

"A FEA e a USP respeitam os direitos autorais deste trabalho. Nós acreditamos que a melhor proteção contra o uso ilegítimo deste texto é a publicação online. Além de preservar o conteúdo motiva-nos oferecer à sociedade o conhecimento produzido no âmbito da universidade pública e dar publicidade ao esforço do pesquisador. Entretanto, caso não seja do interesse do autor manter o documento online, pedimos compreensão em relação à iniciativa e o contato pelo e-mail bibfea@usp.br para que possamos tomar as providências cabíveis (remoção da tese ou dissertação da BDTD)."

ANTÓNIO MARCOS DE OLIVEIRA COSTA

**CÁLCULO DO VALOR EM RISCO DE CARTEIRAS COM GRANDES POSIÇÕES:
UMA ABORDAGEM BASEADA NA TEORIA DOS VALORES EXTREMOS**

MESTRADO PROFISSIONALIZANTE EM MODELAGEM MATEMÁTICA EM FINANÇAS

USP / SÃO PAULO

2007

ANTÓNIO MARCOS DE OLIVEIRA COSTA

**CÁLCULO DO VALOR EM RISCO DE CARTEIRAS COM GRANDES POSIÇÕES:
UMA ABORDAGEM BASEADA NA TEORIA DOS VALORES EXTREMOS**

Dissertação apresentada à Banca Examinadora da Universidade de São Paulo, como exigência parcial para obtenção do título de MESTRE em Modelagem Matemática em Finanças, sob orientação do Prof. Dr. Henrique Von Dreifus.

USP / SÃO PAULO

2007

Banca Examinadora

Ao meu pai, António.

Agradecimentos

Ao concluir este trabalho, percebi que havia uma série de pessoas que haviam colaborado para o seu desenvolvimento.

Em primeiro lugar, gostaria de agradecer à minha esposa, Lucy, pela sua compreensão em relação às várias horas que precisei me dedicar aos estudos logo no início do nosso casamento.

Gostaria também de agradecer ao meu orientador, professor Dreifus, pelo constante apoio ao meu desenvolvimento acadêmico.

Não poderia esquecer dos meus professores do mestrado. A eles agradeço pelas inúmeras discussões que tivemos, normalmente após o horário das aulas. Todas elas muito me auxiliaram a desenvolver o meu entendimento a respeito dos diversos temas abordados durante o curso. Gostaria de agradecer em especial aos professores Pedro Paulo Schirmer, Adilson Simonis e Edson de Faria pela paciência, durante e após o final das aulas, em entrar em discussões que muito me auxiliaram a entender melhor diversos temas.

A todos os meus colegas do mestrado agradeço pela companhia e pela ajuda sempre que precisei.

Por fim, gostaria de agradecer à Bolsa de Mercadorias & Futuros, BM&F, onde tive a oportunidade de participar da discussão de diversas questões de interesse prático para o mercado, as quais eram tratadas com base em sólidos fundamentos acadêmicos e serviam de estímulo para que eu continuasse me desenvolvendo nos meus estudos.

Resumo

Em geral, o Valor em Risco – ou simplesmente VaR – de uma carteira é estimado para um horizonte de um dia e os preços utilizados nesta estimativa correspondem ao nível do mercado após decorrido tal prazo. Por outro lado, no caso do tamanho da carteira ser grande em relação aos volumes usualmente negociados no mercado, dois efeitos podem ocorrer dependendo da estratégia adotada ao se tentar zerá-la: (i) no caso de se procurar liquidá-la em um curto espaço de tempo, desloca-se os preços do mercado de modo a equilibrar a demanda com a oferta adicional que se gerou, ou (ii) alonga-se o tempo total para realizar tal liquidação, executando-a de forma paulatina. No primeiro caso, ou os ativos serão comprados a preços mais altos ou vendidos a preços mais baixos que aqueles normalmente utilizados no cálculo do VaR. No segundo caso, o horizonte de tempo utilizado para se zerar a carteira é maior que aquele utilizado no cálculo usual do VaR. Neste trabalho, utilizaremos uma abordagem de valores extremos para estimar a perda potencial total a que uma carteira está exposta, ou seja, a perda potencial considerando-se que ela é mantida durante um determinado horizonte de tempo – de um dia, por exemplo – e que, a partir deste momento, realiza-se a sua liquidação de acordo com a liquidez do mercado, encerrando-se apenas uma fração da carteira a cada dia. Tal perda potencial será estimada com um nível de confiança elevado – 99.9%, por exemplo – o qual corresponde a eventos de crise.

Palavras-chave: valor-em-risco, teoria dos valores extremos, liquidez.

Abstract

Usually, a portfolio's Value at Risk (VaR) is estimated over a target horizon of one day and the prices used in this estimation are those which can be seen in the financial market after this period of time. On the other hand, if the size of the analyzed portfolio is large compared to the usual traded volumes, two effects can happen when we try to unwind the positions: (i) if we try to unwind them during a very short period of time, supply-demand curves are shifted so that the supply and the demand reach a new equilibrium, or (ii) we do it gradually and over a longer period of time. In the first case, the assets will be bought at higher prices or sold at lower prices than those they are usually traded at in the market. In the second case, the time spent to unwind the positions is longer than the one usually adopted to calculate the portfolio's VaR. This paper uses an extreme value approach in order to estimate the total potential loss which a portfolio is exposed to, that is, the potential loss considering that it is kept for a specific period of time – of one day, for example – and, after that, its positions are unwind according to the market's liquidity. Hence, only a portfolio's fraction is unwind at a day. Such potential loss is estimated at a high confidence level – of 99.9%, for instance – which corresponds to a crisis event.

Keywords: value-at-risk, extreme value theory, liquidity.

1. Introdução	9
2. Descrição do Modelo	10
2.1. Motivações do Modelo de Chavez-Demoulin, Davison e McNeil (CDM)	10
2.2. Revisão da Análise de Peaks-over-Threshold (POT)	10
2.3. A Teoria Básica de Processos de Ponto	13
2.4. O Modelo de Chavez-Demoulin, Davison e McNeil (CDM)	17
2.5. Inclusão do Custo da Liquidação Fracionada de Carteiras no Modelo CDM (VaRF)	19
2.6. Cálculo do VaRF	24
3. Aplicação do Modelo às Ações Negociadas na BOVESPA	27
3.1. Descrição das Estratégias de Liquidação	27
3.2. Calibragem do Modelo e Análise dos Resultados	28
4. Considerações Finais	35
Apêndice A – Propriedades Assintóticas dos Estimadores de Máxima Verossimilhança (EMV)	37
Referências	39

1. Introdução

Segundo Jorion(2001), o Valor em Risco – ou simplesmente VaR – de uma carteira é a máxima perda que uma ela pode sofrer dentro de um dado horizonte de tempo e com um dado nível de confiança. Muito trabalho tem sido desenvolvido na busca do aprimoramento da estimativa do VaR em decorrência da sucessão de crises as quais o mercado financeiro enfrentou, principalmente a partir de meados de 1997 com a crise da Ásia. Para níveis de confiança elevados – os quais correspondem a perdas com baixa probabilidade de ocorrência porém da magnitude daquelas ocorridas em tais crises – a Teoria dos Valores Extremos (TVE) tem tido especial atenção por possibilitar que se modele apenas a distribuição das maiores perdas potenciais de uma carteira, o que viabiliza uma maior precisão na estimativa do seu VaR em relação a técnicas que buscam modelar a distribuição completa dos resultados da carteira.

Neste trabalho, buscaremos estimar o Valor em Risco com Liquidação Fracionada – ou simplesmente VaRF – de uma carteira, o qual definiremos como a máxima perda da carteira, com um dado nível de confiança, da qual se inicia a liquidação após um determinado período de tempo, sendo esta realizada de forma fracionada – ou seja, pouco a pouco de forma a não deslocar significativamente as curvas de oferta e demanda do mercado – ao longo de um ou mais dias. Tal perda potencial será estimada com um nível de confiança elevado – 99.9%, por exemplo – o qual corresponde a eventos de crise. A importância da inclusão da liquidação fracionada, durante a liquidação da carteira, no cálculo do VaR é de grande relevância para dimensionarmos a real magnitude da perda potencial a qual ela está exposta. Tomemos, por exemplo, uma carteira que contenha exclusivamente 10 lotes de um único contrato, do qual usualmente são negociados 10.000 lotes por dia com cerca de 2.000 negócios de tamanho (número de contratos em cada negócio) aproximadamente iguais. Trata-se então de uma posição pequena em relação ao tamanho do mercado, a qual provavelmente poderá ser liquidada em um curto período de tempo. Se, por outro lado, a carteira contivesse, ao invés dos 10 lotes, 20.000 lotes do contrato em questão, provavelmente seriam necessários alguns dias para que ela fosse completamente liquidada sem que a influência nos níveis de preço do mercado fosse substancial. Assim, a carteira com 20.000 lotes estará exposta às oscilações do mercado durante um período bem superior que a de 10 lotes. Este será o efeito que estimaremos neste trabalho. Para tanto, faremos um pequeno ajuste na formulação feita por Chavez-Demoulin, Davison e McNeil (2005) para a utilização de um processo de ponto com a Teoria dos Valores Extremos (TVE) na estimação do Valor em Risco (VaR). Vale destacar que, caso se deseje liquidar uma carteira que possua uma posição A que serve de *hedge* (proteção) para outra posição B, sendo a posição A muito mais líquida que a B, apesar de ser possível se liquidar rapidamente a posição A, pode-se desejar fazê-lo de modo gradual de modo a não deixar posição B desprotegida.

As próximas seções apresentam o modelo utilizado neste trabalho e em seguida aplicam os resultados a carteiras compostas de ações negociadas na Bolsa de Valores de São Paulo – BOVESPA. Vale destacar que a escolha de carteiras contendo apenas ações foi arbitrária, pois a metodologia se aplica a carteiras mais complexas que podem, por exemplo, conter tanto posições em renda fixa quanto em renda variável.

2. Descrição do Modelo

Esta seção apresenta a adaptação do modelo desenvolvido por Chavez-Demoulin, Davison e McNeil (CDM) de modo a contemplar o custo da liquidação fracionada de carteiras no cálculo do Valor em Risco (VaR), gerando a medida de risco a qual denominaremos VaRF.

Apresenta também o ferramental necessário para o correto entendimento do modelo. Já os conceitos utilizados para a estimação dos seus parâmetros com base na maximização da logverossimilhança constam do apêndice A.

2.1. *Motivações do Modelo de Chavez-Demoulin, Davison e McNeil (CDM)*

Apesar da existência de uma vasta literatura que trata da estimação do VaR, Chavez-Demoulin, Davison e McNeil (2005) buscaram um modelo que possuísse as seguintes características:

- Pudesse ser usado na estimativa do VaR para altos níveis de confiança; e
- Identificasse períodos com agrupamentos de volatilidade.

