Valéria Bittencourt, muito mais do que Secretária da pós-graduação, sempre preocupada com o desempenho dos alunos, pela simpatia e disponibilidade.

À Beatriz Vaz de Melo Mendes e Ana Katarina Telles de Novaes Campêlo, membros da banca examinadora, pelas sugestões oferecidas para o engrandecimento desta dissertação.

A todas pessoas que direta ou indiretamente colaboraram para o desenvolvimento desta dissertação e àqueles que, mesmo à distância, torceram pelo meu sucesso.

A todos, meu muito obrigado.

Resumo

Cada vez mais a análise de séries temporais tem estado presente no mercado financeiro, auxiliando a obtenção de previsões e a tomada de decisão. Na presença de eventos extremos, as perdas são sempre muito grandes e em alguns casos podem levar os investidores à falência. Algumas medidas foram criadas para mensurar esse risco, porém elas não levam em conta o fato de que outras variáveis podem ser relevantes à mensuração. Engle & Manganelli (2004) propuseram uma nova metodologia para medir os riscos de investimento, o modelo **CAViaR**, que, além de permitir a inclusão dessas variáveis, permite avaliar a qualidade do ajuste. O principal objetivo desta dissertação é avaliar a nova metodologia proposta por Engle & Manganelli (2004) conjuntamente com as metodologias existentes, nos casos dos índices da Bolsa de Valores de São Paulo – IBOVESPA e da Petrobrás. Por outro lado, verificamos a qualidade da estimação dos parâmetros através de simulação de Monte Carlo, de onde pudemos concluir que a qualidade da estimação está ligada à escolha dos parâmetros iniciais e que, em geral, faz-se necessária a consideração de um grande número de vetores de estimativas iniciais.

Abstract

Time series analysis has been increasingly used in Finance. In the presence of extreme events, losses can turn out to be very large and can lead to bankruptcy. Some techniques have been proposed in the literature such a risk, but they do not take account the impact of additional variables. Engle & Manganelli (2004) proposed the class of **CAViaR** models to account for the effects of such variables. In this thesis, we evaluate **CAViaR** models to more established models in the literature. Parameter estimation is analyzed by Monte Carlo simulation. The results show that the choice of initial values can be quite importante when obtaining point estimates for the parameters that index the model.

Sumário

Introdução	1
1. Modelagem de Séries Financeiras	3
1.1. Conceitos básicos	3
1.2. Modelos de Box–Jenkins	6
1.2.1 O modelo auto-regressivo	7
1.2.2 O modelo de médias móveis	18
1.2.3 O modelo auto-regressivo e de médias móveis	20
1.3. Modelos heteroscedásticos	21
1.3.1 O modelo auto-regressivo com heteroscedasticidade condicional	22
1.3.2 O modelo ARCH generalizado	26
1.3.3 O modelo GARCH integrado	29
1.4. Modelos de longa dependência	30
2. Valor em Risco	31
2.1. RiskMetrics	33

2.2. Abordagem econométrica.....	35
2.3. Estimação quantílica	37
2.4. Valor em Risco em eventos extremos	38
2.4.1. Modelagem de eventos extremos	38
2.4.2. Cálculo de VaR pela TVE	42
3. Valor em Risco Auto-regressivo Condicional	43
4. Simulações.....	50
4.1. Cenário I	50
4.2. Cenário II.....	59
4.3. Cenário III.....	65
5. Aplicações	75
5.1. O caso da Petrobrás.....	75
5.2. O caso do IBOVESPA.....	83
Conclusões.....	90
Referências	92

Introdução

A análise de séries temporais tem estado cada vez mais presente na análise de mercados financeiros. A metodologia proposta por Box & Jenkins (1970) e as contribuições posteriores de diversos autores auxiliam os investidores a realizar previsões e, com isso, decidir o momento correto para investir ou vender suas ações.

Porém, no caso de ocorrência de eventos extremos, os modelos de séries temporais em geral não conseguem prever a magnitude das perdas, fazendo com que os investidores percam muito dinheiro, o que pode levar alguns deles à falência.

Tipicamente, os eventos extremos apresentam probabilidade de ocorrência muito baixa, fato que faz com que alguns investidores não se preocupem com sua ocorrência. Por exemplo, em 24 de outubro de 1929 – data marcada como a quinta-feira negra, quando da quebra da Bolsa de Nova Iorque –, a economia norte-americana entrou em colapso e, conseqüentemente, o mesmo ocorreu com a economia mundial. Em decorrência à quebra, o crescimento alemão foi paralisado, visto que, à época, a Alemanha era dependente dos investimentos norte-americanos; as economias inglesa e francesa foram abaladas, devido à completa impossibilidade alemã de pagar as indenizações de guerra; e colapso do comércio mundial, gerando uma maior restrição da produção internacional. O “crack” da bolsa nova-iorquina fez com que os investidores tivessem grandes perdas.

A solução para o problema dos eventos extremos no mercado financeiro só surgiu na década de 90 e foi a forma encontrada por algumas instituições financeiras para minimizar suas perdas. A ferramenta criada por essas instituições consegue mensurar o risco do investidor considerando os movimentos de mercado. A essa medida, eles deram o nome de Valor em Risco ou, simplesmente, **VaR**. Com isso, as instituições passaram a saber o montante que deveriam ter guardado, caso algum evento catastrófico ocorresse, sem que elas corressem o risco de encerrar suas atividades.

Existem diversas formas de se calcular o risco das carteiras de investimento. Cada uma delas apresenta vantagens e desvantagens com relação às demais. Contudo, nenhuma delas leva em conta a inclusão de outras variáveis. A inclusão de variáveis adicionais para explicar o comportamento de uma determinada carteira pode ajudar a compreender o mecanismo gerador da mesma e, com isso, fazer com que o investidor esteja mais próximo da realidade. Quando não incluímos variáveis significativas, podemos estar subestimando (ou superestimando) o montante que pode ser perdido caso tais eventos ocorram.

Pensando nisso, Engle & Manganelli (2004) propuseram uma nova forma de mensurar o risco do investidor e a ela deram o nome de Valor em Risco Auto-regressivo Condicional ou, simplesmente, **CAViaR**. Essa nova metodologia é bastante diferenciada das existentes anteriormente, visto que nela não são feitas suposições a respeito da distribuição dos retornos da carteira. Nessa metodologia, em vez de modelar a distribuição dos retornos, como faz a maioria dos métodos de cálculo de **VaR** anteriores, modela-se diretamente os quantis da distribuição, não havendo, portanto, suposições a respeito da sua distribuição.

Outra vantagem da metodologia **CAViaR** é a possibilidade de se utilizar diversas especificações e a escolha da melhor forma fica diretamente relacionada à carteira em estudo.

Além de apresentar essa nova metodologia, Engle & Manganelli (2004) apresentam um teste para verificar a adequabilidade do modelo. O teste de Quantis Dinâmicos tem a capacidade de verificar se o modelo considerado é adequado para a obtenção do **VaR**, o que até então não era possível concluir, visto que os modelos antecessores não permitiam esse tipo de avaliação.

Modelagem de Séries Financeiras

A análise de séries temporais financeiras está relacionada com a identificação de alguma relação existente entre os dados que possa ser modelada a fim de se conseguir gerar previsões.

Após identificada a relação, é possível entender o mecanismo gerador da série e, com isso, produzir previsões acuradas de valores futuros, um dos principais objetivos do estudo de séries temporais.

Nas seções seguintes definiremos os conceitos básicos necessários para o entendimento do problema e apresentaremos as técnicas mais utilizadas na análise de séries financeiras.

1.1 Conceitos básicos

Considere o processo estocástico $\{Y_t, t \in \Pi\}$, onde Π é o conjunto índice que controla a evolução do tempo; geralmente, $\Pi = \{1, 2, \dots\}$, podendo ser contínuo em alguns casos. Um processo estocástico é uma família de variáveis aleatórias e a uma determinada realização desse processo dá-se o nome de série temporal.

Um conceito importante na análise de séries temporais é o de estacionariedade. Uma série y_t é dita ser fortemente estacionária se a distribuição conjunta de $\{y_{t_1}, y_{t_2}, \dots, y_{t_k}\}$ for idêntica à de $\{y_{t_1+h}, y_{t_2+h}, \dots, y_{t_k+h}\}$, para todo h , onde k é um número inteiro positivo e arbitrário e $\{t_1, t_2, \dots, t_k\}$ é uma coleção de k números inteiros positivos. Em outras palavras, a distribuição conjunta de $\{y_{t_1}, y_{t_2}, \dots, y_{t_k}\}$ é invariante no espaço de tempo. Esta é uma condição muito forte e raramente observada em dados reais.

Um outro conceito é o de estacionariedade fraca, sob o qual a média de y_t e a covariância entre y_t e y_{t-k} são invariantes no tempo, ou seja, y_t é fracamente

estacionária se $E(y_t) = \mu$ e se a função de autocovariância, $\text{Cov}(y_t, y_{t-k}) = \gamma_k$, depende exclusivamente de k , para todo k e para todo t . Em palavras, a série y_t flutua com variação constante em torno de um nível constante. Assume-se ainda que os dois primeiros momentos de y_t são finitos. Este conceito é muitas vezes referido na literatura como estacionariedade de segunda ordem.

Assim, se y_t for fortemente estacionária e seus dois primeiros momentos forem finitos, então y_t também será fracamente estacionária. Porém, o contrário, em geral, não é verificado. Contudo, se y_t for normalmente distribuída, estacionariedade fraca implicará estacionariedade forte. Neste trabalho, quando usarmos o termo estacionariedade estaremos nos referindo à estacionariedade fraca.

Estacionariedade é a suposição mais comumente feita em modelos de séries temporais. Porém, muitas vezes, as séries em estudo são não-estacionárias, ou seja, não satisfazem uma ou ambas as condições de estacionariedade apresentadas. Esta violação pode ser decorrente da presença de tendência, sazonalidade, períodos de volatilidade ou mudanças na estrutura da série. Geralmente o problema é solucionado tomando-se diferenças sucessivas da série original ou funções da mesma.

A d -ésima diferença de y_t é definida por

$$\Delta^d y_t = (1 - B)^d y_t,$$

onde B é chamado de operador retroativo e $B^k y_t = y_{t-k}$. Assim, a primeira diferença de y_t é dada por

$$\Delta y_t = (1 - B)y_t = y_t - y_{t-1},$$

enquanto que a segunda diferença de y_t é

$$\Delta^2 y_t = (1 - B)^2 y_t = (1 - 2B + B^2)y_t = y_t - 2y_{t-1} + y_{t-2}.$$

Em geral, basta tomar $d = 1$ ou $d = 2$ para que $\Delta^d y_t$ se torne estacionária.

O coeficiente de correlação entre duas variáveis aleatórias X e Y é definido como

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sqrt{E[(X - \mu_X)^2] E[(Y - \mu_Y)^2]}},$$

onde μ_X e μ_Y são, respectivamente, as médias de X e Y . Este coeficiente mede a intensidade da relação linear existente entre X e Y , onde $-1 \leq \rho_{X,Y} \leq 1$ e $\rho_{X,Y} = \rho_{Y,X}$. Não há correlação entre X e Y se $\rho_{X,Y} = 0$. No caso em que X e Y são normalmente distribuídas, então $\rho_{X,Y} = 0$ se e somente se as duas variáveis são independentes. Esta medida pode ser consistentemente estimada por

$$\hat{\rho}_{X,Y} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2 \sum_{t=1}^T (y_t - \bar{y})^2}},$$

onde $\bar{x} = \sum_{t=1}^T x_t/T$ e $\bar{y} = \sum_{t=1}^T y_t/T$ são as médias amostrais de X e Y , respectivamente.

Quando o interesse reside na dependência entre y_t e seus valores defasados y_{t-k} , o conceito de correlação é chamado de autocorrelação e é denotado por ρ_k , o qual, sob a suposição de estacionariedade, é função somente de k . Especificamente,

$$\rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\sqrt{\text{Var}(y_t)\text{Var}(y_{t-k})}} = \frac{\text{Cov}(y_t, y_{t-k})}{\text{Var}(y_t)} = \frac{\gamma_k}{\gamma_0}.$$

Aqui, utilizamos a propriedade de estacionariedade que $\text{Var}(y_t) = \text{Var}(y_{t-k})$. Da definição, temos que $\rho_0 = 1$, $\rho_{-k} = \rho_k$ e $-1 \leq \rho_k \leq 1$. Assim, não há correlação serial se, para todo $k > 0$, $\rho_k = 0$.

A função de autocorrelação amostral é dada por

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}. \quad [1.1]$$

Sob algumas condições, $\hat{\rho}_k$ é um estimador consistente de ρ_k . Por exemplo, se $\{y_t\}$ for uma seqüência independente e identicamente distribuída e $E(y_t) < \infty$, então $\hat{\rho}_k$ será assintoticamente normal com média zero e variância $1/T$, ou seja,

$$\sqrt{T} \hat{\rho}_k \xrightarrow{d} \mathcal{N}(0, 1).$$

Uma série temporal a_t é dita ser ruído branco se a_t é uma seqüência de variáveis aleatórias não-correlacionadas com média zero e variância finita. Em particular, quando a_t é normalmente distribuída, diz-se que a mesma é um ruído branco gaussiano. As autocorrelações de um ruído branco são zero para todas as defasagens diferentes de $k = 0$.

Em modelos financeiros é comum a utilização de retornos para a obtenção de previsões e realização de ajustes. Definiremos a seguir os dois principais tipos de retornos utilizados na literatura. Seja y_t o valor de um ativo no instante de tempo t . Definimos retornos simples de um período como

$$1 + R_t = \frac{y_t}{y_{t-1}} \quad \text{ou, ainda,} \quad R_t = \frac{y_t}{y_{t-1}} - 1,$$

o qual mede a variação do valor do ativo entre $t-1$ e t , em termos percentuais. O logaritmo natural dos retornos simples é chamado de retorno composto continuamente ou, simplesmente, log-retorno, dado por

$$r_t = \log(1 + R_t) = \log\left(\frac{y_t}{y_{t-1}}\right) = \log(y_t) - \log(y_{t-1}) = p_t - p_{t-1},$$

onde $p_t = \log(y_t)$ e r_t corresponde aproximadamente ao percentual da variação da posição.

Considerando k períodos, temos que os retornos simples e os log-retornos são dados por

$$\begin{aligned} 1 + R_t[k] &= \frac{y_t}{y_{t-1}} \times \frac{y_{t-1}}{y_{t-2}} \times \frac{y_{t-2}}{y_{t-3}} \times \dots \times \frac{y_{t-k+1}}{y_{t-k}} = \frac{y_t}{y_{t-k}}, \quad \text{e} \\ r_t[k] &= \log(1 + R_t[k]) = \log\left(\frac{y_t}{y_{t-1}}\right) + \dots + \log\left(\frac{y_{t-k+1}}{y_{t-k}}\right) \\ &= r_t + r_{t-1} + \dots + r_{t-k+1}, \end{aligned} \quad [1.2]$$

respectivamente. A expressão [1.2] pode ainda ser reescrita como

$$r_t[k] = \log(y_t) - \log(y_{t-k}).$$

É comum assumir que r_t é uma série fracamente estacionária. Exemplos da utilização de retornos simples e log-retornos, bem como suas propriedades, são discutidos em Tsay (2002) e em Morettin & Toloi (2004).

1.2 Modelos de Box–Jenkins

Na análise de séries temporais, em geral, utiliza-se o conjunto de observações originais, y_t . Entretanto, quando y_t é uma variável financeira, como, por exemplo, preços de ativos, é mais comum que se utilize r_t ao invés de y_t , por duas razões: retornos são livres de escala de medida e são mais fáceis de tratar do que séries de preços, pois possuem propriedades estatísticas mais atrativas. Assim, iremos construir a classe de modelos de Box–Jenkins utilizando r_t , mas todas as definições apresentadas são igualmente válidas se considerarmos y_t .

A classe de modelos mais comumente utilizada na análise de séries temporais é a dos modelos auto-regressivos integrados e de médias móveis ou, simplesmente, ARIMA(p, d, q), que foi proposta por Box & Jenkins (1970). Sua forma geral é dada por

$$(1 - B)^d r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + a_t - \sum_{j=1}^q \theta_j a_{t-j},$$

onde $\{a_t\}$ é um ruído branco e p , d e q são números inteiros não-negativos que representam, respectivamente, as ordens auto-regressiva, de integração e de médias móveis. Este modelo pode ser reescrito em função de operadores retroativos:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d r_t = \phi_0 + (1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad [1.3]$$

ou, ainda,

$$\Phi(B)\Delta^d r_t = \phi_0 + \Theta(B),$$

onde $\Phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ e $\Theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$.

Alguns casos particulares do modelo ARIMA(p, d, q) merecem destaque. Quando $d = q = 0$, temos o modelo auto-regressivo de ordem p , AR(p). O modelo ARIMA(0, 0, q) é conhecido como modelo de médias móveis de ordem q ou, simplesmente, MA(q). Se r_t é estacionária, não é necessário que se tome diferenças e, conseqüentemente, $d = 0$. A este modelo, dá-se o nome de modelo auto-regressivo e de médias móveis, ARMA(p, q). As propriedades desses modelos também valem para os casos em que $d \neq 0$, conhecidos como modelos integrados, desde que se considere a série $z_t = \Delta^d r_t$, em vez de r_t .

1.2.1 O modelo auto-regressivo

O modelo AR(p) é obtido de [1.3] com $d = q = 0$, ou seja,

$$(1 - \phi_1 B - \dots - \phi_p B^p)r_t = \phi_0 + a_t$$

ou, simplesmente,

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + a_t. \quad [1.4]$$

Estudaremos a seguir as propriedades dos modelos AR, inicialmente, para o caso mais simples, o modelo AR(1), dado por

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t. \quad [1.5]$$

Se assumimos que a série é estacionária, então $E(r_t) = \mu$, $\text{Var}(r_t) = \gamma_0$ e $\text{Cov}(r_t, r_{t-\ell}) = \gamma_\ell$, onde μ e γ_0 são constantes e γ_ℓ é função que depende apenas de ℓ . Assim, temos que

$$E(r_t) = E(\phi_0 + \phi_1 r_{t-1} + a_t) = \phi_0 + \phi_1 E(r_{t-1}).$$

Sob estacionariedade, $E(r_t) = E(r_{t-1}) = \mu$ e, conseqüentemente,

$$\mu = \phi_0 + \phi_1 \mu, \quad \text{ou seja,} \quad \mu = \frac{\phi_0}{1 - \phi_1}.$$

Note que a média de r_t existe se $\phi_1 \neq 1$ e é nula se $\phi_0 = 0$. Logo, no processo AR(1), ϕ_0 é relacionado com a média de r_t . Fazendo $\phi_0 = (1 - \phi_1)\mu$, podemos escrever o modelo AR(1) como

$$\begin{aligned} r_t &= (1 - \phi_1)\mu + \phi_1 r_{t-1} + a_t \\ &= \mu - \phi_1 \mu + \phi_1 r_{t-1} + a_t = \mu + \phi_1(r_{t-1} - \mu) + a_t. \end{aligned}$$

Logo,

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + a_t. \quad [1.6]$$

Tomando o valor esperado do quadrado de [1.6], obtemos

$$\begin{aligned} E[(r_t - \mu)^2] &= E\left\{[\phi_1(r_{t-1} - \mu) + a_t]^2\right\} \\ &= E[\phi_1^2(r_{t-1} - \mu)^2 + a_t^2 + 2\phi_1(r_{t-1} - \mu)a_t] \\ &= \phi_1^2 E[(r_{t-1} - \mu)^2] + E(a_t^2) + 2\phi_1 E[(r_{t-1} - \mu)a_t]. \end{aligned}$$

Da independência entre r_{t-1} e a_t , temos que

$$E[(r_{t-1} - \mu)a_t] = E(r_{t-1}a_t) - \mu E(a_t) = E(r_{t-1})E(a_t) - \mu E(a_t) = 0.$$

Assim,

$$E[(r_t - \mu)^2] = \phi_1^2 E[(r_{t-1} - \mu)^2] + E(a_t^2).$$

Como $E[(r_t - \mu)^2] = \text{Var}(r_t)$, temos que

$$\text{Var}(r_t) = \phi_1^2 \text{Var}(r_{t-1}) + \sigma_a^2,$$

onde σ_a^2 é a variância do ruído branco. Dado que $\text{Var}(r_t) = \text{Var}(r_{t-1})$, devido à suposição de estacionariedade, obtemos

$$\text{Var}(r_t) = \frac{\sigma_a^2}{1 - \phi_1^2},$$

com $\phi_1^2 < 1$, visto que $\text{Var}(r_t)$ é não-negativa. Conseqüentemente, a estacionariedade do modelo AR(1) implica que $-1 < \phi_1 < 1$, que a média e a variância de r_t são finitas e, pela desigualdade de Cauchy-Schwarz, que todas as autocovariâncias são finitas. Resumindo, para que o modelo AR(1) em [1.5] seja estacionário, é necessário e suficiente que $|\phi_1| < 1$.

Multiplicando [1.6] por a_t , tomando o valor esperado e usando novamente a independência entre a_t e r_{t-1} , obtemos

$$\begin{aligned} E[(r_t - \mu)a_t] &= E\{a_t[\phi_1(r_{t-1} - \mu) + a_t]\} \\ &= E[\phi_1(r_{t-1} - \mu)a_t] + E(a_t^2) = E(a_t^2) = \sigma_a^2. \end{aligned} \quad [1.7]$$

Se multiplicarmos [1.6] por $(r_{t-\ell} - \mu)$, tomarmos o valor esperado e utilizarmos o resultado [1.7], obteremos

$$\gamma_\ell = \begin{cases} \phi_1 \gamma_1 + \sigma_a^2, & \text{se } \ell = 0, \\ \phi_1 \gamma_{\ell-1}, & \text{se } \ell > 0, \end{cases}$$

pois $\gamma_{-\ell} = \gamma_\ell$. Portanto, para o modelo AR(1) estacionário em [1.5], temos que a função de autocorrelação de r_t é dada por

$$\rho_\ell = \phi_1 \rho_{\ell-1}, \quad \text{para } \ell > 0;$$

mas $\rho_0 = 1$, logo podemos escrever $\rho_\ell = \phi_1^\ell$. Ou seja, a função de autocorrelação do modelo AR(1) decai exponencialmente com ϕ_1 .

Outro modelo é o AR(2), o qual é dado por

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t. \quad [1.8]$$

Analogamente ao que foi feito para o modelo AR(1), é fácil verificar que

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2},$$

onde $\phi_1 + \phi_2 \neq 1$, para que $E(r_t) < \infty$, conseqüentemente r_t seja um processo estacionário. Fazendo $\phi_0 = \mu(1 - \phi_1 - \phi_2)$ em [1.8], obtemos

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t. \quad [1.9]$$

Multiplicando [1.9] por $(r_{t-\ell} - \mu)$, tomando o valor esperado e considerando a independência entre a_t e $r_{t-\ell}$, temos

$$\begin{aligned} E[(r_t - \mu)(r_{t-\ell} - \mu)] &= E\{(r_{t-\ell} - \mu)[\phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t]\} \\ &= \phi_1 E[(r_{t-1} - \mu)(r_{t-\ell} - \mu)] + \\ &\quad + \phi_2 E[(r_{t-2} - \mu)(r_{t-\ell} - \mu)] + E[a_t(r_{t-\ell} - \mu)] \\ &= \phi_1 \text{Cov}(r_{t-1}, r_{t-\ell}) + \phi_2 \text{Cov}(r_{t-2}, r_{t-\ell}). \end{aligned}$$

Ou seja,

$$\gamma_\ell = \phi_1 \gamma_{\ell-1} + \phi_2 \gamma_{\ell-2}, \quad \ell > 0. \quad [1.10]$$

O caso em que $\ell = 0$ é obtido repetindo-se o procedimento acima e utilizando o quadrado de [1.9], ou seja,

$$\begin{aligned} \gamma_0 &= E[(r_t - \mu)^2] = E\{(r_t - \mu)[\phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t]\} \\ &= \phi_1 E[(r_{t-1} - \mu)(r_t - \mu)] + \phi_2 E[(r_{t-2} - \mu)(r_t - \mu)] + E[a_t(r_t - \mu)] \\ &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + E[a_t(r_t - \mu)]. \end{aligned}$$

Vimos em [1.7] que $E[a_t(r_t - \mu)] = \sigma_a^2$ para o modelo AR(1). Usando o mesmo raciocínio chegamos a

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_a^2$$

ou, ainda,

$$1 = \phi_1 \rho_1 + \phi_2 \rho_2 + \frac{\sigma_a^2}{\gamma_0}$$

e, finalmente,

$$\gamma_0 = \frac{\sigma_a^2}{1 - \phi_1 \rho_1 - \phi_2 \rho_2}.$$

Dividindo [1.10] por γ_0 , obtemos a função de autocorrelação:

$$\rho_\ell = \phi_1 \rho_{\ell-1} + \phi_2 \rho_{\ell-2}, \quad \ell > 0. \quad [1.11]$$

Assim, para o modelo AR(2) estacionário, temos que

$$\rho_\ell = \begin{cases} 1, & \text{se } \ell = 0, \\ \phi_1/(1 - \phi_2), & \text{se } \ell = 1, \\ \phi_1\rho_{\ell-1} + \phi_2\rho_{\ell-2}, & \text{se } \ell > 1, \end{cases}$$

pois $\rho_1 = \phi_1\rho_0 + \phi_2\rho_{-1} = \phi_1 + \phi_2\rho_1$.

A autocorrelação em [1.11] nos diz que a função de autocorrelação do modelo AR(2) satisfaz a equação

$$(1 - \phi_1 B - \phi_2 B^2)\rho_\ell = 0,$$

a qual determina as propriedades da função de autocorrelação do modelo AR(2) e também o comportamento das previsões de r_t . A condição de estacionariedade para este modelo é que as raízes de $(1 - \phi_1 B - \phi_2 B^2) = 0$ estejam fora do círculo unitário.

Generalizando os resultados obtidos anteriormente, chegamos ao modelo AR(p), apresentado em [1.4], cuja média e variância são dadas, respectivamente, por

$$E(r_t) = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p} \quad \text{e} \quad \text{Var}(r_t) = \frac{\sigma_a^2}{1 - \phi_1\rho_1 - \dots - \phi_p\rho_p},$$

onde $\sum_{i=1}^p \phi_i \neq 1$.

A função de autocovariância é dada por

$$\gamma_\ell = \phi_1\gamma_{\ell-1} + \dots + \phi_p\gamma_{\ell-p}, \quad \text{para } \ell > 0,$$

e a função de autocorrelação é

$$\rho_\ell = \phi_1\rho_{\ell-1} + \dots + \phi_p\rho_{\ell-p}, \quad \text{para } \ell > 0. \quad [1.12]$$

Tomando $\ell = 1, 2, \dots, p$, em [1.12], obtemos as equações de Yule-Walker:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2\rho_1 + \dots + \phi_p\rho_{p-1} \\ \rho_2 &= \phi_1\rho_1 + \phi_2 + \dots + \phi_p\rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \dots + \phi_p. \end{aligned}$$

O modelo AR(p) satisfaz a equação

$$(1 - \phi_1 B - \dots - \phi_p B^p)\rho_\ell = 0, \quad \text{para } \ell > 0.$$

Se as raízes do polinômio $(1 - \phi_1 B - \dots - \phi_p B^p) = 0$ estiverem fora do círculo unitário (no plano complexo), o modelo será estacionário.

Na prática, a ordem p do modelo auto-regressivo é desconhecida. Contudo, ela pode ser determinada de duas maneiras. A primeira utiliza o conceito de autocorrelação parcial e a segunda utiliza funções penalizadoras.

Seja $\phi_{k\ell}$ o ℓ -ésimo coeficiente do modelo $\text{AR}(k)$, de tal forma que ϕ_{kk} seja o último coeficiente. Já vimos que

$$\rho_\ell = \phi_{k1}\rho_{\ell-1} + \phi_{k2}\rho_{\ell-2} + \cdots + \phi_{kk}\rho_{\ell-k}, \quad \text{para } \ell = 1, 2, \dots, k.$$

Reescrevendo essas equações em forma matricial, obtemos

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{pmatrix}.$$

Resolvendo essas equações sucessivamente para $k = 1, 2, 3, \dots$, obtemos

$$\begin{aligned} \phi_{11} &= \rho_1, \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}. \end{aligned}$$

Em geral,

$$\phi_{kk} = \frac{|\mathbf{P}_k^*|}{|\mathbf{P}_k|}, \quad [1.13]$$

onde $\mathbf{P}_k$ é a matriz de autocorrelações e $\mathbf{P}_k^*$ é a matriz $\mathbf{P}_k$ com a última coluna substituída pelo vetor de autocorrelações. A ϕ_{kk} dá-se o nome de função de autocorrelação parcial de ordem k .

Pode-se demonstrar que, em modelos $\text{AR}(p)$, a função de autocorrelação parcial comporta-se da seguinte maneira: para $k \leq p$, $\phi_{kk} \neq 0$; caso contrário, ϕ_{kk} é igual a zero. Assim, será necessário estimar a função de autocorrelação parcial do processo AR. Uma maneira de fazê-lo consiste em estimar modelos auto-regressivos de ordens $p = 1, 2, \dots$ e tomar as estimativas do último coeficiente ϕ_{kk} de cada ordem.

Outra maneira consiste em substituir nas equações de Yule-walker do modelo $\text{AR}(k)$ as funções de autocorrelação ρ_ℓ por suas estimativas, ou seja,

devemos substituir as autocorrelações em [1.12] por suas estimativas como calculadas em [1.1], isto é,

$$\hat{\rho}_\ell = \phi_{k1}\hat{\rho}_{\ell-1} + \cdots + \phi_{kk}\hat{\rho}_{\ell-k},$$

e resolver essas equações para $\ell = 1, 2, \dots$. Assim, temos que ϕ_{kk} mede a correlação remanescente entre r_t e r_{t-k} depois de eliminadas as influências de $r_{t-1}, \dots, r_{t-k+1}$.

Sob a suposição de que o modelo é um $AR(p)$, as funções de autocorrelação parcial estimadas de ordem $p+1, p+2, \dots$ têm variância (Box & Pierce, 1970; McLeod, 1978)

$$\text{Var}(\hat{\phi}_{kk}) \simeq \frac{1}{T}, \quad \text{para } k \geq p+1.$$

Para T suficientemente grande, $\hat{\phi}_{kk}$ tem distribuição aproximadamente normal, o que permite a construção de intervalos de confiança para ϕ_{kk} .

Contudo, esta estratégia de identificação é bastante deficiente, como mostrou Hjorth (1987). Nesse trabalho, o autor procedeu à identificação visual de 500 processos $ARMA(1,1)$, dos quais apenas 28 (5,6%) foram corretamente identificados.

Outra forma de identificação de modelos é feita através da utilização do Critério de Informação de Akaike (Akaike, 1974), que para o modelo $AR(k)$ é definido como

$$AIC(k) = -2 \log \hat{\mathcal{L}}_k + 2k = \log \hat{\sigma}_k^2 + \frac{2k}{T}, \quad [1.14]$$

onde $\hat{\mathcal{L}}_k$ é a estimativa da função de verossimilhança maximizada e $\hat{\sigma}_k^2$ é a estimativa de máxima verossimilhança de σ_a^2 para o modelo com k parâmetros.

Na prática, calcula-se o valor de $AIC(k)$ para $k = 1, 2, \dots, P$, onde P é um número inteiro positivo pré-especificado. O modelo selecionado será aquele que fornecer o menor valor de AIC . Note que o último termo de [1.14] é uma função penalizadora e que a inclusão de novos parâmetros aumenta o valor da verossimilhança, fazendo com que $-2 \log \hat{\mathcal{L}}_k$ diminua. Assim, a função penalizadora faz com que modelos super-parametrizados não sejam selecionados. É importante destacar que diferentes funções penalizadoras resultam em diferentes critérios.

O Critério de Informação de Akaike Corrigido (Hurvich & Tsai, 1989) é dado por

$$AICC(k) = \log \hat{\sigma}_k^2 + \frac{2kT}{T - k - 1}. \quad [1.15]$$

Esta medida é assintoticamente equivalente ao AIC , mas tende a funcionar melhor quando T é pequeno. Outro critério é o Critério de Informação Bayesiano, também conhecido como Critério de Informação de Schwarz (Schwarz, 1978), dado por

$$BIC(k) = \log \hat{\sigma}_k^2 + \frac{k \log T}{T}. \quad [1.16]$$

Algumas considerações têm que ser feitas sobre a seleção de modelos por meio de uma função penalizadora. Em modelos puramente auto-regressivos, a seleção de modelos por AIC e $AICC$ é assintoticamente eficiente (Choi, 1982). Suspeita-se que o mesmo ocorra com BIC , mas essa propriedade ainda não foi provada.

É fácil perceber que a inclusão de novos parâmetros é mais fortemente penalizada quando se utiliza BIC em vez de AIC , para tamanhos amostrais maiores do que oito observações, pois

$$\begin{aligned} k \log T &> 2k \\ \log T &> 2 \\ T &> e^2 = 7.389. \end{aligned}$$

Consideraremos agora o processo de estimação dos parâmetros do modelo $AR(p)$. Primeiramente, apresentaremos o método de mínimos quadrados. Em seguida, apresentamos o método de máxima verossimilhança e uma discussão acerca dos dois métodos.

Considere o modelo $AR(p)$ em [1.4]. O método de mínimos quadrados condicionais, que inicia na $(p+1)$ -ésima observação, é utilizado para obtenção das estimativas dos parâmetros desconhecidos, $\phi_1, \dots, \phi_p$ e σ_a^2 . Especificamente, condicionados às primeiras p observações, nós temos

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_p r_{t-p} + a_t, \quad \text{para } t = p+1, \dots, T.$$

Assim, o modelo ajustado é

$$\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} + \dots + \hat{\phi}_p r_{t-p},$$

cujos resíduos associados são

$$\hat{a}_t = r_t - \hat{r}_t.$$

As estimativas de $\phi = (\phi_1, \dots, \phi_p)$ e σ_a^2 são obtidas através da minimização da função

$$S_1(\phi) = \sum_{t=p+1}^T (r_t - \phi_0 - \phi_1 r_{t-1} - \dots - \phi_p r_{t-p})^2 = \sum_{t=p+1}^T (r_t - \hat{r}_t)^2 = \sum_{t=p+1}^T a_t^2.$$

Ou seja, $\hat{\phi}$ é escolhido como a solução do sistema de equações não-lineares

$$\begin{aligned} \frac{\partial S_1(\phi)}{\partial \phi_0} &= -2 \sum_{t=p+1}^T (r_t - \phi_0 - \phi_1 r_{t-1} - \dots - \phi_p r_{t-p}) = 0 \\ \frac{\partial S_1(\phi)}{\partial \phi_1} &= -2 \sum_{t=p+1}^T (r_t - \phi_0 - \phi_1 r_{t-1} - \dots - \phi_p r_{t-p}) r_{t-1} = 0 \\ &\vdots \\ \frac{\partial S_1(\phi)}{\partial \phi_p} &= -2 \sum_{t=p+1}^T (r_t - \phi_0 - \phi_1 r_{t-1} - \dots - \phi_p r_{t-p}) r_{t-p} = 0. \end{aligned}$$

A partir da sequência $\{\hat{a}_t\}$ obtemos

$$\hat{\sigma}_a^2 = \frac{S_1(\hat{\phi})}{T - 2p - 1}.$$

Outra forma de obter as estimativas de mínimos quadrados é utilizando o método exato, o qual consiste em minimizar

$$S_2(\phi) = \sum_{t=1}^T (r_t - \phi_0 - \phi_1 r_{t-1} - \cdots - \phi_p r_{t-p})^2 = \sum_{t=1}^T (r_t - \hat{r}_t)^2 = \sum_{t=1}^T a_t^2.$$

Neste método, as p observações iniciais entram no processo de estimação como parâmetros desconhecidos e que necessitam ser estimados, fazendo com que a convergência nesse caso se dê de forma mais lenta.

Um outro método que pode ser utilizado é o método da máxima verossimilhança, o qual consiste em maximizar a função de distribuição conjunta. Sob a suposição de normalidade dos a_t , temos que a função de verossimilhança de $a_1, \dots, a_T$ é

$$\mathcal{L}(\phi, \sigma_a^2) = (2\pi\sigma_a^2)^{-\frac{T}{2}} \exp \left\{ -\sum_{t=1}^T a_t^2 / 2\sigma_a^2 \right\}. \quad [1.17]$$

Não é difícil verificar que o vetor ϕ que maximiza [1.17] também maximiza a função de log-verossimilhança $\ell(\phi, \sigma_a^2)$, dada por

$$\ell(\phi, \sigma_a^2) = \log \mathcal{L}(\phi, \sigma_a^2) = -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma_a^2 - \frac{1}{2\sigma_a^2} \sum_{t=1}^T a_t^2,$$

ou, ainda, por

$$\begin{aligned} \ell(\phi, \sigma_a^2) = & -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma_a^2 - \\ & - \frac{1}{2\sigma_a^2} \sum_{t=1}^T (r_t - \phi_0 - \phi_1 r_{t-1} - \cdots - \phi_p r_{t-p})^2. \end{aligned} \quad [1.18]$$

Os estimadores de máxima verossimilhança de ϕ e σ_a^2 são obtidos através da solução do sistema de equações não-lineares provenientes das primeiras derivadas de [1.18], ou seja, de

$$\begin{aligned} \frac{\partial \ell(\phi, \sigma_a^2)}{\partial \phi_0} &= \frac{1}{\sigma_a^2} \sum_{t=1}^T (r_t - \phi_0 - \phi_1 r_{t-1} - \cdots - \phi_p r_{t-p}) = 0 \\ \frac{\partial \ell(\phi, \sigma_a^2)}{\partial \phi_i} &= \frac{1}{\sigma_a^2} \sum_{t=1}^T (r_t - \phi_0 - \phi_1 r_{t-1} - \cdots - \phi_p r_{t-p}) r_{t-i} = 0, \quad i = 1, \dots, p, \\ \frac{\partial \ell(\phi, \sigma_a^2)}{\partial \sigma_a^2} &= -\frac{T}{2\sigma_a^2} + \frac{1}{4\sigma_a^4} \sum_{t=1}^T (r_t - \phi_0 - \phi_1 r_{t-1} - \cdots - \phi_p r_{t-p})^2 = 0. \end{aligned}$$

Ansley & Newbold (1980), em um estudo de simulação de Monte Carlo, avaliaram o comportamento dos métodos de estimação apresentados. Para o caso de modelos auto-regressivos, os resultados obtidos por esses autores sugerem que o método da máxima verossimilhança e o de mínimos quadrados condicionais são semelhantes, enquanto o outro método tem desempenho pouco inferior aos demais.

