

Testes para avaliação das previsões do *value-at-risk* e *expected shortfall*

(Assessing value-at-risk and expected shortfall predictions)

Jaime Enrique Lincovil[†]

Chang Chiann[‡]

Resumo

A avaliação das previsões de medidas de risco, como o *value-at-risk* (VaR) e o *expected shortfall* (ES), é uma atividade muito relevante para as instituições financeiras. Os testes que realizam esta avaliação foram introduzidos com o objetivo de verificar a eficiência destas previsões. Neste trabalho comparamos o poder empírico de novas classes de testes de avaliação presentes na literatura. Além disso, utilizamos estes procedimentos para avaliar a eficiência das previsões geradas pelo método *Historical Simulation* e métodos baseado na distribuição generalizada de Pareto. Para avaliar previsões do VaR, a classe de testes Geométrico-VaR destacou por ter um poder empírico maior que os outros testes nos cenários simulados. Isto evidencia as vantagens de trabalhar com durações discretas e incluir covariáveis nos modelos de testes. Para avaliar previsões do ES, aqueles baseados em distribuições condicionais do retorno ao VaR mostraram um alto poder para grandes tamanhos amostrais. Adicionalmente, mostramos que o método baseado na distribuição generalizada de Pareto utilizando durações e covariáveis resulta um ótimo desempenho nas previsões do VaR e do ES segundo a avaliação feita pelos testes considerados.

Palavras-chave: Cobertura condicional; Expected shortfall; Poder empírico; Teste de avaliação; Value-at-risk

Código JEL: C12, C15.

Abstract

Evaluating forecasts of risk measures, such as *value-at-risk* (VaR) and *expected shortfall* (ES), is an important process for financial institutions. Backtesting procedures were introduced to assess the efficiency of these forecasts. In this paper, we compare the empirical power of new classes of backtesting, for VaR and ES, from the statistical literature. Further, we employ these procedures to evaluate the efficiency of the forecasts generated by both the Historical Simulation method and two methods based on the Generalized Pareto Distribution. To evaluate VaR forecasts, the empirical power of the Geometric-VaR class of backtesting was, in general, higher than that of other tests in the simulated scenarios. This supports the advantages of using defined time periods and covariates in the test procedures. On the other hand, to evaluate ES forecasts, backtesting methods based on the conditional distribution of returns to the VaR performed well with large sample sizes. Additionally, we show that the method based on the generalized Pareto distribution using durations and covariates has optimal performance in forecasts of VaR and ES, according to backtesting.

Keywords: Backtesting; Conditional coverage; Empirical power; Expected shortfall; Value-at-risk

JEL Code: C12, C15.

1. Introdução

No contexto do mercado financeiro, as decisões de compra ou venda de ativos financeiros devem ser tomadas segundo o risco que elas implicam. Portanto, a medição do risco dessas resoluções é uma atividade muito relevante. O *value-at-risk* (VaR) e o *expected shortfall* (ES) são medidas muito utilizadas pelas instituições financeiras para mensurar o risco de mercado, uma vez que elas resumem o risco em uma única medida numérica, a qual é facilmente comunicável para as instituições financeiras.

Tanto o VaR como o ES para um retorno futuro são previstos através de um modelo ajustado utilizando os retornos passados. Uma boa estimação do modelo de previsão apresenta grandes dificuldades devido ao fato

Submitted on April 9, 2019. Revised on December 11, 2019. Accepted on December 11, 2019. Published online in December 2019.
Editor in charge: Marcelo Fernandes.

[†]Departamento de Estatística – Universidade de São Paulo, Brazil. E-mail: outliersj@gmail.com

[‡]Departamento de Estatística – Universidade de São Paulo, Brazil. E-mail: chang@ime.usp.br

que, em geral, o comportamento dos retornos é difícil de modelar. As complicações da modelagem são os fatos estilizados da série de retornos como: o agrupamento de volatilidade, variância dinâmica dependente do tempo, caudas pesadas da distribuição entre outros. Dadas estas dificuldades, é necessário avaliar o desempenho das previsões do VaR e do ES sob uma determinada definição de eficiência.

No contexto regulamentativo, o Comitê de Supervisão Bancária de Basiléia (*Basel Committee on Banking Supervision* (BCBS)) é a principal organização reguladora de estratégias para o controle de risco dos bancos. Em 1996, o BCBS promulgou a regulação que recomenda o uso generalizado do VaR (BCBS, 1996). Sob esta regulamentação, os bancos podem gerenciar suas próprias previsões do VaR. Recentemente, o BCBS sugeriu substituir o VaR pelo ES, no intuito de mitigar as fraquezas encontradas na implementação do VaR em situações extremas como: a falta de informação do volume da perda potencial e a violação da propriedade de coerência, para determinar os requisitos de capital econômico no contexto do gerenciamento do risco de mercado (BCBS, 2013). Pela necessidade de conhecer as propriedades teóricas e empíricas das medidas de risco utilizadas surge a necessidade de avaliar as previsões.

A avaliação pode ser realizada através de procedimentos denominados como *backtestings*, os quais testam a hipótese de eficiência de uma sequência de previsões. No caso do VaR, o critério de eficiência é derivado do critério de intervalos de previsão eficientes introduzido por Christoffersen (1998). No caso do ES, apresentaremos o critério sugerido por Acerbi & Szekely (2014). É importante destacar que tais testes avaliam o conjunto de previsões realizadas e não o ajuste do modelo de previsão.

Uma das opções presentes na literatura para determinar o desempenho dos *backtestings* é comparar o respectivo poder empírico via simulação de Monte Carlo. Neste trabalho, comparamos o poder empírico (em amostras finitas) de novos testes em diferentes cenários simulados, uma vez que não há muitos estudos na literatura que comparam o poder empírico de tais testes num mesmo cenário. Neste trabalho, os procedimentos de avaliação do VaR considerados são: (i) o teste do *Generalized Moment Method* (GMM) proposto por Candelon, Colletaz, Hurlin & Tokpavi (2010), (ii) o teste *Conditional Autoregressive VaR* (CAViaR) em versão de modelo linear generalizado implementado por Berkowitz, Christoffersen & Pelletier (2011), (iii) o teste Geométrico–VaR (GVaR) apresentado por Pelletier & Wei (2015). Para a avaliação do ES, consideramos dois testes propostos por Acerbi & Szekely (2014) e também o teste baseado na distribuição truncada em Righi & Ceretta (2013). Na aplicação para dados reais, empregamos os métodos *Historical Simulation* (HS), *Peaks–Over–Threshold* (POT) e *Duration based Peaks–Over–Threshold* (DPOT) nas previsões do VaR e do ES para uma série do Ibovespa, tais previsões foram avaliadas utilizando os testes citados anteriormente. Empregamos a linguagem R (R Core Team, 2018) para desenvolver as aplicações computacionais.

Como os principais resultados obtidos neste trabalho, o estudo de simulação do poder empírico do VaR mostrou que a inclusão de previsões passadas do VaR como covariáveis dentro do modelo de regressão discreto para as durações incrementa o poder dos testes. Isto foi evidenciado no alto poder dos testes CAViaR e GVaR. Por outro lado, o poder dos testes do ES mostrou uma forte dependência da diferença entre a previsão do ES sob a hipótese nula e a previsão deste sob a hipótese alternativa. Na aplicação, mostramos que o método DPOT (a extensão do POT) gerou previsões eficientes, segundo os testes, nos casos em que POT não conseguiu. Além disso, o método HS, embora simples, também gerou bons resultados.

O presente trabalho é organizado da seguinte forma. Na Seção 2 apresentamos o conceito do VaR, violações do VaR, durações entre violações do VaR e a definição do ES. Na Seção 3 descrevemos brevemente os métodos de previsão HS, POT e DPOT. Nas Seções 4 e 5 expomos o principal tópico deste trabalho, isto é, os testes de avaliação do VaR e do ES. Nas Seções 6 e 7 apresentamos os resultados das simulações, uma aplicação e as avaliações das previsões do VaR e do ES, respectivamente. Finalmente, na Seção 8 discutimos as principais conclusões do trabalho.

2. VaR, violações do VaR, durações e o ES

Seja P_t o preço de um ativo no instante t e $r_t = \log(P_t) - \log(P_{t-1})$ o correspondente log–retorno (ou simplesmente *retorno*) ao instante t . Para um $R_t = (P_t - P_{t-1})/P_{t-1}$ pequeno, é sabido que $r_t \approx R_t$, isto é, o retorno r_t representa aproximadamente a variação de preços de ativo do tempo $t - 1$ ao t . O VaR ao tempo t com probabi-

lidade $p \in (0,1)$ é o p -quantil $\inf\{l : F_t(l) \geq p\}$, em que F_t é a função de distribuição acumulada do retorno r_t . Em particular, a definição formal da previsão do VaR para uma posição comprada para distribuições contínuas é dada a seguir.