Assim, observaram que o trabalho de McNeil e Frey (2000) possuía estas duas características. Tratava-se de um estudo na qual o VaR era estimado condicionalmente através de um processo de dois estágios em que modelos GARCH eram utilizados na previsão da volatilidade e técnicas de TVE eram aplicadas aos resíduos da análise GARCH. Por outro lado, perceberam que os resultados da TVE obtidos através deste modelo eram muito influenciados pelo ajuste do GARCH a todo o conjunto de dados, sendo que tal ajuste era realizado no primeiro estágio do processo.

Diante desta questão, Chavez-Demoulin, Davison e McNeil observaram que um dos princípios das estatísticas de valores extremos é que se deve permitir aos dados mais extremos “falarem por si mesmos”. Assim, eles partiram para o desenvolvimento de um modelo que possuísse o verdadeiro espírito da TVE, ou seja, onde esta teoria fosse aplicada diretamente aos dados e sem a necessidade do ajuste de um modelo de séries temporais intermediário. Com este objetivo, eles adaptaram um modelo clássico da TVE: o POT¹. É importante observar que o POT descreve o aparecimento de extremos em dados independentes e identicamente distribuídos (iid), não sendo diretamente aplicável quando há a ocorrência de agrupamentos de volatilidade.

2.2. *Revisão da Análise de Peaks-over-Threshold (POT)*

Esta seção apresenta uma revisão sobre os conceitos do POT que serão aplicados neste trabalho.

¹ A sigla POT vem do inglês *Peaks-over-Threshold*, que significa picos acima de um limiar.

Seja Z_1, \dots, Z_l uma seqüência de variáveis aleatórias independentes e identicamente distribuídas (iid) de uma distribuição contínua $F_Z(z)$ com limite superior z_+ , o qual pode ser infinito. Seja N_u a variável aleatória correspondente à quantidade desses Z_i 's que excede a um alto limiar u . Assuma, ainda, que tais excessos ocorrem em instantes de tempo discretos T_i e que possuam um valor (tamanho do excesso) $W_j = Z_j - u$. Conforme apresentado por Leadbetter (1991), quando $l \rightarrow \infty$ e $u \rightarrow z_+$ de tal forma que $l\{1 - F_Z(u)\} \rightarrow \lambda > 0$ pode-se garantir que os tempos re-escalados destes eventos possam ser modelados como processos homogêneos de Poisson com parâmetro λ . Condicionando $N_u = n$, os tamanhos dos excessos W_1, \dots, W_n constituem uma amostra iid da Distribuição Generalizada de Pareto (GPD²), da qual a função distribuição acumulada (FDA) apresentamos a seguir:

$$G_{\xi, \sigma}(w) = \begin{cases} 1 - \left(1 + \xi \frac{w}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{w}{\sigma}\right), & \xi = 0 \end{cases} \quad (1)$$

Os parâmetros da GPD possuem as seguintes restrições:

- (i) $\sigma > 0$;
- (ii) $w \geq 0$, se $\xi \geq 0$; e
- (iii) $0 \leq w \leq -\sigma/\xi$, se $\xi < 0$.

Como podemos observar nos trabalhos de Pickands (1975), Davinson e Smith (1990), Embrechts, Klüppelberg e Mikosch (1997), o fato de sob tais hipóteses obtermos uma GPD é uma consequência direta de resultados-limite da TVE.

Enquanto o parâmetro σ serve apenas para ajustar a escala, o valor do parâmetro ξ , conhecido como índice de valor extremo (EVI³) identifica claramente as características da GPD:

- (i) $\xi > 0$

Este caso corresponde às distribuições de caudas pesadas como a de Pareto, a t-Student, Cauchy, Burr, LogGama, e de Fréchet.

- (ii) $\xi = 0$

Aqui se enquadram as distribuições das quais as caudas decaem exponencialmente como a Normal, Exponencial, Gama e LogNormal.

² A sigla GDP refere-se, em inglês, a *Generalized Pareto Distribution*.

³ Do inglês *Extreme Value Index*.

(iii) $\xi < 0$

Nesta categoria encontram-se as distribuições de caudas curtas com um valor máximo – e portanto finito – na cauda direita como a Uniforme e a Beta.

A escolha do limiar u também influencia nas estimativas de σ e ξ . Em linhas gerais, u deve ser suficientemente alto para tornar razoavelmente estáveis as estimativas de ξ e $\sigma^* = \sigma - u\xi$. Por outro lado, a escolha de um valor muito elevado para u faz com que haja poucos pontos na amostra Z_1, \dots, Z_l que excedam tal limiar, o que prejudica a qualidade das estimativas de σ e ξ . Assim, o valor de u mais adequado é aquele que melhor pondera tais características, proporcionando um melhor equilíbrio entre o viés e a variância dos estimadores de σ e ξ . A escolha ótima de u não será contemplada no presente estudo, estando maiores detalhes acerca deste assunto disponíveis nos trabalhos de Coles (2001) e Matthys e Beirlant (2003).

A partir da equação (1), podemos obter a função densidade de probabilidade correspondente:

$$g_{\xi, \sigma}(w) = \begin{cases} \frac{1}{\sigma} \left(1 + \xi \frac{w}{\sigma}\right)^{-\left(1 + \frac{1}{\xi}\right)}, & \xi \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{w}{\sigma}\right) & \xi = 0 \end{cases} \quad (2)$$

Vale destacar que, tanto na equação (1) quanto na (2), o caso em que $\xi = 0$ pode ser obtido a partir do caso $\xi \neq 0$ tomando-se o limite quando ξ tende a zero ($\xi \rightarrow 0$). Assim, a partir deste ponto, não mais consideraremos o caso particular em que $\xi = 0$. Consequentemente, partindo da equação (2), podemos obter a logverossimilhança para a estimativa dos parâmetros σ e ξ :

$$\ell(\sigma, \xi) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^n \log \left(1 + \xi \frac{w_j}{\sigma}\right) \quad (3)$$

Os estimadores de máxima verossimilhança de σ e ξ são não-regulares pois, como observaram Davison (1984a), Davison (1984b) e Smith (1985), a estatística de classificação não é assintoticamente normal se $\xi \leq -1/2$. Para $\xi \leq -1$ a estatística possui média infinita, de tal forma que as expansões de Taylor que usualmente são adotadas não fornecem estimadores consistentes. Por outro lado, nas aplicações de finanças é usual que as distribuições dos dados apresentem caudas pesadas ($\xi > 0$), situação na qual os estimadores de máxima verossimilhança possuem as suas propriedades usuais de consistência e eficiência.

Conforme indicado por Hosking e Wallis (1987), os estimadores $\hat{\sigma}$ e $\hat{\xi}$ de máxima verossimilhança σ e ξ podem ser obtidos numericamente utilizando o algoritmo de Newton-Raphson.

Retornemos, agora, à distribuição do número de excessos N_u , dos pontos da amostra Z_1, \dots, Z_I , em relação ao limiar u . De acordo com o exposto no início desta seção, podemos assumi-la com sendo de Poisson com parâmetro λ . Assim, segundo Ross (1987), temos:

$$P(N_u = n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad (4)$$

Assumindo que a quantidade e o tamanho dos excessos W_1, \dots, W_n sejam independentes, e lembrando que $g_{\xi, \sigma}(w)$ é a função densidade de probabilidade da GPD, a logverossimilhança global será dada por:

$$\ell(\lambda, \sigma, \xi) = \log \left[P(N_u = n) \prod_{j=1}^n g_{\xi, \sigma}(w_j) \right] \quad (5)$$

Substituindo (2) e (4) em (5), obtemos a expressão para a logverossimilhança global:

$$\ell(\lambda, \sigma, \xi) = -n \log \lambda - \lambda - \log n! - n \log \sigma - \left(1 + \frac{1}{\xi} \right) \sum_{j=1}^n \log \left(1 + \xi \frac{w_j}{\sigma} \right) \quad (6)$$

2.3. A Teoria Básica de Processos de Ponto

Esta seção apresenta a teoria básica sobre processos de ponto a qual será utilizada no modelo de Chavez-Demoulin, Davison e McNeil (CDM) o qual será apresentado na seção a seguir. Ao leitor interessado em se aprofundar no assunto, sugerimos a leitura de Coles (2001) e de Cox e Isham (1979).

Um processo de ponto sobre um conjunto Ω é uma regra estocástica para a ocorrência e posição de eventos pontuais. Em um contexto de modelagem, com Ω representando um período de tempo, um processo de ponto pode ser usado por exemplo para descrever a ocorrência de tempestades, terremotos ou, no caso da TVE, o instante em que ocorrem excessos acima de um determinado limiar u . A partir do modelo, a probabilidade de ocorrência de eventos (tempestades, terremotos ou excessos acima de um determinado limiar u) pode ser calculada. Pode-se também calcular, dada a ocorrência de um evento, o tempo médio de espera até a ocorrência do próximo. Por exemplo, um processo de ponto bidimensional pode ser utilizado para descrever a posição de fraturas em uma placa de vidro.

Uma maneira de caracterizar as propriedades estatísticas de um processo de ponto é definir um conjunto $N(A)$ de variáveis aleatórias com valores inteiros e não negativos, para cada $A \subset \Omega$, tal que $N(A)$ é o número de pontos no conjunto A . Especificar de uma maneira consistente a distribuição probabilística de cada $N(A)$ determina as características do processo de ponto, o qual chamaremos de N . Assim, podemos definir:

$$\Lambda(A) = E[N(A)] \quad (7)$$

$\Lambda(A)$ fornece o número de pontos esperado em qualquer subconjunto A de Ω e é denominado medida de intensidade do processo. Assumindo $A = [a_1, x_1] \times \dots \times [a_k, x_k] \subset \mathfrak{R}^k$, e dado que ela exista, podemos definir a função densidade (ou função intensidade) do processo:

$$\lambda(x) = \frac{\partial^k \Lambda(A)}{\partial x_1 \dots \partial x_k} \quad (8)$$

O processo de ponto canônico é o processo homogêneo de Poisson unidimensional. Com parâmetro $\lambda > 0$, este processo é um processo em $A \subset \mathfrak{R}$ que satisfaz:

- (i) Para todo $A = [t_1, t_2] \subset \Omega$, $N(A) \sim Poi(\lambda(t_2 - t_1))$ – distribuição de Poisson com parâmetro $\lambda(t_2 - t_1)$; e
- (ii) Para todos os sub-conjuntos A e B tais que $A \cap B = \emptyset$, $N(A)$ e $N(B)$ são variáveis aleatórias independentes.