No mesmo trabalho, esses autores consideraram modelos AR(2) tais que as raízes de $(1 - \phi_1 B - \phi_2 B^2) = 0$ fossem reais ou complexas. Novamente, os resultados sugerem equivalência entre os métodos da máxima verossimilhança e mínimos quadrados condicionais e uma leve inferioridade de mínimos quadrados exatos em horizontes de previsão curtos. Porém, quando os parâmetros estão próximos da fronteira de estacionariedade, os resultados favorecem o método da máxima verossimilhança. Contudo, para previsões em horizontes mais longos e na proximidade da fronteira de estacionariedade, a estimação pelos três métodos é deficiente, havendo, ainda assim, uma leve vantagem do método da máxima verossimilhança.

A precisão dos estimadores pode ser verificada através de intervalos de confiança. Os estimadores de máxima verossimilhança têm distribuição assintótica normal

$$\hat{\boldsymbol{\phi}} \xrightarrow{d} \mathcal{N}(\boldsymbol{\phi}, \mathbf{V}),$$

onde

$$\mathbf{V} = 2\sigma_a^2 \begin{pmatrix} \frac{\partial^2 S_2(\boldsymbol{\phi})}{\partial^2 \phi_0} & \cdots & \frac{\partial^2 S_2(\boldsymbol{\phi})}{\partial \phi_0 \partial \phi_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 S_2(\boldsymbol{\phi})}{\partial \phi_0 \partial \phi_p} & \cdots & \frac{\partial^2 S_2(\boldsymbol{\phi})}{\partial^2 \phi_p} \end{pmatrix}. \quad [1.19]$$

Pode-se mostrar que o estimador de máxima verossimilhança de σ_a^2 é

$$\hat{\sigma}_a^2 = \frac{S_2(\hat{\boldsymbol{\phi}})}{T}, \quad [1.20]$$

e que para T suficientemente grande $\hat{\sigma}_a^2$ e $\hat{\boldsymbol{\phi}}$ são não-correlacionados. Substituindo σ_a^2 por [1.20] na equação [1.19] e calculando as segundas derivadas, obtemos $\hat{\mathbf{V}}$, ou seja, as estimativas das variâncias dos estimadores e das covariâncias entre eles; a partir delas é possível construir intervalos de confiança para os parâmetros. Para os modelos AR(1) e AR(2), essas variâncias são aproximadamente

$$\text{Var}(\hat{\phi}_1) \simeq \frac{1 - \phi_1^2}{T}$$

e

$$\text{Var}(\hat{\phi}_2) = \text{Var}(\hat{\phi}_2) \simeq \frac{1 - \phi_2^2}{T},$$

respectivamente.

Embora a estimação pontual funcione bem, a estimação intervalar tipicamente não é precisa, sobretudo na proximidade da fronteira de estacionariedade. Ansley & Newbold (1980) mostram que a distribuição exata dos

estimadores pode estar bem distante da distribuição assintótica, o que torna a utilização de testes t para verificar a significância dos parâmetros pouco confiável.

Após a identificação do modelo e da estimação dos parâmetros, é necessário que seja investigada a adequabilidade do modelo. Uma alternativa é verificar se as funções de autocorrelação amostrais dos resíduos pertencem ao intervalo de confiança assintótico de nível 95%. Outra alternativa é testar conjuntamente a significância das primeiras m autocorrelações residuais, o que pode ser feito através de um teste de Portmanteau.

Box & Pierce (1970) sugeriram um teste para as autocorrelações dos resíduos que, apesar de não detectar quebras específicas no comportamento do ruído branco, pode identificar se esses valores são muito altos em termos absolutos. Esses autores formularam um teste global envolvendo as m primeiras funções de autocorrelação dos resíduos $\rho_{\hat{a}}$, no qual deseja-se testar a hipótese

$$H_0 : \rho_{\hat{a},1} = \rho_{\hat{a},2} = \dots = \rho_{\hat{a},m} = 0.$$

Admitindo que o modelo estimado é um $AR(p)$, Box & Pierce (1970) mostraram que, sob a hipótese H_0 , a estatística

$$Q_{BP} = T \sum_{k=1}^m \hat{\rho}_{\hat{a},k}^2$$

tem distribuição assintótica χ^2 com $m - p$ graus de liberdade, propriedade que permite concluir pela má qualidade do ajuste quando o valor observado para Q_{BP} exceder o valor crítico associado ao nível de significância adotado.

Segundo Tsay (2002), estudos de simulação sugerem que se considere $m \approx \log(T)$, enquanto Venables & Ripley (2002) recomendam que seja considerado $m = 10$. A rejeição da hipótese nula sugere que o modelo estimado não é apropriado para descrever a série temporal em questão.

Um teste melhorado foi apresentado por Ljung & Box (1978), baseado na estatística

$$Q_{LB} = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_{\hat{a},k}^2}{T-k}, \quad [1.21]$$

que também converge, porém mais rapidamente, para a distribuição χ^2 com $m - p$ graus de liberdade. Por esse fato torna-se mais conveniente utilizar a estatística Q_{LB} quando a amostra não é muito grande.

Como já afirmamos, a obtenção de previsões é um dos principais objetivos em séries temporais, portanto, após a identificação, estimação e validação do modelo, deve-se passar para essa nova etapa.

Para o modelo $AR(p)$ da equação [1.4], suponha que no instante h estejamos interessados em prever o valor de $r_{h+\ell}$, onde $\ell \geq 1$. Dá-se o nome de origem e horizonte de previsão a h e ℓ , respectivamente. Seja $\hat{r}_h(\ell)$ a previsão de $r_{h+\ell}$, a qual dá-se o nome de previsão ℓ -passos à frente de r_t com origem em h . Para a obtenção das previsões utilizaremos as seguintes propriedades:

$$r_{h+\ell} = \begin{cases} \hat{r}_h(\ell), & \text{se } h + \ell > T, \\ r_{h+\ell}, & \text{se } h + \ell \leq T, \end{cases}$$

e

$$a_{h+\ell} = \begin{cases} 0, & \text{se } h + \ell > T, \\ a_{h+\ell}, & \text{se } h + \ell \leq T. \end{cases}$$

A previsão um passo adiante é dada por

$$\hat{r}_h(1) = E(r_{h+1}|r_h, r_{h-1}, \dots) = \phi_0 + \sum_{i=1}^p \phi_i r_{h+1-i},$$

cujo erro de previsão associado é

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1},$$

o qual possui variância

$$\text{Var}[e_h(1)] = \text{Var}(a_{h+1}) = \sigma_a^2.$$

Se a_t for gaussiano, o intervalo de confiança de nível 95% para o erro de previsão um passo à frente de r_{h+1} será

$$\hat{r}_h(1) \pm 1.96 \times \sigma_a.$$

A previsão dois passos à frente é dada por

$$r_{h+2} = \phi_0 + \phi_1 r_{h+1} + \phi_2 r_h + \dots + \phi_p r_{h+2-p} + a_{h+2}.$$

Assim,

$$\hat{r}_h(2) = E(r_{h+2}|r_h, r_{h-1}, \dots) = \phi_0 + \phi_1 \hat{r}_h(1) + \phi_2 r_h + \dots + \phi_p r_{h+2-p},$$

o erro associado sendo

$$e_h(2) = r_{h+2} - \hat{r}_h(2) = \phi_1 [r_{h+1} - \hat{r}_h(1)] + a_{h+2} = a_{h+2} + \phi_1 a_{h+1}.$$

A variância do erro de previsão dois passos adiante é, portanto,

$$\text{Var}[e_h(2)] = (1 + \phi_1^2) \sigma_a^2.$$

Note que $\text{Var}[e_h(2)] \geq \text{Var}[e_h(1)]$.

Da mesma forma, obtemos

$$r_{h+\ell} = \phi_0 + \phi_1 r_{h+\ell-1} + \dots + \phi_p r_{h+\ell-p} + a_{h+\ell}$$

e

$$\hat{r}_h(\ell) = E(r_{h+\ell}|r_h, r_{h-1}, \dots) = \phi_0 + \sum_{i=1}^p \phi_i \hat{r}_h(\ell - i),$$

onde $\hat{r}_h(\ell - i) = r_{h+i}$, se $i \leq 0$, podendo ser obtida recursivamente. Pode-se mostrar que

$$\lim_{\ell \rightarrow \infty} \hat{r}_h(\ell) = E(r_t).$$

A propriedade acima é referida na literatura como “reversão à média”.

1.2.2 O modelo de médias móveis

Outra classe de modelos é a dos modelos de médias móveis. O modelo de médias móveis de ordem q , $MA(q)$, é dado por

$$r_t = \phi_0 + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}. \quad [1.22]$$

É fácil observar que o modelo [1.22] é sempre estacionário, visto que trata-se de uma combinação linear de ruídos brancos, os quais possuem os dois primeiros momentos finitos e invariantes no tempo. Temos que

$$\begin{aligned} E(r_t) &= E(\phi_0 + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}) \\ &= E(\phi_0) + E(a_t) - \theta_1 E(a_{t-1}) - \cdots - \theta_q E(a_{t-q}) = \phi_0. \end{aligned}$$

A variância de r_t é dada por

$$\begin{aligned} \text{Var}(r_t) &= \text{Var}(\phi_0 + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}) \\ &= \text{Var}(a_t) + \theta_1^2 \text{Var}(a_{t-1}) + \cdots + \theta_q^2 \text{Var}(a_{t-q}) \\ &= (1 + \theta_1^2 + \cdots + \theta_q^2) \sigma_a^2 = \left(1 + \sum_{i=1}^q \theta_i^2\right) \sigma_a^2. \end{aligned} \quad [1.23]$$

Aqui, utilizamos o fato de a_t ser ruído branco.

Assuma, por simplicidade, que $\phi_0 = 0$. A função de autocovariância de r_t é obtida multiplicando a equação [1.22] por $r_{t-\ell}$ e tomando o valor esperado. Assim, devido à independência entre $r_{t-\ell}$ e a_t , tem-se que

$$\begin{aligned} E(r_t r_{t-\ell}) &= E\left[(a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q})(a_{t-\ell} - \theta_1 a_{t-\ell-1} - \cdots - \theta_q a_{t-\ell-q})\right] \\ &= E(a_t a_{t-\ell}) - \sum_{i=1}^q \theta_i E(a_{t-i} a_{t-\ell}) - \sum_{j=1}^q \theta_j E(a_t a_{t-\ell-j}) \\ &\quad + \sum_{i=1}^q \sum_{j=1}^q \theta_i \theta_j E(a_{t-i} a_{t-\ell-j}), \end{aligned}$$

onde

$$E(a_t a_{t-\ell}) = \begin{cases} \sigma_a^2, & \text{se } \ell = 0, \\ 0, & \text{se } \ell \neq 0. \end{cases}$$

Assim,

$$\gamma_\ell = \begin{cases} (-\theta_\ell + \theta_1\theta_{\ell+1} + \cdots + \theta_{q-\ell}\theta_q)\sigma_a^2, & \text{se } \ell = 1, 2, \dots, q, \\ 0, & \text{se } \ell > q. \end{cases} \quad [1.24]$$

De [1.23] e [1.24], chegamos à função de autocorrelação

$$\rho_\ell = \begin{cases} \left(\frac{-\theta_\ell + \theta_1\theta_{\ell+1} + \cdots + \theta_{q-\ell}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} \right), & \text{se } \ell = 1, 2, \dots, q, \\ 0, & \text{se } \ell > q. \end{cases}$$

Conseqüentemente, o modelo MA(q) é relacionado linearmente com as suas primeiras q defasagens e, portanto, é um modelo de memória finita. Assim, a função de autocorrelação é útil na identificação da ordem do modelo MA, de forma que se $\rho_\ell = 0$, para $\ell > q$, temos o modelo MA(q). As funções de autocorrelação parciais são obtidas da mesma forma que para os modelos AR apresentada em [1.13] e são diferentes de zero para todo ℓ .

Outra forma de identificação é por meio dos métodos baseados em uma função penalizadora. Assim, os critérios AIC, AICC e BIC apresentados em [1.14], [1.15] e [1.16], respectivamente, podem ser utilizados da mesma forma que para os modelos AR. Contudo, para modelos puramente MA, AIC e AICC não são assintoticamente eficientes.

O processo de estimação por mínimos quadrados para os modelos MA é feito de forma equivalente à adotada para modelos AR. O método de mínimos quadrados condicional assume que, para $t \leq 0$, $a_t = 0$. Assim, para o modelo MA(q) temos a seqüência

$$\begin{aligned} a_1 &= r_1 - \phi_0 \\ a_2 &= r_2 - \phi_0 + \theta_1 a_1 \\ a_3 &= r_3 - \phi_0 + \theta_1 a_2 + \theta_2 a_1 \\ &\vdots \\ a_q &= r_q - \phi_0 + \theta_1 a_{q-1} + \theta_2 a_{q-2} + \cdots + \theta_q a_1. \end{aligned} \quad [1.25]$$

No método de mínimos quadrados exatos, a seqüência [1.25] é adicionada ao modelo como parâmetros adicionais que serão estimados conjuntamente com os demais parâmetros. Em grandes amostras, as estimativas produzidas pelos dois métodos são muito próximas, conforme mostram Ansley & Newbold (1980) para os casos dos modelos MA(1) e MA(2).

A previsão do modelo MA(q) em [1.23] com origem em h e horizonte 1 é dada por

$$r_{h+1} = \phi_0 + a_{h+1} - \theta_1 a_h - \cdots - \theta_q a_{h-q+1}.$$

Tomando o valor esperado condicionado à informação disponível em h , obtemos

$$\hat{r}_h(1) = E(r_{h+1} | r_h, r_{h-1}, \dots) = \phi_0 - \theta_1 a_h - \cdots - \theta_q a_{h-q+1}$$

e, portanto,

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}.$$

Conseqüentemente,

$$\text{Var}[e_h(1)] = \sigma_a^2.$$

1.2.3 O modelo auto-regressivo e de médias móveis

O modelo auto-regressivo e de médias móveis, ARMA, combina os modelo AR e MA discutidos anteriormente em uma estrutura compacta, na qual o número de parâmetros é mantido reduzido.

Fazendo $d = 0$ em [1.3], chegamos ao modelo ARMA(p, q)

$$(1 - \phi_1 B - \cdots - \phi_p B^p)r_t = \phi_0 + (1 - \theta_1 B - \cdots - \theta_q B^q)a_t. \quad [1.26]$$

É importante ressaltar que para modelos ARMA(k, k) os polinômios $(1 - \phi_1 B - \cdots - \phi_p B^k)$ e $(1 - \theta_1 B - \cdots - \theta_q B^k)$ não devem ser coincidentes, visto que o cancelamento desses dois termos faz com que o processo se reduza a um ruído branco.

A média do processo é obtida através do valor esperado de [1.26], de modo que

$$\begin{aligned} E(r_t) &= E(\phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}) \\ &= \phi_0 + \phi_1 E(r_{t-1}) + \cdots + \phi_p E(r_{t-p}). \end{aligned}$$

Sob estacionariedade, temos que $E(r_t) = \mu$, para todo t . Assim,

$$\mu = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}.$$

Note que a média do processo ARMA(p, q) é exatamente a mesma do processo AR(p).

Assumindo novamente, por simplicidade, que $\phi_0 = 0$ e multiplicando o modelo [1.26] por $r_{t-\ell}$, tomando o valor esperado e usando o fato que a_t é independente de $r_{t-\ell}$ e de $a_{t-\ell}$, para $\ell \geq 1$, obtemos a função de autocovariância do processo:

$$\begin{aligned} \gamma_\ell &= E(r_t r_{t-\ell}) \\ &= E[(\phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q})r_{t-\ell}]. \end{aligned}$$

Ou seja,

$$\begin{aligned} \gamma_\ell &= \phi_1 \gamma_{\ell-1} + \cdots + \phi_p \gamma_{\ell-p} + \text{Cov}(r_{t-\ell}, a_t) - \theta_1 \text{Cov}(r_{t-\ell}, a_{t-1}) - \cdots \\ &\quad - \theta_q \text{Cov}(r_{t-\ell}, a_{t-q}), \end{aligned}$$

onde $\text{Cov}(r_{t-\ell}, a_t)$ é a covariância cruzada entre $r_{t-\ell}$ e a_t . Sabemos que, se $\ell > 0$, $\text{Cov}(r_{t-\ell}, a_t) = 0$; caso contrário, $\text{Cov}(r_{t-\ell}, a_t) \neq 0$. Assim,

$$\gamma_\ell = \phi_1 \gamma_{\ell-1} + \cdots + \phi_p \gamma_{\ell-p}, \quad \text{para } \ell > q. \quad [1.27]$$

De [1.27] obtemos a função de autocorrelação de r_t :

$$\rho_\ell = \phi_1 \rho_{\ell-1} + \cdots + \phi_p \rho_{\ell-p}, \quad \text{para } \ell > q.$$

A função de autocorrelação parcial de um modelo ARMA(p, q) pode ser obtida da mesma forma que nos casos anteriores e se comporta como a de um processo puramente MA.

Os processos de identificação, estimação e validação são também idênticos aos apresentados para os modelos AR e MA. No caso da validação, as estatísticas Portmanteau têm distribuição assintótica χ^2_{m-p-q} .

Ansley & Newbold (1980) consideraram a qualidade da estimação em modelos ARMA(1, 1) e observaram que o método da máxima verossimilhança, apesar de ser pouco preciso quando os valores reais de ϕ_1 e θ_1 estão próximos do cancelamento, produz previsões mais precisas que os métodos de mínimos quadrados.

As previsões um passo à frente para o modelo ARMA(p, q) são dadas por

$$\hat{r}_h(1) = E(r_{h+1} | r_h, r_{h-1}, \dots) = \phi_0 + \sum_{i=1}^p \phi_i r_{h+1-i} - \sum_{j=1}^q \theta_j a_{h+1-j},$$

cujos erros de previsão associados são

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}.$$

Logo, $\text{Var}[e_h(1)] = \sigma_a^2$.

Para previsões de horizonte ℓ , temos

$$\hat{r}_h(\ell) = \phi_0 + \sum_{i=1}^p \phi_i \hat{r}_h(\ell - i) - \sum_{j=1}^q \theta_j a_h(\ell - j),$$

onde $a_h(\ell - j) = 0$, se $h + \ell - j > T$; caso contrário, $a_h(\ell - j) = a_{h+\ell-j}$.

1.3 Modelos Heteroscedásticos

Consideraremos a seguir alguns modelos utilizados para modelar a volatilidade de uma série de retornos. O termo volatilidade significa variância condicional da série, que freqüentemente em finanças evolui no tempo. Os modelos apresentados na seção anterior não são adequados para este caso, pois consideram que a variância se mantém constante no tempo.

A volatilidade não é observada diretamente e pode apresentar algumas características. Por exemplo, pode haver aglomerados de volatilidade, ou seja, a volatilidade pode ser elevada durante um período de tempo e baixa fora desse intervalo.

Para este tipo de modelagem, é importante considerar a média e a variância de r_t condicionadas às informações passadas $\mathcal{F}_{t-1} = \{a_1, \dots, a_{t-1}\}$, ou seja,

$$\begin{cases} \mu_t = E(r_t|\mathcal{F}_{t-1}), \\ \sigma_t^2 = \text{Var}(r_t|\mathcal{F}_{t-1}) = E[(r_t - \mu_t)^2|\mathcal{F}_{t-1}], \end{cases} \quad [1.28]$$

onde tipicamente $\mathcal{F}_{t-1}$ consiste de funções lineares dos retornos passados.

1.3.1 O modelo auto-regressivo com heteroscedasticidade condicional

O modelo auto-regressivo com heteroscedasticidade condicional, ARCH, foi introduzido por Engle (1982). A idéia básica é que os componentes da sequência $\{a_t\}$ não são correlacionados serialmente, mas são dependentes e essa dependência pode ser descrita por uma função quadrática de seus valores passados. Alguns autores utilizam o termo choque aleatório para se referir a a_t .

Especificamente, o modelo ARCH(r) assume que

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_r a_{t-r}^2, \end{aligned} \quad [1.29]$$

onde $\{\varepsilon_t\}$ é uma sequência independente e identicamente distribuída com média zero e variância unitária, $\alpha_0 > 0$ e $\alpha_i \geq 0$, para $i > 0$ e as variáveis aleatórias a_t e ε_t são independentes.

Os coeficientes α_i devem satisfazer algumas condições de regularidade para garantir que a variância condicional de a_t seja finita. Na prática, assume-se que ε_t tem distribuição normal ou t -Student padronizada.

Da estrutura do modelo temos que valores elevados de a_t tendem a ser seguidos por valores elevados. É importante destacar que alguns autores utilizam h_t para denotar a variância condicional em [1.29]. Neste caso, temos que $a_t = \sqrt{h_t} \varepsilon_t$.

A média condicional de a_t dado $\mathcal{F}_{t-1}$ é

$$E(a_t|\mathcal{F}_{t-1}) = E(\sigma_t \varepsilon_t|\mathcal{F}_{t-1}) = E(\sigma_t|\mathcal{F}_{t-1})E(\varepsilon_t|\mathcal{F}_{t-1}) = \sigma_t \times 0 = 0.$$

Conseqüentemente, a média incondicional de a_t é

$$E(a_t) = E(E(a_t|\mathcal{F}_{t-1})) = E(0) = 0. \quad [1.30]$$

A variância condicional de a_t dado $\mathcal{F}_{t-1}$ é dada por

$$\begin{aligned} \text{Var}(a_t|\mathcal{F}_{t-1}) &= E(a_t^2|\mathcal{F}_{t-1}) = E(\sigma_t^2 \varepsilon_t^2|\mathcal{F}_{t-1}) = E(\sigma_t^2|\mathcal{F}_{t-1})E(\varepsilon_t^2|\mathcal{F}_{t-1}) \\ &= \sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i a_{t-i}^2. \end{aligned}$$

A variância incondicional de a_t é, portanto,

$$\text{Var}(a_t) = E(a_t^2) = E\{E(a_t^2|\mathcal{F}_{t-1})\} = E(\alpha_0 + \sum_{i=1}^r \alpha_i a_{t-i}^2) = \alpha_0 + \sum_{i=1}^r \alpha_i E(a_{t-i}^2),$$

a qual, sob estacionariedade, se reduz a

$$\text{Var}(a_t) = \alpha_0 + \sum_{i=1}^r \alpha_i E(a_t^2) = \alpha_0 + \sum_{i=1}^r \alpha_i \text{Var}(a_t) = \frac{\alpha_0}{1 - \sum_{i=1}^r \alpha_i}; \quad [1.31]$$

conseqüentemente, $0 \leq \sum_{i=1}^r \alpha_i < 1$.

A função de autocovariância de a_t , para $\ell \geq 1$, é dada por

$$\begin{aligned} \gamma_\ell &= \text{Cov}(a_t, a_{t+\ell}) = E(a_t a_{t+\ell}) = E[E(a_t a_{t+\ell}|\mathcal{F}_{t+\ell-1})] \\ &= E[a_t E(a_{t+\ell}|\mathcal{F}_{t+\ell-1})] = E[a_t E(\sigma_{t+\ell} \varepsilon_{t+\ell}|\mathcal{F}_{t+\ell-1})] \\ &= E[a_t E(\sigma_{t+\ell}|\mathcal{F}_{t+\ell-1}) E(\varepsilon_{t+\ell}|\mathcal{F}_{t+\ell-1})] = E[a_t \sigma_{t+\ell} E(\varepsilon_{t+\ell}|\mathcal{F}_{t+\ell-1})] = 0, \end{aligned}$$

pois ε_t possui média zero. Ou seja, a_t é uma seqüência de variáveis aleatórias não-correlacionadas com média zero e variância dada pela expressão [1.31].

A curtose é uma medida que pode nos auxiliar na identificação da presença de volatilidade. Essa medida é obtida da razão entre o quarto momento de a_t e o quadrado de sua variância. Assim, temos que, para o caso do modelo ARCH(1), o quarto momento de a_t condicionado a $\mathcal{F}_{t-1}$ é dado por

$$E(a_t^4|\mathcal{F}_{t-1}) = E(\sigma_t^4 \varepsilon_t^4|\mathcal{F}_{t-1}) = E(\sigma_t^4|\mathcal{F}_{t-1}) E(\varepsilon_t^4|\mathcal{F}_{t-1}) = \sigma_t^4 E(\varepsilon_t^4|\mathcal{F}_{t-1}) = 3\sigma_t^4,$$

pois, sob normalidade de ε_t , $E(\varepsilon_t^4|\mathcal{F}_{t-1}) = 3$.

Assim, o quarto momento de a_t é dado por

$$\begin{aligned} E(a_t^4) &= E\{E(a_t^4|\mathcal{F}_{t-1})\} = E(3\sigma_t^4) = 3E[(\alpha_0 + \alpha_1 a_{t-1}^2)^2] \\ &= 3E(\alpha_0^2 + 2\alpha_0\alpha_1 a_{t-1}^2 + \alpha_1^2 a_{t-1}^4) = 3[\alpha_0^2 + 2\alpha_0\alpha_1 E(a_{t-1}^2) + \alpha_1^2 E(a_{t-1}^4)] \\ &= 3[\alpha_0^2 + 2\alpha_0\alpha_1 E(a_t^2) + \alpha_1^2 E(a_t^4)] = 3[\alpha_0^2 + 2\alpha_0\alpha_1 \text{Var}(a_t) + \alpha_1^2 E(a_t^4)] \\ &= 3 \left[\alpha_0^2 + 2\alpha_0\alpha_1 \left(\frac{\alpha_0}{1 - \alpha_1} \right) + \alpha_1^2 E(a_t^4) \right] = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}. \end{aligned}$$

Logo, a curtose de a_t é dada por

$$K(a_t) = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \times \frac{(1 - \alpha_1)^2}{\alpha_0^2} = \frac{3(1 - \alpha_1^2)}{(1 - 3\alpha_1^2)}.$$

Note que $K(a_t) > 3$, o que mostra que as caudas da distribuição de a_t são mais pesadas do que as da distribuição normal. Esse fato é referido em finanças como excesso de curtose. Como dissemos anteriormente, a curtose nos auxilia na detecção da volatilidade, pois se há excesso de curtose, então há volatilidade.

Uma maneira simples de construir um modelo ARCH consiste em três etapas:

- [1] construir um modelo econométrico, por exemplo um modelo ARMA para remover a dependência linear (caso exista) da série de retornos e obter seus resíduos a fim de testar o efeito ARCH;
- [2] especificar a ordem do modelo ARCH e estimar os parâmetros desconhecidos;
- [3] verificar a adequabilidade do modelo ARCH considerado.

Essas etapas serão discutidas a seguir.

Um modelo ARMA deve ser escolhido e dele deve-se obter os resíduos correspondentes. A série a_t^2 é então utilizada para verificar a heteroscedasticidade condicional, onde a_t é definida como a seqüência dos resíduos do modelo ARMA.

Para testar o efeito ARCH pode-se utilizar dois testes. O primeiro deles é o de Ljung–Box, definido em [1.21], no qual, ao invés de utilizarmos a série y_t utilizaremos a_t^2 . Contudo, diferentemente do que acontece nos modelos ARMA, concluimos pela adequabilidade do modelo se não rejeitamos a hipótese nula, ou seja, se existe correlação serial nos quadrados dos resíduos do modelo ARMA.

O segundo teste é o proposto por Engle (1982), conhecido como teste dos multiplicadores de Lagrange, o qual é equivalente ao teste F de $H_0 : \alpha_i = 0$, $i = 1, \dots, r$, no modelo de regressão linear

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_r a_{t-r}^2 + e_t, \quad t = r + 1, \dots, T.$$

A estatística de teste é

$$LM = TR^2,$$

onde R^2 é o coeficiente de determinação da regressão. A estatística LM tem distribuição assintótica qui-quadrado com r graus de liberdade.

Assim como nos modelos AR, a função de autocorrelação parcial pode ser usada para sugerir a ordem do modelo ARCH, só que neste caso devemos considerar a função de autocorrelação de a_t^2 . Vale destacar que se a_t segue um processo ARCH(r), a_t^2 segue um processo AR(r).

Os estimadores dos parâmetros do modelo ARCH são obtidos através do método da máxima verossimilhança. Sob a suposição de normalidade dos a_t , a função de verossimilhança de um modelo ARCH(r) é

$$\begin{aligned} f(a_1, \dots, a_T | \boldsymbol{\alpha}) &= f(a_T | \mathcal{F}_{T-1}) f(a_{T-1} | \mathcal{F}_{T-2}) \cdots f(a_{r+1} | \mathcal{F}_r) f(a_1, \dots, a_m | \boldsymbol{\alpha}) \\ &= \prod_{t=r+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left[-\frac{a_t^2}{2\sigma_t^2} \right] \times f(a_1, \dots, a_m | \boldsymbol{\alpha}), \end{aligned}$$

onde $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_r)'$, $f(a_1, \dots, a_m | \boldsymbol{\alpha})$ sendo a função de densidade conjunta de $a_1, \dots, a_r$. A forma exata de $f(a_1, \dots, a_m | \boldsymbol{\alpha})$ é muito complicada e é freqüentemente substituída pela função de verossimilhança *a priori*, especialmente quando o tamanho da amostra é suficientemente grande, o que resulta

no método de máxima verossimilhança condicional

$$f(a_{r+1}, \dots, a_T | \boldsymbol{\alpha}, a_1, \dots, a_r) = \prod_{t=r+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left[-\frac{a_t^2}{2\sigma_t^2} \right],$$

onde σ_t^2 pode ser avaliada recursivamente. A função de log-verossimilhança é, portanto,

$$\ell(a_{r+1}, \dots, a_T | \boldsymbol{\alpha}, a_1, \dots, a_r) = \sum_{t=r+1}^T -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2) - \frac{1}{2} \frac{a_t^2}{\sigma_t^2}. \quad [1.32]$$

As estimativas obtidas através da maximização de [1.32] são chamadas de estimativas de máxima verossimilhança condicional sob normalidade. Para casos de não-normalidade, veja Tsay (2002).

Para verificar a adequabilidade do modelo, considere os choques padronizados

$$\tilde{a}_t = \frac{a_t}{\sigma_t},$$

os quais são variáveis aleatórias independentes e identicamente distribuídas com distribuição normal padrão ou t -Student padronizada. Deve-se calcular a estatística de Ljung–Box em [1.21] para $\tilde{a}_t$ e verificar a sua significância. Além disso, os coeficientes de assimetria e curtose, bem como um gráfico normal de probabilidades (QQ–Plot), podem ser utilizados para verificar a validade da suposição de normalidade.

Também é possível verificar, por meio do teste LM em $\tilde{a}_t^2$, se ainda existe heteroscedasticidade condicional nos resíduos do modelo.

Após concluídas as etapas de identificação, estimação e verificação da adequabilidade do modelo considerado, as previsões poderão ser obtidas. A previsão um passo à frente com origem em h para σ_{h+1}^2 é

$$\hat{\sigma}_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \dots + \alpha_r a_{h+1-r}^2,$$

enquanto que as previsões para dois e ℓ passos adiante são dadas por, respectivamente,

$$\hat{\sigma}_h^2(2) = \alpha_0 + \alpha_1 \hat{\sigma}_h^2(1) + \alpha_2 a_h^2 + \dots + \alpha_r a_{h+2-r}^2$$

e

$$\hat{\sigma}_h^2(\ell) = \alpha_0 + \sum_{i=1}^r \alpha_i \hat{\sigma}_h^2(\ell-1),$$

onde $\hat{\sigma}_h^2(\ell-i) = a_{h+\ell-i}^2$ se $(\ell-i) \leq 0$.

1.3.2 O modelo ARCH generalizado

Assim como um modelo ARMA pode ser mais parcimonioso do que modelos puramente AR ou MA, os modelos GARCH propostos por Bollerslev (1986), em geral, possuem menos parâmetros do que os modelos ARCH.

Para a série de log-retornos r_t , assumamos que a média do modelo pode ser descrita por um modelo ARMA. Seja $a_t = r_t - \mu_t$ o log-retorno corrigido pela média, então a_t segue um processo GARCH(r, s) se

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^r \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2, \end{aligned} \quad [1.33]$$

onde $\{\varepsilon_t\}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média zero e variância unitária, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ e $\sum_{i=1}^m (\alpha_i + \beta_i) < 1$, com $m = \max(r, s)$, $\alpha_i = 0$, se $i > r$ e $\beta_j = 0$, se $j > s$. A restrição sobre $(\alpha_i + \beta_i)$ implica que a variância incondicional é finita, enquanto a variância condicional evolui no tempo. Assumimos novamente que a distribuição de ε_t é normal padrão ou t -Student. Note que se $s = 0$ a equação [1.33] se resume ao modelo ARCH(r) em [1.29].

Seja $\nu_t = a_t^2 - \sigma_t^2 = \sigma_t^2 \varepsilon_t^2 - \sigma_t^2 = \sigma_t^2 (\varepsilon_t^2 - 1)$, logo $\sigma_t^2 = a_t^2 - \nu_t$. Se assumirmos que ε_t tem distribuição normal padrão, então $\varepsilon_t^2 \sim \chi_1^2$ e, assim, temos que ν_t tem variância condicional não constante igual a $2\sigma_t^2$. Substituindo σ_t^2 em [1.33], podemos reescrever o modelo como

$$a_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_j) a_{t-i}^2 + \nu_t - \sum_{j=1}^s \beta_j \nu_{t-j}. \quad [1.34]$$

Note que o modelo em [1.34] é um modelo ARMA(m, s) para a série a_t^2 .

A variável ν_t é uma diferença martingal, pois

$$\begin{aligned} E(\nu_t) &= E\{E(\nu_t | \mathcal{F}_{t-1})\} = E\{E[\sigma_t^2 (\varepsilon_t^2 - 1) | \mathcal{F}_{t-1}]\} \\ &= E\{E(\sigma_t^2 | \mathcal{F}_{t-1}) E[(\varepsilon_t^2 - 1) | \mathcal{F}_{t-1}]\} = E\{\sigma_t^2 [E(\varepsilon_t^2 | \mathcal{F}_{t-1}) - 1]\} \\ &= E\{\sigma_t^2 (1 - 1)\} = 0 \end{aligned}$$

e

$$\begin{aligned} \text{Cov}(\nu_t, \nu_{t+\ell}) &= E(\nu_t \nu_{t+\ell}) = E\{E(\nu_t \nu_{t+\ell} | \mathcal{F}_{t+\ell-1})\} \\ &= E\{E(\nu_t | \mathcal{F}_{t+\ell-1}) E(\nu_{t+\ell} | \mathcal{F}_{t+\ell-1})\} \\ &= E\{\nu_t E[\sigma_{t+\ell}^2 (\varepsilon_{t+\ell}^2 - 1) | \mathcal{F}_{t+\ell-1}]\} \\ &= E\{\nu_t E(\sigma_{t+\ell}^2 | \mathcal{F}_{t+\ell-1}) E[(\varepsilon_{t+\ell}^2 - 1) | \mathcal{F}_{t+\ell-1}]\} \\ &= E\{\nu_t \sigma_{t+\ell}^2 E[(\varepsilon_{t+\ell}^2 - 1) | \mathcal{F}_{t+\ell-1}]\} = E\{\nu_t \sigma_{t+\ell}^2 \times 0\} = 0, \end{aligned}$$

para $\ell \geq 1$.