Definição 2.1 Definimos a previsão $\text{VaR}_{t|t-1}(p)$ feita no tempo $t-1$ com horizonte de uma unidade de tempo e com probabilidade de cobertura p , como o p -quantil que satisfaz

$$p = \Pr(r_t < \text{VaR}_{t|t-1}(p) | \Omega_{t-1}),$$

em que Ω_{t-1} é um conjunto que contém informações até o tempo $t-1$.

Esta previsão do VaR pode ser considerada como uma medida de variação negativa máxima dos retornos. Neste caso, quando o retorno tem uma variação negativa maior que aquela prevista pela previsão do VaR é entendido que a variação dos retornos violou o limite estabelecido pelo VaR. Este fato é conhecido na literatura como *violação* da previsão do VaR. Definimos a variável indicadora da violação do VaR como:

$$I_t = \begin{cases} 1, & \text{se } r_t < \text{VaR}_{t|t-1}(p), \\ 0, & \text{caso contrário.} \end{cases}$$

Além disso, definimos a *duração* da i -ésima violação do VaR por:

$$D_i = t_i - t_{i-1},$$

em que t_i denota o instante da i -ésima violação do VaR. Para $i = 1$ assumimos que $t_0 = 0$.

Outra medida de risco muito utilizada atualmente é o ES. O ES foi introduzido por Rappoport (1993) e logo Artzner, Delbaen, Eber & Heath (1997), Artzner, Delbaen, Eber & Heath (1999) estudaram formalmente suas propriedades. Formalmente o $\text{ES}(p)$ associado ao $\text{VaR}(p)$ é definido por $\text{ES}(p) = 1/(1-p) \int_p^1 \text{VaR}(u) du$. De forma simplificada, o ES nada mais é que a perda esperada dado que o retorno viola o VaR, isto é $\text{ES}(p) = \mathbb{E}[r_t | r_t < \text{VaR}(p)]$ (ver Lemma 2.16 em McNeil, Frey & Embrechts (2005)). A definição da previsão do ES a um passo a frente é dada a seguir.

Definição 2.2 Definimos a previsão $\text{ES}_{t|t-1}(p)$ feita no tempo $t-1$ com horizonte de uma unidade de tempo associada com a previsão $\text{VaR}_{t|t-1}(p)$, que satisfaz $p = \Pr(r_t < \text{VaR}_{t|t-1}(p) | \Omega_{t-1})$ para alguma informação passada Ω_{t-1} , como sendo a esperança condicional:

$$\text{ES}_{t|t-1}(p) = \mathbb{E}[r_t | r_t \leq \text{VaR}_{t|t-1}(p)].$$

A seguir apresentamos os métodos de previsão do VaR e do ES utilizados neste trabalho.

3. Métodos HS, POT e DPOT para a previsão do VaR e do ES

O método HS é um procedimento simples de previsão o qual assume que os processos financeiros históricos se repetem (ver Perignon & Smith (2008)). Aplicando esta suposição, o método HS estima o VaR para o instante t através do percentil da série $\{r_s\}_{s=0}^{t-1} = \{r_0, \dots, r_{t-1}\}$, isto é:

$$\widehat{\text{VaR}}_{t|t-1}^{\text{HS}}(p) = \text{percentil} \{ \{r_s\}_{s=0}^{t-1}, 100p \}.$$

Além disso, o método HS prevê o ES por meio da seguinte média condicional:

$$\widehat{\text{ES}}_{t|t-1}^{\text{HS}}(p) = \frac{1}{N^*} \sum_{s \leq t-1} r_s I_s^*,$$

em que I_s^* é a indicadora da violação do $\widehat{\text{VaR}}_{s|s-1}^{\text{HS}}(p)$ e $N^* = \sum_{s \leq t-1} I_s^*$. Dada a rapidez de cálculo e a facilidade de implementação, este método é amplamente utilizado.

O método POT de previsão do VaR considera o ajuste de uma Distribuição Generalizada de Pareto (DGP) baseada nos excessos dos retornos com respeito a um limiar $\mu > 0$. Seja $y_t = r_t - \mu > 0$ (para $r_t > 0$) o excesso do retorno r_t com respeito a μ e $F_\mu(y_t)$ a função de distribuição acumulada de y_t . Para um μ suficientemente grande, $F_\mu(y_t)$ pode ser aproximada pela função de distribuição acumulada da DGP (ver Embrechts, Klüppelberg & Mikosch (2013)), com parâmetros de escala $\sigma > 0$ e de forma $\gamma \in \mathbb{R}$, dada por:

$$G_{\gamma, \sigma}(y_t) = \begin{cases} 1 - (1 + \gamma y_t / \sigma)^{-1/\gamma}, & \text{se } \gamma \neq 0, \\ 1 - \exp(-y_t / \sigma), & \text{se } \gamma = 0. \end{cases}$$

O suporte da DGP é $y \geq 0$ quando $\gamma \geq 0$ e $0 \leq y \leq -\sigma/\gamma$ quando $\gamma < 0$.

Para um limiar $\mu > 0$ fixo, a previsão do VaR com probabilidade de cobertura p e horizonte de uma unidade temporal obtida pelo método POT segue:

$$\widehat{\text{VaR}}_{t|t-1}^{\text{POT}}(p) = \mu + \frac{\hat{\sigma}}{\hat{\gamma}} \left(\left(\frac{M}{Lp} \right)^{\hat{\gamma}} - 1 \right),$$

em que M é o número de excessos acima de μ em uma amostra com L retornos e $\hat{\sigma}$ e $\hat{\gamma}$ são os respectivos estimadores de máxima verossimilhança (ver Tsay (2005)). Uma simples interpretação do $\widehat{\text{VaR}}_{t|t-1}^{\text{POT}}(p)$ em função de μ e $\hat{\sigma}$ segue: μ atua como um intercepto fixo, $\hat{\sigma}$ estima o impacto de $1/\hat{\gamma}((M/Lp)^{\hat{\gamma}} - 1)$ na previsão $\widehat{\text{VaR}}_{t|t-1}^{\text{POT}}(p)$, em quanto $\hat{\gamma}$ regula a escala do impacto da proporção de excessos dos retornos com respeito a μ . Baseado em Coles, Bawa, Trenner & Dorazio (2001), a previsão do ES baseada na abordagem POT é dada por:

$$\widehat{\text{ES}}_{t|t-1}^{\text{POT}}(p) = \frac{1}{1 - \hat{\gamma}} \left(\widehat{\text{VaR}}_{t|t-1}^{\text{POT}}(p) + \hat{\sigma} - \hat{\gamma} \times \mu \right).$$

Neste caso, $\widehat{\text{ES}}_{t|t-1}^{\text{POT}}(p)$ cresce (ou decresce) junto com $\hat{\sigma}$. Esta metodologia exerce um bom desempenho quando considera-se uma subamostra extrema de acordo com μ . Contudo, Santos & Alves (2013) evidenciam empiricamente que as previsões geradas pelo método POT provocam *agrupamentos de violações do VaR* em períodos de alta volatilidade dos retornos. Com o objetivo de contornar este problema, propõem incorporar como covariáveis as durações entre os tempos de excessos dos retornos no método DPOT (Duration based POT). DPOT propõe modelar o parâmetro de escala (variando no tempo) por $\sigma_t = \alpha / (d_{t,\lambda})^c > 0$, em que $\alpha > 0$, $c > 0$, $\lambda \in \mathbb{N}_+$ e $d_{t,\lambda}$ é a duração entre o tempo t e o instante anterior no qual ocorreu o primeiro dos λ excessos passados. Neste modelo, o impacto de σ_t ocorre na direção contrária da duração, isto é, σ_t decresce conforme o aumento da duração dos λ excessos passados. O impacto de $d_{t,\lambda}$ é regulada pelo parâmetro c . Em consequência, o VaR e o ES variam em conformidade com $\hat{\sigma}_t$ em ambas as abordagens. A previsão do VaR com probabilidade de cobertura p e horizonte de uma unidade temporal obtida pelo método DPOT é dada por:

$$\widehat{\text{VaR}}_{t|t-1}^{\text{DPOT}}(p) = \mu + \frac{\hat{\alpha}}{\hat{\gamma}(d_{t,\lambda})^c} \left(\left(\frac{M}{Lp} \right)^{\hat{\gamma}} - 1 \right).$$

Os parâmetros λ e c são constantes determinadas segundo critérios empíricos. Além disso, a previsão do ES baseada no método DPOT é dada por:

$$\widehat{\text{ES}}_{t|t-1}^{\text{DPOT}}(p) = \frac{\widehat{\text{VaR}}_{t|t-1}^{\text{DPOT}}(p)}{1 - \hat{\gamma}} + \frac{(\hat{\alpha}/d_{t,\lambda})^c - \hat{\gamma} \times \mu}{1 - \hat{\gamma}}.$$

O foco deste trabalho é comparar o poder empírico dos testes para avaliação da eficiência das previsões e não nos métodos de previsão do VaR e do ER. Para mais detalhes sobre previsão ver Mancini & Trojani (2011), Abad, Benito & López (2014), Nadarajah, Zhang & Chan (2014), Krause & Paoletta (2014), Nieto & Ruiz (2016), Trucíos, Hotta & Ruiz (2017), Ferraty & Quintela-Del-Río (2019) e Taylor (2019).