Neste caso, teremos a partir de (7) que:

$$\Lambda(A) = E[N(A)] = \lambda(t_2 - t_1) \quad (9)$$

e, a partir de (8) e (9), obtemos:

$$\lambda(t) = \lambda \quad (10)$$

Se considerarmos agora um processo não-homogêneo de Poisson unidimensional, a condição (ii) permaneceria inalterada, porém a condição (i) deveria ser substituída por:

- (i') Para todo $A = [t_1, t_2] \subset \Omega$, $N(A) \sim Poi(\Lambda(A))$ – distribuição de Poisson com parâmetro $\Lambda(A)$, onde

$$\Lambda(A) = \int_{t_1}^{t_2} \lambda(t) dt \quad (11)$$

Vale destacar que, enquanto o processo homogêneo de Poisson gera pontos distribuídos de forma homogênea ao longo do tempo, pode-se possibilitar uma maior concentração de pontos em determinados instantes de tempo ao se considerar um processo não-homogêneo de Poisson. Justamente esta característica foi explorada por Chavez-Demoulin, Davison e McNeil (2005) para o tratamento dos agrupamentos de volatilidade através do modelo CDM.

Consideremos agora uma família de processos não-homogêneos de Poisson com função intensidade $\lambda(\cdot; \theta)$ da qual, sujeito à validade do modelo, desejamos estimar o parâmetro θ a partir de uma amostra t_1, \dots, t_n em uma região $A \subset \mathcal{R}$. A informação contida nos dados é que os pontos ocorreram em determinados instantes da região A e em nenhum outro fora desta região. A verossimilhança, portanto, deve ser calculada utilizando toda essa informação. Com esse objetivo, definimos:

- (i) $I_i = [t_i, t_i + \delta_i]$, para $i = 1, \dots, n$, um conjunto de pequenos intervalos de tempo em torno das observações; e
- (ii) $\Psi = \Omega \setminus \bigcup_{i=1}^n I_i$.

Como $N(A) \sim Poi(\Lambda(A))$, temos segundo Ross (1987) que:

$$P[N(I_i) = x] = \exp[-\Lambda(I_i; \theta)] \times \frac{[\Lambda(I_i; \theta)]^x}{x!}$$

Logo

$$P[N(I_i) = 1] = \exp[-\Lambda(I_i; \theta)] \times \Lambda(I_i; \theta) \quad (12)$$

onde

$$\Lambda(I_i; \theta) = \int_{t_i}^{t_i + \delta_i} \lambda(t; \theta) dt \approx \lambda(t_i; \theta) \times \delta_i \quad (13)$$

Substituindo (13) em (12), obtemos:

$$P[N(I_i) = 1] \approx \exp[-\lambda(t_i; \theta) \times \delta_i] \times \lambda(t_i; \theta) \times \delta_i \approx \lambda(t_i; \theta) \times \delta_i \quad (14)$$

Pois, como δ_i é pequeno, temos que

$$\exp[-\Lambda(I_i; \theta)] \approx \exp[-\lambda(t_i; \theta) \times \delta_i] \approx 1$$

Além disso,

$$P[N(\Psi) = 0] = \exp[-\Lambda(\Psi; \theta)] \approx \exp[-\Lambda(\Omega; \theta)] \quad (15)$$

pois os δ_i são pequenos. Assim, a verossimilhança será dada por:

$$L(\theta; t_1; \dots; t_n) = P[N(\Psi) = 0, N(I_1) = 1, N(I_2) = 1, \dots, N(I_n) = 1]$$

Como os conjuntos $\Psi, I_1, I_2, \dots, I_n$ são disjuntos, temos que a verossimilhança será dada por:

$$L(\theta; t_1; \dots; t_n) = P[N(\Psi) = 0] \times P[N(I_1) = 1] \times P[N(I_2) = 1] \times \dots \times P[N(I_n) = 1]$$

ou ainda

$$L(\theta; t_1; \dots; t_n) = P[N(\Psi) = 0] \times \prod_{i=1}^n P[N(I_i) = 1] \quad (16)$$

Substituindo (14) e (15) em (16), obtemos:

$$L(\theta; t_1; \dots; t_n) = \exp[-\Lambda(\Omega; \theta)] \times \prod_{i=1}^n [\lambda(t_i; \theta) \times \delta_i]$$

que pode ser reescrito como

$$L(\theta; t_1; \dots; t_n) = \exp[-\Lambda(\Omega; \theta)] \times \prod_{i=1}^n \lambda(t_i; \theta) \times \prod_{i=1}^n \delta_i$$

Como os δ_i são constantes, podemos dividir a equação anterior por $\prod_{i=1}^n \delta_i$ e considerar a verossimilhança, para efeito de maximização, como:

$$L(\theta; t_1; \dots; t_n) = \exp[-\Lambda(\Omega; \theta)] \times \prod_{i=1}^n \lambda(t_i; \theta)$$

Tirando o logaritmo da equação anterior obtemos a logverossimilhança:

$$\ell(\theta; t_1; \dots; t_n) = -\Lambda(\Omega; \theta) + \sum_{i=1}^n \log[\lambda(t_i; \theta)] \quad (17)$$

Mas

$$\Lambda(\Omega; \theta) = \int_{\Omega} \lambda(t; \theta) dt \quad (18)$$

Assim, substituindo (18) em (17), obtemos a expressão final para a logverossimilhança:

$$\ell(\theta; t_1; \dots; t_n) = -\int_{\Omega} \lambda(t; \theta) dt + \sum_{i=1}^n \log[\lambda(t_i; \theta)] \quad (19)$$

A equação (19) será utilizada no modelo CDM para a estimação da distribuição dos instantes de ocorrência dos excessos em relação ao limiar u .

2.4. O Modelo de Chavez-Demoulin, Davison e McNeil (CDM)

Na abordagem tradicional de POT, assume-se que os termos da série $\{Z_t\}$ sejam independentes. Se, por outro lado, tomarmos $\{Z_t\}$ como sendo a série de retornos de uma ação, esta hipótese dificilmente será realista já que tais retornos geralmente são dependentes. Torna-se então necessária a confecção de um novo modelo capaz de considerar tal dependência, o qual foi desenvolvido por Chavez-Demoulin, Davison e McNeil (2005) e que denominamos de CDM. Este modelo está descrito a seguir.

Considere o processo (T_i, W_i) , $i \in Z$, que descreve os excessos W_i ao longo do tempo T_i o qual, quando observado sobre o período de tempo $(0, t_0]$, gera a amostra $(T_1, W_1), \dots, (T_n, W_n)$. Seja H_t toda a história do processo até o instante t , ou seja, H_t é a sigma-álgebra gerada pelo processo até o instante t . Assim, para calcular a verossimilhança dos dados observados, devemos notar que a densidade conjunta $(T_1, W_1), \dots, (T_n, W_n)$ dada H_0 pode ser escrita como:

$$\prod_{i=1}^n f_{T_i, W_i | H_{t_{i-1}}}(t_i, w_i | H_{t_{i-1}})$$

Como também está disponível a informação de que o próximo excesso $(n+1)$ em relação ao limiar u ocorrerá após o instante t_0 , ou seja, $T_{n+1} > t_0$, esta informação também precisará ser incorporada na densidade conjunta para que obtenhamos a verossimilhança:

$$L = \left[\prod_{i=1}^n f_{T_i, W_i | H_{t_{i-1}}}(t_i, w_i | H_{t_{i-1}}) \right] \times P(T_{n+1} > t_0 | H_{t_n}) \quad (20)$$

Assumindo a independência condicional entre os instantes em que ocorreram as violações e a magnitude das mesmas, dada toda a informação disponível até o instante imediatamente anterior, temos:

$$f_{T_i, W_i | H_{t_{i-1}}}(t_i, w_i | H_{t_{i-1}}) = f_{T_i | H_{t_{i-1}}}(t_i | H_{t_{i-1}}) \times f_{W_i | H_{t_{i-1}}}(w_i | H_{t_{i-1}}) \quad (21)$$

Substituindo (21) em (20), obtemos:

$$L = \left[\prod_{i=1}^n f_{T_i | H_{t_{i-1}}} (t_i | H_{t_{i-1}}) \times P(T_{n+1} > t_0 | H_{t_n}) \right]^A \times \left[\prod_{i=1}^n f_{W_i | H_{t_{i-1}}} (w_i | H_{t_{i-1}}) \right]^B \quad (22)$$

onde os termos rotulados de A e B representam respectivamente as contribuições dos instantes de tempo em que ocorreram os excessos e das magnitudes destes excessos.

Tomando o logaritmo da equação (22), obtemos a logverossimilhança:

$$\ell = \left[\log P(T_{n+1} > t_0 | H_{t_n}) + \sum_{i=1}^n \log f_{T_i | H_{t_{i-1}}} (t_i | H_{t_{i-1}}) \right]^A + \left[\sum_{i=1}^n \log f_{W_i | H_{t_{i-1}}} (w_i | H_{t_{i-1}}) \right]^B \quad (23)$$

Vamos primeiramente analisar o termo A . Seja $N(t, s)$ o número de excessos ocorridos durante o intervalo de tempo $(t, s]$ e N o seu respectivo processo de contagem com função intensidade $\lambda_H(t)$. Assim, considerando a equação (19), podemos expressar a contribuição do termo A da equação (23) para a logverossimilhança ℓ como:

$$\ell^A = \left[- \int_0^{t_0} \lambda_H(t) dt + \sum_{i=1}^n \log[\lambda_H(t_i)] \right]^A \quad (24)$$

Chavez-Demoulin, Davison e McNeil (2005) destacam que, na prática, é difícil especificar uma forma plausível e ao mesmo tempo tratável para $\lambda_H(t)$ e sugerem, para séries temporais financeiras, um processo auto-excitado onde

$$\lambda_H(t) = \mu + \sum_{j: t_j < t} \omega(t - t_j) \quad (25)$$

onde μ é uma constante positiva e $\omega(t)$ é não-negativo para $t > 0$ e zero caso contrário. Esta função intensidade faz com que os eventos ocorram em aglomerados, pois a ocorrência de um evento aumenta a taxa de chegada dos próximos. Tipicamente assume-se $\omega(t)$ monótona decrescente, de tal forma que os eventos mais recentes afetem mais a intensidade atual que aqueles que ocorreram em um passado mais distante. Chavez-Demoulin, Davison e McNeil (2005) adotaram para $\omega(t)$ a forma geral sugerida por Ogata (1988):

$$\omega(t - t_j; w_j) = \frac{\psi \exp(\beta w_j)}{(t - t_j + \gamma)^\rho} \quad (26)$$

onde $\rho, \gamma, \psi, \beta, \mu > 0$. Nesta formulação, o aumento da intensidade em função da ocorrência de um evento no passado depende não apenas do instante de tempo no qual ele ocorreu como também da sua magnitude. Além disso, Chavez-Demoulin, Davison e McNeil (2005) perceberam que a adoção de $\rho = 1$ não alterava de modo significativo a

verossimilhança nas aplicações em séries temporais financeiras que eles vinham realizando e optaram por efetivamente adotar esta versão mais simples do modelo.