A média condicional de a_t dado $\mathcal{F}_{t-1}$ é

$$E(a_t|\mathcal{F}_{t-1}) = E(\sigma_t \varepsilon_t|\mathcal{F}_{t-1}) = E(\sigma_t|\mathcal{F}_{t-1})E(\varepsilon_t|\mathcal{F}_{t-1}) = \sigma_t E(\varepsilon_t|\mathcal{F}_{t-1}) = 0.$$

Assim como nos modelos ARCH, a esperança incondicional de a_t é igual a zero, pois

$$E(a_t) = E\{E(a_t|\mathcal{F}_{t-1})\} = E(0) = 0.$$

A variância condicional de a_t é dada por

$$\text{Var}(a_t|\mathcal{F}_{t-1}) = E(a_t^2|\mathcal{F}_{t-1}) = E(\sigma_t^2 \varepsilon_t^2|\mathcal{F}_{t-1}) = E(\sigma_t^2|\mathcal{F}_{t-1})E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2.$$

A variância incondicional de a_t é

$$\begin{aligned} \text{Var}(a_t) &= E(\sigma_t^2) = E \left[\alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_j) a_{t-i}^2 + \nu_t - \sum_{j=1}^s \beta_j \nu_{t-j} \right] \\ &= \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_j) E(a_{t-i}^2) + E(\nu_t) + \sum_{j=1}^s \beta_j E(\nu_{t-j}) \\ &= \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_j) \text{Var}(a_{t-i}). \end{aligned}$$

Sob estacionariedade, $\text{Var}(a_t) = \text{Var}(a_{t-i})$ e, portanto,

$$\text{Var}(a_t) = \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_j) \text{Var}(a_t) = \frac{\alpha_0}{1 - \sum_{i=1}^m (\alpha_i + \beta_j)}$$

cujo denominador é sempre positivo, pois $\sum_{i=1}^m (\alpha_i + \beta_i) < 1$. A função de autocovariância é dada por

$$\begin{aligned} \gamma_\ell &= \text{Cov}(a_t, a_{t+\ell}) = E(a_t a_{t+\ell}) = E\{E(a_t a_{t+\ell}|\mathcal{F}_{t+\ell-1})\} \\ &= E\{a_t E(a_{t+\ell}|\mathcal{F}_{t+\ell-1})\} = E(a_t \times 0) = 0. \end{aligned}$$

Considere o modelo GARCH(1,1)

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned} \tag{1.35}$$

onde $0 \leq \alpha_1, \beta_1 \leq 1$ e $(\alpha_1 + \beta_1) < 1$. Nota-se que volatilidades altas são precedidas de retornos ou volatilidades também elevadas, o que gera os aglomerados de volatilidade presentes em séries financeiras.

O quarto momento condicional de a_t para o modelo GARCH(1,1) é

$$E(a_t^4|\mathcal{F}_{t-1}) = E(\sigma_t^4 \varepsilon_t^4|\mathcal{F}_{t-1}) = E(\sigma_t^4|\mathcal{F}_{t-1})E(\varepsilon_t^4|\mathcal{F}_{t-1}) = \sigma_t^4 E(\varepsilon_t^4|\mathcal{F}_{t-1}) = 3\sigma_t^4,$$

sob normalidade de ε_t . Logo,

$$E(a_t^4) = E\{E(a_t^4|\mathcal{F}_{t-1})\} = E(3\sigma_t^4) = 3E(\sigma_t^4),$$

mas

$$\begin{aligned} E(\sigma_t^4) &= E[(\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^2] \\ &= E(\alpha_0^2 + 2\alpha_0\alpha_1 a_{t-1}^2 + 2\alpha_0\beta_1 \sigma_{t-1}^2 + \alpha_1^2 a_{t-1}^4 + \beta_1^2 \sigma_{t-1}^4 + 2\alpha_1\beta_1 a_{t-1}^2 \sigma_{t-1}^2). \end{aligned}$$

Sob estacionariedade, temos que

$$E(\sigma_t^4) = \alpha_0^2 + 2\alpha_0\alpha_1 E(a_t^2) + 2\alpha_0\beta_1 E(\sigma_t^2) + \alpha_1^2 E(a_t^4) + \beta_1^2 E(\sigma_t^4) + 2\alpha_1\beta_1 E(a_t^2 \sigma_t^2),$$

que pode ser simplificada, usando as propriedades a seguir:

- (i) $E(\sigma_t^2) = E(a_t^2 - \nu_t) = E(a_t^2) = \text{Var}(a_t)$;
- (ii) $E(a_t^2 \sigma_t^2) = E(\sigma_t^4 \varepsilon_t^2) = E(\sigma_t^4) E(\varepsilon_t^2) = E(\sigma_t^4)$;
- (iii) $E(a_t^4) = E(\sigma_t^4 \varepsilon_t^4) = E(\sigma_t^4) E(\varepsilon_t^4) = 3E(\sigma_t^4)$.

A propriedade (iii) utiliza a suposição de normalidade em ε_t .

Assim, temos que

$$\begin{aligned} E(\sigma_t^4) &= \alpha_0^2 + 2\alpha_0\alpha_1 \text{Var}(a_t) + 2\alpha_0\beta_1 \text{Var}(a_t) + 3\alpha_1^2 E(\sigma_t^4) + \beta_1^2 E(\sigma_t^4) + 2\alpha_1\beta_1 E(\sigma_t^4) \\ &= \alpha_0^2 + (2\alpha_0\alpha_1 + 2\alpha_0\beta_1) \text{Var}(a_t) + (3\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1) E(\sigma_t^4) \\ &= \frac{\alpha_0^2 + 2\alpha_0(\alpha_1 + \beta_1) \text{Var}(a_t)}{1 - (3\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1)} = \frac{\alpha_0^2 + 2\alpha_0(\alpha_1 + \beta_1) \left(\frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \right)}{1 - (3\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1)} \\ &= \frac{\alpha_0^2 [1 - (\alpha_1 + \beta_1)] + 2\alpha_0^2(\alpha_1 + \beta_1)}{[1 - (3\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1)] [1 - (\alpha_1 + \beta_1)]} \\ &= \frac{\alpha_0^2 [1 + (\alpha_1 + \beta_1)]}{[1 - (3\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1)] [1 - (\alpha_1 + \beta_1)]}. \end{aligned}$$

Portanto, a curtose de a_t é dada por

$$\begin{aligned} K(a_t) &= \frac{3\alpha_0^2 [1 + (\alpha_1 + \beta_1)]}{[1 - (3\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1)] [1 - (\alpha_1 + \beta_1)]} \times \frac{[1 - (\alpha_1 + \beta_1)]^2}{\alpha_0^2} \\ &= \frac{3 [1 - (\alpha_1 + \beta_1)^2]}{[1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2]}, \end{aligned}$$

a qual, assim como nos modelos ARCH, é sempre superior a 3.

As previsões para processos GARCH podem ser obtidas de modo similar às dos modelos ARMA anteriormente apresentados. Por simplicidade, vamos considerar o modelo GARCH(1,1) dado por [1.35]. A previsão da volatilidade um passo adiante com origem em h , admitindo α_0 , α_1 e β_1 conhecidos, é dada por

$$\hat{\sigma}_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2,$$

onde a_h^2 e σ_h^2 são conhecidos, visto que σ_t^2 pode ser estimada por

$$\sigma_t^2 = \frac{1}{T} \sum_{i=1}^T a_i^2.$$

Para previsões em horizontes maiores, consideramos $a_t^2 = \sigma_t^2 \varepsilon_h^2$ e reescrevemos a equação [1.35] como

$$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 + \alpha_1\sigma_t^2(\varepsilon_t^2 - 1) = \sigma_{t+1}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 + \alpha_1\nu_t.$$

Fazendo $t = h + 1$, temos

$$\sigma_{h+2}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_{h+1}^2 + \alpha_1\nu_{h+1}.$$

Dado que $E(\nu_{h+1}|\mathcal{F}_h) = 0$, então a previsão dois passos à frente é

$$\hat{\sigma}_t^2(2) = \alpha_0 + (\alpha_1 + \beta_1)\hat{\sigma}_h^2(1).$$

A previsão ℓ passos adiante satisfaz a equação

$$\hat{\sigma}_h^2(\ell) = \alpha_0 + (\alpha_1 + \beta_1)\hat{\sigma}_h^2(\ell - 1), \quad \ell > 1. \quad [1.36]$$

Note que o resultado acima é exatamente o mesmo do modelo ARMA(1,1), onde o polinômio auto-regressivo é $1 - (\alpha_1 + \beta_1)B$. Além disso, quando ℓ tende a infinito, temos que a previsão da volatilidade converge para a variância incondicional de a_t , isto é,

$$\sigma_h^2(\ell) \longrightarrow \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \quad \text{quando } \ell \rightarrow \infty.$$

Para os modelos GARCH, os processos de estimação, identificação e verificação são similares aos dos modelos ARCH.

1.3.3 O modelo GARCH integrado

Se o polinômio auto-regressivo do modelo GARCH em [1.33] possuir uma raiz unitária, então o modelo GARCH é dito ser integrado, também conhecido como modelo IGARCH. O modelo IGARCH(1,1) é dado por

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + (1 - \beta_1)a_{t-1}^2 + \beta_1\sigma_{t-1}^2, \end{aligned}$$

onde $0 < \beta_1 < 1$ e $\{\varepsilon_t\}$ é ruído branco. A previsão da volatilidade um passo adiante com origem em h pode ser obtida fazendo $\alpha_1 + \beta_1 = 1$ em [1.36], de onde obtemos

$$\sigma_h^2(\ell) = \sigma_h^2(1) + (\ell - 1)\alpha_0, \quad \ell \geq 1. \quad [1.37]$$

É importante ressaltar que se $\alpha_0 = 0$, então as previsões da volatilidade em [1.37] se resumem a $\sigma_h^2(1)$, para qualquer horizonte de previsão.

Outros modelos têm sido bastante explorados na literatura de finanças, como, por exemplo, os modelos GARCH-M, GARCH exponencial (EGARCH) e GARCH com limiar (TGARCH). Maiores detalhes sobre esses modelos podem ser encontrados em Tsay (2002).

1.4 Modelos de Longa Dependência

Como vimos anteriormente, a metodologia de Box–Jenkins assume estacionariedade na série, comportamento caracterizado pelo decaimento geométrico da correspondente função de autocorrelação. Funções de autocorrelação com padrão de decaimento mais lento fornecem indícios de não-estacionariedade, exigindo a utilização de testes para identificar a natureza da tendência. Se a tendência é estocástica, define-se, então, o parâmetro d do modelo que determina quantas vezes a série deverá ser diferenciada para obter uma série de modo que se estabilize a tendência nela existente.

Muitas séries temporais apresentam longa persistência, ou seja, o tempo necessário para que o efeito de um determinado choque aleatório seja totalmente dissipado é elevado. Nesse caso, os modelos ARIMA se tornam inadequados. Granger & Joyeux (1980) propuseram uma classe que permite a modelagem de séries com essa característica. A proposta desses autores permite que o parâmetro d do modelo ARIMA seja fracionário e, até mesmo, negativo. Com isso, o parâmetro d passa a modelar a longa dependência dos dados, enquanto os demais parâmetros modelam a curta dependência. Essa classe é conhecida como modelos auto-regressivos e de médias móveis com integração fracionária, ARFIMA.

A idéia básica parte de um passeio aleatório que se transforma num ruído branco após a aplicação de diferenças de primeira ordem. Quando utilizamos diferenças fracionárias, de ordem 0.5, por exemplo, deveremos obter um processo de características intermediárias entre a não-estacionariedade e a estacionariedade. As diferenças inteiras de primeira ordem aplicam o operador $(1 - B)$ à série y_t , fornecendo a nova série $z_t = y_t - y_{t-1}$. De forma geral, para tornar o conceito operacional, pode-se usar a expansão binomial para qualquer número real $d > -1$ (Mills, 1999):

$$\Delta^d = (1 - B)^d = \sum_{i=0}^d \binom{d}{i} (-B)^i,$$

onde

$$\binom{d}{i} = \frac{d}{i} \frac{d-1}{i-1} \cdots \frac{d-i+1}{1}.$$

Assim,

$$\Delta^d = 1 - dB + \frac{d(d-1)}{2!} B^2 - \frac{d(d-1)(d-2)}{3!} B^3 + \dots$$

Na prática, é necessário estimar o parâmetro d . Vários estimadores do parâmetro d do modelo ARFIRMA foram propostos na literatura; veja, por exemplo, Geweke e Porter–Hudak (1983) e Reisen (1994).

Valor em Risco

As instituições financeiras, visando à prevenção de grandes perdas inesperadas, desenvolveram, na década de 90, uma ferramenta para medir o risco de suas carteiras de investimento. Essa medida é conhecida como Valor em Risco (**VaR**) e tem sido amplamente utilizada na gestão de riscos de carteiras de investimento. O **VaR** pode ser usado por instituições financeiras para avaliar seus riscos ou por um comitê regulador para definir requerimentos marginais. Em ambos os casos, **VaR** é usado para assegurar que instituições financeiras, após um evento catastrófico, possam continuar suas atividades sem que sejam levadas à falência.

Existem vários tipos de risco em mercados financeiros. Alguns exemplos são os riscos de crédito, de liquidez e de mercado. O **VaR** é uma medida associada ao risco de mercado e é uma estimativa do valor que uma instituição pode perder devido a movimentos de mercado durante um determinado período; para uma exposição geral do **VaR**, ver Duffie & Pan (1997) e Jorion (1997).

Do ponto de vista de uma instituição financeira, **VaR** pode ser definido como a perda máxima de uma posição financeira durante um dado período de tempo com uma dada probabilidade. Então, **VaR** é uma medida de perda associada a um evento raro (ou extraordinário) sob condições normais de mercado. Alternativamente, do ponto de vista de um comitê regulador, **VaR** pode ser definido como a perda mínima sob circunstâncias de mercado extraordinárias. Ambas as definições conduzem à mesma medida, ainda que os conceitos sejam aparentemente diferentes.

Existem duas abordagens diferentes para **VaR**, cada uma associada a um tipo diferente de posição financeira. A primeira delas utiliza um tipo de posição conhecida como posição financeira comprada, ou seja, possuir determinado ativo. O outro tipo de posição é conhecido como posição financeira vendida. Esta operação envolve a venda de ativos que não se possui. Ou seja, o vendedor aluga o ativo e depois é obrigado a comprar o mesmo número de cotas ou ações alugadas. Conseqüentemente, ele lucra com a diminuição do preço das ações,

enquanto que na outra abordagem lucra-se dinheiro com o aumento do valor das mesmas.

A seguir, definiremos **VaR** sob uma estrutura probabilística. Suponha que, no instante t , estejamos interessados no risco de uma posição financeira para os próximos m períodos. Seja $\Delta V(m)$ a mudança no valor dos bens de uma posição do tempo t ao tempo $t + m$, ou seja,

$$\Delta V(m) = V(t + m) - V(t).$$

Denotamos a função de distribuição acumulada de $\Delta V(m)$ por $F_m(x)$. Definimos **VaR** de uma posição comprada no horizonte de tempo m com probabilidade p tal que

$$p = P[\Delta V(m) \leq \text{VaR}] = F_m(\text{VaR}), \quad [2.1]$$

ou seja, $\text{VaR} = F_m^{-1}(p)$. Perdas ocorrem quando $\Delta V(m) < 0$.

Alternativamente, **VaR** pode ser interpretado como segue. Com probabilidade $(1 - p)$, a perda potencial encontrada pelo portador de uma posição financeira vendida sob o horizonte m é menor ou igual ao **VaR**. Neste caso, o investidor de uma posição vendida sofre perda quando o valor dos bens aumenta, isto é, quando $\Delta V(m) > 0$. Então, o **VaR** é definido como

$$p = P[\Delta V(m) \geq \text{VaR}] = 1 - P[\Delta V(m) < \text{VaR}] = 1 - F_m(\text{VaR}).$$

As definições anteriores mostram que **VaR** está relacionado com o comportamento das caudas de $F_m(x)$. Para posições compradas, a cauda esquerda é importante. Já uma posição vendida enfoca a cauda direita de $F_m(x)$. Note que a definição [2.1] também vale para uma posição vendida se utilizamos a distribuição de $-\Delta V(m)$. Portanto, discutiremos métodos para cálculo de **VaR** usando apenas posições compradas.

Para alguma função de distribuição acumulada $F_m(x)$ e uma probabilidade p , tal que $0 < p < 1$, a quantidade

$$x_p = \inf\{x | F_m(x) \geq p\}$$

é chamada de p -ésima função quantílica de $F_m(x)$. Se a distribuição de $\Delta V(m)$, em [2.1], é conhecida, então **VaR** é simplesmente o p -ésimo quantil de $\Delta V(m)$, ou seja, $\text{VaR} = x_p$. Porém, na prática, $F_m(x)$ é desconhecida. Estudos de estimação de **VaR** são essencialmente relacionados com a estimação de $F_m(x)$ e/ou seus quantis e, especialmente, com o comportamento da cauda de $F_m(x)$.

Em aplicações práticas, o cálculo de **VaR** envolve vários fatores:

- i. a probabilidade de interesse p . Por exemplo, $p = 0.01$ ou $p = 0.05$;
- ii. o horizonte m , que pode ser determinado por um comitê regulador. Por exemplo, $m = 1$ dia ou $m = 10$ dias;
- iii. a frequência dos dados, que não necessariamente coincide com o horizonte m . Frequentemente são utilizadas observações diárias;
- iv. a função de distribuição acumulada $F_m(x)$ ou seus quantis;
- v. a quantidade investida ou o valor da carteira marcado pelo mercado.

Dentre esses fatores, $F_m(x)$ é o foco da modelagem econométrica. Diferentes métodos de estimação de $F_m(x)$ conduzem a diferentes abordagens para o cálculo de VaR.

É importante ressaltar que VaR definido em [2.1] é medido em unidades monetárias. Conseqüentemente, VaR indicará o quanto se pode perder em valores da moeda investida.

A seguir, discutiremos algumas abordagens para o cálculo de VaR. Exemplos para cada uma delas podem ser obtidos em Tsay (2002) e Morettin (2005).

2.1 RiskMetrics

J.P. Morgan desenvolveu uma metodologia para cálculo de VaR denominada RiskMetrics™ (Longerstaey & More, 1995). RiskMetrics assume que os log-retornos diários de uma carteira de ativos, dada a informação passada, seguem distribuição normal condicional com média e variância condicionais μ_t e σ_t^2 , respectivamente. Então, se $\mathcal{F}_{t-1}$ é o conjunto de informações disponíveis em $t - 1$, o RiskMetrics assume que

$$r_t | \mathcal{F}_{t-1} \sim \mathcal{N}(\mu_t, \sigma_t^2).$$

Em sua forma mais simples, o método assume que $\mu_t = 0$ e que σ_t^2 evolui de acordo com o modelo

$$\sigma_t^2 = \alpha \sigma_{t-1}^2 + (1 - \alpha) r_{t-1}^2, \quad 0 < \alpha < 1. \quad [2.2]$$

Portanto, $r_t = \sigma_t \varepsilon_t$ é um processo IGARCH(1,1) sem “drift” e ε_t é um ruído branco gaussiano de variância unitária e independente de σ_t .

De [1.2], podemos escrever

$$r_t[k] | \mathcal{F}_t \sim \mathcal{N}(0, \sigma_t^2[k]),$$

onde a volatilidade $\sigma_t^2[k]$ pode ser estimada através dos métodos de modelagem de séries financeiras discutidos no Capítulo 1. Segue-se que

$$\begin{aligned} \sigma_t^2[k] &= \text{Var}(r_t[k] | \mathcal{F}_t) = \sum_{i=1}^k \text{Var}(r_{t+i} | \mathcal{F}_t) \\ &= \sum_{i=1}^k \text{Var}(\sigma_{t+i} \varepsilon_{t+i} | \mathcal{F}_t) = \sum_{i=1}^k E(\sigma_{t+i}^2 \varepsilon_{t+i}^2 | \mathcal{F}_t) \\ &= \sum_{i=1}^k E(\sigma_{t+i}^2 | \mathcal{F}_t) E(\varepsilon_{t+i}^2 | \mathcal{F}_t) = \sum_{i=1}^k E(\sigma_{t+i}^2 | \mathcal{F}_t). \end{aligned}$$

Usando $r_{t-1} = \sigma_{t-1}\varepsilon_{t-1}$, a volatilidade em [2.2] pode ser reescrita como

$$\begin{aligned}\sigma_t^2 &= \alpha\sigma_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2\varepsilon_{t-1}^2 \\ &= \alpha\sigma_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2\varepsilon_{t-1}^2 + \sigma_{t-1}^2 - \sigma_{t-1}^2 \\ &= \sigma_{t-1}^2 + (1-\alpha)\varepsilon_{t-1}^2\sigma_{t-1}^2 - (1-\alpha)\sigma_{t-1}^2 \\ &= \sigma_{t-1}^2 + (1-\alpha)(\varepsilon_{t-1}^2 - 1)\sigma_{t-1}^2.\end{aligned}$$

Em particular, temos

$$\sigma_{t+i}^2 = \sigma_{t+i-1}^2 + (1-\alpha)(\varepsilon_{t+i-1}^2 - 1)\sigma_{t+i-1}^2, \quad i = 2, \dots, k.$$

Como $E[(\varepsilon_{t+i-1}^2 - 1)|\mathcal{F}_t] = 0$, $i \geq 2$, obtemos

$$E(\sigma_{t+i}^2|\mathcal{F}_t) = E(\sigma_{t+i-1}^2|\mathcal{F}_t), \quad i = 2, \dots, k. \quad [2.3]$$

Para previsões da volatilidade um passo à frente, de [2.2] obtemos que

$$\hat{\sigma}_t^2(1) = \alpha\sigma_t^2 + (1-\alpha)r_t^2.$$

Portanto, de [2.3] chegamos a $\text{Var}(r_{t+i}|\mathcal{F}) = \sigma_{t+1}^2$ e, então, $\sigma_t^2[k] = k\sigma_{t+1}^2$. Assim, temos que

$$r_t[k]|\mathcal{F}_t \sim \mathcal{N}(0, k\sigma_{t+1}^2).$$

Conseqüentemente, sob o modelo IGARCH(1,1) em [2.2], a variância condicional de $r_t[k]$ é proporcional ao horizonte k .

Se consideramos uma posição comprada, então perdas ocorrem quando há uma grande queda nos preços, fato que faz com que grandes log-retornos negativos ocorram. Assim, se a probabilidade em [2.1] é fixada em 5%, isto é, $p = 0.05$, o RiskMetrics usa $-1.65\sigma_{t+1}$ para medir o risco da carteira, ou seja, utiliza o quantil de nível 5% da distribuição normal de média zero e variância σ_{t+1}^2 . Em geral, o sinal negativo, que indica perda, é ignorado.

Desta forma, o VaR de um e de k períodos são dados, respectivamente, por

$$\text{VaR} = 1.65 \times \sigma_{t+1} \times Q$$

e

$$\text{VaR}[k] = 1.65 \times \sqrt{k}\sigma_{t+1} \times Q = \sqrt{k}\text{VaR},$$

onde Q é o montante investido na carteira de ativos. Esta regra é conhecida como *regra da raiz quadrada do tempo*.

Uma vantagem do RiskMetrics é a facilidade de compreensão e aplicação. Contudo, quando os log-retornos têm caudas pesadas, violando, portanto, a suposição de normalidade, subestima-se o valor de VaR.

A regra da raiz quadrada do tempo é consequência do modelo utilizado pelo RiskMetrics. Se a suposição de que $\mu_t = 0$ não for válida, a regra, como veremos a seguir, também não será válida. Considere o modelo

$$\begin{aligned}r_t &= \mu_t + \sigma_t\varepsilon_t, \quad \mu_t \neq 0, \\ \sigma_t^2 &= \alpha\sigma_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2\varepsilon_{t-1}^2,\end{aligned}$$

onde ε_t é um ruído branco gaussiano. Então,

$$r_{t+1}|\mathcal{F}_t \sim \mathcal{N}(\mu, \sigma_{t+1}^2)$$

e o p -ésimo quantil utilizado no cálculo de VaR um passo à frente é dado por

$$\mu - z_p \sigma_{t+1},$$

onde z_p é o p -ésimo quantil da distribuição normal padrão. Para k períodos, o p -ésimo quantil é dado por

$$k\mu - z_p \sqrt{k} \sigma_{t+1} \neq \sqrt{k}(\mu - z_p \sigma_{t+1}).$$

A regra também falha se o modelo da volatilidade não é um IGARCH(1,1) sem “drift”.

2.2 Abordagem econométrica

Uma abordagem geral para cálculo de VaR é o uso dos modelos de séries temporais, vistos no Capítulo 1. Para uma série de log-retornos, r_t , é possível estimar a média através dos modelos da família ARIMA, enquanto que a volatilidade pode ser estimada utilizando-se modelos da família GARCH. É importante ressaltar que aqui, diferentemente do que ocorre no RiskMetrics, r_t segue um processo ARMA–GARCH.

Um modelo de séries temporais para r_t pode ser escrito como

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + a_t - \sum_{j=1}^q \theta_j a_{t-j}, \quad [2.4]$$

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^r \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2. \end{aligned} \quad [2.5]$$

Este modelo é chamado ARMA(p, q)–GARCH(r, s).

As equações [2.4] e [2.5] são, respectivamente, as equações da média e da volatilidade de r_t . Estas equações podem ser utilizadas na obtenção de previsões da média e variância condicionais de r_t um passo adiante assumindo-se que os parâmetros são conhecidos. Assim, temos

$$\begin{aligned} \hat{r}_t(1) &= \phi_0 + \sum_{i=1}^p \phi_i r_{t-i+1} + a_{t+1} - \sum_{j=1}^q \theta_j a_{t-j+1}, \\ \hat{\sigma}_t^2(1) &= \alpha_0 + \sum_{i=1}^r \alpha_i a_{t-i+1}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j+1}^2. \end{aligned}$$

Se assumirmos que ε_t é gaussiano, então

$$r_{t+1}|\mathcal{F}_t \sim \mathcal{N}(\hat{r}_t(1), \hat{\sigma}_t^2(1)),$$

cujos quantis podem ser facilmente obtidos para o cálculo de **VaR** através de

$$\hat{r}_t(1) - z_p \hat{\sigma}_t(1).$$

Se assumirmos que ε_t segue a distribuição t -Student padronizada com n graus de liberdade, t_n^* , então o p -ésimo quantil de t_n^* , representado por $t_n^*(p)$, será dado por $\hat{r}_t(1) - t_n^*(p) \hat{\sigma}_t(1)$.

A relação existente entre os quantis da distribuição t -Student com n graus de liberdade, $t_n(p)$, e $t_n^*(p)$ é

$$\begin{aligned} p &= P[t_n \leq t_n(p)] = P\left(\frac{t_n}{\sqrt{n/(n-2)}} \leq \frac{t_n(p)}{\sqrt{n/(n-2)}}\right) \\ &= P\left(t_n^* \leq \frac{t_n(p)}{\sqrt{n/(n-2)}}\right), \end{aligned}$$

$n > 2$. Logo, o quantil utilizado para calcular **VaR** de um período é dado por

$$\hat{r}_t(1) - \frac{t_n(p) \hat{\sigma}_t(1)}{\sqrt{n/(n-2)}}.$$

Suponha que no instante t estejamos interessados em obter o valor de **VaR** k períodos adiante. Se o retorno r_t segue o modelo apresentado em [2.4] e [2.5], então a média e a variância condicional de $r_t[k]$ dado o conjunto $\mathcal{F}_t$ podem ser obtidas por meio dos métodos de previsão apresentados no capítulo anterior.

A média condicional $E(r_t[k]|\mathcal{F}_t)$ pode ser obtida através dos métodos de previsão dos modelos ARIMA. Especificamente, temos que

$$\hat{r}_t[k] = \hat{r}_t(1) + \dots + \hat{r}_t(k),$$

onde $\hat{r}_t(i)$ é a previsão i passos à frente de r_t , ou seja, de r_{t+i} . Usando a representação MA infinita (filtro linear)

$$r_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$$

do modelo ARMA em [2.4], podemos escrever o erro de previsão i passos à frente como

$$e_t(i) = r_{t+i} - \hat{r}_t(i) = a_{t+i} + \psi_1 a_{t+i-1} + \dots + \psi_{i-1} a_{t+1}.$$

Logo,

$$\begin{aligned} e_t[k] &= e_t(1) + e_t(2) + \dots + e_t(k) \\ &= a_{t+1} + (a_{t+2} + \psi_1 a_{t+1}) + \dots + \sum_{i=0}^{k-1} \psi_i a_{t+k-i} \\ &= a_{t+k} + (1 + \psi_1) a_{t+k-1} + \dots + \left(\sum_{i=0}^{k-1} \psi_i \right) a_{t+1}, \end{aligned}$$

onde $\psi_0 = 1$.

A previsão da volatilidade do retorno k períodos adiante é a variância condicional de $e_t[k]$ dado $\mathcal{F}_t$. Assim,

$$\begin{aligned}\text{Var}(e_t[k]|\mathcal{F}_t) &= \text{Var}(a_{t+k}|\mathcal{F}_t) + (1 + \psi_1)^2 \text{Var}(a_{t+k-1}|\mathcal{F}_t) + \cdots \\ &\quad + \left(\sum_{i=0}^{k-1} \psi_i \right)^2 \text{Var}(a_{t+1}|\mathcal{F}_t) \\ &= \hat{\sigma}_t^2(k) + (1 + \psi_1)^2 \hat{\sigma}_t^2(k-1) + \cdots + \left(\sum_{i=0}^{k-1} \psi_i \right)^2 \hat{\sigma}_t^2(1),\end{aligned}$$

onde $\hat{\sigma}_t^2(i)$ é a previsão i passos à frente da volatilidade com origem em t . Se a volatilidade segue o modelo GARCH em [2.5], então ela pode ser obtida recursivamente como apresentado no Capítulo 1.

2.3 Estimação quantílica

A estimação quantílica ou empírica, oferece uma forma não-paramétrica para o cálculo de VaR. Nesta abordagem não se faz nenhuma suposição a respeito da distribuição dos retornos. Existem dois métodos de estimação quantílica. O primeiro utiliza quantis empíricos diretamente, enquanto que o segundo utiliza técnicas de regressão quantílica. Este último será discutido no próximo capítulo.

Sejam $r_1, r_2, \dots, r_T$ os retornos da carteira e assuma que os retornos são variáveis aleatórias independentes e identicamente distribuídas com função densidade de probabilidade $f(x)$ e função de distribuição acumulada $F(x)$. Considere as estatísticas de ordem $r_{(1)}, r_{(2)}, \dots, r_{(T)}$ e seja x_p o p -ésimo quantil de $F(x)$, ou seja, $x_p = F^{-1}(p)$. Assuma que $f(x_p) \neq 0$, então a estatística de ordem $r_{(\ell)}$ é assintoticamente normal com média x_p e variância $p(1-p)/[T \times f(x_p)^2]$ (Cox & Hinkley, 1974). Ou seja,

$$r_{(\ell)} \sim \mathcal{N}\left(x_p, \frac{p(1-p)}{T \times f(x_p)^2}\right), \quad \ell = p \times T.$$

Podemos utilizar $r_{(\ell)}$ para estimar o quantil x_p . Em geral, ℓ é não-inteiro. Neste caso, devemos utilizar uma simples interpolação para obter a estimativa do quantil. Se $p_i < p < p_{i+1}$, onde $p_i = i/T$ e $i = 1, \dots, T$, o quantil x_p pode ser estimado por

$$\hat{x}_p = \frac{p_{i+1} - p}{p_{i+1} - p_i} r_{(i)} + \frac{p - p_i}{p_{i+1} - p_i} r_{(i+1)}.$$

As principais vantagens da utilização do VaR pelo método de quantis empíricos são a simplicidade e a ausência de suposições sobre a distribuição dos

retornos. Entretanto, esta abordagem possui várias desvantagens. Primeiro, assume que a distribuição dos retornos permanece inalterada durante todo o período. Dado que **VaR** se relaciona principalmente com as caudas da distribuição dos retornos, esta suposição implica que a perda prevista não pode ser maior do que a histórica, o que não ocorre na prática. Segundo, para quantis extremos (quando p é próximo a zero ou um) os quantis empíricos não são eficientes para os quantis teóricos. Terceiro, a estimação direta dos quantis não leva em consideração o efeito de variáveis exógenas que são relevantes ao estudo. Em aplicações reais, **VaR** obtido pelos quantis empíricos podem servir como um limite inferior para o **VaR** verdadeiro.

2.4 Valor em risco em eventos extremos

2.4.1 Modelagem de eventos extremos

A Teoria dos Valores Extremos (TVE) é um ramo da probabilidade que estuda o comportamento estocástico de extremos associados a um conjunto de variáveis aleatórias $r_{(1)}, r_{(2)}, \dots, r_{(T)}$, independentes e identicamente distribuídas com função de distribuição $F(x)$. Dentro da denominação de extremos estão o mínimo $r_{(1)} = \min(r_1, \dots, r_T)$, o máximo $r_{(T)} = \max(r_1, \dots, r_T)$, estatísticas de ordem extremas e excessos além (ou abaixo) de limiares elevados (ou baixos). Uma questão importante é que as propriedades das distribuições desses extremos são determinadas pelas caudas inferior e superior de $F(x)$.

Nas definições que seguem, vamos enfatizar a utilização do mínimo, visto que sua utilização tem mais relevância no cálculo de **VaR** para posições compradas. Entretanto, a teoria apresentada é também aplicável a posições vendidas, pois o máximo pode ser obtido do mínimo através de

$$r_{(T)} = -\min(-r_1, \dots, -r_T).$$

A distribuição de $r_{(1)}$, denotada por $F_{\min}(x)$, é dada por

$$\begin{aligned} F_{\min}(x) &= P(r_{(1)} \leq x) = 1 - P(r_{(1)} > x) \\ &= 1 - P(r_1 > x, r_2 > x, \dots, r_T > x) \\ &= 1 - \prod_{t=1}^T P(r_t > x) = 1 - P(r_t > x)^T \\ &= 1 - [1 - P(r_t \leq x)]^T = 1 - [1 - F(x)]^T. \end{aligned}$$

Na prática, $F(x)$ é desconhecida e, conseqüentemente, $F_{\min}(x)$ também é desconhecida. Entretanto, quando $T \rightarrow \infty$, $F_{\min}(x)$ se torna degenerada, ou seja, $F_{\min}(x) \rightarrow 0$, quando $x \rightarrow -\infty$ e $F_{\min}(x) \rightarrow 1$, quando $x \rightarrow \infty$, quando o tamanho da amostra tende a infinito. Portanto, a TVE se preocupa

com a busca das seqüências $\{\beta_T\}$ e $\{\alpha_T\}$, com $\alpha_T > 0$, tais que a distribuição de

$$r_{(1)}^* = (r_{(1)} - \beta_T) / \alpha_T$$

convirja para uma distribuição não-degenerada quando $T \rightarrow \infty$. Sob a suposição de independência, a distribuição limite do mínimo normalizado $r_{(1)}^*$ é dada por

$$F_{\min}^*(x) = \begin{cases} 1 - \exp \left[-(1 + \xi x)^{1/\xi} \right], & \text{se } \xi \neq 0, \\ 1 - \exp[-\exp(x)], & \text{se } \xi = 0, \end{cases} \quad [2.6]$$

para $x < -1/\xi$, se $\xi < 0$ e para $x > -1/\xi$, se $\xi > 0$. O caso em que $\xi = 0$ é dado pelo limite quando $\xi \rightarrow 0$. O parâmetro ξ é chamado de parâmetro de forma e determina o comportamento da cauda da distribuição limite. Esta condição é conhecida como Teorema de Fisher–Tippett (Fisher & Tippett, 1928).

A distribuição limite em [2.6] é chamada de distribuição de valores extremos generalizada (GEV) para o mínimo (Jenkinson, 1955). A distribuição limite $F_{\min}(x)$ segue uma das três distribuições:

- i. $\xi = 0$, distribuição de Gumbel:

$$F_{\min}^*(x) = 1 - \exp[-\exp(x)], \quad x \in \mathbb{R}.$$

- ii. $\xi < 0$, distribuição de Fréchet:

$$F_{\min}^*(x) = \begin{cases} 1 - \exp \left[-(1 + \xi x)^{1/\xi} \right], & \text{se } x < -1/\xi, \\ 1, & \text{caso contrário.} \end{cases}$$

- iii. $\xi > 0$, distribuição de Weibull:

$$F_{\min}^*(x) = \begin{cases} 1 - \exp \left[-(1 + \xi x)^{1/\xi} \right], & \text{se } x > -1/\xi, \\ 0, & \text{caso contrário.} \end{cases}$$

A função densidade da distribuição de valores extremos generalizada para o mínimo é dada por

$$f_{\min}^*(x) = \begin{cases} (1 + \xi x)^{1/\xi - 1} \exp[-(1 + \xi x)^{1/\xi}], & \text{se } \xi \neq 0, \\ \exp[x - \exp(x)], & \text{se } \xi = 0. \end{cases} \quad [2.7]$$

A Figura 2.1 apresenta o comportamento das densidades dos três tipos de distribuição de valores extremos, onde $\xi = 0.75$ para a distribuição de Weibull e $\xi = -0.43$ para a distribuição de Fréchet.

A Figura 2.2 apresenta o comportamento das distribuições de Fréchet (a) e Weibull (b) para diversos valores de ξ . Note que, à medida em que $\xi \rightarrow 0$, ambas as distribuições tendem à distribuição de Gumbel.