4. Backtesting do VaR

Neste trabalho, consideraremos a definição de eficiência utilizada em trabalhos como Berkowitz et al. (2011) e Pelletier & Wei (2015), entre muitos outros. A definição da hipótese de eficiência é elaborada de forma diferente de acordo com o modelo de cada teste, embora o critério de eficiência pode ser o mesmo. O conceito de eficiência é basicamente uma adaptação da definição de intervalos de previsão eficientes introduzida por Christoffersen (1998).

Definição 4.1 *Seja $r_1, \dots, r_{t_0-1}, r_{t_0}, \dots, r_T$ uma série de retornos e r_{t_0}, \dots, r_T uma pós-amostra. Dizemos que a sequência $\text{VaR}_{t_0|t_0-1}(p), \dots, \text{VaR}_{T|T-1}(p)$ é eficiente se*

$$p = \Pr(r_t < \text{VaR}_{t|t-1}(p) | \Omega_{t-1}) = \Pr(I_t = 1 | \Omega_{t-1}), \forall t \in \{t_0, \dots, T\}.$$

Pelo Lema 1 em Christoffersen (1998), a sequência $\text{VaR}_{t_0|t_0-1}(p), \dots, \text{VaR}_{T|T-1}(p)$ é eficiente se e somente se a sequência $I_s | \Omega_{s-1} \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$ para $s \in \{t_0, \dots, T\}$, ou equivalentemente, se $D_i | \Omega_{t_i-1} \stackrel{\text{i.i.d.}}{\sim} \text{Geométrica}(p)$ para $i \in \{1, \dots, N\}$, em que t_i é o tempo da i -ésima violação. Isto é, uma sequência de previsões do VaR é eficiente se as respectivas violações acontecem de forma independente, com probabilidade de cobertura constante p ao longo do processo de previsão. Uma sequência eficiente segundo a Definição 4.1 é dita ter uma *cobertura condicional correta* (CC correta). Consequentemente, a hipótese (nula) da CC correta é definida por H_{CC} : “ $I_s | \Omega_{s-1} \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$ para $s \in \{t_0, \dots, T\}$ ” ou seu equivalente em termos das durações. Neste trabalho, consideramos as informações passadas até tempo $t-1$ dadas por $\Omega_{1,t-1} = \{I_{t-1}, \dots, I_1\}$ e $\Omega_{2,t-1} = \{I_{t-1}, \dots, I_1, r_{t-1}, \dots, r_1\}$.

4.1 Testes baseados no método dos momentos generalizados

O teste GMM para testar H_{CC} com respeito a $\Omega_{1,t}$, apresentado em Candelon et al. (2010), utiliza o polinômio ortogonal da distribuição Geométrica de uma determinada ordem. A seguir, definimos o $(j+1)$ -ésimo termo deste polinômio.

Definição 4.2 *O $(j+1)$ -ésimo termo do polinômio ortogonal associado à distribuição Geométrica com probabilidade de sucesso p avaliado na duração observada $D = d$ é definido pela seguinte relação recursiva:*

$$M_{j+1}(d, p) = \frac{(1-p)(2j+1) + p(j-d+1)}{(j+1)\sqrt{1-p}} M_j(d, p) - \left(\frac{j}{j+1}\right) M_{j-1}(d, p),$$

para $j \in \mathbb{N}$, com $M_{-1}(d, p) = 0$ e $M_0(d, p) = 1$. Assim, o polinômio ortogonal de ordem k associado à distribuição Geométrica é dado por $\sum_{j=1}^k M_j(d, p)$.

Seja D_1, \dots, D_N uma sequência de durações entre violações do VaR e seja $\mathbb{E}[M_{j+1}(D_i, p)]$ a esperança do polinômio aleatório $M_{j+1}(D_i, p)$. A abordagem GMM utiliza a seguinte propriedade dos termos do polinômio $\sum_{j=1}^k M_j(d, p)$ como segue: se a sequência $D_1 | \Omega_{t_1-1}, \dots, D_N | \Omega_{t_N-1}$ é i.i.d. com distribuição Geométrica(p), então $\mathbb{E}[M_{j+1}(D_i, p)] = 0, \forall j \in \mathbb{N}, \forall i = 1, \dots, N$ (Candelon et al., 2010, p.320). Sob estas condições de momentos, a hipótese H_{CC} com respeito ao polinômio ortogonal de ordem k é formulada como:

$$H_{CC} : \mathbb{E}[M_j(D_i, p)] = 0, \forall j \in \{1, \dots, k\}, \forall i \in \{1, \dots, N\}.$$

Se assumimos que o processo das durações $\{D_i : i \geq 1\}$ é estacionário e ergódico, então sob H_{CC} a estatística do teste $J_{CC}(k)$ tem a seguinte propriedade assintótica:

$$L_1 = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{M}(D_i, p) \right)^\top \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{M}(D_i, p) \right) \xrightarrow{\mathcal{D}} \chi_k^2,$$

em que $\mathbf{M}(D_i, p) = (M_1(D_i, p), \dots, M_k(D_i, p))^\top$ e $\xrightarrow{\mathcal{D}}$ denota a convergência em distribuição. Nesta abordagem, a hipótese alternativa não é especificada. Ver Hansen (1982) para mais detalhes.

4.2 Teste CAViaR

Seguindo a Engle & Manganelli (2004), Berkowitz et al. (2011) consideram um modelo logístico para a probabilidade de cobertura, em que $I_t | \Omega_{2,t-1} \sim \text{Bernoulli}(\pi_t)$, $\pi_t = \Pr(I_t = 1 | \Omega_{2,t-1}) = (1 + e^{-\eta_t})^{-1}$, $\eta_t = \delta_0 + \delta_1 I_{t-1} + \delta_2 \text{VaR}_{t|t-1}(p)$ é o preditor linear e $\delta = (\delta_0, \delta_1, \delta_2) \in \mathbb{R}^3$.

Sob H_{CC} , π_t não depende de I_{t-1} nem de $\text{VaR}_{t|t-1}(p)$. Ou seja, $H_{CC} : \pi_t = p, \forall t \in \{t_0, \dots, T\}$. As hipóteses do teste utilizando os parâmetros em δ são dadas por:

$$\begin{cases} H_{CC} : \delta_0 = \log\left(\frac{p}{1-p}\right) \text{ e } \delta_1 = 0 \text{ e } \delta_2 = 0, \\ H_A : \delta_0 \neq \log\left(\frac{p}{1-p}\right) \text{ ou } \delta_1 \neq 0 \text{ ou } \delta_2 \neq 0. \end{cases}$$

O teste da razão de verossimilhança generalizada é empregado, cuja estatística L_2 é dada por:

$$L_2 = -2 \left(\mathcal{L}_1(\delta^0; \tilde{I}_T) - \mathcal{L}_1(\hat{\delta}; \tilde{I}_T) \right) \xrightarrow{\mathcal{D}} \chi_3^2,$$

em que $\tilde{I}_T = (I_{t_0}, \dots, I_T)$, $\delta^0 = \log(p/(1-p))$, $\delta^0 = (\delta_0, 0, 0)$, $\hat{\delta} = (\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2)$ é o estimador de máxima verossimilhança de δ e \mathcal{L}_1 é respectiva função de log-verossimilhança.

4.3 Testes Geométrico-VaR (GVaR)

Hass (2005) propõe utilizar a distribuição Weibull discreta tipo II caracterizada em Rinne (2008) para modelar as durações das violações, cuja função de risco é dada por $\lambda_d = \pi d^{b-1}$. Posteriormente, Pelletier & Wei (2015) propõem adicionar a previsão do VaR como covariável na função de risco através do seguinte modelo:

$$\begin{aligned} \lambda_d^i &= \Pr(I_{t_{i-1}+d} = 1 | I_{t_{i-1}+d-1} = 0, \dots, I_{t_{i-1}+1} = 0, I_{t_{i-1}} = 1, \mathbf{r}_i(d)) \\ &= \pi d^{b-1} \exp\{\beta \text{VaR}_{t_{i-1}+d}(p)\}, \end{aligned} \quad (1)$$

em que $\beta \geq 0$, $b > 1$, $\pi \in (0, 1)$ e t_{i-1} é o instante da $(i-1)$ -ésima violação, $t_i = t_{i-1} + d$ é o instante em que ocorreria a i -ésima violação do VaR com uma duração igual a d e $\mathbf{r}_i(d) = \{r_{t_{i-1}+d-1}, r_{t_{i-1}+d-2}, \dots\}$. Deste modo, λ_d^i mede a probabilidade de ocorrer i -ésima violação da previsão $\text{VaR}_{t_{i-1}+d}(p)$ depois de d períodos.