Consideremos, agora, a parte B da equação (23), que corresponde aos excessos em relação ao limiar u condicionais aos instantes em que tais excessos ocorreram. Chavez-Demoulin, Davison e McNeil (2005) adotaram, neste caso, uma GDP com parâmetro de cauda fixo $\xi_j = \xi$ e parâmetro de escala definido através de um processo autorregressivo de primeira ordem $\log \sigma_j = a + b \times w_j$, ou seja:

$$W_i | W_{i-1} = w_{i-1} \sim GDP_{\xi, \exp(a+b \times w_{i-1})} \quad (27)$$

onde os parâmetros a, b e ξ podem ser estimados maximizando

$$\ell^B = \left[\log f_{W_1}(w_1) + \sum_{i=2}^n \log f_{W_i | W_{i-1}}(w_i | w_{i-1}) \right]^B \quad (28)$$

Por outro lado, Chavez-Demoulin, Davison e McNeil (2005) também ajustaram a versão mais simples do modelo com parâmetro de escala fixo $\sigma_j = \sigma$ e obtiveram bons resultados – embora de qualidade inferior a aqueles obtidos com adoção do modelo autorregressivo para σ_j .

2.5. Inclusão do Custo da Liquidação Fracionada de Carteiras no Modelo CDM (VaRF)

Considere uma carteira C com exposições financeiras F_1, \dots, F_m , nos seus m ativos A_1, \dots, A_m respectivamente, no dia t_0 – correspondente à data de hoje – e tomados com base nos seus respectivos preços de fechamento desta data. Imagine uma outra carteira C_t que, no dia t , possua estas mesma exposições financeiras em cada um deste m ativos A_1, \dots, A_m . Considere agora a seguinte estratégia para zerar a carteira C_t , a partir do dia t , a qual denominaremos estratégia Φ :

- (i) Mantem-se inalteradas as quantidades dos m ativos que compõem a carteira C_t durante os primeiros d_0 dias de negociação $(t+1, \dots, t+d_0)$ imediatamente posteriores ao dia t ; e
- (ii) Após o dia d_0 , inicia-se efetivamente a zeragem da carteira, a qual é realizada ao longo dos $(d_{\max} - d_0)$ dias de negociação seguintes $(t+(d_0+1), \dots, t+d_{\max})$.

Suponha que $q_{1,k}, \dots, q_{m,k}$, para $k = (d_0+1), \dots, d_{\max}$, sejam os percentuais, tomados em relação às quantidades de cada um dos m ativos que compoem a carteira C_t no dia t , que

serão zeradas no preço de fechamento do k -ésimo dia dentre os $(d_{\max} - d_0)$ dias citados em (ii). Considere, ainda, que os preços (de fechamento) destes ativos ao final deste mesmo dia sejam $p_1(t+k), \dots, p_m(t+k)$ respectivamente. O financeiro obtido neste dia em decorrência de tal zeragem será então dado por:

$$F_k(t) = \sum_{s=1}^m \left(q_{s,k} \times F_s \times \frac{p_s(t+k)}{p_s(t)} \right).$$

Como este financeiro foi gerado a partir de um financeiro inicial, no dia t , de

$$F_k = \sum_{s=1}^m (q_{s,k} \times F_s),$$

o log-retorno referente a tal zeragem será de

$$r_k(t) = \log \left(\frac{F_k(t)}{F_k} \right) \quad (29)$$

Como o financeiro inicial da carteira C_t é, por construção, independente de t e dado por

$$F = \sum_{k=d_0+1}^{d_{\max}} F_k,$$

o log-retorno total da carteira será, após finalizada a sua zeragem, dado por:

$$R(t) = \log \left(\frac{\sum_{k=d_0+1}^{d_{\max}} F_k(t)}{F} \right) \quad (30)$$

O nosso interesse é, justamente, estimar menor valor do log-retorno total da carteira $R(t)$, obtido através da equação (30), com um elevado nível de confiança.

Vale destacar que, para a cada log-retorno, podemos calcular o respectivo retorno

$$Retorno(t) = \exp(R(t)) - 1$$

e o respectivo resultado total da carteira como

$$Resultado(t) = F \times Retorno(t).$$

Para facilitar o entendimento do que foi exposto, vamos citar alguns exemplos:

Exemplo 1:

Consideremos uma carteira formada por ativos muito líquidos e em quantidades pequenas em relação ao volume usualmente negociado no mercado. Assim, será necessário apenas $d_{\max} - d_0 = 1$ dia para zerá-la. Caso se deseje efetivamente começar a zeragem hoje, devemos tomar $d_0 = 0$.

Exemplo 2:

Tomemos agora alguns dos ativos da carteira do exemplo (1) como sendo pouco líquidos, de modo que sejam necessários $d_{\max} - d_0 = 5$ dias para zerá-los sem que se influencie substancialmente os preços negociados no mercado. Tal qual no exemplo (1), caso se deseje efetivamente começar a zeragem hoje, continuaremos tomando $d_0 = 0$.

Exemplo 3:

Imaginemos, agora, que os ativos da carteira do exemplo (1) posições compradas em contratos futuros de boi gordo negociados na Bolsa de Mercadorias & Futuros (BM&F), a qual gostaria de estimar o volume de margem necessário para a cobertura de uma eventual inadimplência. Como os ajustes financeiros referentes a um determinado dia somente são pagos no dia seguinte, esta potencial inadimplência também será detectada após decorrido $d_0 = 1$ dia, quando esta Bolsa poderá iniciar a zeragem da carteira. Como, por hipótese, esta carteira é muito líquida, será necessário apenas $d_{\max} - d_0 = 1$ dia para zerá-la. Logo $d_{\max} = 2$ dias.

Exemplo 4:

Caso, tal qual no exemplo (2), alguns dos contratos do exemplo (3) fossem pouco líquidos, a BM&F necessitaria por exemplo de $d_{\max} - d_0 = 5$ dias para zerar a carteira. Como tal zeragem somente ocorreria após a inadimplência ter sido detectada, teríamos $d_0 = 1$ dia. Assim, $d_{\max} = 6$ dias.

Exemplo 5:

Neste exemplo, tomaremos por base uma carteira hipotética C formada por posições em compradas em $m = 2$ ações:

- (i) 800 ações A_1 com preço hoje de 20 reais; e
- (ii) 100 ações A_2 com preço hoje de 10 reais.

Assim, o financeiro da carteira em cada uma das ações será de:

- (i) $F_1 = 800 \times 20 = 16000$ reais na ação A_1 ; e
- (ii) $F_2 = 100 \times 10 = 1000$ reais na ação A_2 .

Consequentemente, o financeiro total da carteira será $F = F_1 + F_2 = 16000 + 1000 = 17000$ reais.

Uma possível estratégia para zerá-la é aguardarmos $d_0 = 1$ dia e, daqui a $d_0 + 1 = 2$ dias vendermos 25% das ações A_1 (=200 ações) e 100% das ações A_2 (=100 ações). Finalmente, daqui a $d_{\max} = d_0 + 2 = 3$ dias, os 75% ações A_1 ainda restantes (=600 ações) seriam zeradas.

Vamos, agora, calcular o log-retorno referente à zeragem da carteira C_t , com $t = \text{hoje menos 1 ano}$. Suponhamos, então, que o preço de fechamento das ações A_1 e A_2 , há um ano atrás, fossem de 15 e 5 reais respectivamente. Imaginemos ainda que no dia seguinte tais preços passassem para 16 e 6 reais, no próximo atingissem 17 e 7 reais, em seguida para 18 e 8 reais e, finalmente, chegassem a 19 e 9 reais. Em tal situação, teríamos os seguintes financeiros obtidos em decorrência da zeragem:

$$F_2(t) = \sum_{s=1}^2 \left(q_{s,2} \times F_s \times \frac{p_s(t+2)}{p_s(t)} \right) = 25\% \times 16000 \times \frac{17}{15} + 100\% \times 1000 \times \frac{7}{5} = 5933.33 \text{ reais}$$

$$F_3(t) = \sum_{s=1}^2 \left(q_{s,3} \times F_s \times \frac{p_s(t+3)}{p_s(t)} \right) = 75\% \times 16000 \times \frac{18}{15} + 0\% \times 1000 \times \frac{8}{5} = 14400.00 \text{ reais}$$

Deste modo, o log-retorno seria dado por

$$R(t) = \log \left(\frac{\sum_{k=d_0+1}^{d_{\max}} F_k(t)}{F} \right) = \log \left(\frac{5933.33 + 14400.00}{17000.00} \right) = 0.1790 = 17.90\%$$

Poderemos, então, calcular o Valor em Risco com Liquidação Fracionada (VaRF) da carteira, o qual definiremos como a máxima perda da carteira, com um dado nível de confiança, dada uma estratégia Φ de zeragem a qual, por hipótese, não impacta significativamente o nível dos preços (que estejam sendo negociados no mercado) dos seus ativos.

Embora os cálculos para a obtenção do VaRF sejam realizados em termos de log-retornos, é bastante simples convertê-los para retornos ou valores financeiros:

$$VaRF_{\text{Retorno}}(t) = \exp(VaRF_{\text{Log-Retorno}}(t)) - 1$$

$$VaRF_{\text{Financeiro}}(t) = F \times VaRF_{\text{Retorno}}(t).$$

Em linhas gerais, para calcular o VaRF, serão feitos pequenos ajustes no modelo CDM e aplicaremos esta nova versão do modelo à série de log-retornos $R(t)$ definida na equação (30). Vale destacar que o modelo CDM trata o principal problema da série $R(t)$: a dependência temporal evidenciada nos agrupamentos dos excessos sobre o limiar u .

Assim, faremos dois ajustes no modelo CDM, os quais indicaremos a seguir.

Primeiro Ajuste – Equações (25) e (26)

Sejam, para $k = (d_0 + 1), \dots, d_{\max}$, as seguintes expressões:

$$r_k^{\text{aux}}(t) = \log \left(\frac{\sum_{s=1}^m \left(q_{s,k} \times F_s \times \frac{p_s(t)}{p_s(t-1)} \right)}{F_k} \right)$$

$$r_k^+(t) = \text{Max}[r_k^{\text{aux}}(t); 0],$$

onde $\text{Max}[x; y]$ indica o maior dentre os valores x e y .

$$r_k^-(t) = r_k^+(t) - r_k^{\text{aux}}(t)$$

$$\bar{r}_k(t) = \eta \times r_k^+(t) + (1 - \eta) \times r_k^-(t) \quad (31)$$

$$\bar{R}_k(t) = \begin{cases} \bar{r}_k(t) + \varphi \times \bar{R}_k(t-1), & \text{se } t > 0 \\ \bar{r}_k(0), & \text{se } t = 0 \end{cases}$$

Adotaremos, no lugar das equações (25) e (26), a seguinte expressão para a função intensidade $\lambda_{\text{H}}(t)$:

$$\lambda_{\text{H}}(t) = \mu + \sum_{k=d_0+1}^{d_{\max}} (\beta_k \times \bar{R}_k(t)) \quad (32)$$

onde $\bar{R}_k(t)$ é obtido a partir das equações (31). Além disso, devemos ter $0 \leq \eta \leq 1$, $0 \leq \varphi < 1$, $\mu \geq 0$ e $\beta_{d_0+1}, \dots, \beta_{d_{\max}} \geq 0$.