Na gestão de riscos, o principal interesse reside na distribuição de Fréchet, a qual inclui as distribuições estáveis e *t*-Student, como casos particulares. A

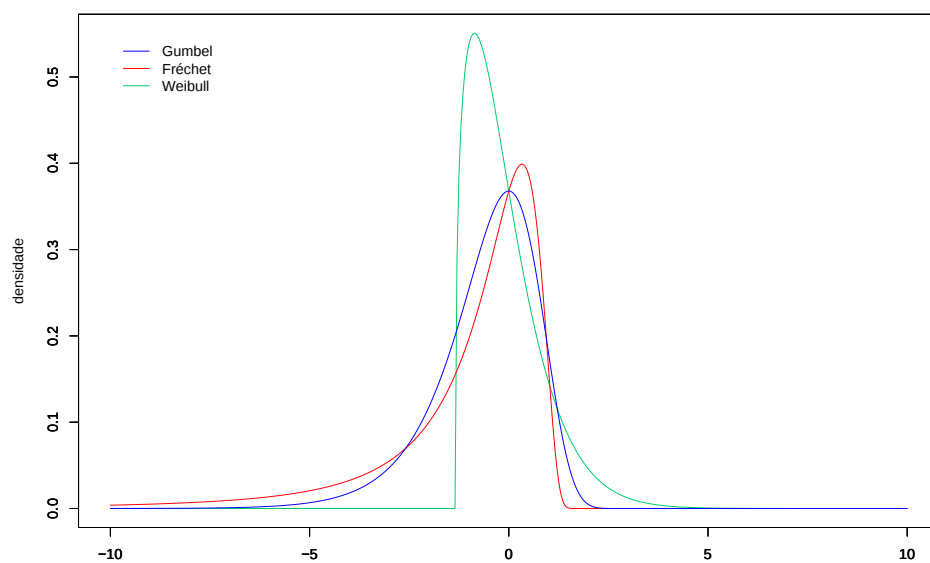


Figura 2.1. Densidades das distribuições de valores extremos, com $\xi = 0.75$ para a Weibull e $\xi = -0.43$ para a Fréchet.

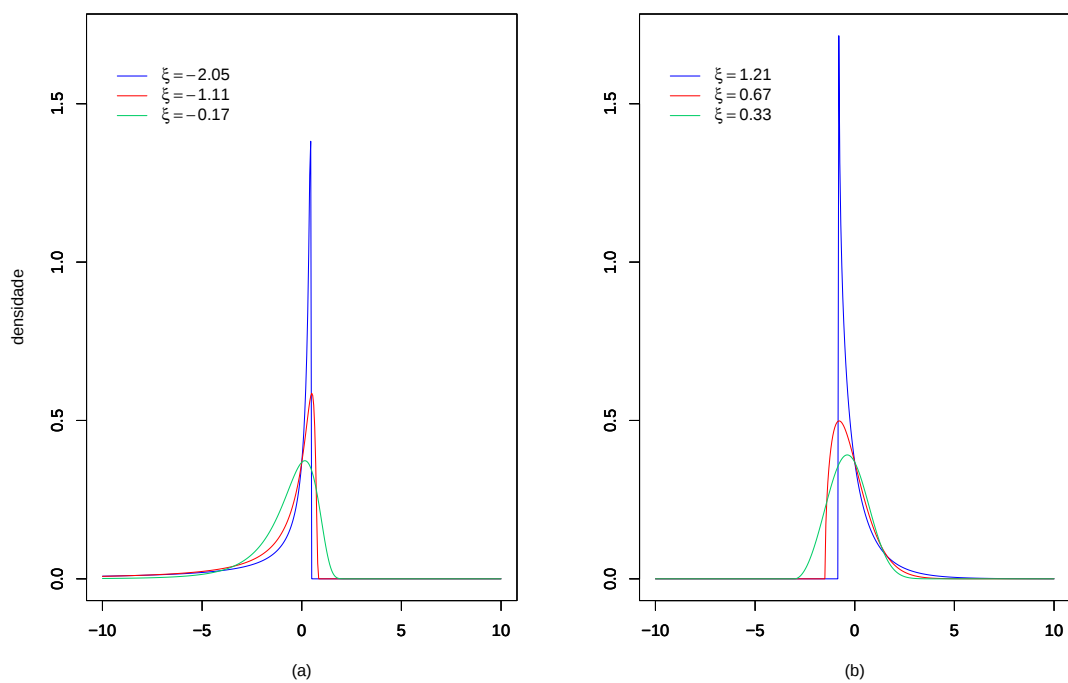


Figura 2.2. Densidades das distribuições de Fréchet (a) e Weibull (b) para diversos valores de ξ .

distribuição de Gumbel consiste em distribuições com caudas mais leves, como as distribuições normal e log-normal.

Alternativamente, podemos reescrever [2.7] com $x = (r - \beta)/\alpha$, onde α e β são, respectivamente, os parâmetros de escala e locação. Assim,

$$f_{\min}^*(x) = \begin{cases} \frac{1}{\alpha} \left(1 + \frac{\xi(r-\beta)}{\alpha}\right)^{1/\xi-1} \exp\left[-\left(1 + \frac{\xi(r-\beta)}{\alpha}\right)^{1/\xi}\right], & \text{se } \xi \neq 0, \\ \frac{1}{\alpha} \exp\left[\frac{(r-\beta)}{\alpha} - \exp\left(\frac{r-\beta}{\alpha}\right)\right], & \text{se } \xi = 0. \end{cases} \quad [2.8]$$

Os parâmetros da distribuição GEV podem ser estimados numericamente pela maximização da função de verossimilhança ou, como já dissemos anteriormente, da função de log-verossimilhança. Assim, quando $\xi = 0$, a função de log-verossimilhança é dada por

$$\ell(r_1, \dots, r_T; \alpha, \beta, \xi = 0) = -T \log(\alpha) + \sum_{i=1}^T \left(\frac{r_i - \beta}{\alpha}\right) - \sum_{i=1}^T \exp\left\{-\left(\frac{r_i - \beta}{\alpha}\right)\right\},$$

e as estimativas de máxima verossimilhança são obtidas através da solução do sistema de equações não-lineares

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= -\frac{T}{\alpha} - \frac{1}{\alpha^2} \sum_{i=1}^T (r_i - \beta) - \frac{1}{\alpha^2} \sum_{i=1}^T (r_i - \beta) \exp\left\{-\left(\frac{r_i - \beta}{\alpha}\right)\right\} = 0, \\ \frac{\partial \ell}{\partial \beta} &= -\frac{T}{\alpha} - \frac{1}{\alpha} \sum_{i=1}^T \exp\left\{-\left(\frac{r_i - \beta}{\alpha}\right)\right\} = 0. \end{aligned}$$

Quando $\xi \neq 0$, a função de log-verossimilhança é dada por

$$\begin{aligned} \ell(r_1, \dots, r_T; \alpha, \beta, \xi) &= -T \log(\alpha) + \left(\frac{1}{\xi} - 1\right) \sum_{i=1}^T \log\left[1 + \frac{\xi(r_i - \beta)}{\alpha}\right] \\ &\quad - \sum_{i=1}^T \left[1 + \frac{\xi(r_i - \beta)}{\alpha}\right]^{\frac{1}{\xi}}, \end{aligned}$$

e a solução do sistema de equações

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= -\frac{T}{\alpha} - \frac{1}{\alpha} \left(\frac{1}{\xi} - 1\right) \sum_{i=1}^T \left[\frac{\xi(r_i - \beta)}{\alpha + \xi(r_i - \beta)}\right] \\ &\quad + \frac{1}{\alpha^2} \sum_{i=1}^T (r_i - \beta) \left[1 + \frac{\xi(r_i - \beta)}{\alpha}\right]^{\frac{1}{\xi}-1} = 0 \\ \frac{\partial \ell}{\partial \beta} &= -\left(\frac{1}{\xi} - 1\right) \sum_{i=1}^T \left[\frac{\xi}{\alpha + \xi(r_i - \beta)}\right] + \frac{1}{\alpha} \sum_{i=1}^T \left[1 + \frac{\xi(r_i - \beta)}{\alpha}\right]^{\frac{1}{\xi}-1} = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \xi} = & -\frac{1}{\xi^2} \sum_{i=1}^T \log \left[1 + \frac{\xi(r_i - \beta)}{\alpha} \right] + \left(\frac{1}{\xi} - 1 \right) \sum_{i=1}^T \frac{r_i - \beta}{\alpha + \xi(r_i - \beta)} \\ & - \frac{1}{\xi} \sum_{i=1}^T \left[1 + \frac{\xi(r_i - \beta)}{\alpha} \right]^{\frac{1}{\xi}} \left\{ \frac{\log \left[1 + \frac{\xi(r_i - \beta)}{\alpha} \right]}{\xi} + \frac{r_i - \beta}{\alpha + (r_i - \beta)} \right\} = 0 \end{aligned}$$

determina as estimativas de máxima verossimilhança.

Para uma exposição mais completa sobre a TVE, o leitor pode consultar Embrechts *et al.* (1997) e Mendes (2004).

2.4.2 Cálculo de VaR pela TVE

Assuma que a amostra em estudo contém T observações; ela será dividida em g grupos de tamanho n . Ou seja, $T = ng$. Se $T = ng + m$, com $m < n$, as primeiras m observações deverão ser desconsideradas. Assim, devemos estimar os parâmetros da GEV para o conjunto de mínimos dos g grupos, tal que $r_{n,i} = \min\{r_{(i-1)n+j}\}$, $1 \leq j \leq n$, $i = 1, \dots, g$. De [2.8] é possível obter o quantil da GEV para uma dada probabilidade p .

Seja p^* a probabilidade de ocorrência de uma perda potencial e r_n^* o p^* -ésimo quantil dos mínimos dos g grupos. Assim,

$$p^* = \begin{cases} 1 - \exp \left[- \left(1 + \frac{\xi_n(r_n^* - \beta_n)}{\alpha_n} \right)^{1/\xi_n} \right], & \text{se } \xi_n \neq 0, \\ 1 - \exp \left[- \exp \left(\frac{r_n^* - \beta_n}{\alpha_n} \right) \right], & \text{se } \xi_n = 0. \end{cases}$$

Tomando $\log(1 - p^*)$, da equação acima obtemos o quantil

$$r_n^* = \begin{cases} \beta_n - \frac{\alpha_n}{\xi_n} \{ 1 - [-\log(1 - p^*)]^{\xi_n} \}, & \text{se } \xi_n \neq 0 \\ \beta_n + \alpha_n \log[-\log(1 - p^*)], & \text{se } \xi_n = 0. \end{cases}$$

Para obter o VaR da série de retornos original r_t , temos que relacionar os quantis desta série com os quantis da série de mínimos. Assim, temos que

$$p^* = P(r_{n,i} \leq r_n^*) = 1 - [1 - P(r_t \leq r_n^*)]^n$$

ou, equivalentemente,

$$1 - p^* = [1 - P(r_t \leq r_n^*)]^n = (1 - p)^n. \quad [2.9]$$

Mais precisamente, para p fixado, o p -ésimo quantil de r_t é r_n^* , se a probabilidade p^* é escolhida com base em [2.9], onde $p^* = P(r_t \leq r_n^*)$. Conseqüentemente, o VaR de uma posição comprada é dado por

$$\text{VaR} = \begin{cases} \beta_n - \frac{\alpha_n}{\xi_n} \{ 1 - [-n \log(1 - p)]^{\xi_n} \}, & \text{se } \xi_n \neq 0, \\ \beta_n + \alpha_n \log[-n \log(1 - p)], & \text{se } \xi_n = 0. \end{cases}$$

Note que utilizamos a notação α_n , β_n e ξ_n , pois para cada escolha de n tem-se estimativas diferentes para os parâmetros.

Para múltiplos períodos temos que VaR pode ser estimado por

$$\text{VaR}[k] = k^{-\xi_n} \text{VaR},$$

onde k é o número de períodos considerados.

Valor em Risco Auto-regressivo Condicional

Engle & Manganelli (2004) propuseram uma nova abordagem para obtenção do Valor em Risco utilizando regressão quantílica na qual, em vez de modelar a distribuição completa dos retornos, modela-se diretamente os seus quantis. Essa nova metodologia permite a utilização de diversas formas funcionais, onde a escolha da melhor delas deve ser determinada pelo conjunto de dados em estudo.

Como já dissemos anteriormente, a volatilidade dos retornos pode apresentar aglomerados, ou seja, os retornos podem ser autocorrelacionados. Consequentemente, o **VaR**, que é ligado ao desvio padrão da distribuição pode apresentar comportamento semelhante.

Uma forma natural de neutralizar essa característica é a utilização de alguma especificação auto-regressiva. Pensando nisso, Engle & Manganelli (2004) propuseram uma especificação quantílica auto-regressiva condicional, a qual chamaram de Valor em Risco auto-regressivo condicional ou, simplesmente, **CAViaR**.

A forma geral da especificação **CAViaR** é

$$\text{VaR}_t = \beta_0 + \sum_{i=1}^p \beta_i \text{VaR}_{t-i} + g(\beta_{p+1}, \dots, \beta_{p+q}; \mathcal{F}_{t-1}). \quad [3.1]$$

Segundo os autores, na maioria dos casos práticos a especificação em [3.1] se reduz a um modelo de primeira ordem do tipo

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + g(\beta_2, r_{t-1}, \text{VaR}_{t-1}). \quad [3.2]$$

O termo auto-regressivo de primeira ordem assegura que **VaR** varia suavemente no tempo, enquanto $g(\beta_2, r_{t-1}, \text{VaR}_{t-1})$ faz a ligação de **VaR**_{*t*} a *r*_{*t*-1}, ou seja, $g(\cdot)$ mede o quanto **VaR** pode variar baseado na nova informação de *r*. Então, $g(\cdot)$ funciona como a curva de impacto das inovações dos modelos GARCH, introduzida por Engle & Ng (1993).

Note que para o processo [3.1] não ser explosivo, as raízes de

$$1 - \beta_1 z - \beta_2 z^2 - \dots - \beta_p z^p = 0$$

devem estar fora do círculo unitário.

Algumas especificações **CAViaR** são:

[E1] Adaptativo:

$$\text{VaR}_t = \text{VaR}_{t-1} + \beta_1 \left\{ [1 - \exp\{k(r_{t-1} + \text{VaR}_{t-1})\}]^{-1} - \theta \right\}. \quad [3.3]$$

Na especificação adaptativa, k é um número finito positivo e quando $k \rightarrow \infty$ o último termo converge para $\beta_1 [I(r_{t-1} \leq -\text{VaR}_{t-1}) - \theta]$ quase certamente. A especificação adaptativa incorpora a seguinte regra: quando o **VaR** é excedido, ele pode ser imediatamente aumentado, mas quando isso não ocorre, ele pode ser diminuído suavemente. Ou seja, se **VaR** é excedido, quando k tende a infinito, acrescenta-se $\beta_1 \times (1 - \theta)$ ao valor do **VaR**; caso contrário, o valor a ser acrescido em **VaR** é igual a $-\theta \times \beta_1$.

[E2] Valor absoluto simétrico:

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 |r_{t-1}|. \quad [3.4]$$

Essa especificação responde simetricamente aos retornos passados, visto que a função $g(\cdot)$ é modular e, assim, as observações de r_t de mesma magnitude, porém de sinais diferentes, são tratadas igualmente.

[E3] Inclinação assimétrica:

$$\begin{aligned} \text{VaR}_t &= \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 \max(r_{t-1}, 0) + \beta_3 \min(r_{t-1}, 0) \\ &= \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 r_{t-1} I(r_{t-1} > 0) + \beta_3 r_{t-1} I(r_{t-1} < 0). \end{aligned} \quad [3.5]$$

A especificação da inclinação assimétrica responde diferentemente aos retornos passados positivos e negativos, a não ser que $\beta_2 = \beta_3$, visto que $g(\cdot)$ considera esses retornos separadamente.

[E4] GARCH(1,1) Indireto:

$$\text{VaR}_t = \sqrt{\beta_0 + \beta_1 \text{VaR}_{t-1}^2 + \beta_2 r_{t-1}^2}. \quad [3.6]$$

Assim como na especificação do valor absoluto simétrico, a especificação GARCH(1,1) indireto responde simetricamente aos retornos passados, visto que a função $g(\cdot)$ é quadrática.

Engle & Manganelli (1999) apresentam outras duas especificações, a saber:

[E5] Adaptativo Simétrico Proporcional:

$$\text{VaR}_t = \text{VaR}_{t-1} + \beta_1 \max(0, |r_{t-1}| - \text{VaR}_{t-1}) - \beta_2 \min(0, |r_{t-1}| - \text{VaR}_{t-1})$$

e

[E6] Valor Absoluto Assimétrico:

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 |r_{t-1} - \beta_3|.$$

Contudo, essas especificações não serão consideradas nesta dissertação.

Ainda segundo Engle & Manganelli (2004), a especificação GARCH(1,1) indireto pode ser corretamente especificada se os dados seguem de fato um processo GARCH(1,1) com erros independentes e identicamente distribuídos. As especificações do valor absoluto simétrico e da inclinação assimétrica podem ser corretamente especificadas por um processo GARCH no qual o desvio padrão, ao invés da variância, é modelado simétrica ou assimetricamente com erros independentes e identicamente distribuídos. Este tipo de modelagem foi introduzido por Taylor (1986) e Schwert (1988) e analisado por Engle (2002). Contudo, as especificações CAViaR são mais gerais do que estes modelos. De fato, as especificações CAViaR podem ser usadas em situações onde a volatilidade é constante, mas a distribuição do erro é não-constante ou em situações onde ambas, volatilidade e distribuição do erro, não são constantes.

Os parâmetros do modelo CAViaR são estimados através dos métodos usuais de estimação em regressão quantílica, técnica introduzida por Koenker & Bassett (1978), a qual consiste da minimização da função

$$CQ = \frac{1}{T} \left\{ \sum_{t:r_t \geq \text{VaR}_t} \theta |r_t - \text{VaR}_t| + \sum_{t:r_t < \text{VaR}_t} (1 - \theta) |r_t - \text{VaR}_t| \right\}, \quad [3.7]$$

que chamamos de critério quantílico. Reescrevendo a equação [3.7], obtemos

$$CQ = \frac{1}{T} \left\{ - \sum_t [I(r_t < \text{VaR}_t) - \theta] (r_t - \text{VaR}_t) \right\}. \quad [3.8]$$

Assim, a θ -ésima regressão quantílica é definida pelo vetor $\hat{\beta}_\theta$ que minimiza [3.8]. Note que devemos substituir VaR_t em [3.8] por uma das quatro especificações CAViaR apresentadas anteriormente. Logo, para o caso da especificação do valor absoluto simétrico, $\hat{\beta}_\theta = \{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\}$ é obtido como o vetor que minimiza o critério quantílico

$$\min_{\beta} \frac{1}{T} \left\{ - \sum_t [I(r_t < \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 |r_{t-1}|) - \theta] (r_t - \beta_0 - \beta_1 \text{VaR}_{t-1} - \beta_2 |r_{t-1}|) \right\}.$$

A classe dos estimadores de regressão quantílica é um caso particular dos estimadores de mínimos absolutos, cujas propriedades têm sido amplamente discutidas na literatura. Koenker & Bassett (1978) mostraram que essa classe é mais robusta do que a dos estimadores de mínimos quadrados quando a distribuição dos erros tem caudas pesadas, o que, como já dissemos, ocorre com frequência em finanças.

White (1994) provou a consistência das estimativas de regressão quantílica não-linear em generalidade suficiente para incluir modelos de séries temporais. A consistência, a normalidade assintótica e a equivalência assintótica dos testes de Wald e escore (também conhecido como teste do multiplicador de Lagrange) realizado a partir de estimativas de mínimos absolutos em modelos dinâmicos não-lineares foi apresentada por Weiss (1991).

Esses autores mostraram que

$$\hat{\beta}_\theta \xrightarrow{p} \beta_\theta,$$

quando $T \rightarrow \infty$, sob certas condições, onde $\hat{\beta}_\theta$ é a solução de

$$\min_{\beta} \frac{1}{T} \left\{ - \sum_t [I(r_t < \text{VaR}_t) - \theta] (r_t - \text{VaR}_t) \right\}.$$

As condições de regularidade estão apresentadas em White (1994) e em Engle & Manganelli (2004).

Seguindo a demonstração de normalidade assintótica apresentada por White (1994), Engle & Manganelli (2004) provaram que

$$\sqrt{T} \mathbf{A}_T^{-1/2} \mathbf{D}_T (\hat{\beta}_\theta - \beta_\theta) \xrightarrow{p} \mathcal{N}(0, \mathbf{I}),$$

onde

$$\begin{aligned} \mathbf{A}_T &= E \left[T^{-1} \theta (1 - \theta) \sum_{t=1}^T \nabla' \text{VaR}_t \nabla \text{VaR}_t \right], \\ \mathbf{D}_T &= E \left[T^{-1} \sum_{t=1}^T h_t(0) \nabla' \text{VaR}_t \nabla \text{VaR}_t \right], \end{aligned}$$

$\nabla \text{VaR}_t = \frac{\partial}{\partial \beta} \text{VaR}_t$ sendo o gradiente de VaR_t e $h_t(x)$ sendo a função de densidade condicional do termo erro, a qual pode ser estimada através de métodos de *kernel*.

Por exemplo, para o caso da especificação do valor absoluto simétrico, ∇VaR_t é o vetor tri-dimensional das primeiras derivadas parciais de β . Note que a expressão [3.4] é recursiva; assim,

$$\begin{aligned} \text{VaR}_t &= \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 |r_{t-1}| \\ &= \beta_0 + \beta_1 (\beta_0 + \beta_1 \text{VaR}_{t-2} + \beta_2 |r_{t-2}|) + \beta_2 |r_{t-1}| \\ &= \beta_0 + \beta_1 [\beta_0 + \beta_1 (\beta_0 + \beta_1 \text{VaR}_{t-3} + \beta_2 |r_{t-3}|) + \beta_2 |r_{t-2}|] + \beta_2 |r_{t-1}|. \end{aligned}$$

As derivadas de VaR_t são

$$\begin{aligned} \frac{\partial \text{VaR}_t}{\partial \beta_0} &= 1 + \beta_1 \frac{\partial \text{VaR}_{t-1}}{\partial \beta_0}, \\ \frac{\partial \text{VaR}_t}{\partial \beta_1} &= \text{VaR}_{t-1} + \beta_1 \frac{\partial \text{VaR}_{t-1}}{\partial \beta_1}, \\ \frac{\partial \text{VaR}_t}{\partial \beta_2} &= |r_{t-1}| + \beta_1 \frac{\partial \text{VaR}_{t-1}}{\partial \beta_2}. \end{aligned}$$

Portanto,

$$\nabla' \text{VaR}_t = \begin{pmatrix} 1 & + \beta_1 \frac{\partial}{\partial \beta_0} \text{VaR}_{t-1} \\ \text{VaR}_{t-1} & + \beta_1 \frac{\partial}{\partial \beta_1} \text{VaR}_{t-1} \\ |r_{t-1}| & + \beta_1 \frac{\partial}{\partial \beta_2} \text{VaR}_{t-1} \end{pmatrix}$$

para o caso da especificação do valor absoluto simétrico.

Para o caso da especificação da inclinação assimétrica, temos que, assim como no caso do valor absoluto simétrico, VaR_t é recursiva; logo,

$$\begin{aligned} \text{VaR}_t &= \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 \max(0, r_{t-1}) + \beta_3 \min(0, r_{t-1}) \\ &= \beta_0 + \beta_1 [\beta_0 + \beta_1 \text{VaR}_{t-2} + \beta_2 \max(0, r_{t-2}) + \beta_3 \min(0, r_{t-2})] \\ &\quad + \beta_2 \max(0, r_{t-1}) + \beta_3 \min(0, r_{t-1}). \end{aligned}$$

Tomando as primeiras derivadas de VaR_t com relação a $\beta_0, \beta_1, \beta_2$ e β_3 , obtemos

$$\begin{aligned} \frac{\partial \text{VaR}_t}{\partial \beta_0} &= 1 + \beta_1 \frac{\partial \text{VaR}_{t-1}}{\partial \beta_0}, \\ \frac{\partial \text{VaR}_t}{\partial \beta_1} &= \text{VaR}_{t-1} + \beta_1 \frac{\partial \text{VaR}_{t-1}}{\partial \beta_1}, \\ \frac{\partial \text{VaR}_t}{\partial \beta_2} &= \beta_1 \frac{\partial \text{VaR}_{t-1}}{\partial \beta_2} + \max(0, r_{t-1}), \\ \frac{\partial \text{VaR}_t}{\partial \beta_3} &= \beta_1 \frac{\partial \text{VaR}_{t-1}}{\partial \beta_3} + \min(0, r_{t-1}). \end{aligned}$$

Na especificação GARCH(1,1) indireto, podemos reescrever o modelo [3.6] de tal forma que

$$\begin{aligned} \text{VaR}_t &= \sqrt{\beta_0 + \beta_1 \text{VaR}_{t-1}^2 + \beta_2 r_{t-1}^2} \\ &= \sqrt{\beta_0 + \beta_1 (\beta_0 + \beta_1 \text{VaR}_{t-2}^2 + \beta_2 r_{t-2}^2) + \beta_2 r_{t-1}^2} \\ &= \sqrt{\beta_0 + \beta_1 [\beta_0 + \beta_1 (\beta_0 + \beta_1 \text{VaR}_{t-3}^2 + \beta_2 r_{t-3}^2) + \beta_2 r_{t-2}^2] + \beta_2 r_{t-1}^2}, \end{aligned}$$

cujas derivadas de VaR_t com relação a β são

$$\begin{aligned} \frac{\partial \text{VaR}_t}{\partial \beta_0} &= \frac{1 + 2\beta_1 \text{VaR}_{t-1} \frac{\partial}{\partial \beta_0} \text{VaR}_{t-1}}{2\text{VaR}_t}, \\ \frac{\partial \text{VaR}_t}{\partial \beta_1} &= \frac{\text{VaR}_{t-1}^2 + 2\beta_1 \text{VaR}_{t-1} \frac{\partial}{\partial \beta_1} \text{VaR}_{t-1}}{2\text{VaR}_t}, \\ \frac{\partial \text{VaR}_t}{\partial \beta_2} &= \frac{r_{t-1}^2 + 2\beta_1 \text{VaR}_{t-1} \frac{\partial}{\partial \beta_2} \text{VaR}_{t-1}}{2\text{VaR}_t}. \end{aligned}$$

Usando o mesmo raciocínio, obtém-se a derivada de VaR_t com relação a β_1 para a especificação adaptativa em [3.3]:

$$\begin{aligned}\frac{\partial \text{VaR}_t}{\partial \beta_1} &= \frac{\partial \text{VaR}_{t-1}}{\partial \beta_1} + \left\{ \frac{1}{[1 + \exp\{k(r_{t-1} + \text{VaR}_{t-1})\}] - \theta} \right\} \\ &\quad - \beta_1 \frac{k \exp\{k(r_{t-1} + \text{VaR}_{t-1})\} \frac{\partial \text{VaR}_{t-1}}{\partial \beta_1}}{\sqrt{1 + \exp\{k(r_{t-1} + \text{VaR}_{t-1})\}}}.\end{aligned}$$

Vale ressaltar que k , na especificação Adaptativa, pode ser estimado conjuntamente com o parâmetro β_1 . Contudo, a idéia por trás dessa especificação é a da simplicidade, motivo pelo qual Engle & Manganelli (2004) recomendam que se considere um valor arbitrário para k .

As matrizes $\mathbf{A}_T$ e $\mathbf{D}_T$ podem ser estimadas, respectivamente, por

$$\begin{aligned}\hat{\mathbf{A}}_T &= T^{-1} \theta (1 - \theta) \sum_{t=1}^T \nabla' \widehat{\text{VaR}}_t \nabla \widehat{\text{VaR}}_t, \\ \hat{\mathbf{D}}_T &= (2T \hat{c}_T)^{-1} \sum_{t=1}^T I(|r_t - \widehat{\text{VaR}}_t| < \hat{c}_T) \nabla' \widehat{\text{VaR}}_t \nabla \widehat{\text{VaR}}_t,\end{aligned}$$

onde

$$\hat{c}_T = T^{-1} \sum_{i=1}^T \frac{\nabla \widehat{\text{VaR}}_i \nabla' \widehat{\text{VaR}}_i}{\widehat{\text{VaR}}_i}$$

é chamado de *bandwidth*. Desta forma, é possível realizar testes de hipóteses sobre os parâmetros do modelo CAViaR. Engle & Manganelli (1999) apresentaram as estatísticas de teste de Wald e escore para o modelo CAViaR, os quais podem ser utilizados para verificar se termos defasados adicionais de VaR e de r devem ser considerados.

Se um modelo estiver corretamente especificado, então $P(r_t < -\text{VaR}_t) = \theta$, para todo t . Isto é equivalente a requerer que a sucessão de funções indicadoras $I(r_t < -\text{VaR}_t)$ seja independente e identicamente distribuída. Conseqüentemente, uma propriedade que qualquer estimativa VaR deve satisfazer é que a seqüência, que possivelmente é autocorrelacionada e até mesmo heteroscedástica, deve ser transformada em uma série independente de funções indicadoras. Um modo natural para testar a validade do modelo de previsão é verificar se a série das funções indicadoras $I(r_t < -\text{VaR}_t)$ é independente e identicamente distribuída.

Existem vários testes para verificar a suposição de independência de uma seqüência de realizações de uma variável aleatória como, por exemplo, o teste *runs* de Mood (1940) e o teste de Ljung-Box, apresentado em [1.21]. Esses testes podem revelar a presença de correlação serial na série de funções indicadoras. Contudo, o fato destas funções indicadoras serem não-correlacionadas serialmente não é o bastante para avaliar o desempenho de uma estimativa VaR. Outro ponto importante é que esses testes não consideram a inclusão de covariáveis.

Engle & Manganelli (2004) propuseram um teste com a capacidade de identificar a presença de correlação serial. Seja

$$Hit_t = I(r_t < -\text{VaR}_t) - \theta.$$

A função Hit_t assume valor $(1 - \theta)$ sempre que r_t é menor que $-\text{VaR}_t$, ou seja, toda vez um *hit* ocorre, e $-\theta$, caso contrário. Claramente o valor esperado de Hit é zero. Adicionalmente, da definição da função quantílica, a esperança condicional de Hit dada qualquer informação conhecida em $t - 1$ também deve ser zero. Uma simples aplicação da lei das esperanças iteradas mostra que Hit deve ser não-correlacionada com qualquer subconjunto pertencente ao conjunto de informações passadas, $\mathcal{F}_{t-1}$:

$$E(Hit_t f_{t-1}) = f_{t-1} E(Hit_t | \mathcal{F}_{t-1}), \quad \forall f_{t-1} \in \mathcal{F}_{t-1}.$$

Em particular, Hit_t é não-correlacionada com $Hit_{t-\ell}$, para todo ℓ , com a sequência VaR_t prevista e com uma constante.

A qualidade do ajuste do modelo CAViaR pode ser verificada através do teste de quantis dinâmicos (Engle & Manganelli, 2004), que é realizado por meio da regressão artificial

$$Hit_t = X'\delta + u_t, \tag{3.9}$$

onde

$$u_t = \begin{cases} -\theta, & \text{com probabilidade } (1 - \theta), \\ (1 - \theta), & \text{com probabilidade } \theta, \end{cases}$$

e entram como regressores termos defasados da função Hit_t . Outras funções podem compor o teste como, por exemplo, variáveis *dummy* anuais.

Em sua proposta, Engle & Manganelli (2004) apresentam o teste para observações amostrais e para observações pós-amostrais, as quais devem ser reservadas para este fim, não sendo utilizadas no processo de estimação.

A estatística de teste de quantis dinâmicos é dada por

$$\frac{\widehat{\delta}' X' X \widehat{\delta}}{\theta(1 - \theta)} \sim \chi_p^2,$$

onde $p - 1$ é o número de covariáveis na regressão artificial em [3.9] e $\widehat{\delta}$ é obtido através do método de mínimos quadrados, com

$$\widehat{\delta} = (X'X)^{-1} X' Hit_t \xrightarrow{d} \mathcal{N}(\delta, \theta(1 - \theta)(X'X)^{-1})$$

pelo Teorema Central do Limite (Engle & Manganelli, 1999). O teste de quantis dinâmicos quando realizado para as observações amostrais verifica se o modelo está bem especificado, e pode servir como um critério de seleção de modelos.

Uma outra forma de avaliar se a especificação escolhida é apropriada é verificando o percentual de *hits* na amostra. Se esse percentual estiver próximo a θ , temos um indicativo de que o modelo é adequado, ou seja,

$$q = \frac{1}{T} \sum_{t=1}^T I(r_t < -\text{VaR}_t) \approx \theta.$$

Como dissemos anteriormente, Engle & Manganelli (2004) sugerem que esses procedimentos sejam realizados nas observações amostrais e pós-amostrais.

Simulações

Neste capítulo investigaremos a precisão das estimativas dos parâmetros do modelo **CAViaR** para as especificações do valor absoluto simétrico (**SAV**), da inclinação assimétrica (**AS**), **GARCH(1,1)** indireto (**IG**) e adaptativa (**ADP**) propostas por Engle & Manganelli (2004).

Diversos cenários foram considerados, nos quais o tamanho amostral e o valor de θ são alterados, a fim de verificar o comportamento das estimativas em cada um deles. Como tamanhos amostrais, consideramos $T = \{200, 500, 1000\}$ e $\theta = \{0.01, 0.05\}$ foram os níveis adotados para a determinação do **VaR**.

Foram consideradas em cada caso 1000 réplicas do experimento, onde cada um dos 1000 vetores de estimativas iniciais foi escolhido de tal forma que seguisse distribuição uniforme entre 0 e 1, conforme sugerem Engle & Manganelli (2004). Na especificação **ADP** utilizamos $k = 10$, mesmo valor considerado por Engle & Manganelli (2004).

No processo de estimação, adotamos a seguinte estratégia: para cada vetor de parâmetros iniciais obteve-se o vetor ótimo, o qual foi avaliado como vetor inicial em um novo processo de minimização. O processo continuou até a obtenção de convergência.

4.1 Cenário I

Seja r_t a série de caudas pesadas apresentada na Figura 4.1 (a) de tamanho $T = 200$, cujo histograma é apresentado em (b), do qual obtemos $\text{VaR}_1 = -\text{Quant}_{0.05}(r_t) = 1.863$. De posse de r_t e VaR_1 obtém-se, recursivamente, a sequência VaR_t , com o auxílio do vetor de estimativas iniciais e, por fim, o valor de CQ , o qual deverá ser minimizado.

O processo de estimação considera a série r_t e os 1000 vetores de estimativas iniciais e para cada um deles minimiza o critério quantílico, sendo o vetor

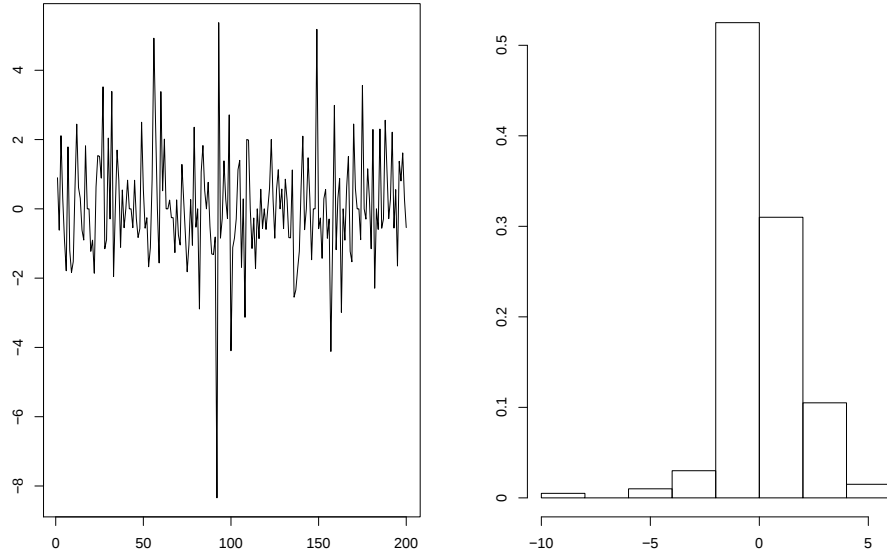


Figura 4.1. Evolução (a) e histograma (b) da série dos log-retornos r_t [$T = 200$].

que proporcionou o menor valor de CQ o vetor ótimo para cada avaliação. Na Figura 4.2 apresentamos os histogramas das estimativas de CQ , β_1 , β_2 e β_3 para o caso da especificação **SAV**, obtidas após o processo de minimização.

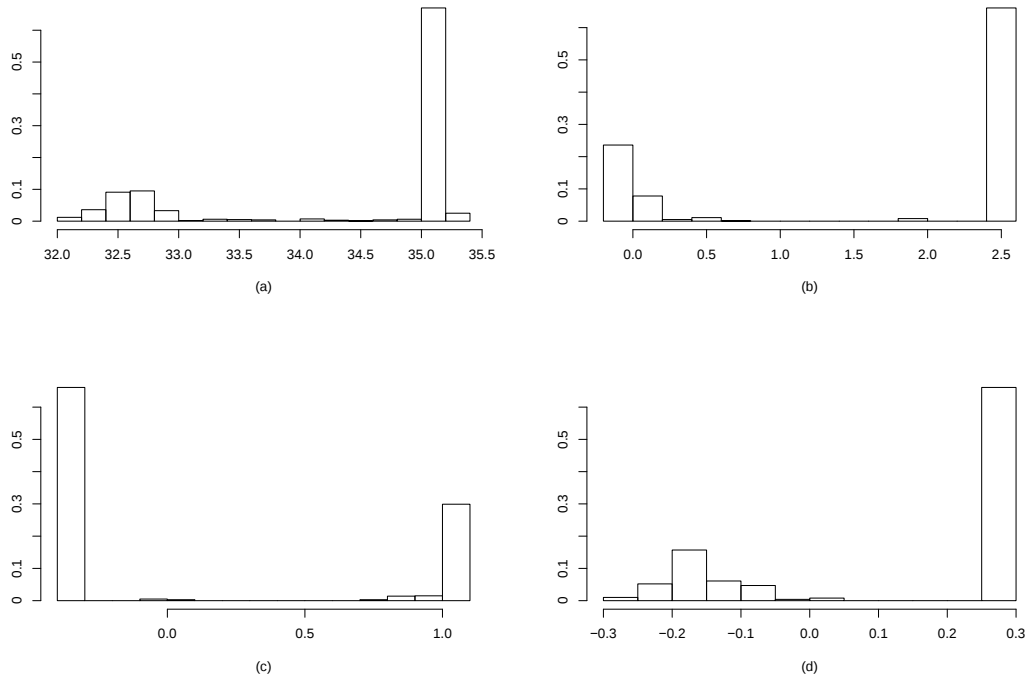


Figura 4.2. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação **SAV** [$T = 200, \theta = 0.05$].