Adicionalmente, neste trabalho consideramos os modelos do tipo logístico e probito para a função de risco das durações. Isto é

$$\lambda_d^i = \frac{\exp\{h(\lambda(d)) + \beta \text{VaR}_{t_{i-1}+d}(p)\}}{1 + \exp\{h(\lambda(d)) + \beta \text{VaR}_{t_{i-1}+d}(p)\}}, \quad (2)$$

$$\lambda_d^i = \Phi\{\Phi^{-1}(\lambda(d)) + \beta \text{VaR}_{t_{i-1}+d}(p)\}, \quad (3)$$

em que $h(u) = \log(u/(1-u))$ e $\Phi(u) = \Pr(Z < u)$ para $Z \sim N(0, 1)$.

Considerando os modelos (1), (2) e (3) a hipótese de CC correta pode ser expressada por $H_{CC} : \lambda_d^i = p, \forall i, \forall d$, ou seja, a probabilidade de ocorrer a i -ésima violação do VaR depois de d períodos é igual a p , não depende de sua duração d e de $\text{VaR}_{t_{i-1}+d}(p)$. As hipóteses do teste são:

$$\begin{cases} H_{CC} : \pi = p \text{ e } b = 1 \text{ e } \beta = 0, \\ H_A : \pi \neq p \text{ ou } b \neq 1 \text{ ou } \beta > 0. \end{cases}$$

Neste caso também é empregado o teste da razão de verossimilhança generalizada cuja estatística L_3 é dada por:

$$L_3 = -2 \left(\mathcal{L}_2(\theta^0; \tilde{D}_N) - \mathcal{L}_2(\hat{\theta}; \tilde{D}_N) \right) \xrightarrow{\mathcal{D}} \chi_3^2,$$

em que $\tilde{D}_N = (D_1, \dots, D_N)$, $\theta^0 = (p, 1, 0)$, $\hat{\theta} = (\hat{\pi}, \hat{b}, \hat{\beta})$ é o estimador de máxima verossimilhança de $\theta = (\pi, b, \beta)$ e \mathcal{L}_2 é a função de log-verossimilhança da distribuição Weibull discreta tipo II considerando uma censura não informativa para a primeira e última duração. O procedimento é denominado como teste Geométrico-VaR (GVaR).

Para os procedimentos apresentados nesta seção, a distribuição assintótica da estatística dos testes pode ser utilizada para calcular o p -valor e tomar decisão sobre H_{CC} . Outra alternativa é estimar o p -valor via métodos de Monte Carlo.

5. Backtesting do ES

Os métodos para a avaliação das previsões do ES que consideramos neste trabalho são do tipo não paramétrico, isto é, eles testam hipóteses que assumem condições gerais sobre uma possível função de distribuição de probabilidade que seguem os retornos.

Seja $K_t(x) = \Pr(r_t < x)$ e $K_t^{[p]}(x) = \Pr(r_t < x | r_t < \text{VaR}(p)) = \min(1, K_t(\cdot)/p)$ a distribuição acumulada e a distribuição caudal acumulada utilizada para realizar as previsões do VaR e do ES, respectivamente. Em termos simples, Acerbi & Szekely (2014) propõem que uma adequada distribuição $K_t^{[p]}$ utilizada para fazer a previsão do ES, deve ser aproximada, ou equivalente, à verdadeira distribuição caudal dos retornos $F_t^{[p]}(\cdot) = \min(1, F_t(\cdot)/p)$. Isto é estabelecido na seguinte definição.

Definição 5.1 Dizemos que a sequência $\text{ES}_{t_0|t_0-1}(p), \dots, \text{ES}_{T|T-1}(p)$ é eficiente se

$$K_t^{[p]} = F_t^{[p]}, \forall t \in \{1, \dots, T\}.$$

Contrariamente à Definição 2.2, esta definição somente considera características populacionais da distribuição dos retornos e não uma avaliação das previsões do ES como a média condicional dos retornos observados da pós-amostra. Caso contrário, as previsões do ES feitas pelo modelo ajustado ($\text{ES}_{t|t-1}^K(p)$) poderiam subestimar (ou superestimar) as previsões geradas do verdadeiro modelo ($\text{ES}_{t|t-1}^F(p)$) ainda que as previsões do VaR ($\text{VaR}_{t|t-1}^K(p)$) fossem iguais àsquelas geradas via F_t ($\text{VaR}_{t|t-1}^F(p)$). Conforme descrição acima e a Definição 2.2, as seguintes hipóteses são estabelecidas para serem testadas.

$$\begin{cases} H_0 : K_t^{[p]} = F_t^{[p]}, \forall t \in \{1, \dots, T\} \\ H_A : \text{ES}_{t|t-1}^F(p) \geq \text{ES}_{t|t-1}^K(p) \text{ para todo } t \text{ e } > \text{ para algum } t, \\ \text{e } \text{VaR}_{t|t-1}^F(p) = \text{VaR}_{t|t-1}^K(p), \forall t \in \{1, \dots, T\}, \end{cases}$$

em que $\text{ES}_{t|t-1}^F(p)$ e $\text{VaR}_{t|t-1}^F(p)$ são as previsões geradas por $F_t^{[p]}$ e $\text{ES}_{t|t-1}^K(p)$ e $\text{VaR}_{t|t-1}^K(p)$ são as previsões geradas por $K_t^{[p]}$, respectivamente. Os seguintes procedimentos foram propostos para testar estas hipóteses. Neste caso, o par de hipóteses são as mesmas para os diferentes testes que apresentaremos a seguir.

5.1 Teste 1

Depois de simples procedimentos algébricos baseados na Definição 2.2, podemos verificar que uma sequência eficiente de previsões $\widehat{\text{ES}}_{1|0}(p), \dots, \widehat{\text{ES}}_{T|T-1}(p)$ deve satisfazer $\mathbb{E} \left[r_t / \widehat{\text{ES}}_{t|t-1}(p) - 1 \mid r_t < \widehat{\text{VaR}}_{t|t-1}(p) \right] = 0$ para cada $t \in \{1, \dots, T\}$. Seja I_1, \dots, I_T uma amostra das violações do VaR correspondente a uma série de retornos r_1, \dots, r_T e $N_T = \sum_{t=1}^T I_t$. A estatística¹ proposta por Acerbi & Szekely (2014) para o Teste 1 é dada por:

$$Z_1 = \frac{1}{N_T} \left[\sum_{t=1}^T \frac{r_t I_t}{\widehat{\text{ES}}_{t|t-1}(p)} \right] - 1.$$

¹Em Acerbi & Szekely (2014), $-\text{VaR}(p) < 0$ e $-\text{ES}(p) < 0$, obtendo uma versão da estatística Z_1 dada por $(1/N_T) \sum_{t=1}^T r_t I_t / (\widehat{\text{ES}}_{t|t-1}(p)) + 1$. O mesmo aplica para as estatísticas Z_2 e Z_3 . Nossa versão é equivalente.

Note que, sob H_0 , $\mathbb{E}[Z_1] = 0$, mas sob H_A , $\mathbb{E}[Z_1] < 0$. Consequentemente, valores de Z_1 próximos de zero indicam evidência a favor de H_0 e evidência em contra caso contrário.

5.2 Teste 2

Rescrevendo o ES dado na Definição 2.2 como uma esperança não condicional, uma correta previsão do ES deve satisfazer $\widehat{\text{ES}}_{t|t-1}(p) = \mathbb{E}[r_t I_t / p]$ para todo $t \in \{1, \dots, T\}$. Analogamente ao caso anterior, Acerbi & Szekely (2014) propõem uma segunda opção de estatística de teste para as anteriores hipóteses dada por:

$$Z_2 = \frac{1}{Tp} \left[\sum_{t=1}^T \frac{r_t I_t}{\widehat{\text{ES}}_{t|t-1}(p)} \right] - 1.$$

Como no Teste 1, sob H_0 , $\mathbb{E}[Z_2] = p^{-1} - 1$, mas sob H_A , $\mathbb{E}[Z_2] < p^{-1} - 1$. Analogamente ao Teste 1, valores de Z_2 próximos de zero indicam evidência a favor de H_0 e evidência em contra no caso contrário.

5.3 Teste 3

Em outra linha, Righi & Ceretta (2013) propõem estudar as flutuações da distribuição padronizada truncada do retorno r_t , isto é, a distribuição da transformação $(r_t - \widehat{\text{ES}}_{t|t-1}(p))I_t / \text{DP}_{t|t-1}$, em que $\text{DP}_{t|t-1}(p) = \sqrt{\widehat{\text{Var}}(r_t | r_t < \text{VaR}_{t|t-1}(p))}$. Portanto, propomos considerar a estatística

$$Z_3 = \frac{1}{N_T} \sum_{t=1}^T \left(\frac{(r_t - \widehat{\text{ES}}_{t|t-1}(p))I_t}{\sqrt{\widehat{\text{Var}}(r_t | r_t < \text{VaR}_{t|t-1}(p))}} \right),$$

para procedimento de teste, em que $\widehat{\text{Var}}(r_t | r_t < \text{VaR}_{t|t-1}(p)) = \sum_{s=t_{a_0}}^{t-1} (r_s - \widehat{\text{ES}}_{t|t-1}(p))^2 I_s / ws$ e t_{a_0} é o tempo do primeiro retorno da amostra utilizada para estimar $\widehat{\text{ES}}_{t|t-1}(p)$. A interpretação dos valores de Z_3 é análoga aos Testes 1 e 2. Para os três *backtesting* do ES apresentados aqui, decidimos sob H_0 usando o p -valor de Monte Carlo descrito na Seção 9.3.