É interessante notarmos que $r_k^+(t)$ e $r_k^-(t)$ são, respectivamente, a parte positiva e negativa (em valor absoluto) do log-retorno da parcela da carteira C_{t-1} , referente ao dia $(t-1)$, que seria zerada no k -ésimo dia de zeragem – dia $(t+k-1)$ – caso os preços de fechamento desse dia coincidisse com os do dia t .

Esta nova formulação, tal qual a do modelo CDM, também é auto-excitada pois, quando ocorrem grandes resultados, positivos ou negativos, da carteira em um dado dia, a função

intensidade $\lambda_{11}(t)$ aumenta em função do aumento de $\bar{R}_k(t)$. Além disso, o efeito dos resultados positivos é diferenciado do efeito dos resultados negativos através do parâmetro η . Por fim, o parâmetro φ faz com que os resultados mais recentes influenciem mais a função intensidade do que os mais antigos.

Adicionalmente, ela incorpora o efeito da zeragem fracionada da carteira durante os $(d_{\max} - d_0)$ dias posteriores ao dia d_0 .

Segundo Ajuste – Equações (27) e (28)

Para calcular o termo B da logverossimilhança indicada na equação (23), adotaremos uma GPD com ξ e σ constantes pois, além de mais simples, fornece bons resultados segundo podemos observar no estudo de Chavez-Demoulin, Davison e McNeil (2005). Assim, tomando por base a equação (3), as equações (27) e (28) serão substituídas por:

$$\ell^B = \left[-n \log \sigma - \left(1 + \frac{1}{\xi} \right) \sum_{j=1}^n \log \left(1 + \xi \frac{w}{\sigma} \right) \right]^B \quad (33)$$

2.6. Cálculo do VaRF

O objetivo de calibrarmos os parâmetros do modelo é podermos calcular o VaRF, ou seja, estimarmos o quantil da distribuição preditiva do log-retorno $R(t)$ da carteira, conforme calculado a partir da equação 30, obtido com base em uma determinada estratégia de zeragem da carteira. Vale destacar que tal estimativa é condicional ao histórico de informação disponível até hoje ($t_0 = T - 1$). Assim, temos:

$$VaRF = VaRF_{\phi}(T) = \inf \{ z \in \mathfrak{R} : F_{R(t)|H_{T-1}}(z) \geq \phi \} \quad (34)$$

Neste trabalho, com será visto mais adiante, tomaremos $\phi = 99.9\%$, ou seja, imaginaremos que adotando tal estratégia de zeragem, esperaremos ter perdas superiores, em log-retorno, ao VaRF em um a cada 1000 dias, ou seja, aproximadamente 1 em cada 4 anos⁴. Assim, podemos imaginar que estamos estimando a perda potencial da carteira em uma situação atípica (porém possível) de estresse do mercado.

Mas

$$P(Z_T > z | H_{T-1}) = P(Z_T - u > z - u | Z_T > u, H_{T-1}) \times P(Z_T > u, H_{T-1}) \quad (35)$$

⁴ Levando-se em conta que em um ano há cerca de 250 dias de negociação em pregão, em quatro anos há cerca de 1000 dias de pregão.

O segundo termo do lado direito da equação (35) pode ser estimado utilizando-se o processo auto-excitado, pois representa a probabilidade condicional de um evento ocorrer no intervalo $(T-1, T]$:

$$1 - P[N(T-1, T) = 0 | H_{T-1}] = 1 - \exp\left(-\int_{T-1}^T \lambda_H(u) du\right) \quad (36)$$

onde $\lambda_H(u)$ é obtido através da equação (32) dado que o modelo tenha sido calibrado, ou seja, uma vez obtidos os estimadores de $\eta, \varphi, \mu, \beta_{d_0+1}, \dots, \beta_{d_{\max}}$ que aparecem nas equações (31) e (32).

O primeiro termo pode ser estimado utilizando-se os parâmetros da GDP obtidos com a otimização dos parâmetros ξ e σ do termo ℓ^B da equação (33).

O valor procurado (VaRF) é a solução da equação

$$P(Z_T > VaRF_\phi(T) | H_{T-1}) = 1 - \varphi. \quad (37)$$

Substituindo as equações (1), (36) e (37) na (35), e lembrando que

$$\begin{aligned} P(Z_T - u > z - u | Z_T > u, H_{T-1}) &= 1 - G_{\xi, \sigma}(w) = 1 - G_{\xi, \sigma}(z - u) \\ &= 1 - \left[1 - \left(1 + \xi \frac{z - u}{\sigma} \right)^{-\frac{1}{\xi}} \right] = \left(1 + \xi \frac{z - u}{\sigma} \right)^{-\frac{1}{\xi}} \end{aligned}$$

temos

$$\left(1 + \xi \frac{z - u}{\sigma} \right)^{-\frac{1}{\xi}} \times \left(1 - \exp\left(-\int_{T-1}^T \lambda_H(u) du\right) \right) = 1 - \varphi.$$

Logo

$$z = VaRF_\phi(T) = \left[\left(\frac{1 - \exp\left(-\int_{T-1}^T \lambda_H(u) du\right)}{1 - \varphi} \right)^\xi - 1 \right] \times \frac{\sigma}{\xi} + u \quad (38)$$

Demoulin, Davison e McNeil (2005) fixaram u de tal forma que cerca de 10% dos dados estivessem acima deste limiar. Neste trabalho, fixaremos u adotando este mesmo critério. Assim, como trabalharemos com uma série com cerca de 1000 dados (quatro anos vezes

250 dados por ano), haverá cerca de $10\% \times 1000 = 100$ pontos para calibrarmos a parte de valores extremos do modelo.

Por fim, vale destacar que toda a teoria foi desenvolvida para modelarmos a cauda direita da distribuição. Ao calcularmos o risco, por outro lado, estamos preocupados com perdas, ou seja, com a cauda esquerda da distribuição dos log-retornos $R(t)$ da equação (30). Por este motivo, toda a análise será realizada multiplicando-se $R(t)$ por (-1) , já que os menores valores de $R(t)$ correspondem aos maiores valores de $(-R(t))$.

3. Aplicação do Modelo às Ações Negociadas na BOVESPA

Esta seção apresenta uma aplicação do modelo ao cálculo do VaRF com 99.9% de confiança de uma carteira com ações negociadas na Bolsa de Valores de São Paulo (BOVESPA) indicadas na tabela a seguir.

Ações		Financeiro (R\$ MM)
Código BOVESPA	Descrição	
ACES4	Acesita PN	100.00
ARCZ6	Aracruz PNB	100.00
BBDC4	Bradesco PNB	100.00
CMIG4	Cemig PN	100.00
ELET3	Eletróbrás ON	100.00
ITAU4	Itaú PN	100.00
PETR4	Petrobrás PN	100.00
SDIA4	Sadia PN	100.00
TNLP4	Telemar PN	100.00
USIM5	Usiminas PNA	100.00
VALE5	Vale do Rio Doce PNA	100.00
VCPA4	VCP PN	100.00
Total		1,200.00

Tabela 1 – Composição da carteira de ações

Será calculado o risco da carteira apresentada na tabela 1 relacionado a duas estratégias de liquidação distintas, cada uma delas levando à obtenção um determinado valor de VaRF.

Os cálculos são realizados supondo-se que hoje seja o último dia de negociação, ou dia de pregão, do ano de 2006 e utilizam as cotações de fechamento das ações de todos os pregões realizados ao longo dos anos de 2003, 2004, 2005 e 2006. Assim, estimaremos o risco da carteira imediatamente após fechamento (final) deste dia.

3.1. Descrição das Estratégias de Liquidação

A tabela a seguir apresenta as duas estratégias de zeragem que serão analisadas neste trabalho.

Ação	Financeiro Hoje	Estratégia 1			Estratégia 2		
		D+1	D+2	D+3	D+1	D+2	D+3
ACES4	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
ARCZ6	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
BBDC4	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
CMIG4	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
ELET3	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
ITAU4	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
PETR4	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
SDIA4	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
TNLP4	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
USIM5	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
VALE5	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%
VCPA4	100.00	0.0%	12.5%	87.5%	0.0%	25.0%	75.0%

Tabela 2 – Estratégias para zeragem da carteira de ações

Na tabela 2, D+1, D+2 e D+3 representam as datas dos três pregões seguintes ao dia de hoje e os percentuais em cada uma destas colunas indicam a parcela da quantidade de cada uma das ações da carteira que será zerada nestes dias. Assim, por exemplo, enquanto na estratégia 1 12.5% das ações de ARCZ6 serão zerados no segundo dia (D+2) e os 87.5% restantes serão zerados no terceiro dia (D+3), na estratégia 2 um quarto das ações (25%) será zerados em (D+2) e os três quartos restantes (75%) em (D+3).

Em ambas as estratégias, a zeragem será realizada do segundo dia em diante. Assim, tanto para a estratégia 1 quanto para a 2, teremos $d_0 = 1$. Além disso, em ambas as estratégias, a carteira será zerada em $d_{\max} = 3$ dias. Assim, deveremos estimar, para cada uma das estratégias, os seguintes parâmetros:

- ξ e σ do termo ℓ^B da equação (33) – parâmetros da GPD; e
- $\eta, \varphi, \mu, \beta_2, \beta_3$ do termo ℓ^A conforme indicado nas equações (24), (31) e (32) – parâmetros referentes à parte auto-excitada do modelo – $\lambda_H(u)$.

3.2. Calibragem do Modelo e Análise dos Resultados

A tabela a seguir apresenta os resultados obtidos com a calibragem do modelo em ambas as estratégias.

	Estratégia	
	1	2
Amostra	988	988
Percentil u	10.00%	10.00%
u	2.91%	2.84%
Pontos GPD	99	99
ksi	0.0421	0.0764
sigma	0.0144	0.0130
LV_B	316.5914	323.2920
eta	0.6470	0.6230
fi	0.9523	0.9462
mi	0.0220	0.0113
Beta2	0.3311	0.4207
Beta3	0.3311	0.4207
LV_A	-324.3255	-323.5025

Tabela 3 – Tamanho da amostra e valores estimados para os parâmetros do modelo

Com base nos valores estimados para os parâmetros dos modelos apresentados na tabela 3, o VaRF associado a cada uma das estratégias, expresso em log-retorno, calculado com a informação disponível até o final do último pregão de 2006, foi calculado através da equação (38). Os valores obtidos foram -9.60% para a estratégia 1 e -9.24% para a estratégia 2.

Vale lembrar que, conforme destacada na seção 2.6, os valores obtidos a partir das informações da tabela 3 devem ser multiplicados por (-1) de modo a fornecerem a estimativa do VaRF.