Note que após o processo de minimização, a função CQ apresenta uma grande dispersão e que a maioria dos valores obtidos se encontram distantes do valor mínimo alcançado. Ou seja, mesmo após o processo de minimização,

as estimativas podem estar distantes do ponto ótimo. É possível observar também que as estimativas de β_i , $i = 1, 2, 3$, apresentam grande variação, inclusive com mudança de sinal no valor da estimativa.

A Figura 4.3 apresenta os histogramas das estimativas obtidas após o processo de minimização para a especificação **AS**, de onde podemos concluir que, assim como no caso da especificação **SAV**, as estimativas de CQ variam muito, ficando na maioria das vezes distantes do mínimo alcançado. Como também ocorreu no caso anterior, percebe-se que há grande variação nas estimativas dos parâmetros, as quais apresentam mudança de sinal, como, por exemplo, no caso de β_4 .

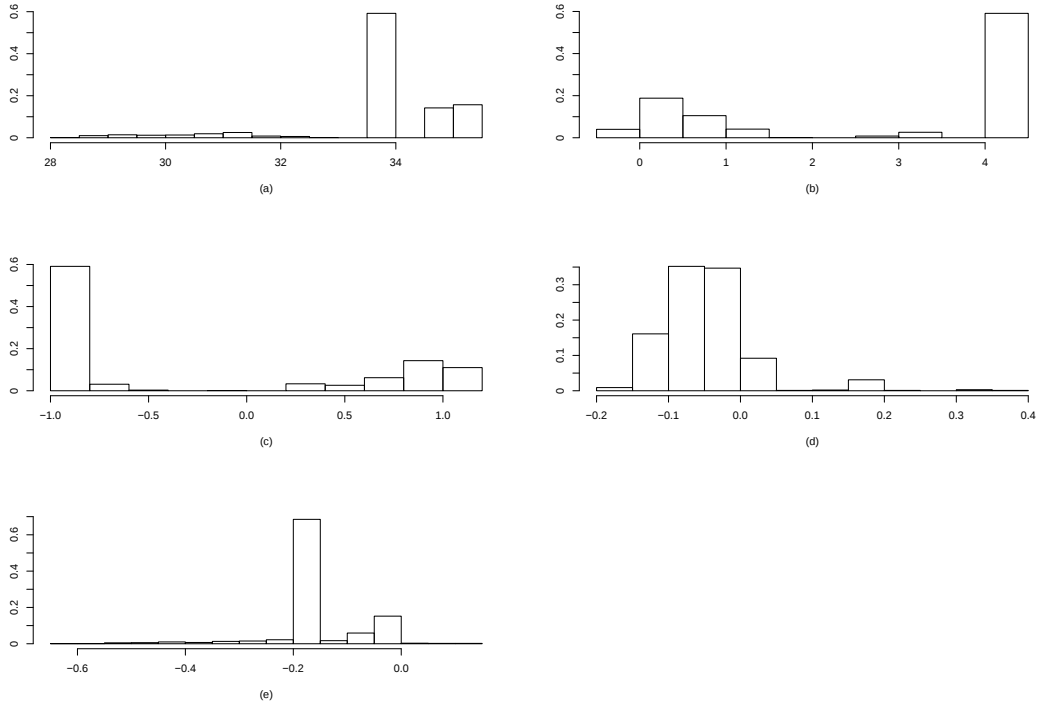


Figura 4.3. Histogramas de CQ , β_1 , β_2 , β_3 e β_4 para a especificação **AS** [$T = 200, \theta = 0.05$].

No caso da especificação **IG** (Figura 4.4) percebe-se a existência de dois aglomerados muito próximos nas estimativas de CQ e que em alguns poucos casos foram obtidos valores inferiores. Assim como nas especificações **SAV** e **AS**, as estimativas dos parâmetros se apresentam muito dispersas e com sinal variável.

Na especificação **ADP** não foram observados os mesmos problemas encontrados nas outras especificações. A Figura 4.5 apresenta os histogramas de CQ e β_1 , de onde podemos concluir que o processo de estimação é mais eficaz do que nas especificações anteriormente avaliadas, tendo em vista que o aglomerado de valores ótimos de CQ está posicionado no início da distribuição de frequência, ou seja, na maioria das vezes, o método obteve valores mínimos

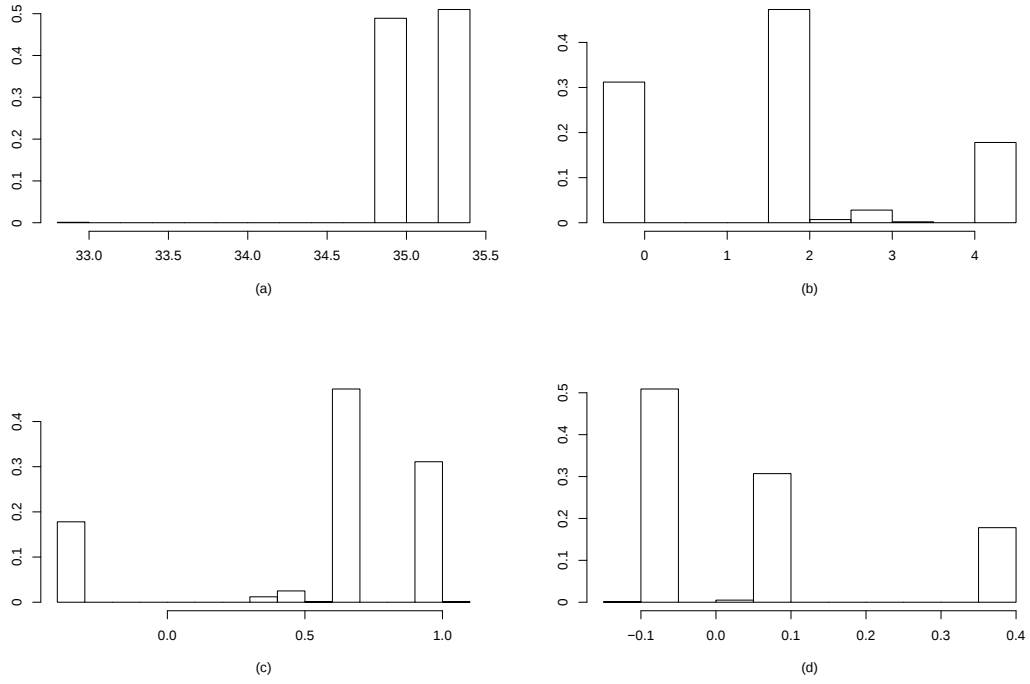


Figura 4.4. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação IG [$T = 200, \theta = 0.05$].

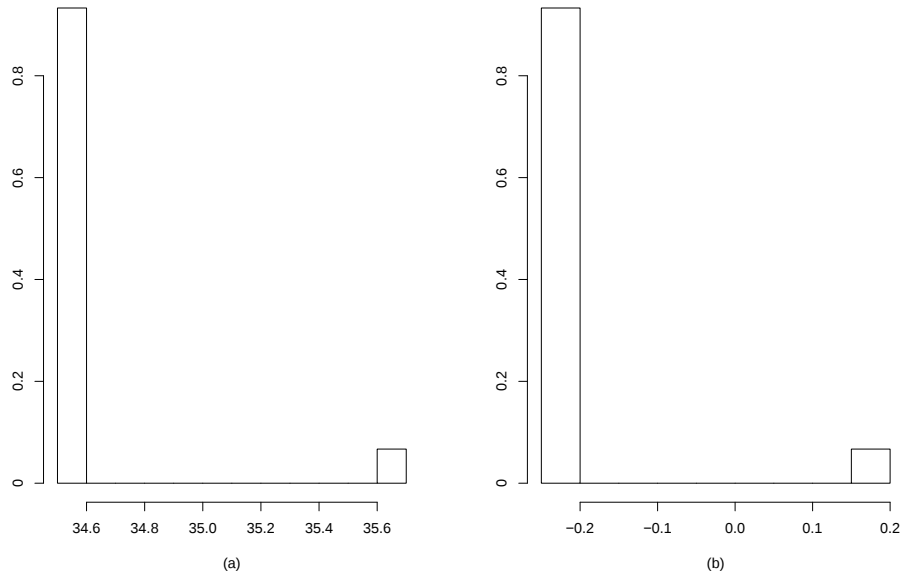


Figura 4.5. Histogramas de CQ e β_1 para a especificação ADP [$T = 200, \theta = 0.05$].

para o critério quantílico. Percebe-se ainda que os valores negativos de β_1 estão diretamente associados com os casos onde CQ foi mínimo.

A Figura 4.6 apresenta as seqüências VaR_t obtidas através dos parâmetros que proporcionaram os menores valores de CQ para as especificações SAV, AS, IG e ADP, das quais apenas na última observa-se um valor superior a 2%, para a observação 200. Note que próximo da 100ª observação, à exceção de ADP, todas as especificações conseguiram captar a queda ocorrida em r_t .

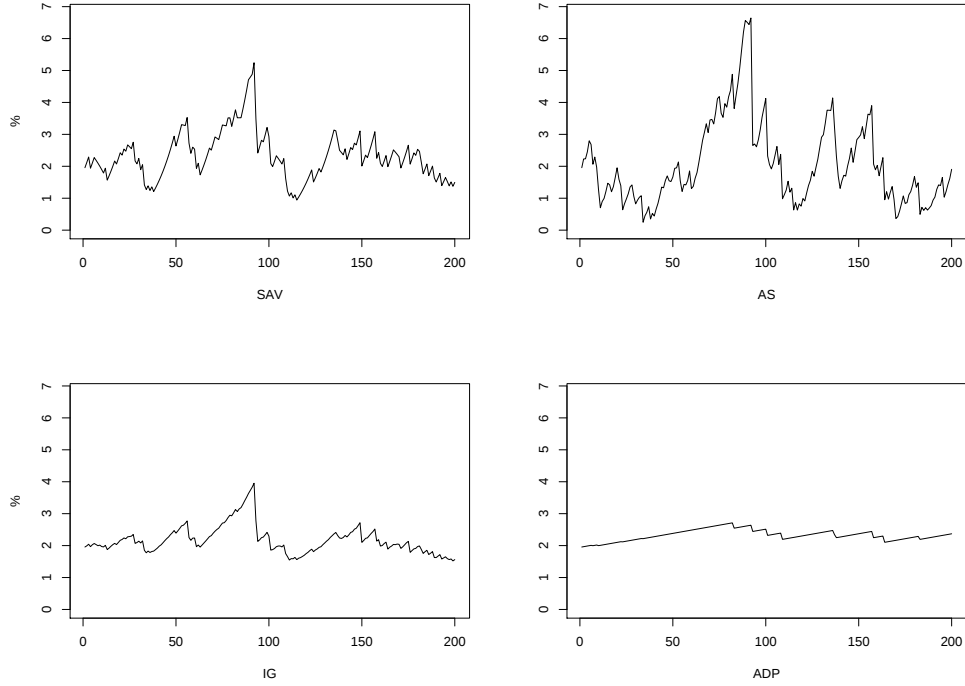


Figura 4.6. Sequências VaR_t ótimas para as especificações SAV, AS, IG e ADP [$T = 200, \theta = 0.05$].

Na Figura 4.7 são apresentadas as séries obtidas pela diferença entre as sequências $\text{VaR}_{\min}$ (ou seja, obtida a partir do vetor de parâmetros para os quais CQ é mínimo) e $\text{VaR}_{\max}$ (CQ máximo), de onde podemos concluir que o caso onde há menor discrepância é o da especificação ADP, fato que pode ser facilmente explicado tendo em vista que é nesse caso onde a distância entre CQ máximo e CQ mínimo é menor.

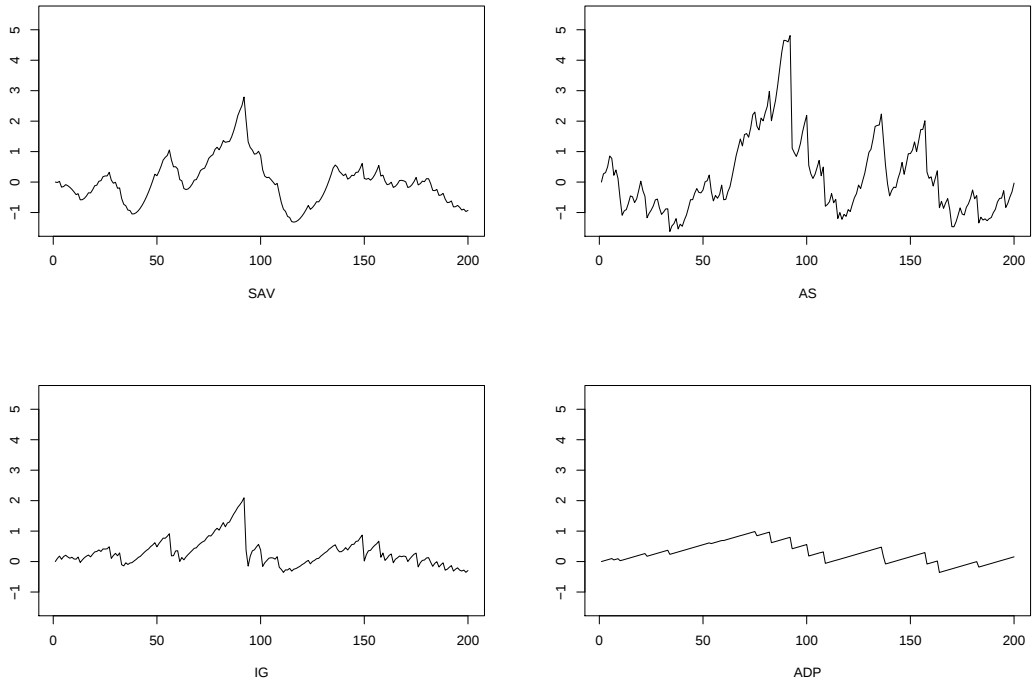


Figura 4.7. Diferença entre $\text{VaR}_{\min}$ e $\text{VaR}_{\max}$ obtidas para especificações SAV, AS, IG e ADP [$T = 200, \theta = 0.05$].

Contudo, à medida em que a diferença entre as estimativas mínima e máxima de CQ aumenta no processo de estimação, percebe-se que a série de discrepâncias entre $\text{VaR}_{\min}$ e $\text{VaR}_{\max}$ também é afetada. A Figura 4.7 nos mostra a importância de chegar a uma estimativa de CQ que seja mínima, pois, como vimos, os resultados em CQ podem ser desastrosos, como, por exemplo, na especificação **AS**. Dentre as quatro especificações foi exatamente em **AS** que a diferença nas estimativas de CQ se apresentou mais elevada (Figura 4.3). Nas demais especificações as discrepâncias flutuaram em torno de zero durante todo o período.

O Quadro 4.1 apresenta algumas estatísticas para as seqüência q_i , $i = 1, \dots, 1000$. De acordo com Engle & Manganelli (2004), há evidência de que o modelo está bem especificado se $q \approx 0.05$. Do Quadro 4.1, observamos que apenas para a especificação **IG** o percentual da seqüência VaR_t ótima é igual ao valor de θ adotado.

Quadro 4.1. Percentuais de *hits* observados para cada uma das especificações utilizadas [$T = 200, \theta = 0.05$].

Estatística	Especificação			
	SAV	AS	IG	ADP
β ótimo	4.500	6.500	5.000	4.000
mínimo	3.000	3.500	4.000	4.000
média	4.779	4.968	4.792	4.202
máximo	6.000	7.000	5.500	5.500
variância	0.202	0.196	0.093	0.150
viés	-0.222	-0.032	-0.208	-0.798
EQM	0.251	0.197	0.137	0.787

Em média, todas as especificações se aproximam de 0.05, mas a que se apresentou menos viesada foi a especificação **AS**, entretanto, a especificação **SAV** foi a que apresentou menor erro quadrático médio.

No caso em que $\theta = 0.01$, os resultados foram semelhantes aos observados para o caso de $\theta = 0.05$ nas especificações **SAV** (Figura 4.8), **AS** (Figura 4.9) e **IG** (Figura 4.10). Contudo, no caso da especificação **ADP**, todas as 1000 estimativas obtidas foram iguais a $CQ = 10.53041$ e $\beta_1 = -1.799$.

A Figura 4.11 apresenta as seqüências VaR_t obtidas através do vetor de parâmetros ótimos que proporcionou menor valor de CQ para cada uma das especificações. Note que, assim como no caso onde $\theta = 0.05$, a queda próxima a r_{100} é captada.

Através da Figura 4.12 é possível observar que há uma grande discrepância entre as seqüências **VaR** baseadas nas estimativas dos parâmetros cujos valores de CQ foram, respectivamente, mínimo e máximo. O gráfico da discrepância para o caso da especificação **ADP** não é apresentado devido ao fato das seqüên-

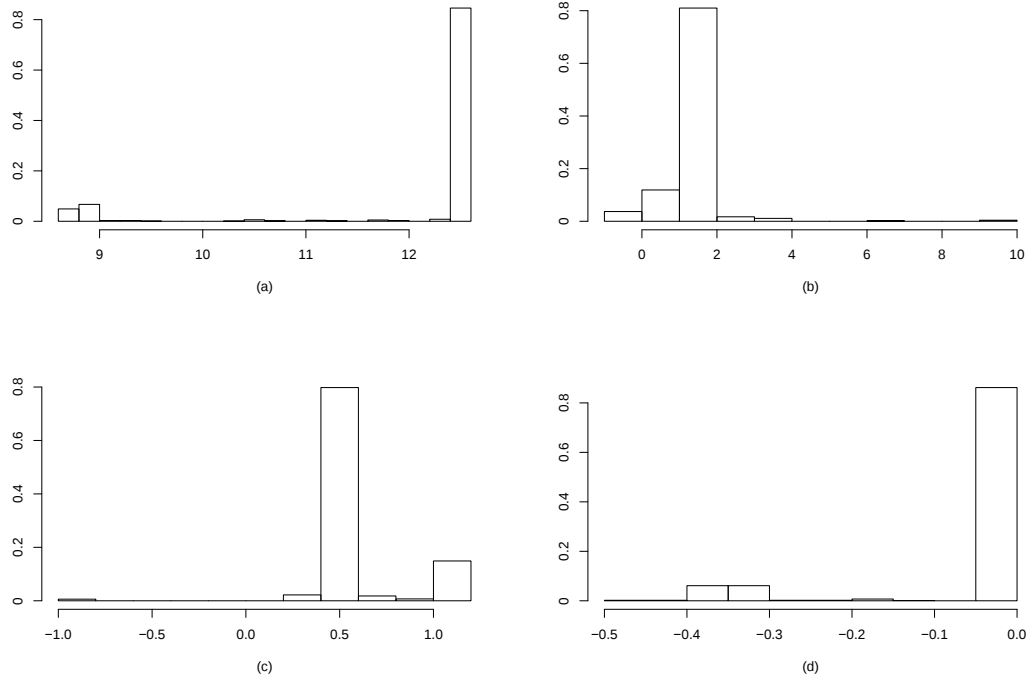


Figura 4.8. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação SAV [$T = 200, \theta = 0.01$].

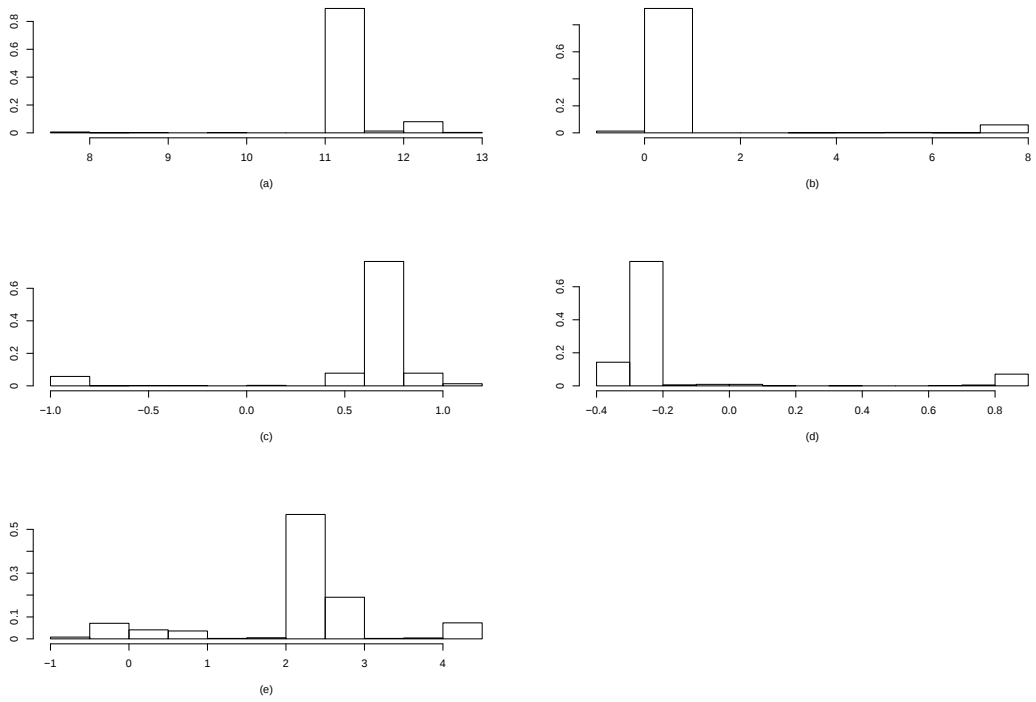


Figura 4.9. Histogramas de CQ , β_1 , β_2 , β_3 e β_4 para a especificação AS [$T = 200, \theta = 0.01$].

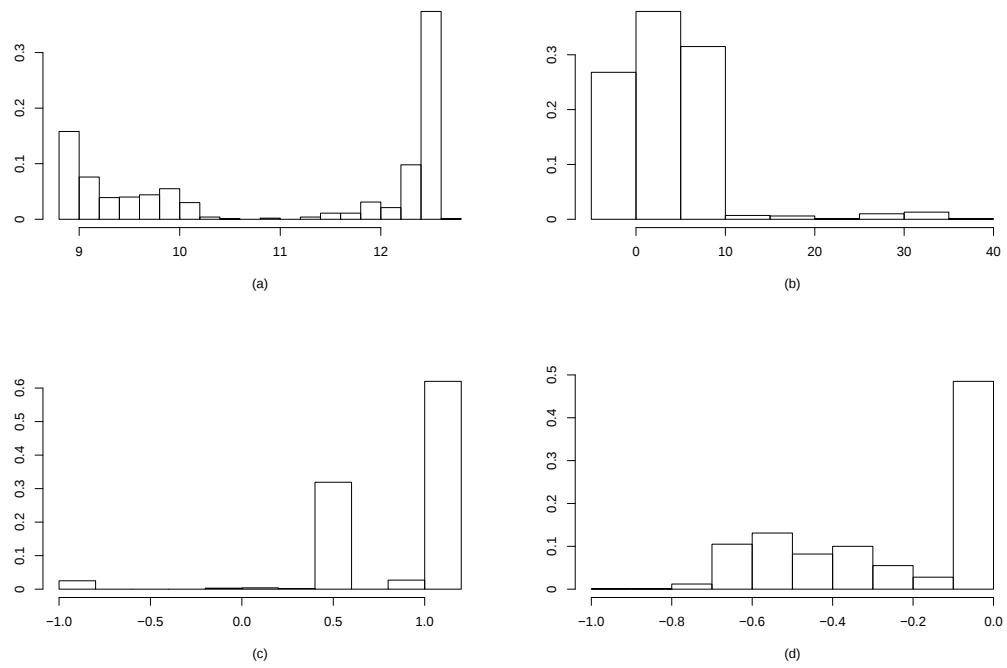


Figura 4.10. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação IG [$T = 200, \theta = 0.01$].

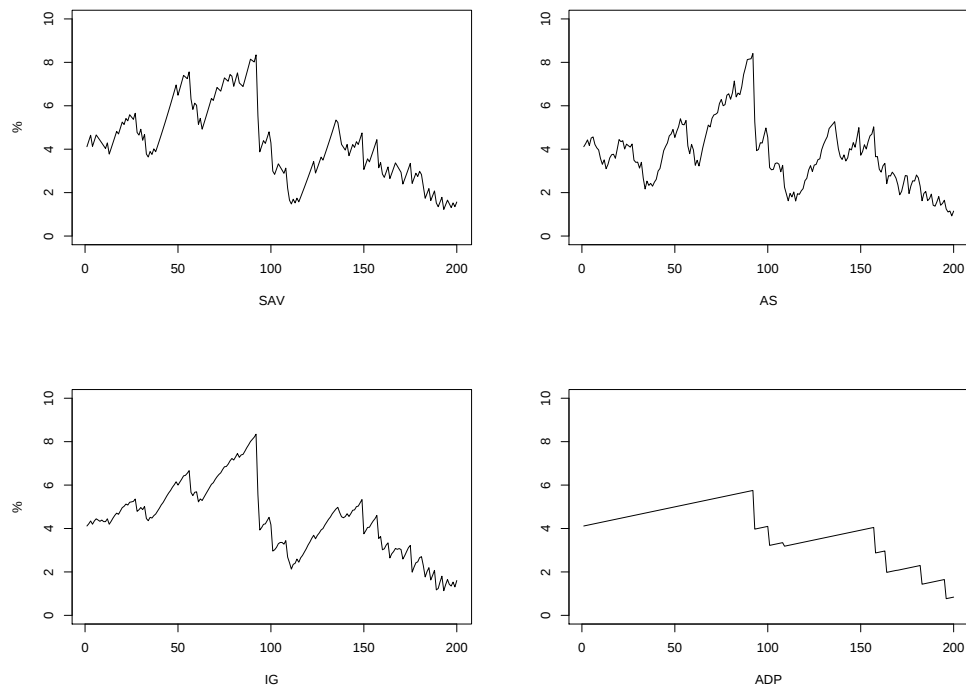


Figura 4.11. Sequências VaR_t ótimas para as especificações SAV, AS, IG e ADP [$T = 200, \theta = 0.01$].

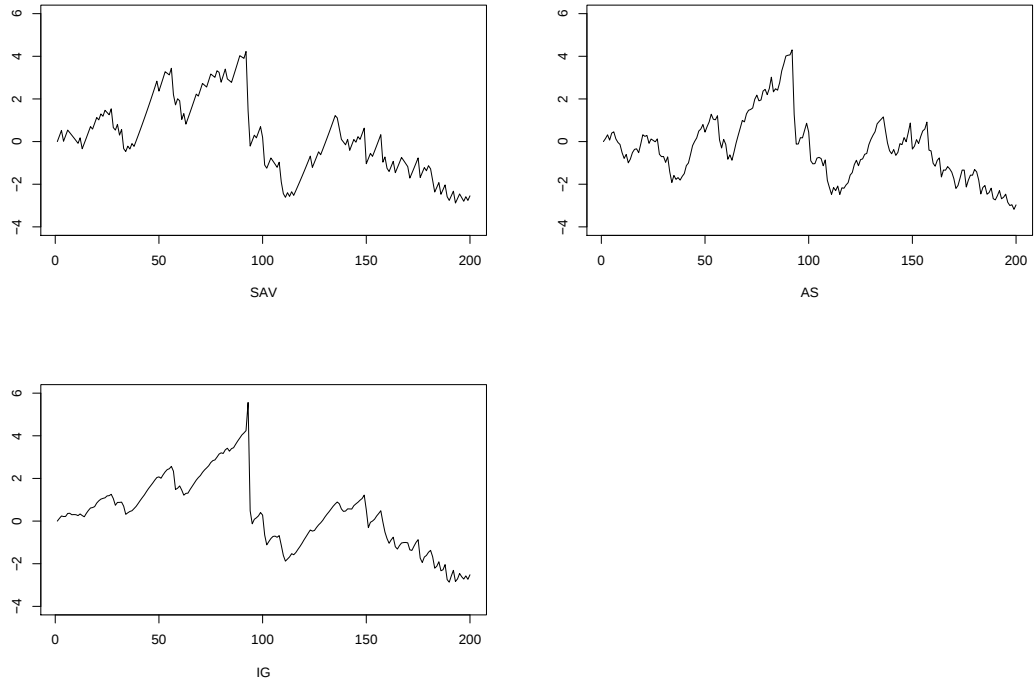


Figura 4.12. Diferença entre $\text{VaR}_{\min}$ e $\text{VaR}_{\max}$ obtidas para especificações SAV, AS e IG [$T = 200, \theta = 0.01$].

cias VaR terem sido iguais nas 1000 avaliações, consequência do fato de todas as estimativas terem sido idênticas.

O Quadro 4.2 apresenta algumas estatísticas relacionadas a q , de onde é possível concluir que a menos viesada das especificações, quando consideramos $T = 200$ e $\theta = 0.01$, é AS. Entretanto, essa especificação apresenta EQM superior ao da especificação SAV.

Quadro 4.2. Percentuais de *hits* observados para cada uma das especificações utilizadas [$T = 200, \theta = 0.01$].

Estatística	Especificação			
	SAV	AS	IG	ADP
β ótimo	1.000	0.500	0.500	1.500
mínimo	0.000	0.000	0.000	—
média	0.942	0.961	0.798	—
máximo	1.500	2.000	1.500	—
variância	0.060	0.154	0.109	—
viés	-0.058	-0.039	-0.202	—
EQM	0.063	0.155	0.150	—

4.2 Cenário II

A Figura 4.13 apresenta a série r_t , de tamanho 500, considerada na obtenção das estimativas dos parâmetros do modelo CAViaR. Note que entre as observações 350 e 400 há uma grande variação na série, fato que irá gerar valores muito elevados de VaR quando da consideração das especificações CAViaR.

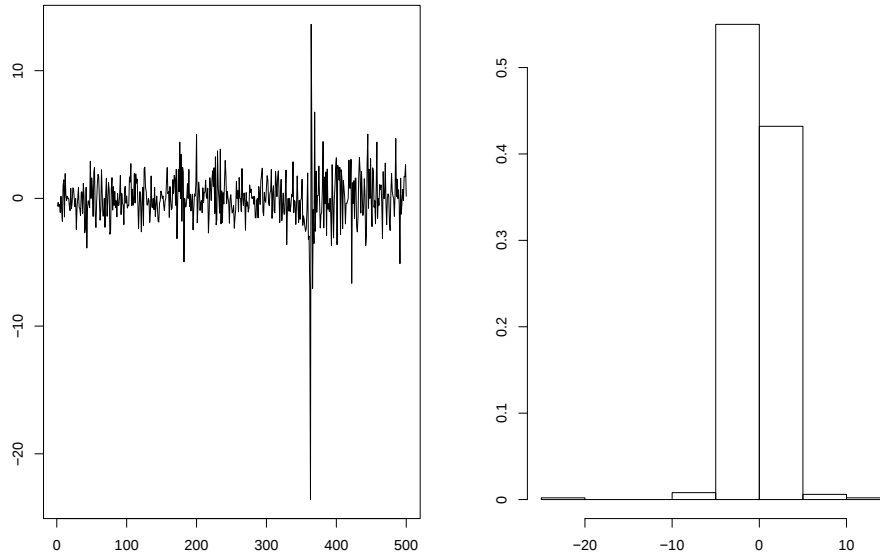


Figura 4.13. Evolução (a) e histograma (b) da série dos log-retornos r_t [$T = 500$].

Diferentemente do ocorrido no caso em que $T = 200$, observa-se que as estimativas de CQ , para o caso da especificação SAV, apresentaram um aglomerado, desta vez posicionado à esquerda do histograma, evidenciando que o método fornece com maior frequência valores de CQ de baixa magnitude, conforme podemos observar na Figura 4.14. Adicionalmente, observa-se que os parâmetros estimados estão diretamente associados ao valor de CQ obtido.

A Figura 4.15 apresenta os histogramas das estimativas da especificação AS, de onde podemos concluir que, assim como na especificação SAV, valores mais baixos de CQ são obtidos com maior frequência.

As estimativas de CQ para o caso da especificação IG apresentam uma peculiaridade. Nota-se da Figura 4.16 que as estimativas de CQ concentram-se na parte esquerda do gráfico, mas cerca de 20% delas estão na parte central (por volta de $CQ = 110$) e que uma quantidade mínima de observações foi superior a esse valor, com máximo correspondendo a aproximadamente $CQ = 114$. As estimativas de β_2 algumas vezes apresentaram troca de sinal.

No caso da especificação ADP, todas as estimativas foram iguais, com $CQ = 111.6112$ e $\beta_1 = 0.3879$. As seqüências VaR_t obtidas através dos vetores de parâmetros ótimos de cada especificação são apresentadas na Figura 4.17.

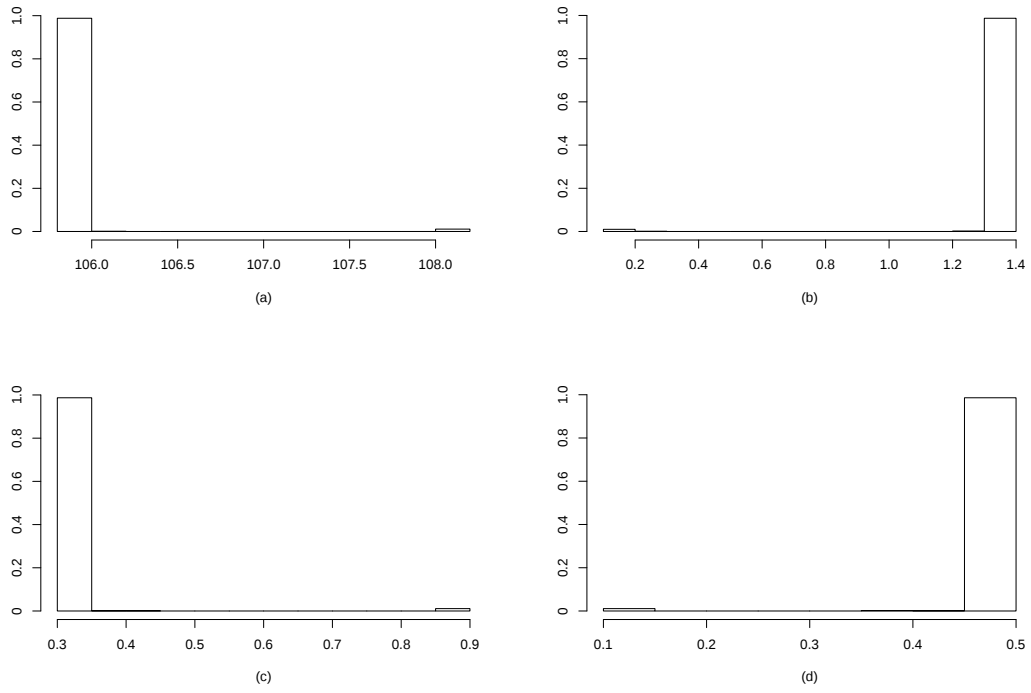


Figura 4.14. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação SAV [$T = 500, \theta = 0.05$].

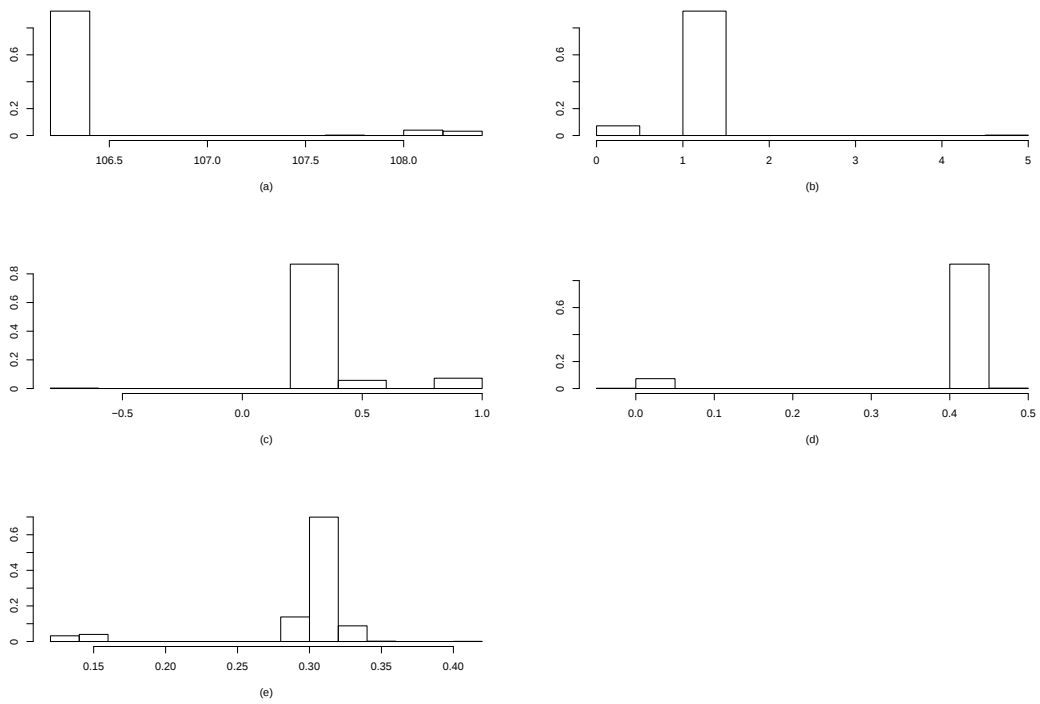


Figura 4.15. Histogramas de CQ , β_1 , β_2 , β_3 e β_4 para a especificação AS [$T = 500, \theta = 0.05$].