6. Poder empírico dos Backtesting

Nesta seção, comparamos o poder empírico dos procedimentos de testes apresentados, isto é, a probabilidade de rejeitar H_A dado que esta de fato é falsa. Para o cálculo analítico do poder, precisamos conhecer a distribuição da estatística de teste sob H_A . Mas, geralmente, as distribuições das estatísticas dos testes apresentados nas Seções 4 e 5 são difíceis de caracterizar analiticamente. Além disso, não é possível garantir a convergência em distribuição da estatística do teste GVAr sob a H_{CC} dado que esta restringe alguns parâmetros na fronteira do espaço paramétrico. Adicionalmente, a literatura evidência que o uso da distribuição assintótica leva a superestimar (ou infraestimar) a taxa do erro tipo I dos testes. Seguindo a tendência de trabalhos anteriores, implementamos testes do tipo de Monte Carlo apresentado por Dufour (2006) para o cálculo de poder empírico contornando os problemas anteriores. A implementação computacional dos testes e estimação do poder foram elaboradas tomando como base as funções fornecidas por Santos (2010).

Nesta seção comparamos o poder empírico dos testes apresentados nas Seções 4 e 5 sob diferentes cenários. Em todos os casos, o poder empírico nada mais é que a frequência de rejeição da respectiva hipótese nula no total de 5000 replicas nas quais são geradas previsões do VaR e do ES não eficientes. Consideramos os tamanhos pós-amostrais (n) de 250, 500, 750 e 1000, um horizonte de previsão de uma unidade, tamanhos amostrais para o ajuste do modelo (ws) de previsão de 250 e 500, um nível de significância $\alpha = 0,1$ e as probabilidades de cobertura (p) de 0,01 e 0,05.

6.1 Poder empírico dos testes de avaliação do VaR

Para o cálculo do poder empírico dos exercícios de *backtesting* do VaR, utilizamos Processos Geradores dos Dados (PGDs) heteroscedásticos, descritos no Apêndice 9.5, para obter a série de retornos. Por exemplo, Christoffersen & Pelletier (2004) apresentam um trabalho interessante comparando o poder dos *backtesting*: Kupiec, CC de Christoffersen, CaViaR e Geométrico. Os PGDs utilizados são os modelos tipo GARCH. Neste trabalho, o PGD1 é o APARCH(1,1) que gera series de retornos com comportamento assimétrico, o PGD2 é o GARCH(1,1)- $t(v)$ com efeitos assimétricos nas inovações e o PGD3 é o AR(3)-TGARCH(1,1) com efeitos assimétricos ponderados na volatilidade. Estes processos heteroscedásticos são amplamente utilizados na literatura devido a sua capacidade de gerar series com agrupamento da volatilidade em intervalos temporais.

Uma vez que os dados sejam gerados, utilizamos o método HS para obter sequências de previsões do VaR não eficientes. Estas previsões geram *agrupamento das violações* do VaR, o qual segundo a literatura é considerado evidência em contra H_{CC} (ver Berkowitz et al. (2011) página 2222 e Pelletier & Wei (2015) página 735. Neste caso, é esperado que os testes de avaliação devem rejeitar H_{CC} . Os passos mais detalhados para obter o p -valor e o poder empírico podem ser encontrados nos Apêndices 9.1 e 9.2, respectivamente.

As seguintes notações são usadas nas tabelas 1 a 6 que mostram os resultados obtidos nas simulações. Os testes GMM baseados nos polinômios de ordem (k) 3 e 5, são denotados por GMM₃ e GMM₅, respectivamente. O teste CAViaR é denotado simplesmente por CAViaR. Os testes Geométrico-VaR baseados nos modelos (1), (2) e (3) são denotados por GVAr₁, GVAr₂ e GVAr₃, respectivamente. A Frequência de Amostras Rejeitadas (FAR) num total de 5000 réplicas para cada teste foi calculada seguindo as especificações dos respectivos autores. Para os testes do tipo GMM e CAViaR aceitamos amostras com pelo menos 2 violações (FAR1) e para os testes do tipo GVAr aceitamos amostras com pelo menos três durações (FAR2). O poder empírico dos testes obtido no estudo de simulação é apresentado nas Tabelas 1-6.

Tabela 1
Poder empírico para os testes de cobertura condicional. PGD1 e $ws = 250$.

	$p = 0,01$				$p = 0,05$			
	$n = 250$	$n = 500$	$n = 750$	$n = 1000$	$n = 250$	$n = 500$	$n = 750$	$n = 1000$
GMM ₃	0,379	0,619	0,685	0,780	0,687	0,866	0,935	0,971
GMM ₅	0,389	0,661	0,741	0,827	0,688	0,898	0,930	0,968
CAViaR	0,599	0,732	0,786	0,871	0,691	0,832	0,804	0,864
GVAr ₁	0,491	0,729	0,845	0,940	0,765	0,932	0,974	0,993
GVAr ₂	0,478	0,732	0,852	0,941	0,749	0,934	0,974	0,992
GVAr ₃	0,525	0,772	0,860	0,944	0,796	0,940	0,975	0,992
FAR1	$4,2 \times 10^{-3}$	0,002	0	0	0,013	0	0	0
FAR2	$5,2 \times 10^{-3}$	0,002	0	0	0,014	0	0	0

Nas Tabelas 1 e 2 observamos o comportamento do poder para o PGD1 e ws igual a 250 e 500. As seguintes observações valem para ambas as tabelas. Não distinguimos uma diferença notória do desempenho dos testes do tipo GMM. No caso dos procedimentos GVAr, notamos que quando é considerado o modelo (3) o poder é levemente maior. Adicionalmente, em ambas as tabelas, quando $n = 250$ o teste CAViaR foi mais poderoso, mas para os outros casos notamos que os testes GVAr têm um melhor desempenho para ambos os valores de ws . Observamos também, que o comportamento do poder não depende de ws .

Na Tabela 3, o teste CAViaR foi mais poderoso para todos os tamanhos amostrais quando $ws = 250$, mas para $ws = 500$ e partir de $n = 500$ os testes GVAr foram mais poderosos. A Tabela 4 mostra que o teste GVAr para o modelo (1) é mais poderoso a medida que n cresce. Finalmente, tanto na Tabela 3 quanto na Tabela 4, não distinguimos uma grande diferença entre o desempenho dos testes GMM e também que para $n = 250$ o teste CAViaR foi mais poderoso.

O comportamento do poder para o PGD3 são similares ao PGD2. Na Tabela 5, o teste CAViaR foi mais poderoso para $ws = 250$ e $n = 250, 500, 750$. Para $ws = 500$, a classe de testes GVAr teve um melhor desempenho

Tabela 2
Poder empírico para os testes de cobertura condicional. PGD1 e $ws = 500$.

	$p = 0,01$				$p = 0,05$			
	$n = 250$	$n = 500$	$n = 750$	$n = 1000$	$n = 250$	$n = 500$	$n = 750$	$n = 1000$
GMM ₃	0,388	0,592	0,673	0,751	0,721	0,902	0,952	0,984
GMM ₅	0,409	0,637	0,729	0,801	0,723	0,929	0,951	0,981
CAViaR	0,553	0,650	0,711	0,782	0,659	0,830	0,803	0,848
GVaR ₁	0,466	0,686	0,803	0,902	0,742	0,925	0,974	0,994
GVaR ₂	0,459	0,688	0,817	0,909	0,724	0,925	0,973	0,994
GVaR ₃	0,489	0,722	0,838	0,910	0,759	0,936	0,978	0,995
FAR1	0,366	0,087	0,010	0,001	0,044	8×10^{-4}	0	0
FAR2	0,370	0,089	0,011	0,001	0,046	8×10^{-4}	0	0

Tabela 3
Poder empírico para os testes de cobertura condicional. PGD2 e $p = 0,01$.

	$ws = 250$				$ws = 500$			
	$n = 250$	$n = 500$	$n = 750$	$n = 1000$	$n = 250$	$n = 500$	$n = 750$	$n = 1000$
GMM ₃	0,226	0,248	0,319	0,406	0,208	0,242	0,302	0,364
GMM ₅	0,223	0,275	0,355	0,462	0,206	0,272	0,343	0,415
CAViaR	0,469	0,497	0,608	0,694	0,385	0,412	0,461	0,527
GVaR ₁	0,374	0,426	0,502	0,584	0,258	0,421	0,508	0,544
GVaR ₂	0,364	0,423	0,502	0,588	0,258	0,422	0,508	0,544
GVaR ₃	0,356	0,416	0,501	0,592	0,257	0,423	0,502	0,555
FAR1	0,153	0,004	0	0	0,282	0,039	0,002	0
FAR2	0,156	0,004	0	0	0,287	0,040	0,002	0

Tabela 4
Poder empírico para os testes de cobertura condicional. PGD2 e $p = 0,05$.