Assim, em termos financeiros, o VaRF associado à estratégia 1 é

$$VaRF_{\phi=99,9\%} = R\$1200MM \times (\exp(-,0960) - 1) = -R\$109.8MM .$$

Já para a estratégia 2 o VaRF em termos financeiros é de

$$VaRF_{\phi=99,9\%} = R\$1200MM \times (\exp(-,0924) - 1) = -R\$105.9MM .$$

As tabelas 4, 5.1 e 5.2 apresentam a matriz de informação, a matriz de covariância e os desvios obtidos tanto para as estimativas dos parâmetros do modelo de valores extremos (GDP) quanto para a parte auto-excitada (processo de ponto). As propriedades assintóticas dos estimadores de máxima verossimilhança utilizadas para a elaboração destas tabelas estão descritas no apêndice A.

Estratégia	Matriz de Informação			Matriz de Covariância			GPD
	ksi	sigma		ksi	sigma		Desvio
1	ksi	163.7	6116.7	ksi	0.0127	-0.0002	0.1127
	sigma	6116.7	439852.2	sigma	-0.0002	0.0000	0.0022
2	ksi	142.6	6233.5	ksi	0.0152	-0.0002	0.1231
	sigma	6233.5	506796.3	sigma	-0.0002	0.0000	0.0021

Tabela 4 – Matriz de informação, matriz de covariância e desvios da calibragem da GDP

Estratégia 1 - Processo de Ponto						
Matriz de Informação						
	eta	fi	mi	Beta2	Beta3	
eta	17.1	-350.2	-237.8	-24.1	-24.1	
fi	-350.2	27274.9	15575.5	1889.2	1889.2	
mi	-237.8	15575.5	10430.8	1145.6	1145.6	
Beta2	-24.1	1889.2	1145.6	138.2	138.2	
Beta3	-24.1	1889.2	1145.6	138.2	138.2	
Matriz de Covariância						Desvio
	eta	fi	mi	Beta2	Beta3	
eta	0.091	0.002	0.004	-0.025	-0.012	0.302
fi	0.002	0.001	0.000	0.000	-0.010	0.027
mi	0.004	0.000	0.001	-0.002	-0.010	0.035
Beta2	-0.025	0.000	-0.002	112589.998	-112589.983	335.544
Beta3	-0.012	-0.010	-0.010	-112589.983	112590.209	335.545

Tabela 5.1 – Matriz de informação, matriz de covariância e desvios da calibragem do processo de ponto da estratégia 1

Estratégia 2 - Processo de Ponto						
Matriz de Informação						
	eta	fi	mi	Beta2	Beta3	
eta	24.4	-423.4	-274.4	-24.8	-24.8	
fi	-423.4	27734.3	15774.5	1699.5	1699.5	
mi	-274.4	15774.5	10652.2	1030.7	1030.7	
Beta2	-24.8	1699.5	1030.7	110.1	110.1	
Beta3	-24.8	1699.5	1030.7	110.1	110.1	
Matriz de Covariância						Desvio
	eta	fi	mi	Beta2	Beta3	
eta	0.064	0.002	0.003	-0.022	-0.022	0.254
fi	0.002	0.001	0.000	-0.006	-0.006	0.027
mi	0.003	0.000	0.001	-0.006	-0.006	0.034
Beta2	-0.022	-0.006	-0.006	14660.233	-14660.077	121.079
Beta3	-0.022	-0.006	-0.006	-14660.077	14660.233	121.079

Tabela 5.1 – Matriz de informação, matriz de covariância e desvios da calibragem do processo de ponto da estratégia 2

Do mesmo modo que, no trabalho de Chavez-Demoulin, Davison e McNeil (2005), os desvios associados às estimativas de alguns dos parâmetros são grandes em relação aos seus valores estimados, neste trabalho também obteve-se desvios elevados. Coles (2001) destaca que menores desvios para os estimadores de máxima verossimilhança usualmente podem ser obtidos a partir da função de perfil da verossimilhança. Por outro lado, conforme poderemos visualizar através dos diagramas Q-Q (quantil empírico x quantil estimado utilizando o modelo), tal qual no trabalho de Chavez-Demoulin, Davison e McNeil (2005), o modelo parece estar bem ajustado.

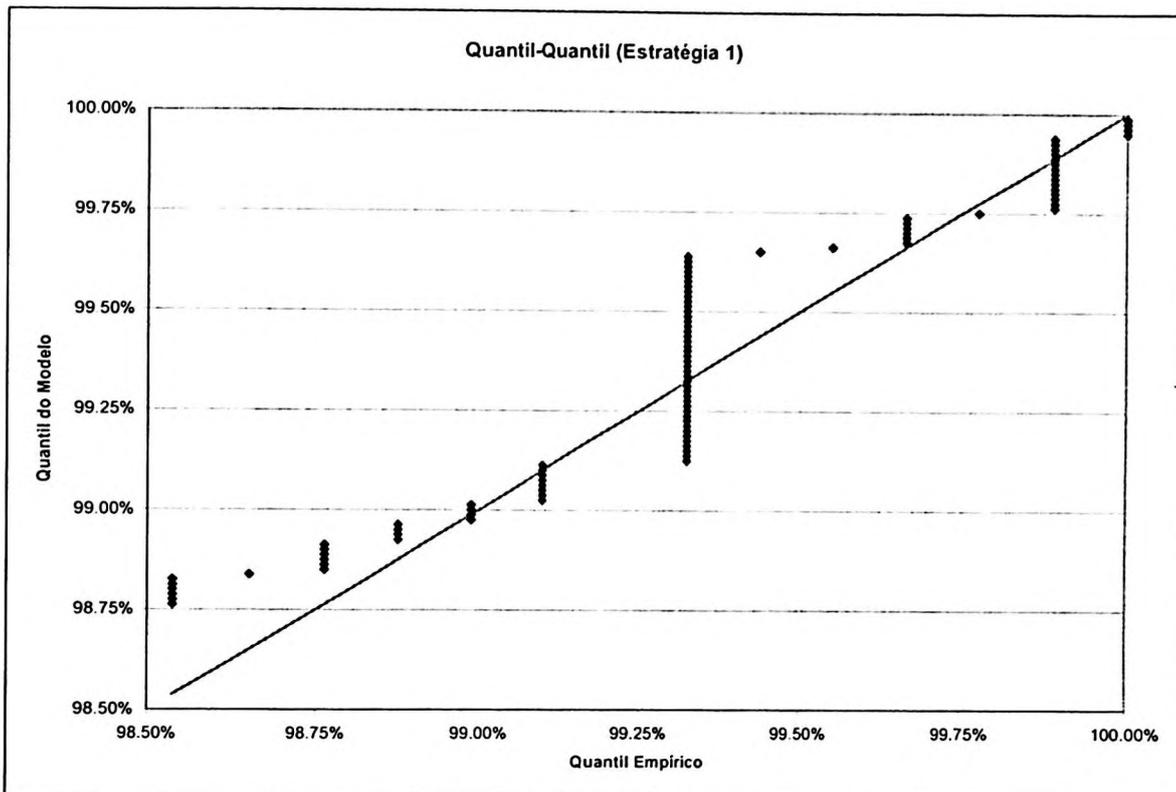


Figura 1.1 – Evolução histórica da estimativa do VaR e do log-retorno da carteira na estratégia 1

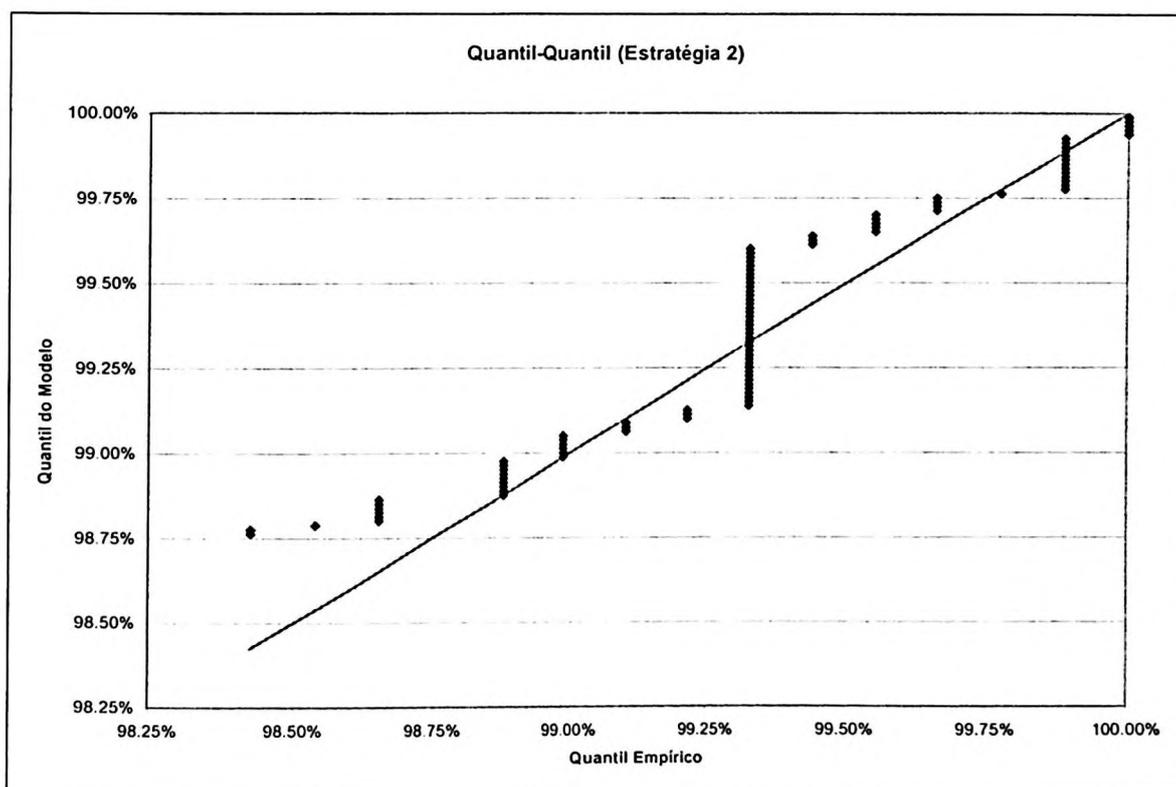


Figura 1.2 – Evolução histórica da estimativa do VaRF e do log-retorno da carteira na estratégia 2

Conforme podemos observar nas figuras 1.1 e 1.2, os quantis estimados pelos modelos, em ambas as estratégias, estão próximos aos quantis empíricos já que todos os pontos (x =quantil empírico, y =quantil do modelo) estão próximos à reta $y=x$. Isto nos leva a crer que ambos os modelos estejam bem ajustados. Vale destacar que os diagramas quantil-quantil foram traçados para quantis elevados os quais correspondem a situações de crise e constituem o foco deste trabalho.

O histórico da evolução da estimativa do VaRF em relação ao log-retorno da carteira, para cada uma das estratégias está apresentado a seguir.