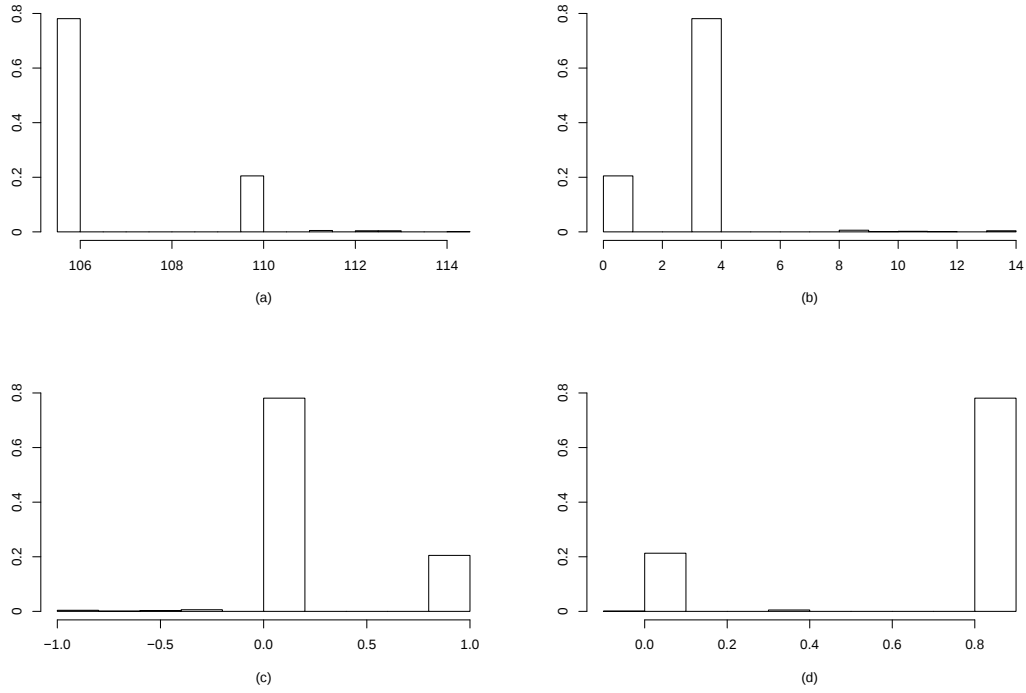


Figura 4.16. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação IG [$T = 500, \theta = 0.05$].

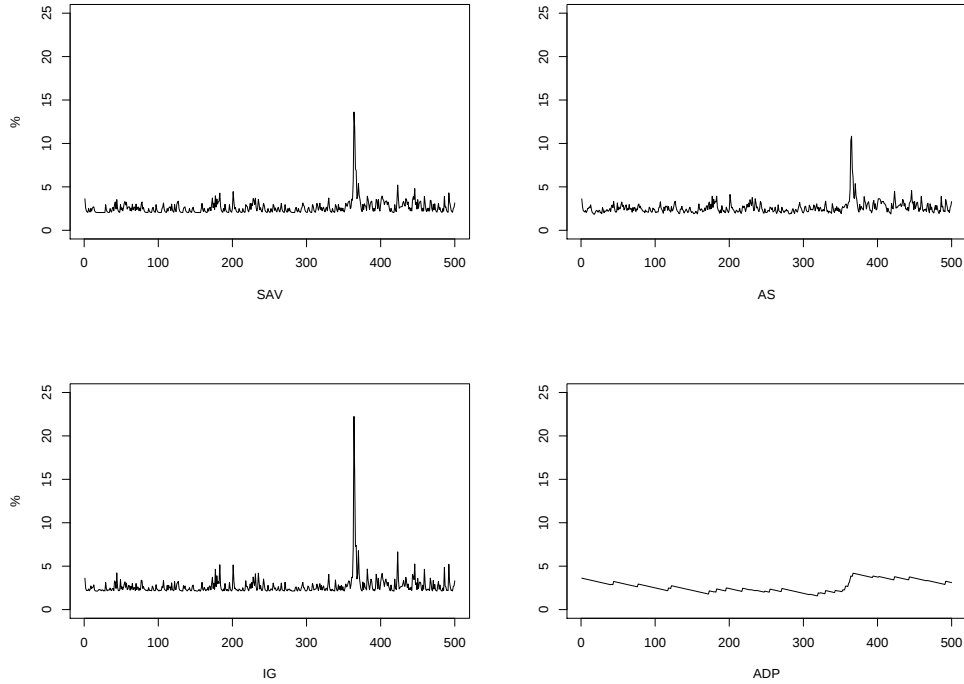


Figura 4.17. Sequências VaR_t ótimas para as especificações SAV, AS, IG e ADP [$T = 500, \theta = 0.05$].

Através da Figura 4.18 podemos observar as discrepâncias entre as seqüências $VaR_{\min}$ e $VaR_{\max}$ para as especificações SAV, AS e IG. Note que todas flutuam em torno de zero, mas entre as observações 350 e 400 as diferenças se apresentam elevadas, como decorrência do salto apresentado em r_t (Figura 4.13) nesse período.

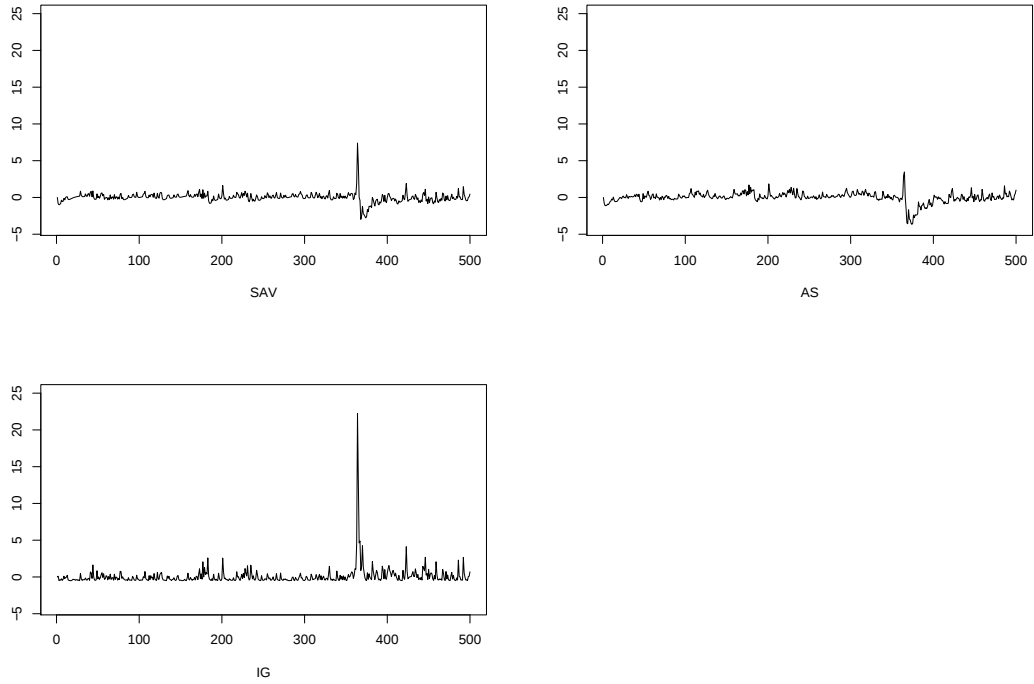


Figura 4.18. Diferença entre $\text{VaR}_{\min}$ e $\text{VaR}_{\max}$ obtidas para especificações SAV, AS e IG $[T = 500, \theta = 0.05]$.

O Quadro 4.3 apresenta algumas medidas de posição e dispersão para os valores de q estimados em cada uma das 1000 seqüências VaR estimadas, de onde podemos concluir que a especificação SAV foi a que apresentou o menor viés, enquanto que a especificação IG apresentou $q = 0.05$ para a seqüência VaR gerada pelo β ótimo.

Quadro 4.3. Percentuais de *hits* observados para cada uma das especificações utilizadas $[T = 500, \theta = 0.05]$.

Estatística	Especificação			
	SAV	AS	IG	ADP
β ótimo	5.200	5.200	5.000	0.00460
mínimo	4.600	4.600	4.400	—
média	5.004	4.976	4.937	—
máximo	5.200	5.400	5.200	—
variância	0.014	0.021	0.020	—
viés	0.004	-0.024	-0.063	—
EQM	0.014	0.022	0.024	—

Quando consideramos $\theta = 0.01$, os resultados obtidos para as especificações SAV, AS e IG foram semelhantes ao caso em que $\theta = 0.05$ e podem ser observados nas Figuras 4.19, 4.20 e 4.21, respectivamente. Porém, no caso da especificação ADP (Figura 4.22), observa-se que as estimativas de CQ concentram-se à direita do gráfico, evidenciando que esta especificação não é tão eficaz quanto às demais.

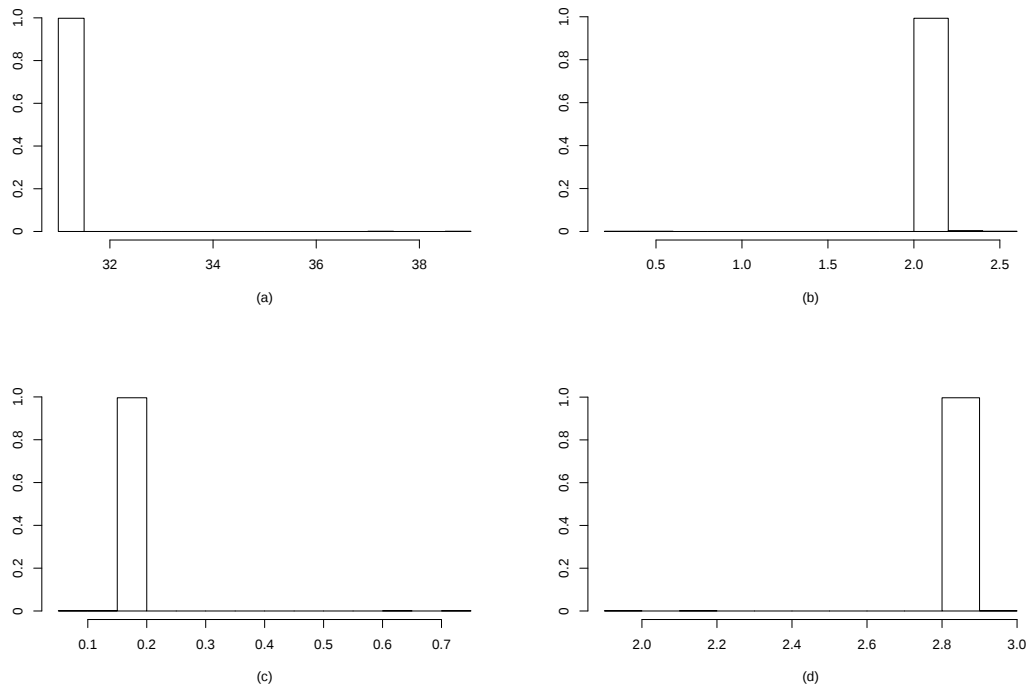


Figura 4.19. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação SAV [$T = 500, \theta = 0.01$].

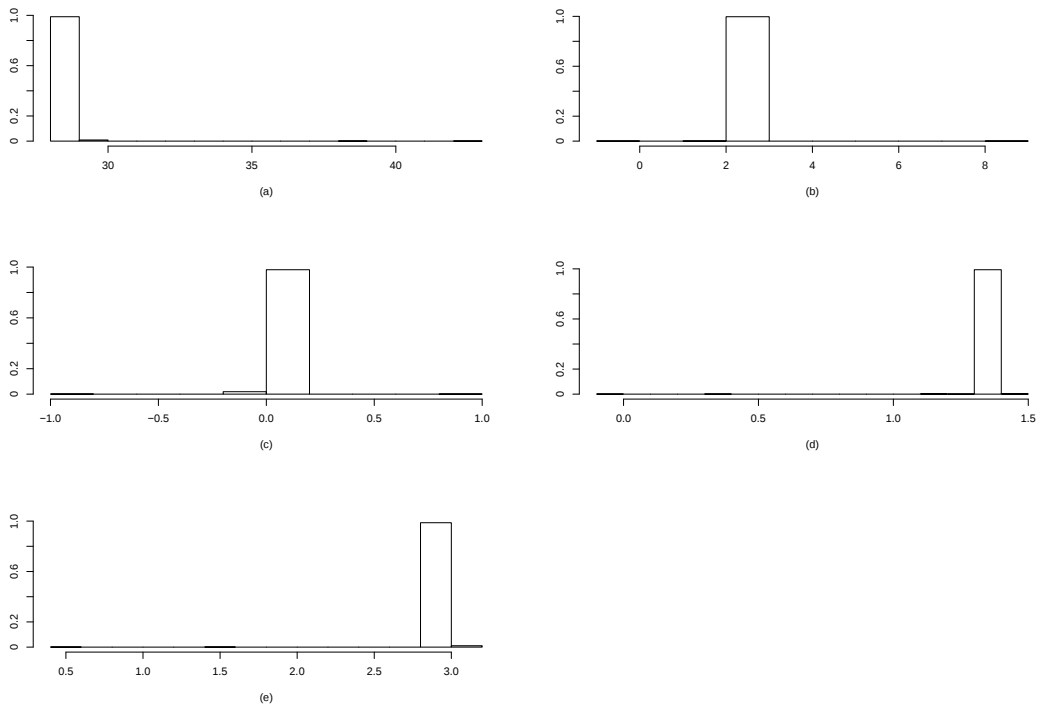


Figura 4.20. Histogramas de CQ , β_1 , β_2 , β_3 e β_4 para a especificação AS [$T = 500, \theta = 0.01$].

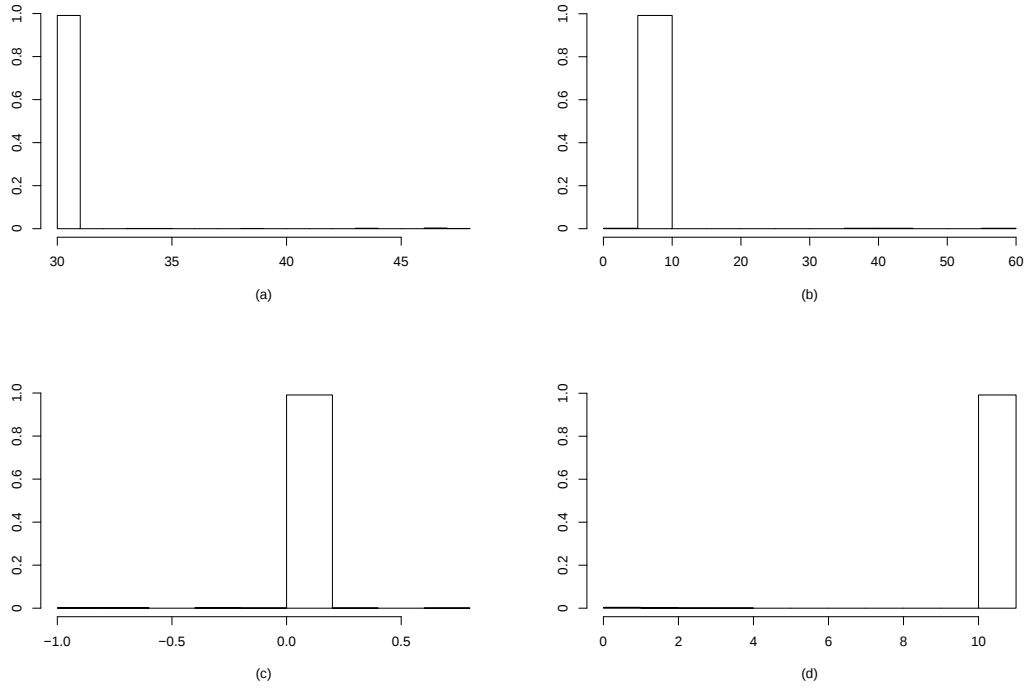


Figura 4.21. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação IG [$T = 500, \theta = 0.01$].

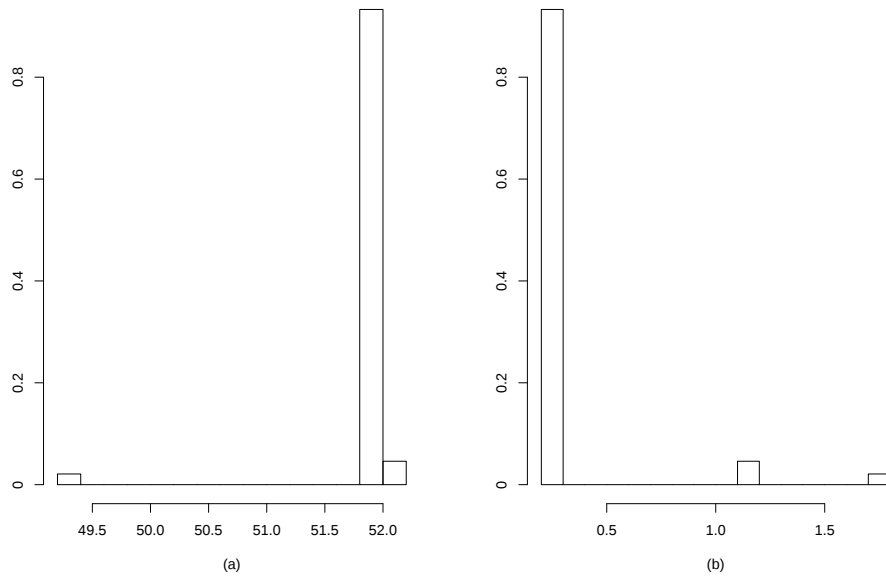


Figura 4.22. Histogramas de CQ e β_1 para a especificação ADP [$T = 500, \theta = 0.01$].

No caso de $\theta = 0.05$, para a especificação **ADP**, obtivemos todas as estimativas com mesmo valor, mas vale destacar que esse fato não indica qualidade do método, visto que este, como já dissemos, está diretamente relacionado com o conjunto de dados em estudo.

A Figura 4.23 apresenta as estimativas de VaR_t cujos valores de CQ foram mínimos, para cada uma das especificações consideradas. Como dissemos, o fato de haver uma grande variação na série por volta das observações 350 e 400 conduziu a valores muito altos de VaR nesse período. Entretanto, a especificação ADP não foi capaz de mensurar a magnitude do risco do investidor, tendo a estimativa VaR correspondente a essa variação alcançado um valor muito abaixo do que o alcançado por seus concorrentes.

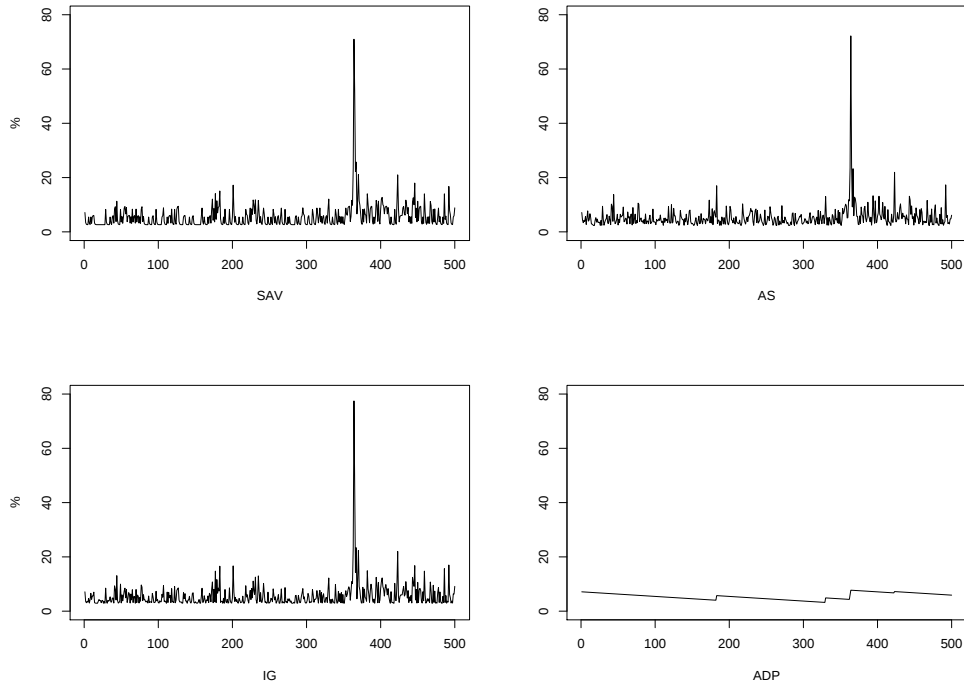


Figura 4.23. Sequências VaR_t ótimas para as especificações SAV, AS, IG e ADP $[T = 500, \theta = 0.01]$.

Percebe-se, da Figura 4.24 apresentada a seguir, que as discrepâncias de cada uma das especificações evoluem em torno de zero e que, à exceção da especificação ADP, todas conseguiram captar a mudança ocorrida na série entre as observações 350 e 400.

Do Quadro 4.4 é possível concluir que, apesar de nenhuma das especificações ter encontrado o percentual de *hits* na amostra quando da consideração do β ótimo, em média, elas se aproximaram de $\theta = 0.01$ e, conseqüentemente, apresentaram baixo viés.

4.3 Cenário III

A Figura 4.25 apresenta a série r_t , de tamanho 1000, considerada na obtenção das estimativas dos parâmetros do modelo CAViaR. Note que entre as observações 1 e 50 há uma grande variação na série, na qual r_t chegou a cerca

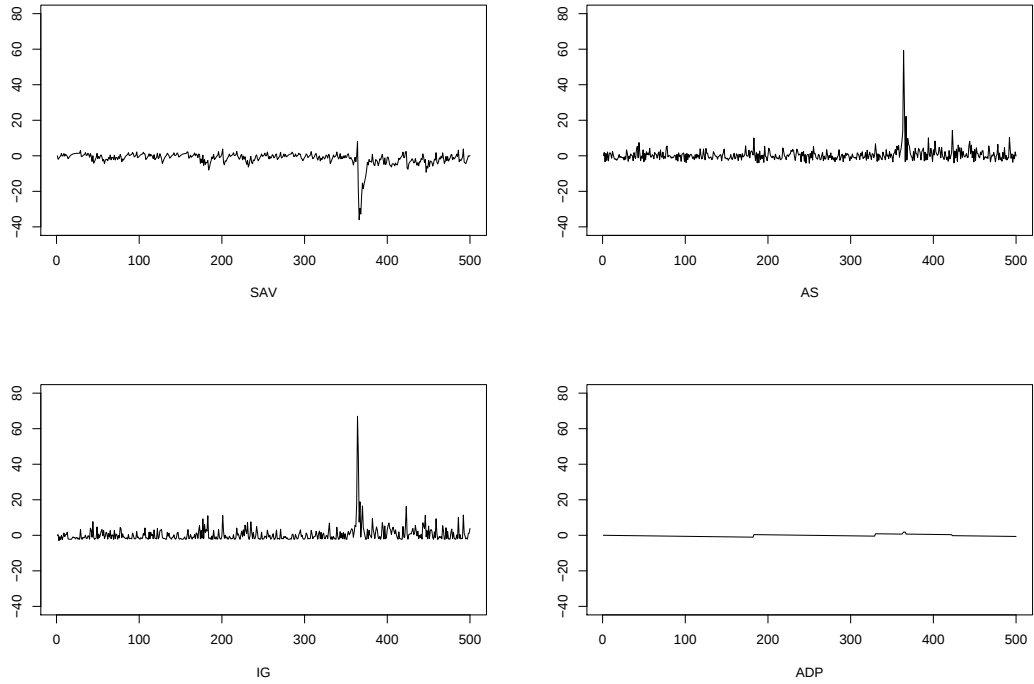


Figura 4.24. Diferença entre $\text{VaR}_{\min}$ e $\text{VaR}_{\max}$ obtidas para especificações SAV, AS, IG e ADP $[T = 500, \theta = 0.01]$.

Quadro 4.4. Percentuais de *hits* observados para cada uma das especificações utilizadas $[T = 500, \theta = 0.01]$.

Estatística	Especificação			
	SAV	AS	IG	ADP
β ótimo	1.200	1.200	0.600	0.800
mínimo	0.600	0.600	0.400	0.600
média	0.973	0.995	0.875	0.613
máximo	1.200	1.600	2.000	0.800
variância	0.019	0.028	0.021	0.003
viés	-0.027	-0.005	-0.125	-0.387
EQM	0.020	0.028	0.036	0.152

de -4 , ou seja, houve uma redução de aproximadamente 4% no índice com relação ao período anterior.

A partir da Figura 4.26 podemos observar o comportamento das estimativas dos parâmetros da especificação SAV para $\theta = 0.05$, de onde é possível concluir que as estimativas de CQ apresentam-se concentradas à direita do gráfico, fato que, como já dissemos, evidencia que o processo de estimação dos parâmetros para a série apresentada na Figura 4.25 não é eficaz. As estimativas de β_1 e β_2 estão concentradas em dois grupos, havendo uma grande distância entre eles. Já as estimativas de β_3 apresentam dois aglomerados bastante próximos e um terceiro mais afastado dos demais.

No caso da especificação AS (Figura 4.27), o processo de estimação se apresenta mais bem sucedido do que para a especificação SAV, tendo em vista que,

na maioria das vezes, valores mais baixos de CQ são obtidos. As estimativas dos parâmetros apresentam relação com o valor de CQ obtido, porém, no caso das estimativas de β_2 e β_3 , observa-se que há mudança de sinal na estimativa. Interpretação semelhante se dá para o caso da especificação IG, apresentada na Figura 4.28.

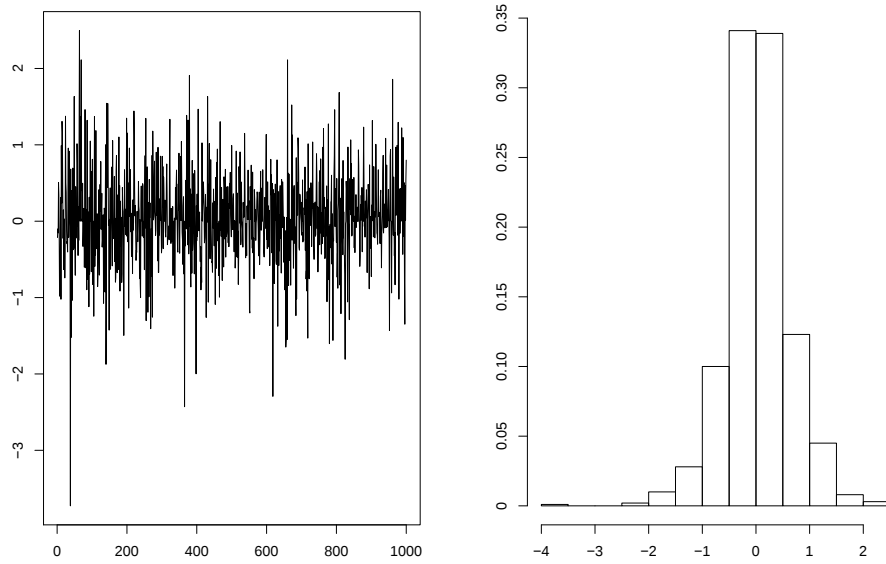


Figura 4.25. Evolução (a) e histograma (b) da série dos log-retornos r_t [$T = 1000$].

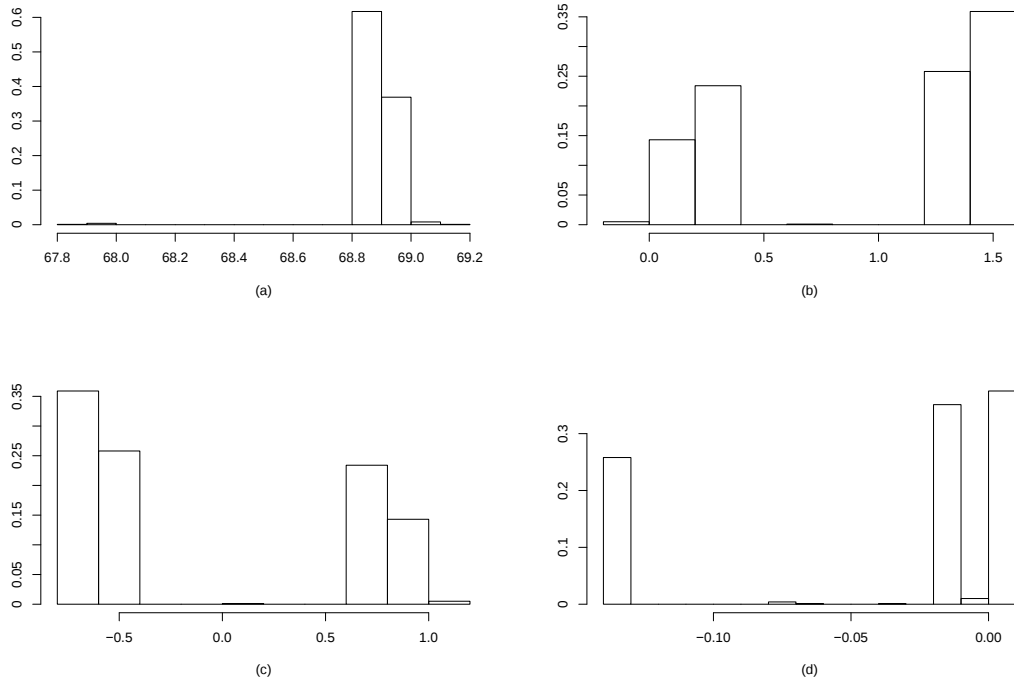


Figura 4.26. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação SAV [$T = 1000, \theta = 0.05$].

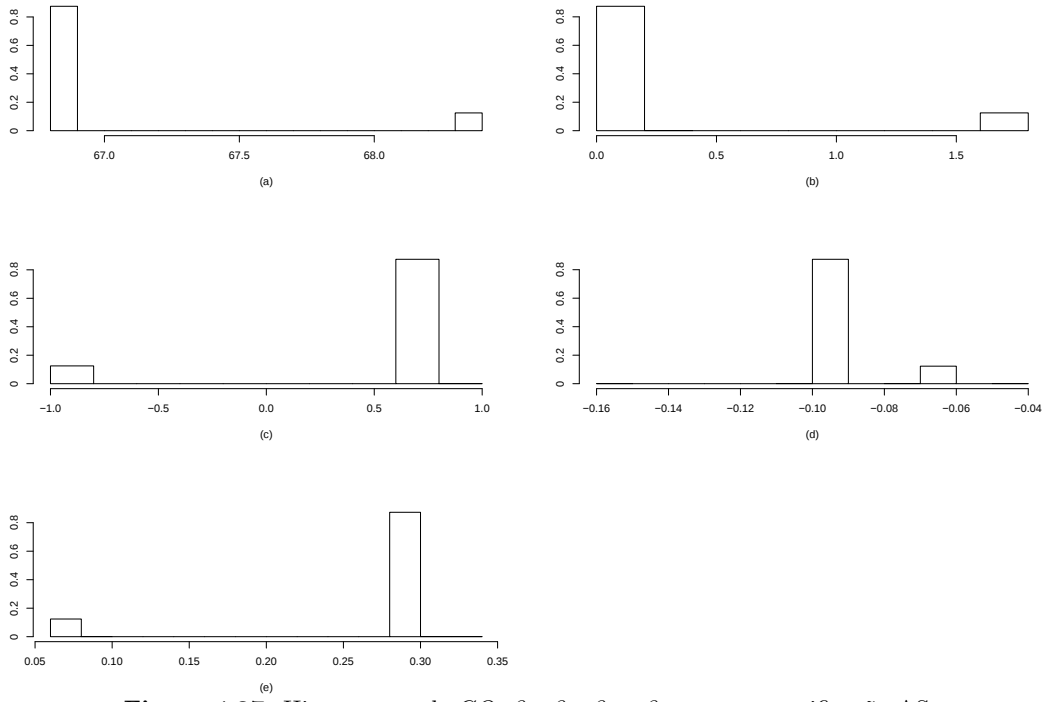


Figura 4.27. Histogramas de CQ , β_1 , β_2 , β_3 e β_4 para a especificação AS
 $[T = 1000, \theta = 0.05]$.

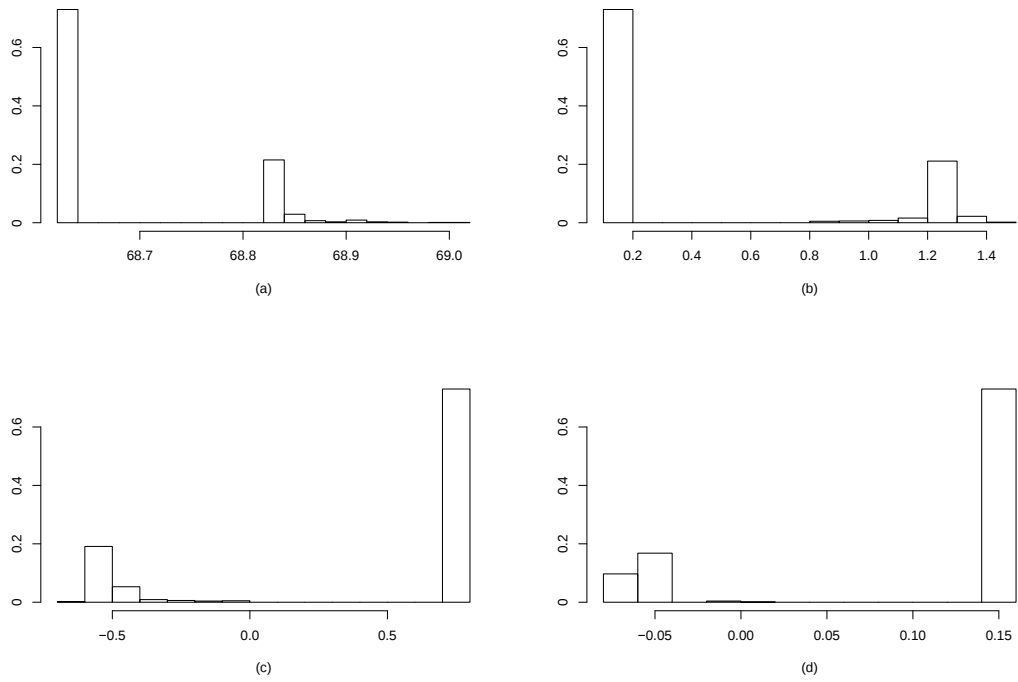


Figura 4.28. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação IG
 $[T = 1000, \theta = 0.05]$.

A Figura 4.29 apresenta os histogramas das 1000 estimativas de CQ e β_1 obtidas para o caso da especificação ADP, de onde é possível concluir que, na maioria das vezes, obteve-se valores mais baixos de CQ e que as estimativas de β_1 estão associadas com o valor de CQ obtido.

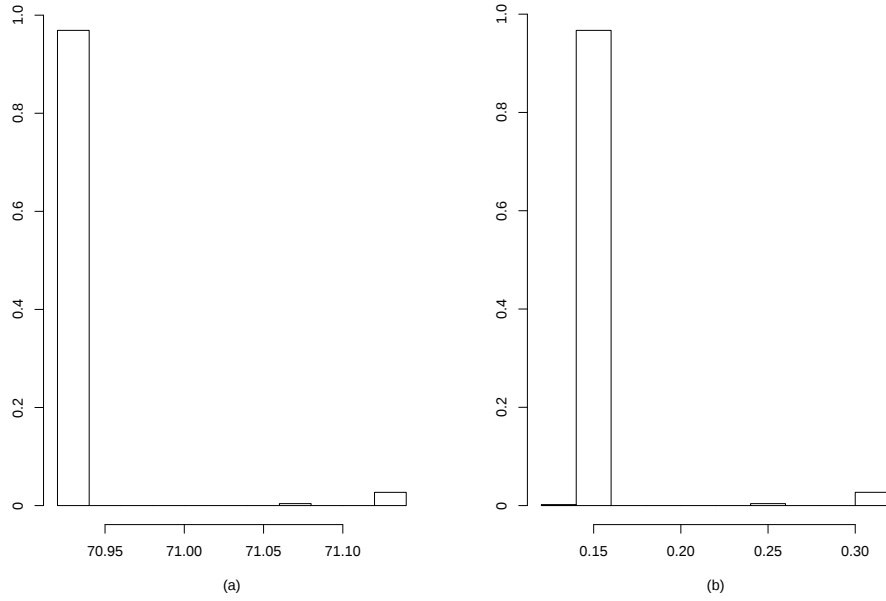


Figura 4.29. Histogramas de CQ e β_1 para a especificação ADP [$T = 1000, \theta = 0.05$].