	$ws = 250$				$ws = 500$			
	$n = 250$	$n = 500$	$n = 750$	$n = 1000$	$n = 250$	$n = 500$	$n = 750$	$n = 1000$
GMM ₃	0,427	0,549	0,637	0,743	0,459	0,599	0,696	0,777
GMM ₅	0,427	0,550	0,633	0,745	0,462	0,600	0,695	0,772
CAViaR	0,503	0,560	0,652	0,719	0,461	0,538	0,566	0,630
GVaR ₁	0,430	0,572	0,703	0,793	0,477	0,621	0,752	0,824
GVaR ₂	0,429	0,567	0,695	0,787	0,477	0,619	0,750	0,820
GVaR ₃	0,430	0,570	0,693	0,790	0,476	0,620	0,748	0,823
FAR1	0,008	0,003	0	0	$3,8 \times 10^{-3}$	0	0	0
FAR2	0,008	0,003	0	0	$3,8 \times 10^{-3}$	0	0	0

e também não notamos grandes diferenças entre os dois testes do tipo GMM. Na Tabela 6, verificamos que os testes do tipo GVaR têm maior poder conforme n aumenta.

Em resumo, notamos que em todas as tabelas o poder empírico dos testes aumenta junto com n . As Tabelas 1 e 2 mostram que em todos os casos os testes GVaR foram mais poderosos quando ws se manteve fixo (exceto no caso $n = 250$ e $p = 0,01$ na Tabela 1). Nas Tabelas 3–6, a probabilidade de cobertura p foi mantida fixa, observamos que na maioria dos casos o teste CAViaR supera aos demais quando $n = 250$. Os testes do tipo MMG tiveram um bom desempenho, mas quase sempre para $n \leq 750$ os testes GVaR tiveram melhor desempenho, o qual indica evidência a favor de trabalhar com durações em vez de indicadores de violações e, além disso,

Tabela 5
Poder empírico para os testes de cobertura condicional. PGD3 e $p = 0,01$.

	$ws = 250$				$ws = 500$			
	$n = 250$	$n = 500$	$n = 750$	$n = 1000$	$n = 250$	$n = 500$	$n = 750$	$n = 1000$
GMM ₃	0,525	0,598	0,697	0,780	0,492	0,599	0,688	0,776
GMM ₅	0,548	0,628	0,740	0,814	0,540	0,648	0,741	0,806
CAViaR	0,662	0,710	0,785	0,849	0,570	0,666	0,713	0,762
GVaR ₁	0,575	0,563	0,765	0,841	0,594	0,607	0,773	0,840
GVaR ₂	0,574	0,562	0,765	0,841	0,593	0,608	0,773	0,840
GVaR ₃	0,584	0,588	0,779	0,857	0,597	0,624	0,781	0,850
FAR1	0,004	2×10^{-4}	0	0	$2,6 \times 10^{-3}$	$1,4 \times 10^{-3}$	2×10^{-4}	0
FAR2	0,202	0,009	2×10^{-4}	0	0,371	0,089	0,011	0,001

Tabela 6
Poder empírico para os testes de cobertura condicional. PGD3 e $p = 0,05$.

	$ws = 250$				$ws = 500$			
	$n = 250$	$n = 500$	$n = 750$	$n = 1000$	$n = 250$	$n = 500$	$n = 750$	$n = 1000$
GMM ₃	0,610	0,780	0,877	0,939	0,659	0,849	0,924	0,960
GMM ₅	0,603	0,763	0,864	0,928	0,645	0,843	0,913	0,956
CAViaR	0,630	0,702	0,752	0,814	0,633	0,737	0,762	0,794
GVaR ₁	0,594	0,788	0,892	0,944	0,652	0,827	0,921	0,961
GVaR ₂	0,593	0,787	0,891	0,944	0,652	0,827	0,921	0,961
GVaR ₃	0,604	0,810	0,906	0,951	0,657	0,836	0,924	0,965
FAR1	0,009	0	0	0	0,04	0,001	0	0
FAR2	4×10^{-4}	0	0	0	0,002	0	0	0

incluir como covariável as previsões do VaR. Em geral, podemos afirmar que a medida que n aumenta os testes do tipo GVaR tendem a ter maior poder empírico comparando com os outros em todas as tabelas.

6.2 Poder empírico dos testes de avaliação do ES

Para obter o poder empírico dos Testes 1–3 na Seção 5, empregamos os mesmos valores de n e p da seção anterior. Neste caso, utilizamos como único PGD a distribuição *t-Student* centrada no zero, como v graus de liberdade e parâmetro de escala igual a 0,5. Obtemos uma sequência de previsões não eficientes do ES como segue: (i) obtemos um valor de $v \sim \text{Exponencial}(1/10) + 2$, (ii) geramos retornos r_1, \dots, r_n e previsões $ES_1(p), \dots, ES_n(p)$ do modelo *t-Student*(v), (iii) finalmente geramos z_1, \dots, z_n , em que $z_i \sim \mathcal{U}(0, \Delta)$ e logo obter $ES'_i(p) = ES_i(p) - z_i$, para $i = 1, \dots, n$. Consideramos valores do Δ iguais a 0,25, 0,5, 1,25, 1,5, 2,25 e 2,5. Logicamente, é esperado que, em quanto Δ e n cresçam, o poder dos testes também deveria crescer. Nas seções 9.3 e 9.4 explicamos detalhadamente como são obtidos o p -valores e poder empírico para estes testes.

A Tabela 7 resume os resultado do estudo de simulação para obter o poder empírico para diferentes valores de Δ , n e p . Inicialmente, para $\Delta = 0,25$ o poder dos testes não é maior do que 0,6, para os ambos valores de p , e este não se cresce fortemente junto com n . Em geral, para $\Delta \geq 0,5$ notamos que, de fato, o poder aumenta junto com n para os ambos valores de p . Neste caso, os Testes 1 e 3 superam com muita vantagem ao Teste 2 obtendo ambos um bom desempenho. Considerando $p = 0,01$, tais procedimentos têm um mínimo de 0,256 e 0,211 para $n = 250$ e um máximo para 0,925 e 0,877 para $n = 1000$. Considerando $p = 0,05$, os testes têm um mínimo de 0,449 e 0,668 para $n = 250$ e um máximo para 0,997 e 0,987 para $n = 1000$.

O estudo de simulação evidência que o poder dos testes estudados aqui depende fortemente do valor do parâmetro de distorção Δ . Em particular, para valores pequenos de Δ , o poder dos testes foi baixo, mas para valores grandes de Δ observamos valores um poder alto. Contudo, o baixo nível do poder do Teste 2 evidencia

seu baixo desempenho para rejeitar H_A .

Tabela 7
Poder empírico dos Testes 1, 2 e 3.

	$p = 0,01$				$p = 0,05$			
	$n = 250$	$n = 500$	$n = 750$	$n = 1000$	$n = 250$	$n = 500$	$n = 750$	$n = 1000$
$(\Delta = 0,25)$								
Teste 1	0,253	0,189	0,214	0,214	0,252	0,328	0,377	0,407
Teste 2	0,118	0,107	0,115	0,111	0,125	0,145	0,152	0,153
Teste 3	0,143	0,118	0,087	0,060	0,536	0,566	0,569	0,586
$(\Delta = 0,5)$								
Teste 1	0,256	0,303	0,347	0,384	0,449	0,584	0,683	0,733
Teste 2	0,111	0,111	0,114	0,113	0,149	0,177	0,211	0,226
Teste 3	0,211	0,151	0,572	0,562	0,668	0,703	0,750	0,772
$(\Delta = 1,25)$								
Teste 1	0,523	0,618	0,685	0,739	0,832	0,929	0,957	0,980
Teste 2	0,157	0,120	0,153	0,164	0,247	0,345	0,429	0,483
Teste 3	0,739	0,738	0,739	0,750	0,864	0,909	0,928	0,944
$(\Delta = 1,5)$								
Teste 1	0,591	0,693	0,777	0,814	0,887	0,959	0,980	0,989
Teste 2	0,170	0,130	0,173	0,181	0,260	0,389	0,502	0,598
Teste 3	0,764	0,770	0,781	0,791	0,895	0,934	0,951	0,960
$(\Delta = 2,25)$								
Teste 1	0,729	0,819	0,874	0,910	0,958	0,988	0,994	0,996
Teste 2	0,190	0,154	0,210	0,230	0,369	0,563	0,705	0,781
Teste 3	0,826	0,827	0,853	0,860	0,949	0,969	0,979	0,983
$(\Delta = 2,5)$								
Teste 1	0,765	0,842	0,896	0,925	0,965	0,989	0,996	0,997
Teste 2	0,213	0,168	0,225	0,257	0,396	0,618	0,764	0,851
Teste 3	0,845	0,840	0,863	0,877	0,950	0,971	0,983	0,987

7. Avaliação das previsões geradas pelos métodos POT e DPOT

Utilizaremos os *backtestings*, apresentados nas Seções 4 e 5, para realizar uma avaliação pós-amostral (ver Tashman (2000)). Avaliaremos as previsões do VaR e do ES geradas pelos métodos HS, POT e DPOT utilizando os retornos diários da série do Ibovespa no período 06/06/2011 a 05/07/2019 (obtidos dos dados históricos do site <http://finance.yahoo.com>). Consideraremos previsões diárias para uma posição comprada, probabilidades de cobertura $p = 0,01, 0,05$, tamanhos amostrais para o ajuste do modelo iguais a $ws = 500, 1000$ e uma mesma pós-amostra de tamanho $n = 1000$.