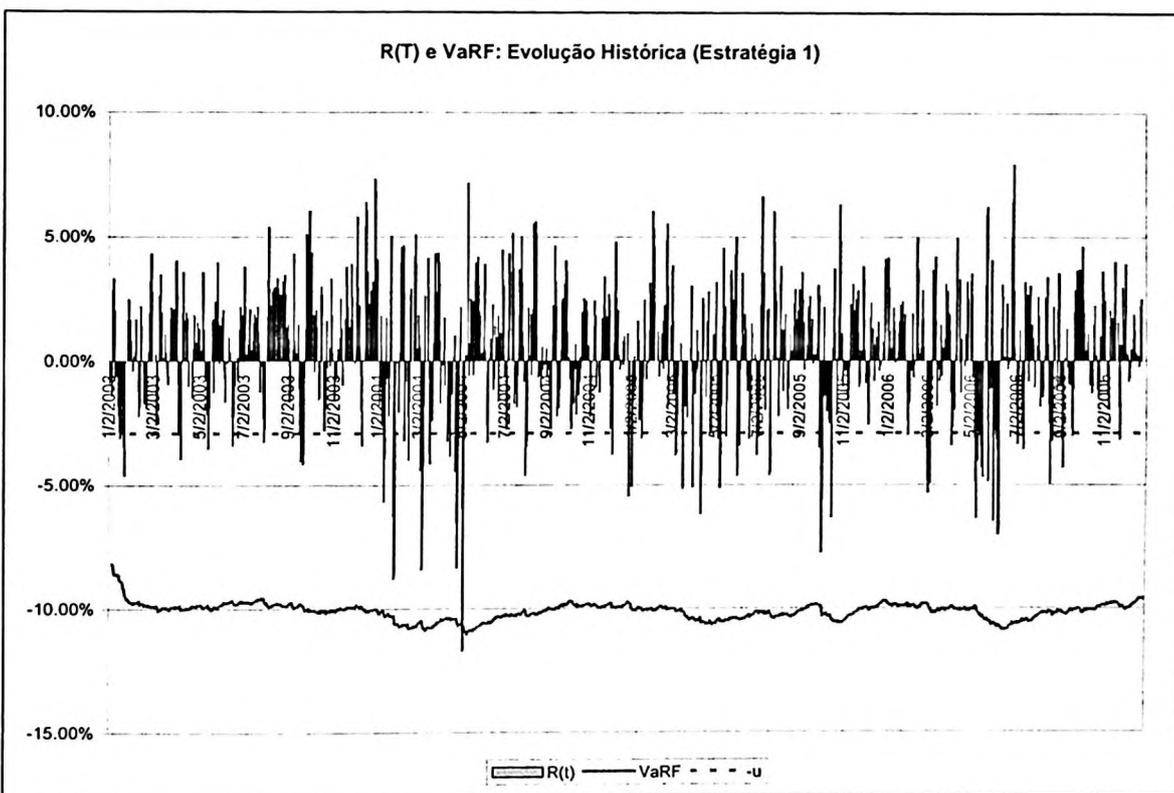


Figura 2.1 – Evolução histórica da estimativa do VaRF e do log-retorno da carteira na estratégia 1

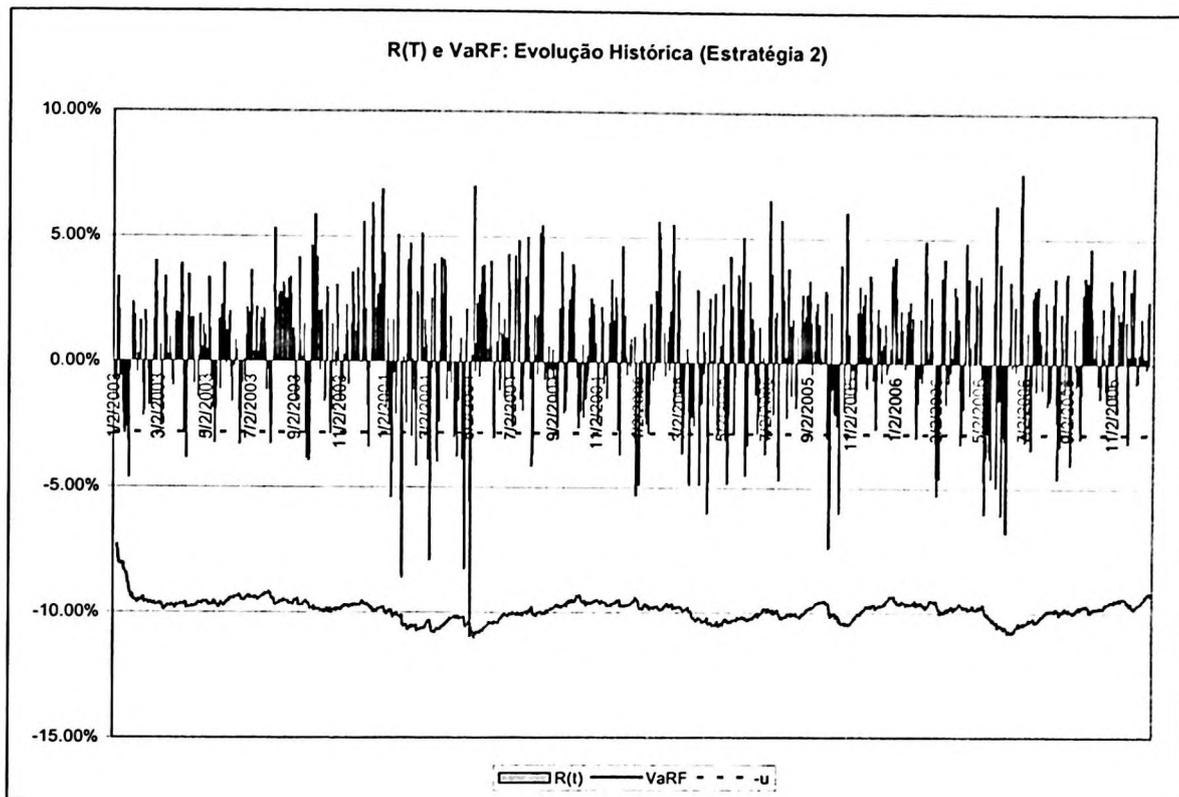


Figura 2.2 – Evolução histórica da estimativa do VaRF e do log-retorno da carteira na estratégia 2

Durante o horizonte de tempo de quatro anos – que corresponde a cerca de 1000 dias de negociação – considerado neste trabalho, houve em cada uma das estratégias um dia em que o log-retorno da carteira excedeu o seu VaRF. Esta era a quantidade de excessos esperada já que, como o VaRF foi calculado com 99.9% de confiança, esperaríamos ter $0.1\% \times 1000 = 1$ violação do VaRF em cada estratégia durante esse período. Além disso, conforme era desejado, a parte auto-excitada do modelo faz com que o VaRF aumente sempre que há agrupamentos de volatilidade e o consequente aumento da probabilidade de ocorrência de perdas maiores que as usuais durante a zeragem da carteira. Tal qual no trabalho de Chavez-Demoulin, Davison e McNeil (2005), os valores iniciais do VaRF são baixos devido à falta inicial de dados porém, logo em seguida, as estimativas do VaRF atingem um nível condizente com o risco da carteira. Tal fato não gera problema algum já que estamos apenas analisando o comportamento histórico do VaRF mas, na realidade, a aplicação prática do modelo sugerido é a estimação de perdas potenciais futuras da carteira caso seja adotada uma das duas estratégias de zeragem abordadas.

Por fim, a figura 3 apresenta o histórico da diferença entre o VaRF estimado para a carteira associado a cada uma das duas estratégias de zeragem.

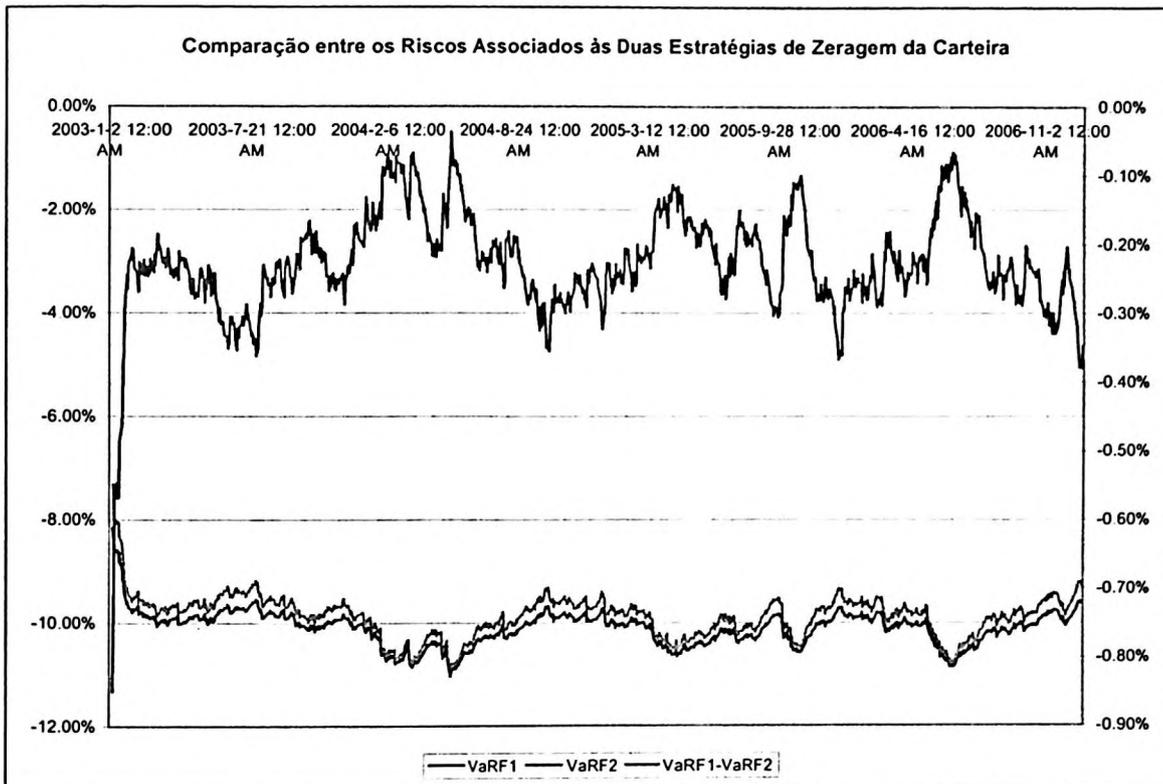


Figura 3 – Comparação entre os riscos associados às duas estratégias de zeragem da carteira

Na figura 3, enquanto os valores de VaRF associados às estratégias de zeragem 1 e 2 são aprenesados no eixo vertical esquerdo, a diferença entre os riscos está indicada no eixo vertical direito.

Conforme pode ser observado, passado o período inicial em que os modelos possuem poucos dados para “alimentá-los”, a adoção da estratégia 1 expõe a carteira a um risco maior que a adoção da 2, sendo que tal diferença de riscos situa-se entre 0.04% e 0.38% durante o período analisado.