As seqüências VaR obtidas a partir dos vetores ótimos que proporcionaram os menores valores de CQ para cada uma das especificações são apresentadas na Figura 4.30. Note que a variação apresentada entre as observações 1 e 50 foi detectada por todas as especificações CAViaR e os valores de VaR associados foram todos próximos. À exceção da especificação AS, os valores de VaR associados foram os maiores de suas respectivas seqüências.

A Figura 4.31 apresenta as diferenças entre as seqüências VaR de CQ mínimo e máximo, de onde é possível concluir que as mesmas apresentam flutuação constante em torno de zero. Porém, nos casos das especificações AS e IG, essa flutuação se dá com variação não constante ao longo do tempo. Já nos casos das especificações SAV e ADP, a série de diferenças evolui no intervalo $[-0.5; 0.5]$.

Na avaliação da adequabilidade do modelo, apenas a especificação ADP apresenta viés elevado, conforme podemos observar no Quadro 4.5. Percebe-se ainda que essa especificação, em nenhum dos 1000 ajustes, alcançou o percentual de *hits* desejado, tendo em vista que o valor máximo obtido foi de 0.047, o qual é inferior a 0.05. As demais especificações apresentaram bom desempenho na identificação de *hits*.

No caso em que $\theta = 0.01$, nota-se comportamento semelhante nos resultados das especificações SAV (Figura 4.32), AS (Figura 4.33) e IG (Figura 4.34),

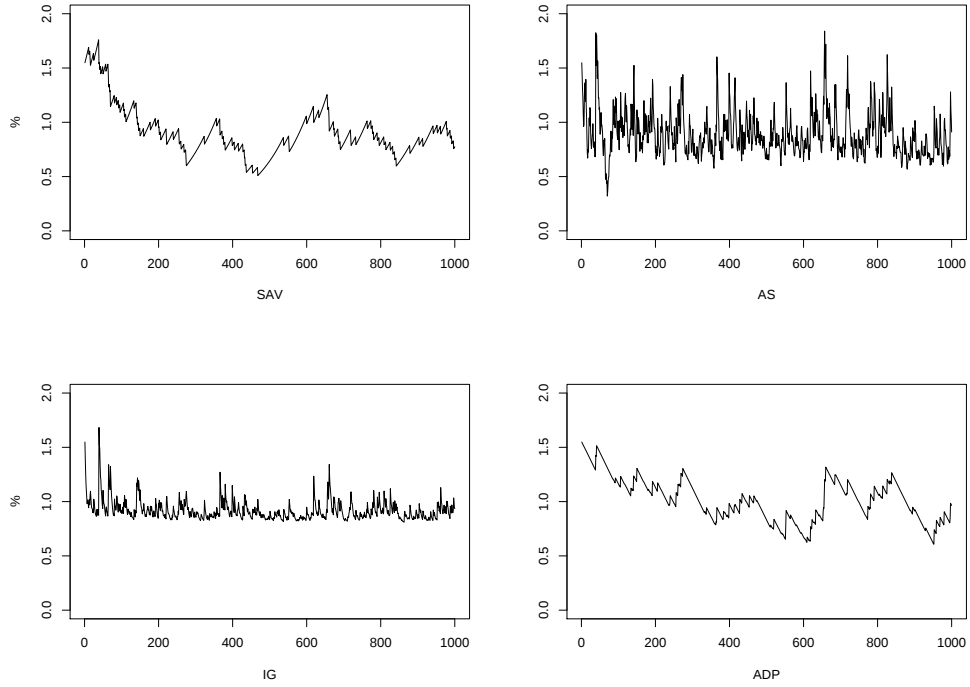


Figura 4.30. Sequências VaR_t ótimas para as especificações SAV, AS, IG e ADP $[T = 1000, \theta = 0.05]$.

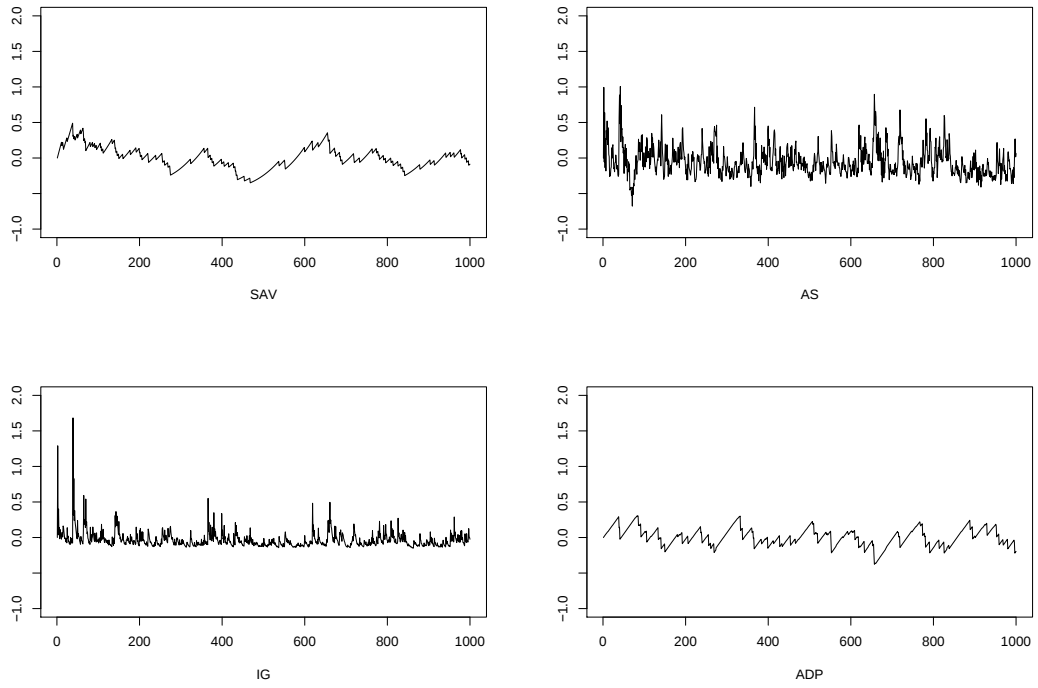


Figura 4.31. Diferença entre $\text{VaR}_{\min}$ e $\text{VaR}_{\max}$ obtidas para especificações SAV, AS, IG e ADP $[T = 1000, \theta = 0.05]$.

Quadro 4.5. Percentuais de *hits* observados para cada uma das especificações utilizadas [$T = 1000, \theta = 0.05$].

Estatística	Especificação			
	SAV	AS	IG	ADP
β ótimo	5.200	4.900	5.000	4.500
mínimo	4.700	4.700	4.400	4.300
média	4.990	4.985	5.001	4.465
máximo	5.200	5.200	5.200	4.500
variância	0.004	0.006	0.005	0.002
viés	-0.010	-0.015	0.001	-0.535
EQM	0.004	0.006	0.005	0.289

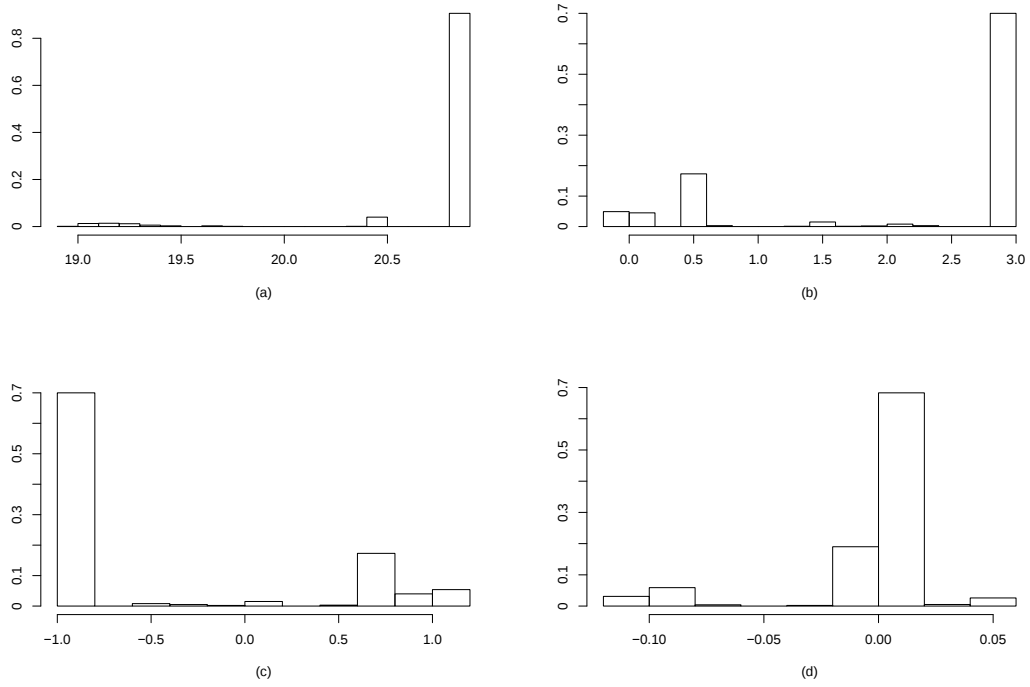


Figura 4.32. Histogramas de CQ , β_1 , β_2 e β_3 para a especificação SAV [$T = 1000, \theta = 0.01$].

nas quais se observa que o processo de estimação de CQ não é eficaz, enquanto que as estimativas dos parâmetros apresentam grande dispersão, havendo, inclusive, mudança no sinal da estimativa para alguns parâmetros. As estimativas de CQ e β_1 da especificação ADP, quando $\theta = 0.01$, foram todas iguais a, respectivamente, 22.5405 e 0.2259.

As Figuras 4.35 e 4.36 apresentam, respectivamente, as seqüências VaR_t ótimas e as discrepâncias entre $\text{VaR}_{\min}$ e $\text{VaR}_{\max}$ para cada uma das especificações consideradas. Da Figura 4.36 podemos concluir que as discrepâncias não oscilam em torno de zero. Esse fato é preocupante, tendo em vista que é possível que haja superestimação do valor de VaR , causando má alocação de recursos.

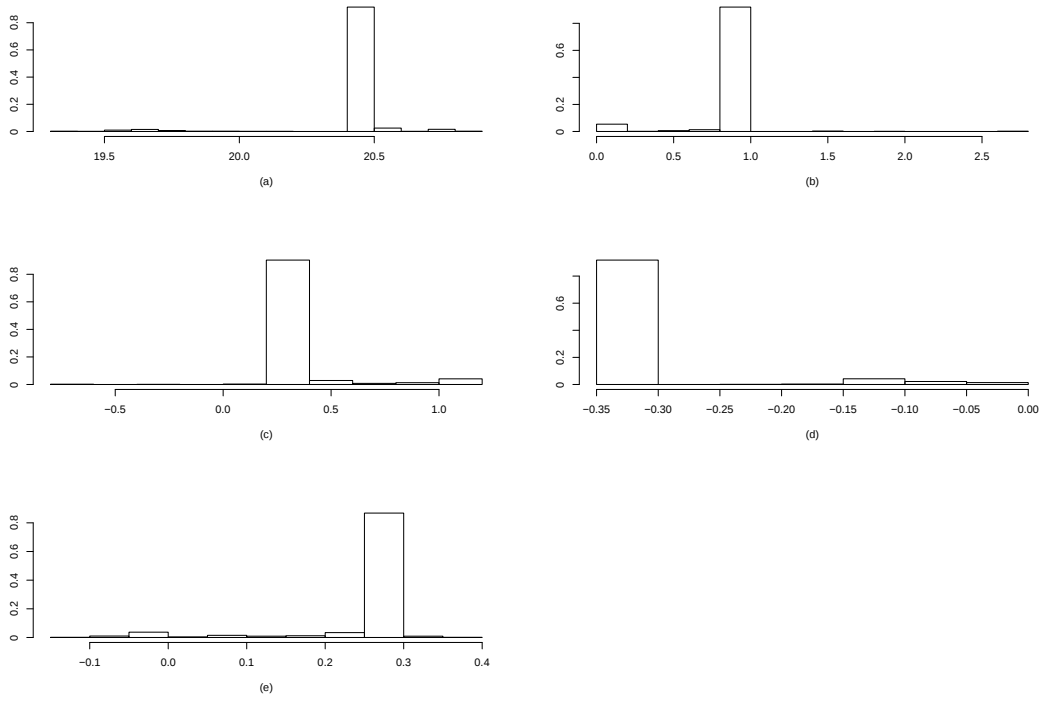


Figura 4.33. Histogramas de CQ , β_1 , β_2 , β_3 e β_4 para a especificação AS $[T = 1000, \theta = 0.01]$.

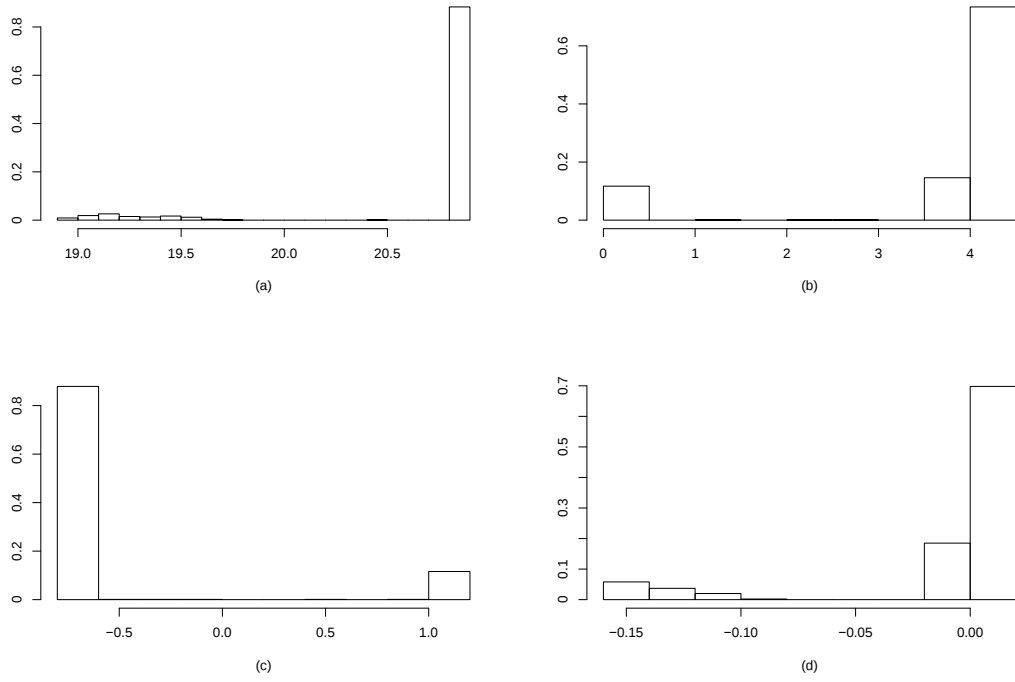


Figura 4.34. Histogramas de CQ (a), β_1 (b), β_2 (c) e β_3 (d) para a especificação IG $[T = 1000, \theta = 0.01]$.

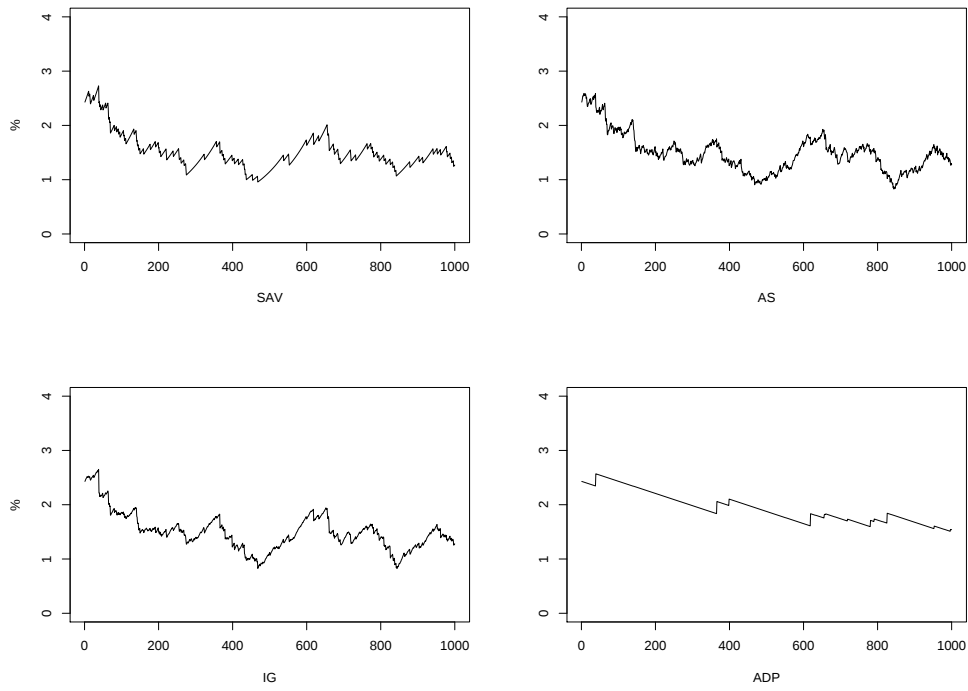


Figura 4.35. Sequências VaR_t ótimas para as especificações SAV (a), AS (b), IG (c) e ADP (d) $[T = 1000, \theta = 0.01]$.

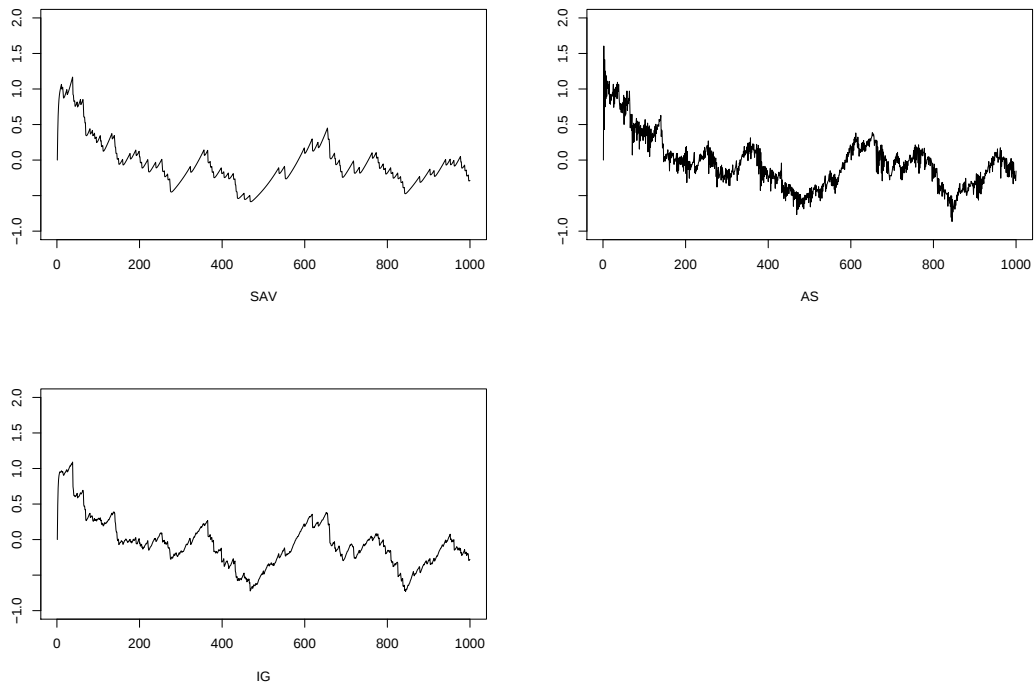


Figura 4.36. Diferença entre $\text{VaR}_{\min}$ e $\text{VaR}_{\max}$ obtidas para especificações SAV (a), AS (b), IG (c) e ADP (d) $[T = 1000, \theta = 0.01]$.

As especificações SAV, AS e IG apresentam baixo viés, conforme podemos observar no Quadro 4.6, sendo IG a única especificação que alcançou exatamente $\theta = 0.01$ para a sequência ótima. Nesse quadro não apresentamos os resultados de ADP devido ao fato das sequências VaR obtidas terem sido idênticas e, conseqüentemente, os percentuais de *hits* também são todos iguais aos obtidos através da sequência de CQ mínimo, onde $q = 0.6\%$.

Quadro 4.6. Percentuais de *hits* observados para cada uma das especificações utilizadas [$T = 1000, \theta = 0.01$].

Estatística	Especificação			
	SAV	AS	IG	ADP
β ótimo	1.100	0.900	1.000	0.600
mínimo	0.600	0.800	0.800	—
média	1.006	0.986	0.965	—
máximo	1.300	1.200	1.100	—
variância	0.003	0.005	0.003	—
viés	0.006	-0.014	-0.035	—
EQM	0.003	0.005	0.004	—

Aplicações

Neste capítulo apresentaremos aplicações da metodologia CAViaR ao mercado financeiro brasileiro. Para isso, consideraremos duas séries: Petrobrás e IBOVESPA. Os dados utilizados são referentes ao período entre 02/01/1995 e 16/09/2005, totalizando 2652 observações diárias, que foram obtidas a partir do **AE Broadcast**, software de acompanhamento da bolsa de valores de São Paulo, gentilmente cedido pela Agência Estado (<http://www.ae.com.br>). As 375 observações finais das duas séries (período de 18/03/2004 a 16/09/2005) foram reservadas para validação das especificações CAViaR, compondo o período que chamamos de pós-amostra. Portanto, as amostras consideradas contêm 2277 observações.

Na análise das séries foram utilizados os seguintes softwares: **RATS**, para cálculo de **VaR** através da metodologia RiskMetrics; **EViews**, para a abordagem econométrica do **VaR**; e **R**, para o cálculo de **VaR** pela metodologia CAViaR e pela TVE.

Na análise através dos modelos CAViaR consideramos $k = 10$ para a especificação adaptativa. Na avaliação da adequabilidade pelo teste de quantis dinâmicos consideramos como regressores os quatro primeiros termos defasados da função Hit_t .

Os códigos criados para cada uma dessas análises, bem como as séries utilizadas, estão disponíveis no endereço <http://box.de.ufpe.br/~tfsales>.

5.1 O caso da Petrobrás

As Figuras 5.1 e 5.2 apresentam as evoluções do índice da Petrobrás entre 02/01/1995 e 16/09/2005 e da série de log-retornos. Note que após serem tomados os log-retornos a distribuição passou a se assemelhar a uma distribuição simétrica, porém com caudas mais pesadas que a distribuição normal,

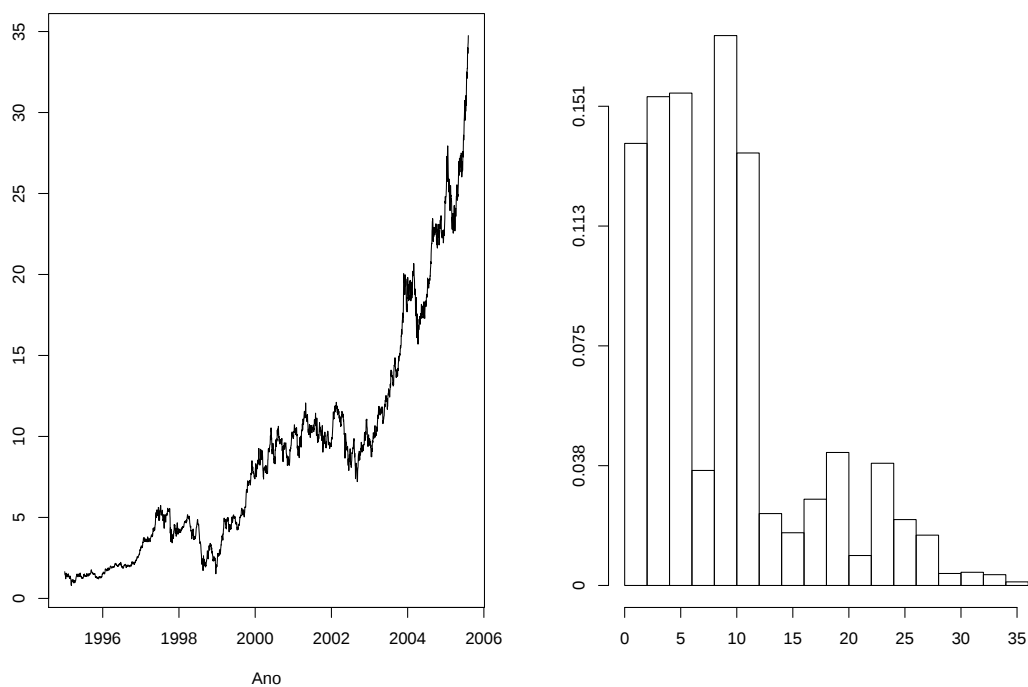


Figura 5.1. Série de Índices Petrobrás entre 02/01/1995 e 16/09/2005.

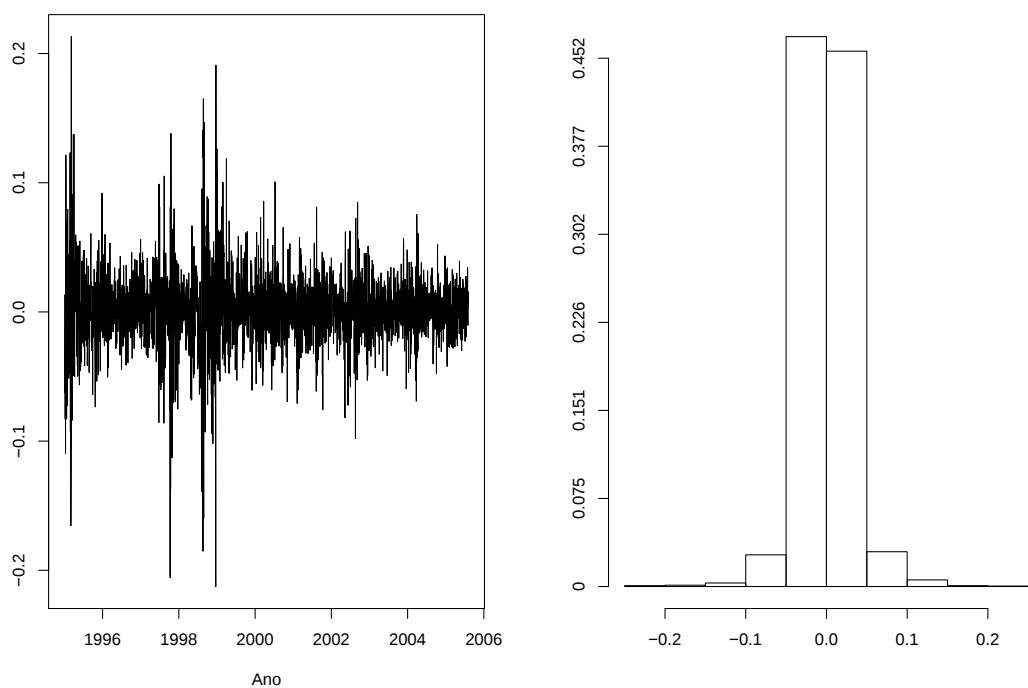


Figura 5.2. Série de log-retornos da Petrobrás entre 03/01/1995 e 16/09/2005.

como podemos observar no histograma dos log-retornos, apresentado na Figura 5.2 (b).

O Quadro 5.1 apresenta algumas estatísticas descritivas da série completa e das séries denominadas de amostra (r_t , $t \in [1, 2277]$) e pós-amostra (r_t , $t \in [2278, 2651]$). Note que as séries amostra e pós-amostra apresentam padrões diferentes; por exemplo, a curtose da primeira é muito elevada, enquanto a da segunda é inferior a 1.3.

Quadro 5.1. Estatísticas descritivas da série de log-retornos da Petrobrás.

Estatística	r_t	r_t , $t \in [1, 2277]$	r_t , $t \in [2278, 2651]$
Mínimo	-0.0923	-0.0923	-0.03004
1º quartil	-0.0058	-0.0062	-0.0046
Média	0.0005	0.0005	0.0007
Mediana	0.0000	0.0000	0.0007
3º quartil	0.0068	0.0073	0.0057
Máximo	0.0925	0.0925	0.0327
Variância	0.0002	0.0002	0.0001
Assimetria	-0.0423	-0.0374	-0.0108
Curtose	7.4939	6.9291	1.2936

[A1] RiskMetrics

Assumindo que a média condicional do modelo é zero, ajustaremos o modelo em [2.2], ou seja,

$$r_t = a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha \sigma_{t-1}^2 + (1 - \alpha) r_{t-1}^2, \quad 0 < \alpha < 1.$$

Os resultados do ajuste estão apresentados no Quadro 5.2, de onde chegamos a

$$\sigma_t^2 = 0.9136 \sigma_{t-1}^2 + 0.0864 r_{t-1}^2.$$

Quadro 5.2. Resultados do ajuste RiskMetrics aos log-retornos da Petrobrás.

	Estimativa	Erro Padrão	p -valor
α	0.9136	0.0075	0.0000
$Q(10)$	0.0009	$Q(20)$	0.0052

A previsão da volatilidade um passo adiante é dada por

$$\hat{\sigma}_{2277}^2(1) = 0.9136 \sigma_{2277}^2 + 0.0864 r_{2277}^2,$$

e do ajuste obtemos $\hat{\sigma}_{2277}^2 = 0.0098$ e $r_{2277} = 0.01678$, logo, $\hat{\sigma}_{2278}^2(1) = 0.000109$. Assim, a perda máxima que o investidor pode sofrer é de 0.01723, ao nível de 5%, pois

$$\text{VaR} = 1.65 \times \sqrt{0.000109} = 1.65 \times 0.01044 = 0.01723.$$

Ou seja, a perda máxima esperada para $t = 2278$ é de 1.723%.

A Figura 5.3 apresenta a volatilidade estimada através da metodologia RiskMetrics para a série de log-retornos da Petrobrás, sendo possível observar quatro grandes picos de volatilidade: em 10/3/1995, durante a crise do México; em 27/10/1997, crise asiática; em 10/9/1998, período inicial da moratória da Rússia; e em 14/1/1999, mudança de câmbio fixo para câmbio flutuante no Plano Real.

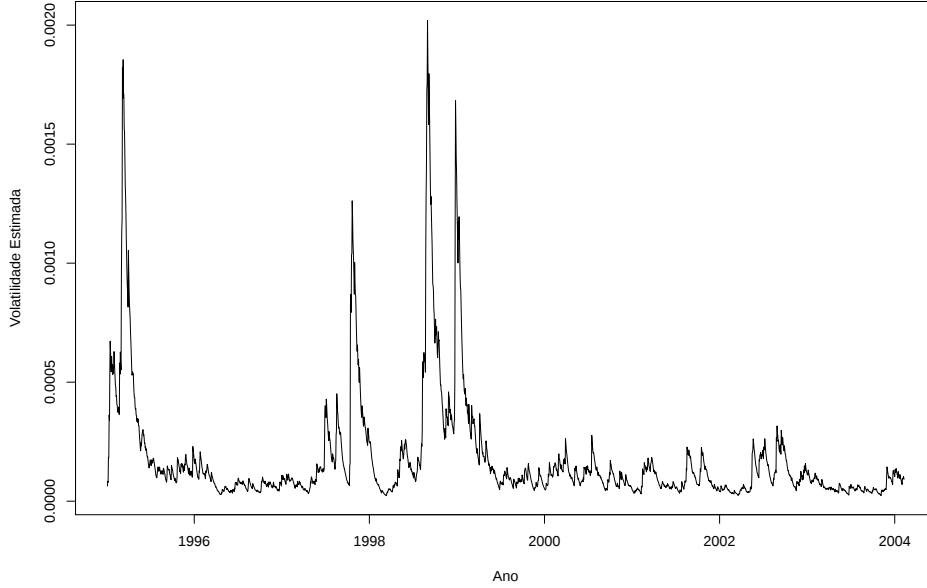


Figura 5.3. Volatilidade estimada da série de log-retornos da Petrobrás através do RiskMetrics.

[A2] Abordagem econométrica

O modelo $\text{ARMA}(p, q)\text{--GARCH}(r, s)$ é dado por

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + a_t - \sum_{j=1}^q \theta_j a_{t-j},$$

$$a_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2.$$

Para a seleção de modelos utilizamos o BIC , que selecionou o modelo $\text{ARMA}(2,2)\text{--GARCH}(1,1)$. Assim, temos que

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2},$$

$$a_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

O Quadro 5.3 apresenta os resultados do ajuste $\text{ARMA}(2,2)\text{--GARCH}(1,1)$, de onde concluímos que todos os parâmetros são significativos. O teste de

Quadro 5.3. Resultados do ajuste ARCH–GARCH aos log-retornos da Petrobrás.

Equação da média			
	Estimativa	Erro Padrão	<i>p</i> -valor
ϕ_0	0.0010	0.0002	0.0000
ϕ_1	1.0758	0.1238	0.0000
ϕ_2	−0.3943	0.1146	0.0006
θ_1	−0.9542	0.1290	0.0000
θ_2	0.2608	0.1189	0.0283
$Q(10)$	0.1142	$Q(20)$	0.4678
Equação da variância			
	Estimativa	Erro Padrão	<i>p</i> -valor
α_0	0.0001	9.49×10^{-7}	0.0000
α_1	0.1393	0.0101	0.0000
β_1	0.8234	0.0138	0.0000
$Q(10)$	0.0062	$Q(20)$	0.0849

Ljung-Box não detectou auto-correlação em a_t (equação da média), tendo detectado nos quadrados dos resíduos do ajuste ARMA(2,2) (equação da variância), justificando assim a utilização do modelo ARMA–GARCH.

Temos que

$$r_t = 0.0010 + 1.0758r_{t-1} - 0.3943r_{t-2} + a_t + 0.9542a_{t-1} - 0.2608a_{t-2},$$

$$\sigma_t^2 = 0.0001 + 0.1393a_{t-1}^2 + 0.8234\sigma_{t-1}^2.$$

Do ajuste, temos que $a_{2277} = 0.0142$ e $\sigma_{2277}^2 = 0.000127$, logo,

$$\hat{\sigma}_{2277}^2(1) = 0.0001 + 0.1393 \times (0.0142)^2 + 0.8234 \times 0.000127 = 0.000234.$$

Assumindo normalidade nos log-retornos, obtemos o Valor em Risco de 5% como sendo

$$\text{VaR} = 1.65 \times \sqrt{0.000234} = 1.65 \times 0.0153 = 0.02525.$$

[A3] Estimação quantílica

Para o cálculo do VaR 5% através da estimação quantílica, se faz necessária a obtenção do quantil de nível 0.05 para a série de log-retornos da Petrobrás. Note que $2277 \times 0.05 = 113.85$ e, portanto, consideramos $p_{113} = 0.04963$, $p_{114} = 0.05007$, $r_{(113)} = -0.019305$ e $r_{(114)} = -0.019189$.

Desta forma, obtivemos

$$\begin{aligned} \text{VaR} &= \left(\frac{0.05007 - 0.05}{0.05007 - 0.04963} \right) \times (-0.01931) + \left(\frac{0.05 - 0.04963}{0.05007 - 0.04963} \right) \times (-0.01919) \\ &= -0.019206. \end{aligned}$$

[A4] VaR pela TVE

Na abordagem da TVE, consideramos $n = 21$, de forma a considerar grupos referentes a aproximadamente um mês comercial e, conseqüentemente, $g = 108$. Assim, temos que as $m = 9$ primeiras observações não foram utilizadas na obtenção dos mínimos. Os mínimos dos 108 grupos formados estão apresentados na Figura 5.4.

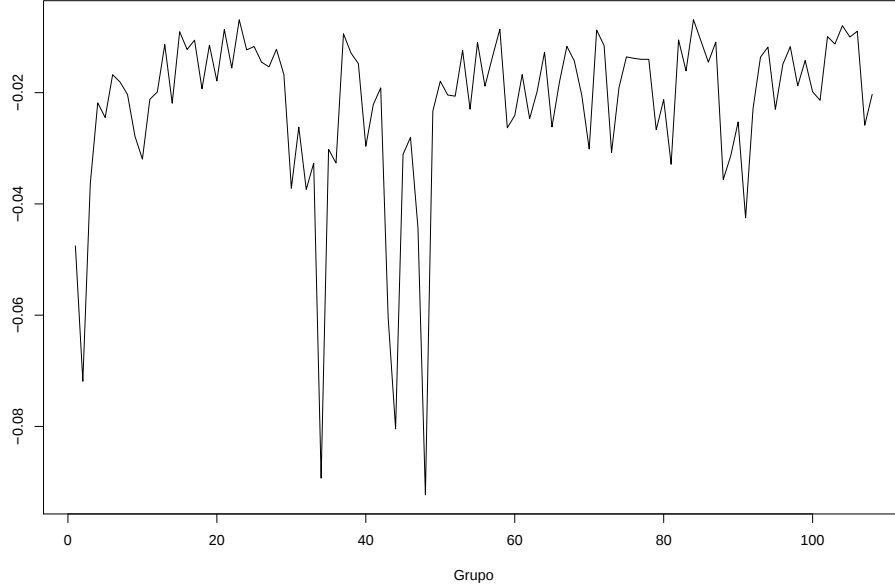


Figura 5.4. Série de mínimos da Petrobrás em grupos de tamanho 21.

O Quadro 5.4 apresenta as estimativas dos parâmetros da distribuição GEV, de onde podemos concluir que a série de mínimos segue distribuição de Fréchet. Portanto, o VaR de 5% deve ser estimado por

$$\text{VaR} = -0.02321 - \left(\frac{0.01140}{-0.55109} \right) \times \{1 - [-21 \log(0.95)]^{-0.55109}\} = -0.02404,$$

ou seja, a perda máxima esperada é de 0.02404 vezes o montante investido.