A aplicação do método POT depende da escolha do limiar μ . Adicionalmente, o método DPOT depende da escolha dos parâmetros c e λ . Denotemos por μ_{th} o limiar tal que o $th \times 100\%$ dos retornos, de uma determinada amostra, são menores que μ_{th} , para $th \in [0, 1]$. Santos & Alves (2013) apresentam uma aplicação em que para $\mu_{0,1}$, $c = 0,75$ e $\lambda = 3$ as previsões tiveram um bom desempenho. Dado que $\mathbb{E}(Y|Y > \mu_{th})$, o valor esperado de DGP, é uma função linear de μ_{th} , um critério informal de escolha do limiar é visualizar um gráfico da vida residual média $(1/n_{th}) \sum_{i=1}^{n_{th}} (x_i - \mu_{th})$, em que $x_1, \dots, x_{n_{th}}$ são as n_{th} observações que excedem μ_{th} , em função de μ_{th} e escolher valores de μ_{th} para os quais a média de excessos é aproximadamente uma função linear de μ_{th} (ver Coles et al. (2001), página 78–79). Utilizamos as implementações no language R de Santos & Alves (2012) para aplicar o método DPOT, o pacote *evir* (Pfaff & McNeil, 2018) para aplicar o método POT e o pacote *PerformanceAnalytics* (Peterson & Peter, 2019) para aplicar o método HS. Escolhemos o valor de μ e c

Tabela 8
Estatísticas descritivas dos retornos

Média	0,0002	1º Quartil	−0,0083
Desvio Padrão	0,0146	3º Quartil	0,0089
Curtose	1,9170	Mínimo	−0,0921
Assimetria	−0,1489	Máximo	0,0638

procurando atingir uma proporção de violações do VaR o mais próximo possível a p . A pós-amostra e todas as previsões são apresentadas nas Figuras 2 e 3.

Na Figura 1(a) observamos duas tendências da série do Ibovespa: um decrescimento até o início do ano 2016 e um crescimento até o presente ano. A Tabela 8 apresenta as estatísticas descritivas dos retornos do Ibovespa (ver Figura 1(b)). Um coeficiente de kurtosis igual a 1,917 é refletido na forma leptocúrtica da densidade estimada dos retornos, o coeficiente de assimetria $-0,1489$ reflete uma assimetria à direita, além disso observamos caudas pesadas da distribuição empírica dos dados, ver Figuras 1(c) e 1(d).

Para a escolha do limiar obtemos os gráficos da média de vida residual (parcial) dos excessos apresentados na Figura 1(e)–(f). Neste gráfico, mostramos a média dos excessos para a amostra total, e parciais para três seções adjacentes e disjuntas da pós-amostra, isto é, o primeiro terço (A1), segundo terço (A2) e terceiro terço (A3). Na Figura 1(e) distinguimos uma tendência linear (aproximada) para $th \leq 0,1$. Na Figura 1(f) observamos uma tendência aproximadamente linear para $0,04 \leq th \leq 0,1$. Escolhemos valores de th dentro desse intervalo de valores de c tais que a Frequência de Violações (FV) de cada sequência² de previsões seja próximo de p . Para decidir sobre H_{CC} consideraremos um nível de significância $\alpha = 0,05$.

Para $ws = 250$, escolhemos $p = 0,01$ e $th = 0,0215$ para o método POT, mas escolhemos $th = 0,08$ e $c = 1$ para o DPOT. Para $p = 0,05$, escolhemos $th = 0,023$ para POT e $th = 0,1$ e $c = 1$ para DPOT. Os retornos e as correspondente previsões do VaR, neste caso, são apresentados na Figura 2 (a)–(d). Para estes parâmetros, os resultados da avaliação são apresentados na Tabela 9. Confirmamos que a frequência de violações é próximo das probabilidades de cobertura. A eficiência da sequência das previsões do VaR geradas pelo método HS não foi rejeitada em nenhum caso. Aquela gerada pelo método POT foi rejeitada três vezes. No caso do método DPOT, sua previsão foi rejeitada somente uma vez. A eficiência de todas as previsões do ES não foram rejeitadas em nenhum caso.

Por outro lado, para $ws = 500$, $p = 0,01$ e $0,05$ os parâmetros $th = 0,025$ para o POT e $th = 0,1$ e $c = 0,3$ para o DPOT bons resultados foram obtidos. Os retornos e as correspondentes previsões do VaR nestes casos são apresentados nas Figuras 3 (a)–(d). Para estes parâmetros, os resultados da avaliação são apresentados na Tabela 10. No caso da previsão do VaR, novamente a eficiência das previsões do HS não foi rejeitada. Como no caso anterior, a eficiência das previsões do ES não foi rejeitada em nenhum caso.

Resumindo, a avaliação mostra que um método simples como o do HS, neste caso, pode ser tão bom quanto outros mais sofisticados como o POT ou o DPOT. O método DPOT mostrou ser mais reflexível que o POT, obtendo menos rejeições, devido a inclusão dos parâmetros adicionais c e λ . Embora sejam previsões para quantidades diferentes, as previsões do ES evidenciaram menos rejeições que as previsões do VaR, o fato poderia ser considerado como uma vantagem do ES sobre o VaR.

²Por exemplo, para uma sequência de previsões $VaR_1(p), \dots, VaR_n(p)$, a FV é a proporção em que ocorrem os eventos $\{r_i < VaR_{i=1}(p)\}_{i=1}^n$.