Já era esperado que o VaRF associado à estratégia 1 fosse superior, em valor absoluto, ao associado à estratégia 2 já que, conforme pode ser observado na tabela 2, uma maior parcela da carteira é liquidada em (D+3) na estratégia 1 em relação à estratégia 2.

4. Considerações Finais

Este trabalho tratou da estimação, com um dado nível de confiança, da perda potencial a que está exposta uma carteira da qual se inicia a liquidação após um determinado período de tempo, sendo esta realizada de forma fracionada – ou seja, pouco a pouco de forma a não deslocar significativamente as curvas de oferta e demanda do mercado – ao longo de um ou mais dias. Como esta perda potencial foi estimada com um nível de confiança elevado, a medida de risco obtida, denominada VaRF, corresponde a perdas potenciais em eventos de crise. Além disso, a estimativa foi feita de forma condicional à informação disponível até o momento em que ela era calculada.

Com este objetivo, partiu-se do trabalho de Chavez-Demoulin, Davison e McNeil (2005), que ao longo do texto denominamos modelo “CDM”, o qual não somente trata de estimativas do VaR para altos níveis de confiança como também identifica períodos com agrupamentos de volatilidade.

O modelo CDM é uma adaptação de um modelo clássico da Teoria dos Valores Extremos (TVE), conhecido como POT⁵, na qual os dados “falam por si mesmos” não havendo, portanto, a necessidade do ajuste de um modelo de séries temporais intermediário – conforme é necessário em outros modelos. Para tanto, o modelo CDM combina um processo de ponto auto-excitado, que estima o instante em que ocorrem os excessos acima de um determinado limiar, com um modelo de POT para estimar a magnitude de tais excessos.

Ao se assumir a independência, condicional a toda a informação disponível até o momento, entre os instantes em que ocorrerão os próximos excessos e a magnitude destes, pode-se ajustar a parte do modelo referente à magnitude dos excessos (modelo de valores extremos) separadamente da parte referente aos instantes de ocorrência destes (processo de ponto). Assim, tal hipótese torna o ajuste do modelo mais simples do que o de um outro em que tais variáveis estejam efetivamente acopladas. Esta é a essência do modelo CDM.

Após algumas adaptações no modelo CDM, chegou-se ao modelo utilizado para a estimação do VaRF. Como aplicação, foi analisado o risco associado a zeragem de uma carteira em duas estratégias distintas. Em ambos os casos, os resultados obtidos estavam dentro do esperado. Assim, o modelo proposto possibilitou a estimativa do VaRF da carteira, associado a cada estratégia de zeragem, com níveis de confiança elevados. Em decorrência de ser auto-excitado, tal qual o modelo CDM, tratou adequadamente os agrupamentos de volatilidade ao estimar valores mais elevados para o VaRF quando tais agrupamentos ocorriam. Deste modo, o modelo proposto pode ser utilizado, por exemplo, para a escolha da estratégia de zeragem de uma dada carteira que a exponha ao menor risco.

⁵ A sigla POT vem do inglês *Peaks-over-Threshold*, que significa picos acima de um limiar.

Um possível aprimoramento para modelo proposto é a incorporação do efeito da zeragem da carteira nos preços praticados no mercado. Um outro, que pode inclusive estar ligado ao anterior, é a inclusão no cálculo do VaRF do efeito do diferencial entre os preços das ofertas para compra e para venda de cada ação.

Apêndice A – Propriedades Assintóticas dos Estimadores de Máxima Verossimilhança (EMV)

Esta seção apresenta as principais propriedades dos estimadores de máxima verossimilhança (EMV) utilizadas neste trabalho. Coles (2001) apresenta um ótimo resumo das propriedades de tais estimadores que são usualmente utilizadas na TVE. Adicionalmente, Heij, Boer, Franses, Kloek, e Dijk (2004) tratam do assunto de forma bastante clara e intuitiva.

Seja $\mathfrak{F} = \{f(x; \theta) : \theta \in \Theta\}$ uma família de funções de densidade de probabilidade conjunta de um conjunto de n observações, não necessariamente independentes, $x = (x_1, x_2, \dots, x_n)$. A verossimilhança é então definida por $L(\theta) = f(x; \theta)$, ou seja, é uma função de $\theta = (\theta_1, \theta_2, \dots, \theta_d)$ fixadas as observações $x = (x_1, x_2, \dots, x_n)$.

No caso particular das observações serem independentes, teremos que $L(\theta) = \prod_{i=1}^n f_i(x_i; \theta)$.

Caso, além de independentes, as observações sejam identicamente distribuídas, teremos que $L(\theta) = \prod_{i=1}^n f(x_i; \theta)$ pois $f_i(x_i; \theta) = f(x_i; \theta)$.

Definimos o estimador de máxima verossimilhança (EMV) $\hat{\theta}_0$ de θ_0 como o valor de θ que maximiza o valor da verossimilhança $L(\theta)$. Como o log é uma função monótona crescente, maximizar a logverossimilhança $\ell(\theta) = \log(L(\theta))$ conduz aos mesmos resultados que a maximização da verossimilhança $L(\theta)$. Caso as observações sejam independentes, teremos que $\ell(\theta) = \sum_{i=1}^n \log f_i(x_i; \theta)$. Caso elas também sejam identicamente distribuídas, chegamos a $\ell(\theta) = \sum_{i=1}^n \log f(x_i; \theta)$.

Como, em geral, os EMVs não podem ser obtidos explicitamente – devendo, portanto, ser obtidos numericamente – é comum se optar pela maximização da log-verossimilhança $\ell(\theta)$ ao invés da verossimilhança $L(\theta)$ simplesmente por questões relacionadas à precisão e tratabilidade numérica. Adicionalmente, é comum utilizar-se ao algoritmo de Newton-Raphson em tal otimização.

Sob condições de regularidade bastante brandas sobre $\ell(\theta)$ – que, segundo Heij, Boer, Franses, Kloek, e Dijk (2004) praticamente são equivalentes ao modelo (isto é, à distribuição de probabilidade conjunta dos dados) ter sido especificada corretamente – os estimadores de máxima verossimilhança possuem várias propriedades estatísticas que são assintoticamente ótimas, algumas das quais estão apresentadas a seguir.

EMV1: Os EMVs são consistentes, isto é, converge em probabilidade para o valor real do parâmetro a medida que o tamanho da amostra aumenta.

$$\lim_{n \rightarrow \infty} P\left[|\hat{\theta}_{0, MV}(n) - \theta_0| < \delta\right] = 1, \forall \delta > 0$$

EMV2: Os EMVs são assintoticamente eficientes, ou seja, $\sqrt{n} \times (\hat{\theta}_{0, MV}(n) - \theta_0)$ possui, dentre todos os estimadores consistentes de θ_0 , a menor matriz de covariância – a razão de se multiplicar $(\hat{\theta}_{0, MV}(n) - \theta_0)$ por \sqrt{n} é obter-se uma matriz de covariância finita e não nula ao se tomar o limite quando n tende a infinito.

EMV3: Os EMVs possuem uma distribuição assintoticamente normal multivariada, conforme indicado a seguir.

$$\hat{\theta}_{0, MV} \xrightarrow{d} N(\mu = \theta_0; \Sigma^2 = (I_E(\theta_0))^{-1}),$$

onde $(I_E(\theta_0))^{-1}$ é o inverso da matriz de informação esperada $I_E(\bar{\theta}_0)$, a qual é dada por:

$I_E(\theta_0) = E\left[-\frac{\partial^2 \ell(\theta_0)}{\partial \theta \partial \theta'}\right]$, ou seja, $[I_E(\theta_0)]_{i,j} = E\left[-\frac{\partial^2 \ell(\theta_0)}{\partial \theta_i \partial \theta_j}\right]$. Assim, a matriz de informação esperada $I_E(\theta_0)$ mede a curvatura esperada da superfície de log-verossimilhança.

Logo, com um intervalo de confiança $(1 - \phi)$ aproximado para θ_i , é dado por:

$$\hat{\theta}_i \pm z_{\frac{\phi}{2}} \sqrt{\Psi_{i,i}(\theta_0)},$$

onde $\Psi(\bar{\theta}_0)$ é o inverso da matriz de informação esperada $I_E(\theta_0)$.

Na prática, substituímos θ_0 pela sua estimativa $\hat{\theta}_0$ no cálculo da matriz de informação e obtemos o intervalo $\hat{\theta}_i \pm z_{\frac{\phi}{2}} \sqrt{\Psi_{i,i}(\hat{\theta}_0)}$ para a estimativa do parâmetro θ_i com $(1 - \phi)$ de confiança.

Referências

- Chavez-Demoulin, V., Davison, A.C., McNeil, A.J (2005): "A point process approach to Value-at-Risk estimation", *Quantitative Finance* 5(2), 227-234.
- Coles, S. G. (2001): "An Introduction to Statistical Modeling of Extreme Values", Springer, London.
- Cox, D. R., and V. S. Isham (1979): "Point Processes", Chapman and Hall, London.
- Davison, A. C. (1984a): "Modelling excesses over high thresholds, with an application", in *Statistical Extremes and Applications*, ed. by J. de Oliveira, pp. 461-482. D. Reidel.
- Davison, A. C. (1984b): "A statistical model for contamination due to long-range atmospheric transport of radionuclides", Ph.D. Thesis, Department of Mathematics, Imperial College of Science and Technology, London.
- Davison, A. C., and R. L. Smith (1990): "Models for exceedances over high thresholds (with discussion)", *Journal of The Royal Statistical Society, B*, 52, 393-442.
- Embrechts, P., C. Klüppelberg, and T. Mikosch (1997): *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
- Heij, C., Boer, P., Franses, P. H., Kloek, T. and Dijk, H. K. (2004): "Econometric Methods with Applications in Business and Economics", Oxford University Press, New York.
- Hosking, J., and Wallis, J. (1987): "Parameter and quantile estimation for the generalized Pareto distribution", *Technometrics*, 29, 339-349.
- Jorion, P. (2001): *Value at Risk: the New Benchmark for Measuring Financial Risk*. McGraw-Hill, New York.
- Leadbetter, M. (1991): "On a basis for Peaks over Threshold modeling", *Statistics and Probability Letters*, 12, 357-362.
- Matthys, G., and Beirlant, J. (2003): "Estimating the Extreme Value Index and High Quantiles with Exponential Regression Models", *Statistica Sinica*, 13, 853-880.
- McNeil, A., and R. Frey (2000): "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach", *Journal of Empirical Finance*, 7, 271-300.
- Ogata, Y. (1988): "Statistical models for earthquake Occurrences and residuals analysis for point processes", *Journal of the American Statistical Association*, 83, 9-27.

Pickands, J. (1975): "Statistical inference using extreme order statistics", *The Annals of Statistics*, 3, 119-131.

Ross, S. M. (1987): "Introduction to Probability and Statistics for Engineer and Scientists", John Wiley & Sons, Berkley.

Smith, R. L. (1985): "Maximum likelihood estimation in a class of nonregular Cases", *Biometrika*, 72, 67-92.