Quadro 5.4. Estimativas dos parâmetros da GEV para a série de mínimos dos log-retornos da Petrobrás.

Parâmetro	α	β	ξ
Estimativa	0.01140	-0.02321	-0.55109
Erro Padrão	3.1×10^{-5}	3.4×10^{-5}	0.01923

[A5] CAViaR

O Quadro 5.5 apresenta as estimativas dos parâmetros do modelo CAViaR, seus respectivos erros-padrão e p -valores associados, o valor do critério quantílico, o percentual de *hits* e o p -valor do teste de quantis dinâmicos para as

observações amostrais e pós-amostrais, obtidas para a série de log-retornos da Petrobrás.

Note que o parâmetro β_2 é sempre significativo independente do nível θ adotado, indicando que aglomerados de volatilidade também ocorrem nas caudas da distribuição dos retornos. No tocante ao percentual de *hits* identificado na amostra, temos que, para $\theta = 0.05$, nas quatro especificações, obteve-se q próximo a 5%; em particular, na especificação IG esse percentual foi igual a 5.01%. Contudo, o desempenho pós-amostra das quatro especificações não foi bom, visto que o melhor desempenho foi verificado para ADP, cujo percentual foi de cerca de 2.7%.

No caso em que $\theta = 0.01$, as especificações AS, IG e ADP identificaram cerca de 1.01% de *hits* na amostra, enquanto a SAV identificou 1.1%. Novamente, o desempenho pós-amostra dessas especificações não foi satisfatório, salvo a especificação ADP, que identificou 1.07% de *hits* na amostra.

Das especificações consideradas, apenas AS e IG não foram rejeitadas pelo teste de quantis dinâmicos, no caso em que $\theta = 0.05$ para as observações amostrais e pós-amostrais. Com $\theta = 0.01$, nenhuma especificação foi rejeitada no período pós-amostra; porém, quando realizado entre as observações amostrais, apenas AS não foi rejeitada.

Quadro 5.5. Estimativas das especificações CAViaR para a séries de log-retornos da Petrobrás.

Estimativas	$\theta = 0.05$				$\theta = 0.01$			
	SAV	AS	IG	ADP	SAV	AS	IG	ADP
β_1	0.00016	0.00134	0.00002	-0.00001	0.00004	0.00290	0.00001	0.00003
Erro Padrão	0.00014	0.00043	0.00001	0.00001	0.00004	0.00120	0.00004	0.00001
p -valor	0.12019	0.00082	0.00249	0.00000	0.17750	0.00720	0.00950	0.00000
β_2	0.99111	0.78836	0.80030	—	0.98870	0.78380	0.60751	—
Erro Padrão	0.00822	0.04423	0.02003	—	0.01300	0.04370	0.06861	—
p -valor	0.00000	0.00000	0.00000	—	0.00000	0.00000	0.00000	—
β_3	0.85151	0.07385	0.35472	—	1.42212	-0.00752	2.24581	—
Erro Padrão	0.00211	0.04843	0.30042	—	0.00667	0.10469	1.15823	—
p -valor	0.00000	0.06366	0.11885	—	0.00000	0.47148	0.0000	—
β_4	—	0.54287	—	—	—	0.76394	—	—
Erro Padrão	—	0.11675	—	—	—	0.08576	—	—
p -valor	—	0.00001	—	—	—	0.00000	—	—
CQ	3.43067	2.96525	3.10144	3.41028	1.15821	0.81254	0.91642	0.13600
q amostra (%)	5.27270	5.05272	5.00879	5.14060	1.09845	1.01050	1.01050	1.01050
q pós-amostra (%)	0.80000	1.60000	2.13333	2.66667	0.00000	0.53333	0.22674	1.06667
DQ amostra	0.00000	0.62843	0.30547	0.00000	0.00000	0.55011	0.00123	0.00000
DQ pós-amostra	0.03385	0.16912	0.34575	0.00000	1.00000	0.99146	0.89124	0.86610

Assim, a especificação escolhida, tanto para $\theta = 0.05$ quanto para $\theta = 0.01$, foi a da inclinação assimétrica (AS), tendo em vista que a mesma não foi

rejeitada pelo teste de quantis dinâmicos. Note que para $\theta = 0.05$ todos os parâmetros de **AS** são significativos ao nível 10%, mas para $\theta = 0.01$ apenas o parâmetro β_3 foi não-significativo. Ou seja, quando $\theta = 0.05$, o modelo responde diferentemente às observações positivas e negativas, tendo em vista que $\beta_3 \neq \beta_4$. Já quando $\theta = 0.01$, as observações positivas não apresentam impacto no cálculo de **VaR**.

Temos, portanto, que o modelo considerado para a obtenção do **VaR** de 5% é dado por

$$\text{VaR}_t = 0.0013 + 0.7884\text{VaR}_{t-1} + 0.0739 \max(r_{t-1}, 0) + 0.5429 \min(r_{t-1}, 0).$$

O impacto em **VaR** do aumento de uma unidade em VaR_{t-1} ocasiona um acréscimo de 0.7884 no **VaR** do próximo período. O aumento de uma unidade em r_t , caso r_t seja positivo, é de apenas 0.0739, enquanto se $r_t < 0$, o aumento em **VaR** é de 0.5429.

A Figura 5.5 apresenta as seqüências **VaR** estimadas para $\theta = 0.05$ (a) e $\theta = 0.01$ (b), de onde obtemos $\text{VaR}_{2277} = 0.01798$, para $\theta = 0.05$. Tendo em vista que $r_{2277} = 0.00540753$, o valor em risco em $t = 2278$ é

$$\begin{aligned} \text{VaR}_{2788} &= 0.0013 + 0.7884\text{VaR}_{2277} + 0.0739\max(r_{2277}, 0) + 0.5429\min(r_{2277}, 0) \\ &= 0.0013 + 0.7884 \times 0.01798 + 0.0739 \times 0.00541 = 0.0159. \end{aligned}$$

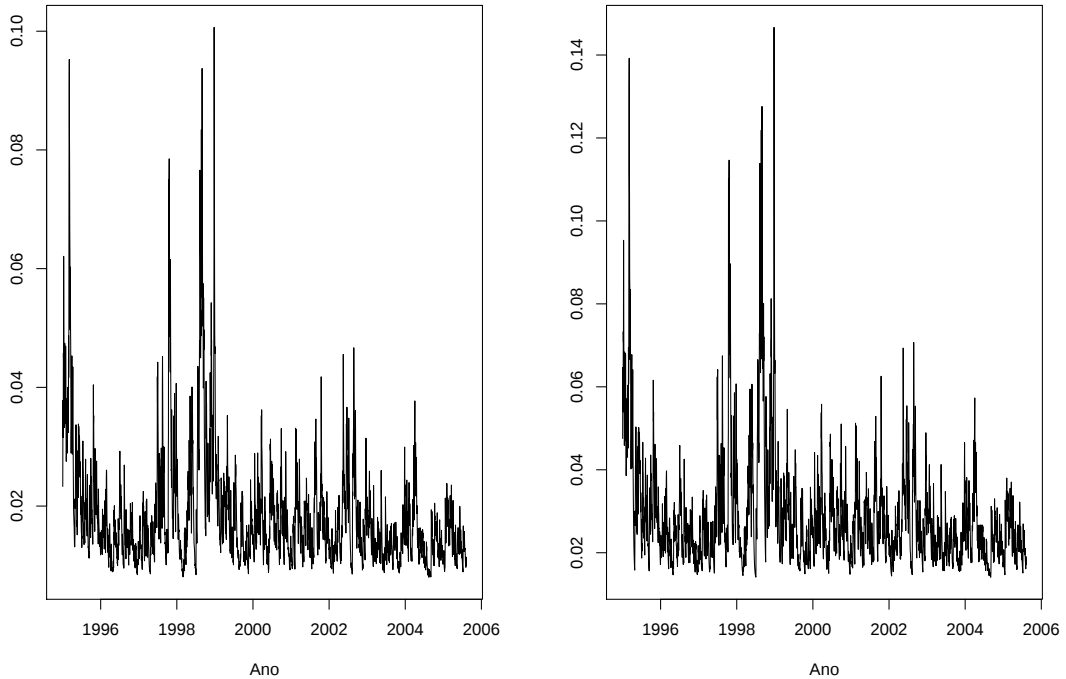


Figura 5.5. Seqüências **VaR** estimadas pela especificação **AS** para os log-retornos da Petrobrás para $\theta = 0.05$ (a) e $\theta = 0.01$ (b).

O Quadro 5.6 apresenta um resumo das estimativas VaR obtidas por cada uma das abordagens consideradas. Note que as estimativas variam de acordo com o método. As metodologias RiskMetrics e CAViaR estimaram o valor de VaR próximo a 1.6%. A estimativa VaR obtida através da estimação quantílica foi um pouco superior, cerca de 2%; enquanto as demais metodologias estimaram em aproximadamente 2.5% o valor de VaR 5%.

Quadro 5.6. Estimativas de VaR 5% para $t = 2278$, Petrobrás.

Abordagem		VaR ₂₂₇₈
Não-Condiciona	Quantílica	1.921%
	TVE	2.404%
Condicional	RiskMetrics	1.723%
	Econométrica	2.525%
	CAViaR	1.590%

5.2 O caso do IBOVESPA

A Figura 5.6 apresenta a evolução dos log-retornos do índice IBOVESPA entre 02/01/1995 e 16/09/2005, de onde é possível observar que em meados de 1998 houve um aumento de mais de 10% no índice IBOVESPA com relação ao período anterior.

O Quadro 5.7 apresenta algumas estatísticas descritivas da série completa e das séries denominadas amostra (r_t , $t \in [1, 2277]$) e pós-amostra (r_t , $t \in [2278, 2651]$). Note que até o terceiro quartil as séries são muito parecidas. Entretanto, os pontos mínimo e máximo de r_t são muito distantes dos observados na pós-amostra. As diferenças também são notadas na assimetria e na curtose das séries.

Quadro 5.7 Estatísticas descritivas da série de log-retornos da Petrobrás.

Estatística	r_t	$r_t, t \in [1, 2277]$	$r_t, t \in [2278, 2651]$
Mínimo	-0.0748	-0.0748	-0.0244
1º quartil	-0.0052	-0.0053	-0.0045
Média	0.0003	0.0003	0.0003
Mediana	0.0006	0.0006	0.0005
3º quartil	0.0061	0.0063	0.0051
Máximo	0.1252	0.1252	0.0224
Variância	0.0001	0.0001	0.00004
Assimetria	0.6075	0.6236	-0.1977
Curtose	13.558	12.9535	0.1848

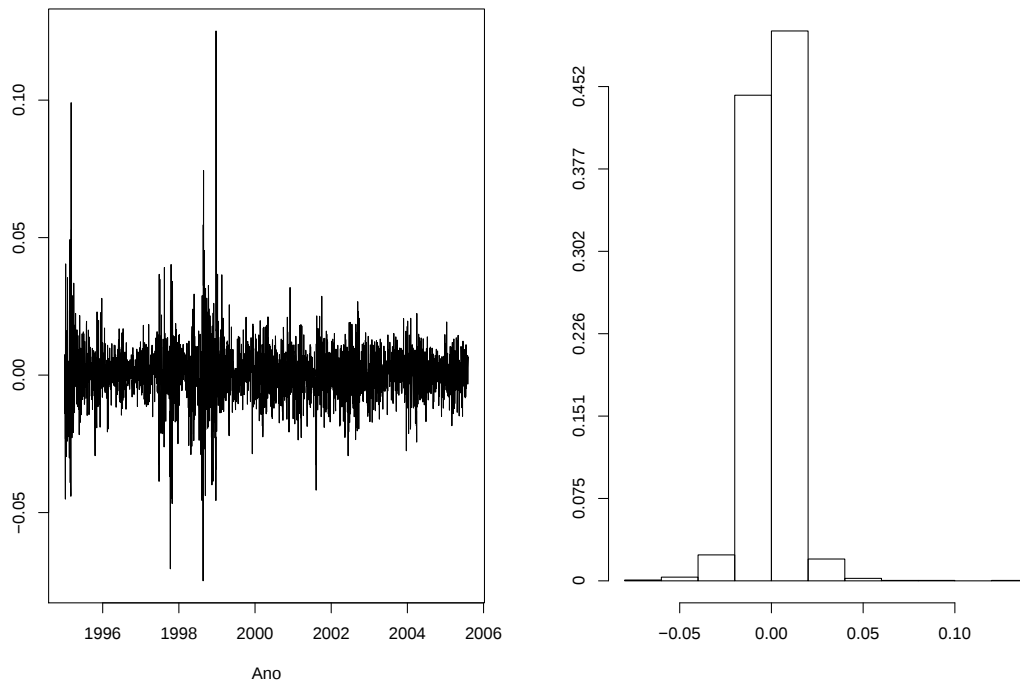


Figura 5.6. Série de log-retornos do IBOVESPA entre 03/01/1995 e 16/09/2005.

[A1] RiskMetrics

Assumindo novamente que a média condicional da série de log-retornos do IBOVESPA é igual a zero, ajustaremos o modelo em [2.2], ou seja,

$$r_t = a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha \sigma_{t-1}^2 + (1 - \alpha) r_{t-1}^2, \quad 0 < \alpha < 1.$$

Os resultados do ajuste estão apresentados no Quadro 5.8, de onde chegamos a

$$\sigma_t^2 = 0.8977 \sigma_{t-1}^2 + 0.1023 r_{t-1}^2,$$

cujo p -valor da estatística de Ljung-Box aplicada aos quadrados dos resíduos $Q(10)$ indica que o modelo não deve ser rejeitado.

Quadro 5.8. Resultados do ajuste RiskMetrics aos log-retornos do IBOVESPA.

	Estimativa	Erro Padrão	p -valor
α	0.8977	0.0048	0.0000
$Q(10)$	0.0023	$Q(20)$	0.0083

Assim, a previsão da volatilidade um passo adiante é dada por

$$\hat{\sigma}_{2277}^2(1) = 0.8977 \sigma_{2277}^2 + 0.1023 r_{2277}^2,$$

e do ajuste obtemos $\hat{\sigma}_{2277} = 0.000134$ e $r_{2277} = 0.0092$; logo, $\hat{\sigma}_{2278}^2(1) = 0.000109$. Assim, a perda máxima que o investidor pode sofrer é de 0.02692, ao nível de 1%, pois

$$\text{VaR} = 2.326 \times \sqrt{0.000109} = 0.02692.$$

A Figura 5.7 apresenta a volatilidade estimada, de onde se observa que os quatro grandes picos de volatilidade identificados para o caso da Petrobrás também se fazem presente no caso do IBOVESPA.

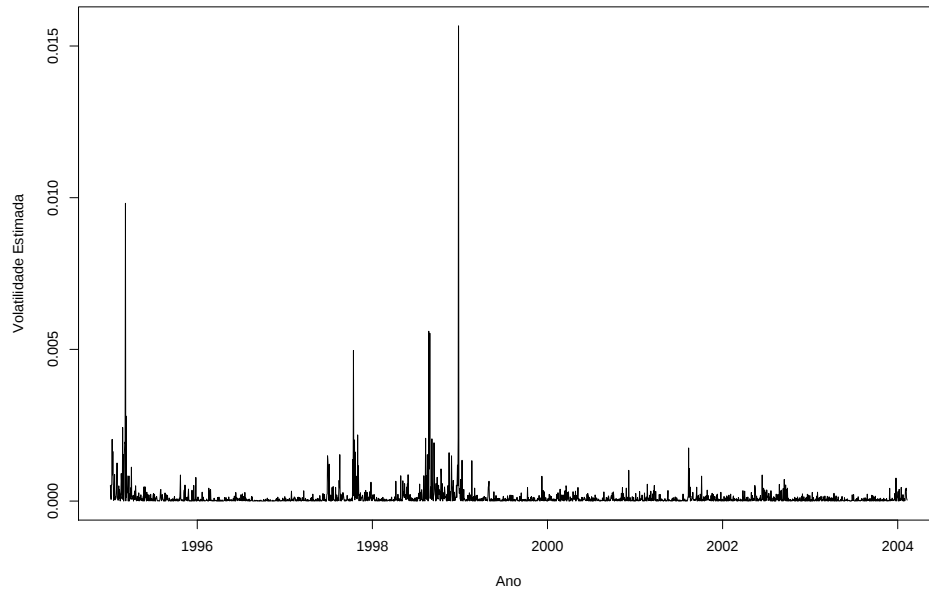


Figura 5.7. Volatilidade estimada da série de log-retornos do IBOVESPA através do RiskMetrics.

[A2] Abordagem econométrica

A minimização do critério BIC levou ao modelo MA(2)–GARCH(1,1), cujos resultados são apresentados no Quadro 5.9, de onde podemos concluir que o parâmetro θ_2 é não-significativo. Entretanto, como dissemos, testes t não são confiáveis em séries temporais, motivo pelo qual o parâmetro foi considerado. O teste de Ljung-Box não detectou autocorrelação em a_t (equação da média) e detectou correlação serial nos quadrados de a_t^2 , os resíduos do ajuste MA(2) (equação da variância), confirmando a validade do modelo.

Quadro 5.9. Resultados do ajuste ARCH–GARCH aos log-retornos da IBOVESPA.

Equação da média			
	Estimativa	Erro Padrão	p -valor
ϕ_0	0.0007	0.0002	0.0001
θ_1	0.0529	0.0232	0.0224
θ_2	−0.0303	0.0209	0.1478
$Q(10)$	0.0983	$Q(20)$	0.2297
Equação da variância			
	Estimativa	Erro Padrão	p -valor
α_0	0.0001	7.34×10^{-7}	0.0000
α_1	0.1515	0.0116	0.0000
β_1	0.8138	0.0157	0.0000
$Q(10)$	0.0360	$Q(20)$	0.0350

Assim, temos que

$$r_t = 0.0007 + a_t - 0.0529a_{t-1} + 0.0303a_{t-2},$$

$$\sigma_t^2 = 0.0001 + 0.1515a_{t-1}^2 + 0.8138\sigma_{t-1}^2.$$

Do ajuste, temos que $a_{2277} = 0.00849$ e $\sigma_{2277}^2 = 0.000119$; logo,

$$\hat{\sigma}_{2277}^2(1) = 0.0001 + 0.1515 \times (0.00849)^2 + 0.8138 \times 0.000119 = 0.00021$$

e, portanto,

$$\text{VaR} = 2.326 \times \sqrt{0.00021} = 2.326 \times 0.01449 = 0.0337.$$

Ou seja, o Valor em Risco é de cerca de 3.4% do montante investido.

[A3] Estimação quantílica

Para o cálculo do VaR 1% através da estimaco quantlica, se faz necessria a obteno do quantil de nvel 0.01 para a srie de log-retornos do IBOVESPA. Temos que $2277 \times 0.01 = 22.77$ e, portanto, consideramos $p_{22} = 0.009662$, $p_{23} = 0.010101$, $r_{(22)} = -0.030111$ e $r_{(23)} = -0.029655$.

De tal forma, que obtivemos

$$\text{VaR} = \left(\frac{0.0101 - 0.01}{0.0101 - 0.0097} \right) \times (-0.030111) + \left(\frac{0.01 - 0.0097}{0.0101 - 0.0097} \right) \times (-0.029655)$$

$$= -0.029759.$$

[A4] VaR pela TVE

Na abordagem da TVE, consideramos $n = 21$ e $g = 108$, assim temos que as $m = 9$ primeiras observaes no foram utilizadas na obteno dos mnimos. Os mnimos dos 108 grupos formados so apresentados na Figura 5.8.

O Quadro 5.10 apresenta as estimativas dos parmetros da distribuio GEV, de onde podemos concluir, assim como no caso dos log-retornos da Petrobrs, que a srie de mnimos segue distribuio de Frchet. Portanto, o VaR de 1% deve ser estimado por

$$\text{VaR} = -0.02139 - \left(\frac{0.01114}{-0.68151} \right) \times \{1 - [-21 \log(0.99)]^{-0.68151}\} = -0.05223,$$

ou seja, a perda mxima esperada  de aproximadamente 5.2% do montante investido.

Quadro 5.10. Estimativas dos parmetros da GEV para a srie de mnimos dos log-retornos da IBOVESPA.

Parmetro	α	β	ξ
Estimativa	0.01114	-0.02139	-0.68151
Erro Padro	2.0×10^{-6}	2.0×10^{-6}	0.00593

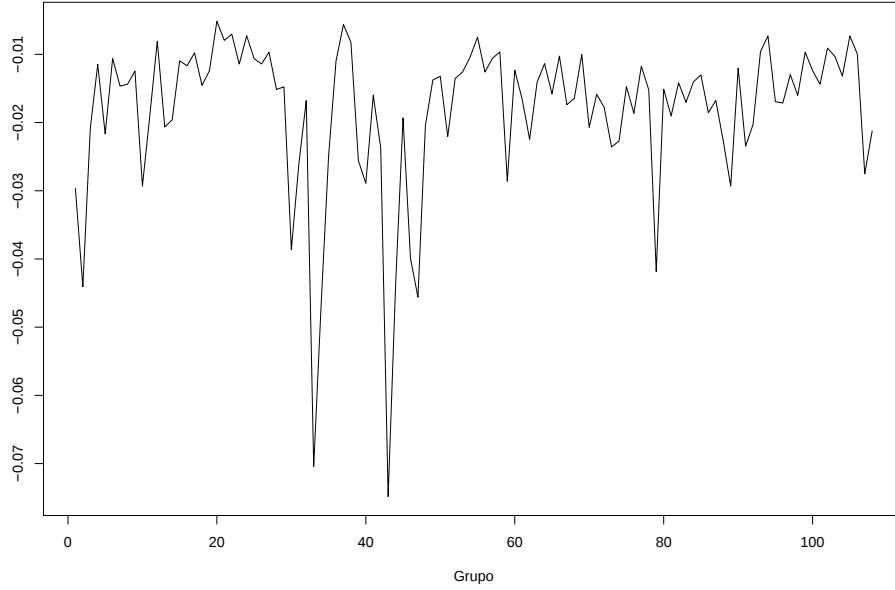


Figura 5.8. Série de mínimos do IBOVESPA em grupos de tamanho 21.

[A4] CAViaR

O Quadro 5.11 apresenta as estimativas dos parâmetros do modelo CAViaR, seus respectivos erros-padrão e p -valores associados, o valor do critério quantílico, o percentual de *hits* e o p -valor do teste de quantis dinâmicos para as observações amostrais e pós-amostrais, obtidas para os log-retornos do IBOVESPA.

No caso do índice IBOVESPA, o termo auto-regressivo (β_2) é sempre significativo, indicando presença de aglomerados de volatilidade também na cauda de r_t .

O percentual de *hits* encontrado nas observações amostrais é sempre próximo ao nível θ adotado para o cálculo de VaR, com destaque para a especificação AS com $\theta = 0.05$, onde $q = 5.01\%$. Contudo, quando consideramos as observações pós-amostrais, o cenário foi bastante diferente. Por exemplo, esperava-se valor próximo a 5% no caso da especificação SAV (com $\theta = 0.05$), mas esta só foi capaz de captar cerca de 1.33% de *hits*. Das quatro especificações a que apresentou melhor desempenho pós-amostral foi ADP, captando 4.53% e 1.33% para os casos em que $\theta = 0.05$ e $\theta = 0.01$, respectivamente. Esse fato é devido às diferenças existentes entre as observações amostrais e pós-amostrais. Se as séries fossem mais homogêneas nos dois períodos, esperar-se-ia um percentual de *hits* mais próximo do nível adotado.

Das quatro especificações apenas AS não foi rejeitada pelo teste de quantis dinâmicos, para as observações amostrais no caso em que $\theta = 0.05$. No caso em que $\theta = 0.01$, apenas a especificação IG não foi rejeitada pelo mesmo teste. Quando consideramos os períodos pós-amostrais, apenas SAV foi rejeitada para os dois níveis. Adicionalmente, a especificação ADP foi rejeitada para $\theta = 0.01$.

Quadro 5.11. Estimativas das especificações CAViaR para a séries de log-retornos da IBOVESPA.

Estimativas	$\theta = 0.05$				$\theta = 0.01$			
	SAV	AS	IG	ADP	SAV	AS	IG	ADP
β_1	0.00017	0.00086	0.00001	-0.00010	-0.00002	0.00254	0.00004	0.00003
Erro Padrão	0.00025	0.00028	0.00001	0.00001	0.00001	0.00160	0.00001	0.00001
p -valor	0.25501	0.00122	0.05225	0.00000	0.44191	0.05431	0.00042	0.00000
β_2	0.98947	0.88201	0.82677	—	0.99992	0.76731	0.75555	—
Erro Padrão	0.01676	0.02548	0.03874	—	0.00062	0.14432	0.02731	—
p -valor	0.00000	0.00000	0.00000	—	0.00000	0.00000	0.00000	—
β_3	0.33846	-0.05013	0.35071	—	0.15733	0.04871	1.16273	—
Erro Padrão	0.00362	0.04339	0.12411	—	0.00601	0.14600	0.79236	—
p -valor	0.00000	0.12395	0.00235	—	0.00000	0.36932	0.07111	—
β_4	—	0.34064	—	—	—	0.80508	—	—
Erro Padrão	—	0.05465	—	—	—	0.61912	—	—
p -valor	—	0.00000	—	—	—	0.94641	—	—
CQ	2.90130	2.47259	2.61591	2.88840	0.91365	0.69831	0.77542	0.91398
q amostra (%)	5.31634	5.00879	5.09666	5.05272	1.01050	1.01050	0.96666	1.01050
q pós-amostra (%)	1.33333	3.73333	3.73333	4.53333	0.80000	0.00000	0.53333	1.33333
DQ amostra	0.00000	0.99342	0.00649	0.00000	0.00402	0.00546	0.53754	0.00000
DQ pós-amostra	0.06159	0.61904	0.60718	0.25172	0.00000	1.00000	0.94053	0.00320

Assim, chegamos a dois modelos: para $\theta = 0.05$, a especificação mais apropriada é AS, enquanto a especificação IG se mostrou mais eficaz para $\theta = 0.01$. Esse fato sugere que os processos geradores dos níveis quantílicos 5% e 1% são distintos, enfatizando a necessidade de se utilizar a metodologia proposta por Engle & Manganelli (2004), visto que as demais técnicas existentes consideram que o processo gerador é único.

O modelo utilizado para o cálculo de VaR de 1%, no caso do IBOVESPA, é dado por

$$\text{VaR}_t = \sqrt{0.00004 + 0.7556\text{VaR}_{t-1}^2 + 1.1627r_{t-1}^2},$$

onde $\text{VaR}_{2277} = 0.01639$ e $r_{2277} = 0.0066$; conseqüentemente, $\text{VaR}_{2278} = 0.0178$. Vale destacar que o acréscimo de uma unidade em r_{t-1}^2 proporciona um aumento de $\sqrt{1.1627} = 1.0783$ em VaR_t .

A Figura 5.5 apresenta as seqüências VaR estimadas para $\theta = 0.05$ (a) e $\theta = 0.01$ (b). Note que os picos de volatilidade identificados através da metodologia RiskMetrics têm impacto no valor de VaR. Ou seja, a metodologia CAViaR conseguiu captar os períodos onde a volatilidade se apresentou mais elevada, elevando também o Valor em Risco.

O Quadro 5.12 apresenta as estimativas VaR de 1% para a série de log-retornos do IBOVESPA de onde é possível concluir que as técnicas anteriormente existentes conduziram a estimativas de VaR bastante superiores quando comparadas com a obtida pela metodologia CAViaR. A estimativa de VaR

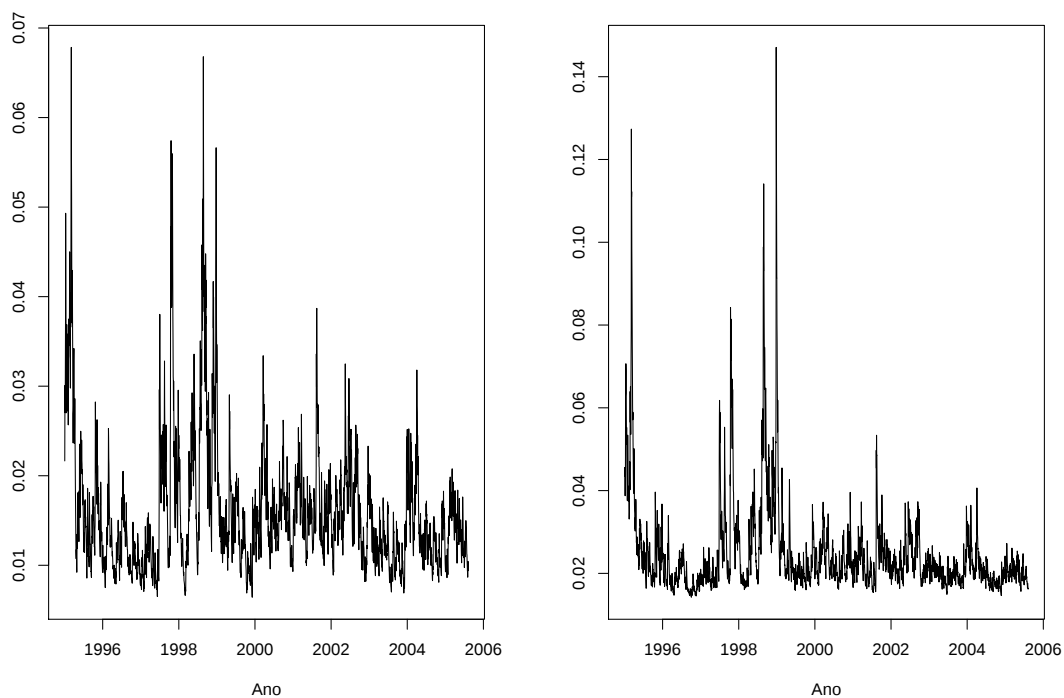


Figura 5.9. Sequências VaR estimadas pelas especificações AS (a) e IG (b) para os log-retornos da IBOVESPA para $\theta = 0.05$ e $\theta = 0.01$.

fornecida pelo modelo CAViaR, mais especificamente pela especificação IG, aproxima-se mais da realidade tendo em vista que foi identificada diferença no processo gerador entre os diferentes níveis considerados, enquanto que as demais técnicas não levam em conta essa mudança.

Quadro 5.12. Estimativas de VaR 1% para $t = 2278$, IBOVESPA.

Abordagem		VaR ₂₂₇₈
Não-Condiciona	Quantílica	2.976%
	TVE	5.223%
Condicional	RiskMetrics	2.692%
	Econométrica	3.370%
	CAViaR	1.780%

Conclusões

O processo de estimação dos modelos CAViaR se comporta de maneira bastante peculiar. Em geral, faz-se necessária a utilização de um número elevado de vetores de parâmetros iniciais até que seja obtido um critério quantílico satisfatório. Em nossas simulações, foi possível constatar que, mesmo após obtida a convergência, as estimativas de CQ podem estar muito distantes.

Outro fato grave é que, além da existência dessa grande distância entre os CQ 's mínimo e máximo alcançados no processo de estimação, as suas estimativas podem estar concentradas nos valores mais elevados de CQ , enquanto uma minoria de estimativas pode estar próxima do mínimo obtido. Esse fato corrobora a necessidade de se considerar um grande número de vetores iniciais.

Foi possível constatar também que as estimativas dos parâmetros podem apresentar grande variação, inclusive com mudança de sinal, fato que faz com que as interpretações dos parâmetros sejam completamente diferentes.

Em todos os cenários considerados as seqüências VaR obtidas foram capazes de detectar as perdas ocorridas em r_t , mas as estimativas de VaR obtidas para essas observações extremas, na maioria da vezes, podem variar bastante. Vale destacar que apenas a especificação ADP não se mostrou eficaz nessa detecção.

Outro ponto que merece destaque é o fato que para uma determinada especificação, considerando o mesmo θ , as estimativas VaR podem variar bastante. Vimos que, em geral, quando calculamos as diferenças entre as seqüências VaR de mínimo e máximo CQ , as diferenças oscilam em torno de zero. Contudo, em alguns casos, vimos que a oscilação se dá um pouco acima do zero, ou seja, quando consideramos a seqüência de CQ máximo estamos subestimando o valor em risco. Esse fato revela a importância de um processo de estimação bem sucedido, tendo em vista que a alocação de recursos pode ser feita de maneira errônea no caso de se obter uma seqüência VaR para um β não-ótimo.

Através da metodologia CAViaR foi possível constatar que os aglomerados de volatilidade também ocorrem nas caudas de distribuição dos log-retornos,

tendo em vista que o parâmetro auto-regressivo se mostrou significativo em todas as especificações.

Das diferentes metodologias para o cálculo de **VaR** de 5% utilizadas, obtivemos estimativas distintas para os dados da Petrobrás. As abordagens econométrica e da teoria do valor extremo estimaram **VaR** em cerca de 2.5%, enquanto nas outras abordagens esse valor foi próximo a 1.6%. No caso do IBOVESPA, onde estimamos **VaR** de 1%, as estimativas diferiram bastante. Nesse caso, a técnica baseada na teoria dos valores extremos estimou **VaR** pouco superior a 5.2%, enquanto a metodologia proposta por Engle & Manganelli (2004) estimou **VaR** em torno de 1.8%.

Além disso, chegamos a dois modelos distintos para o IBOVESPA, um para cada nível θ considerado. Esse fato constitui indício de que a metodologia **CAViaR** é mais eficaz que as demais técnicas consideradas, visto que estas assumem, implicitamente, que o processo gerador é o mesmo em toda a série.

De forma geral, nós sugerimos a utilização dos modelos **CAViaR**, contudo alertamos para que seja considerado um grande número de vetores iniciais no processo de estimação, a fim de minimizar as chances de se obter uma sequência **VaR** que esteja distante da realidade.

Referências

- [1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control*, **19**, 716–723.
- [2] Ansley, C. F. & Newbold, P. (1980). Finite sample properties of estimators of autoregressive moving average models. *Journal of Econometrics*, **13**, 159–183.
- [3] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307–327.
- [4] Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
- [5] Box, G. E. P. & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association*, **65**, 1509–1526.
- [6] Choi, B. S. (1992). *ARMA Model Identification*. New York: Springer-Verlag.
- [7] Cox, D.R. & Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- [8] Duffie, D. & Pan, J. (1997). An overview of value at risk. *Journal of Derivatives*, Spring, 7–48.
- [9] Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer-Verlag.
- [10] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflations. *Econometrica*, **50**, 987–1007.
- [11] Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, **17**, 425–446.
- [12] Engle, R. F. & Manganelli, S. (1999). CAViaR: Conditional Value at Risk by Quantile Regression. *NBER*, Working Paper 7341.

- [13] Engle, R. F. & Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by regression quantiles. *Journal of Business and Economic Statistics*, **22**(4), 367–381.
- [14] Engle, R. F. & Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, **48**, 1749–1778.
- [15] Fisher, R. A. & Tippett, L. H. C. (1922). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.
- [16] Geweke, J. & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221–238.
- [17] Granger, C. W. J. & Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15–29.
- [18] Hjorth, U. (1987). On model selection in the computer age. *Technical Report*. n. LITH-MAT-R-87-08, Linköping University, Suecia.
- [19] Hurvich, C. M. & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, **76**, 297–307.
- [20] Jenkinson, A. F. (1955). The frequency distribution of annual maximum (or minimum) of meteorological events. *Quarterly Journal of the Royal Meteorological Society*, **81**, 158–171.
- [21] Jorion, P. (1997). *Value at Risk: The New Benchmark for Controlling Market Risk*. Chicago: McGraw-Hill.
- [22] Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33–50.
- [23] Longestae, J. & More, L. (1995). *Introduction to RiskMetrics™*, 4^a edição. New York: Morgan Guaranty Trust Company.
- [24] Ljung, G. M. & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **66**, 265–270.
- [25] McLeod, A. I. (1978). On the distribution and application of residual autocorrelations in Box-Jenkins models. *Journal of the Royal Statistical Society B*, **40**, 296–302.
- [26] Mendes, B. V. M. (2004). *Análise de Eventos Extremos*. Rio de Janeiro: Editora E-Papers.
- [27] Mills, T. C. (1999). *The Econometric Modelling of Financial Time Series*. Cambridge: Cambridge University Press.
- [28] Mood, A. (1940). The distribution theory of runs. *Annals of Mathematical Statistics*, **11**, 367–392.
- [29] Morettin, P. A. (2005). *Econometria Financeira: Um Curso em Séries Temporais Financeiras*. São Paulo: Universidade de São Paulo. [disponível para download em <http://www.ime.usp.br/~pam/texto.pdf>]
- [30] Morettin, P. A. & Toloi, C. M. C. (2004). *Análise de Séries Temporais*. São Paulo: Blücher.

- [31] Reisen, V. A. (1994). Estimation of the fractional parameter for the ARIMA(p,d,q) model using the smoothed periodogram. *Journal of Time Series Analysis*, **15**, 335–350.
- [32] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- [33] Schwert, G. W. (1988). Why does stock market volatility change over time?. *Journal of Finance*, **44**, 1115–1153.
- [34] Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester, UK: Wiley.
- [35] Tsay, R. S. (2002). *Introduction of Financial Time Series*. New York: Wiley.
- [36] Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*, 4th ed. New York: Springer-Verlag.
- [37] Weiss, A. (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory*, **7**, 46–68.
- [38] White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.