Figura 1
Gráficos de análise descritiva

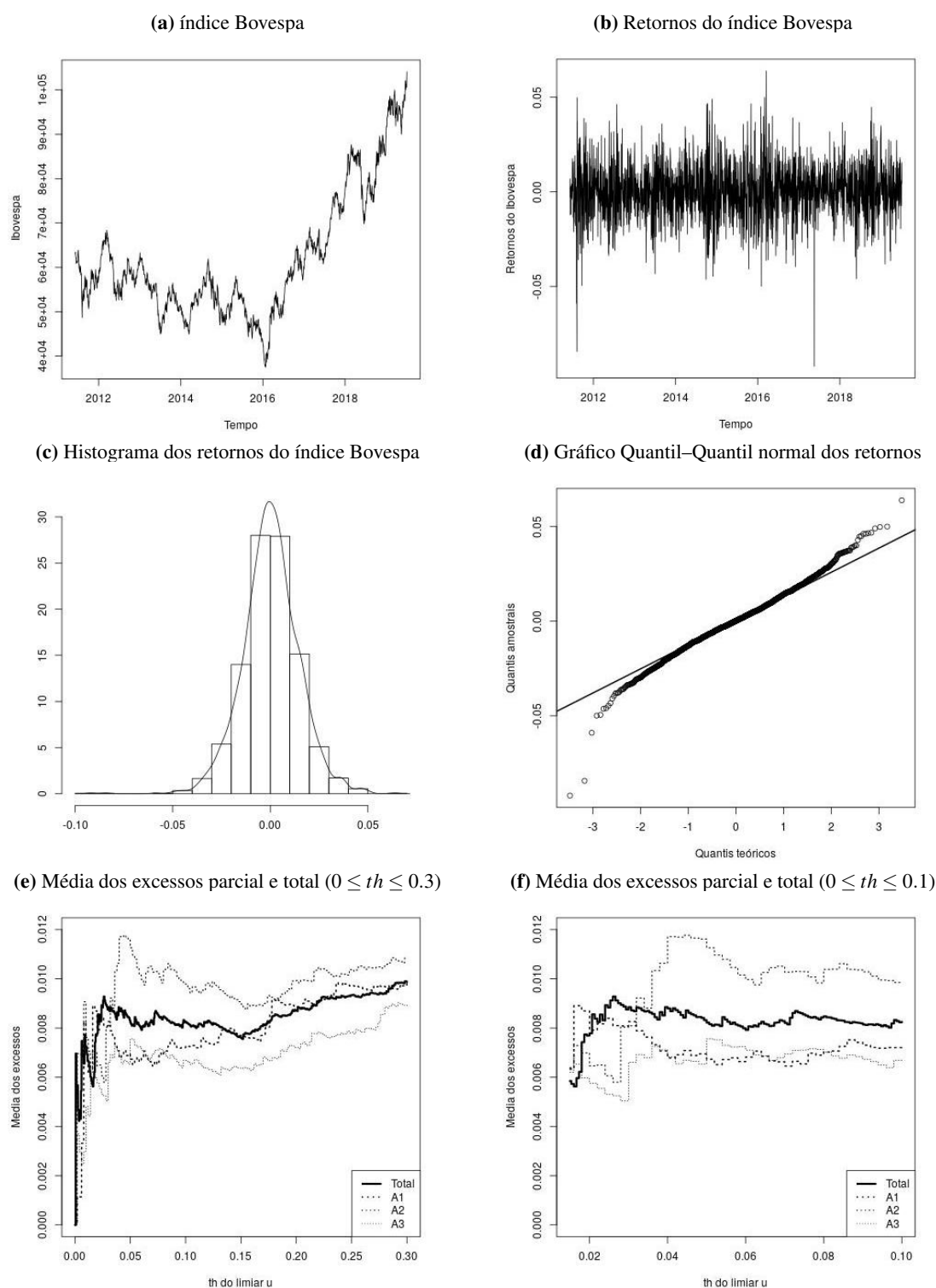
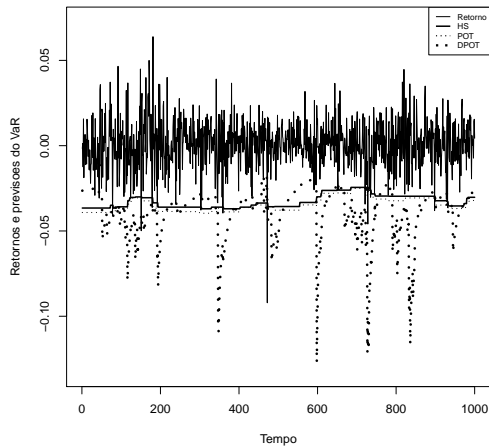
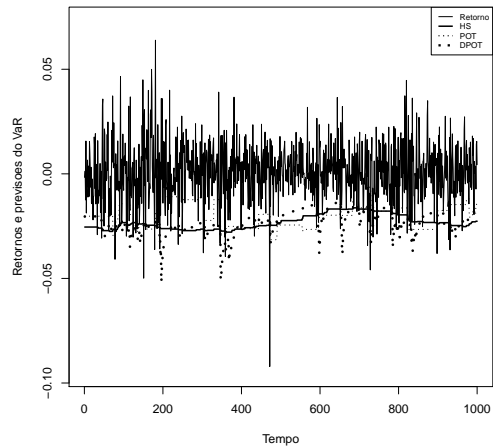


Figura 2
Pós-amostra e previsões do VaR e do ES para $ws = 250$.

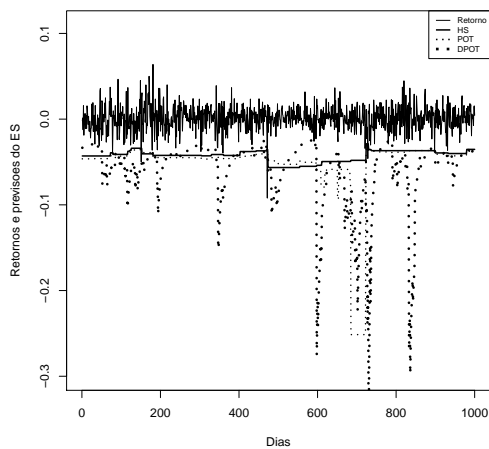
(a) Pós-amostra e previsões do VaR para $p = 0,01$: $th = 0,0215$ para o POT; $th = 0,08$, $c = 1$ e $\lambda = 3$ para o DPOT.



(b) Pós-amostra e previsões do VaR para $p = 0,05$: $th = 0,023$ para o POT; $th = 0,1$ e $c = 1$ para o DPOT.



(c) Pós-amostra e previsões do ES para $p = 0,01$: $th = 0,0215$ para o POT; $th = 0,08$, $c = 1$ e $\lambda = 3$ para o DPOT.



(d) Pós-amostra e previsões do ES para $p = 0,05$: $th = 0,023$ para o POT; $th = 0,1$ e $c = 1$ para o DPOT.

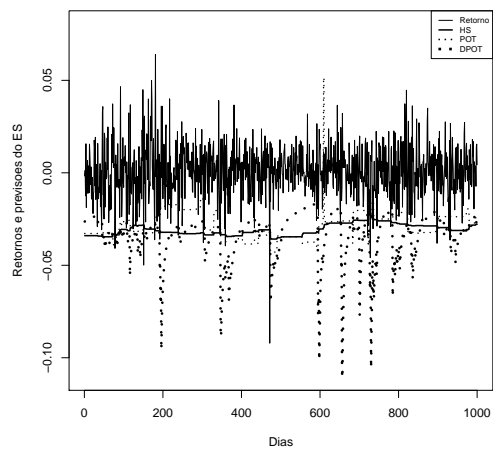
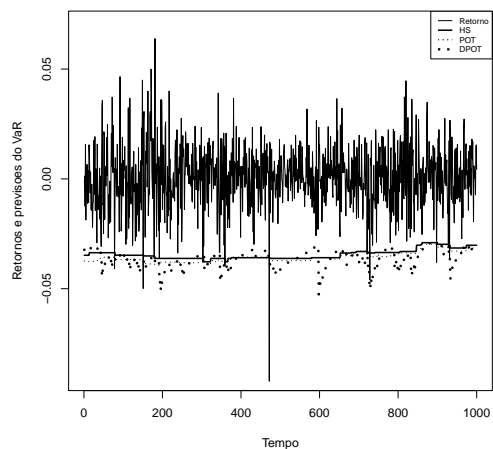


Tabela 9
Avaliação das previsões ($ws = 250$).

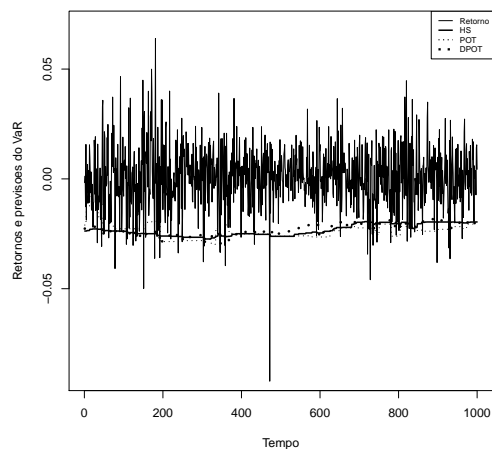
	$p = 0,01$			$p = 0,05$		
	HS	POT	DPOT	HS	POT	DPOT
FV	0,014	0,011	0,013	0,055	0,059	0,055
GMM ₃	0,461	0,719	0,245	0,482	0,090	0,588
CAViaR	0,079	0,036	5e-04	0,164	0,007	0,266
GVaR ₁	0,494	0,580	0,238	0,194	0,046	0,372
Teste 1	0,841	0,783	0,642	0,687	0,517	0,778
Teste 2	0,920	0,789	0,951	0,799	0,880	0,828
Teste 3	0,686	0,663	0,619	0,609	0,504	0,640

Figura 3
Pós-amostra e previsões do VaR e do ES para $ws = 500$.

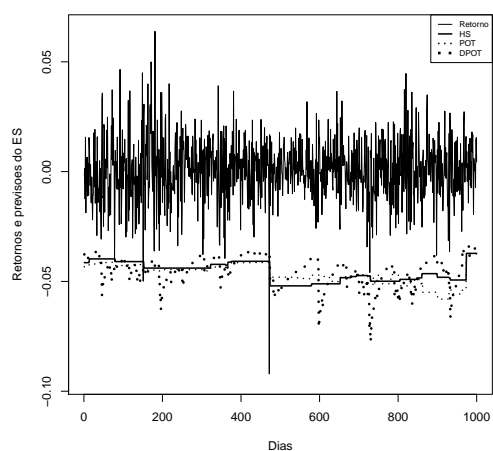
(a) Pós-amostra e previsões do VaR para $p = 0,01$: $th = 0,025$ para o POT; $th = 0,1$ e $c = 0,3$ para o DPOT.



(b) Pós-amostra e previsões do VaR para $p = 0,05$: $th = 0,025$ para o POT; $th = 0,1$ e $c = 0,3$ para o DPOT.



(c) Pós-amostra e previsões do ES para $p = 0,01$: $th = 0,025$ para o POT; $th = 0,1$ e $c = 0,3$ para o DPOT.



(d) Pós-amostra e previsões do VaR para $p = 0,05$: $th = 0,025$ para o POT; $th = 0,1$ e $c = 0,3$ para o DPOT.

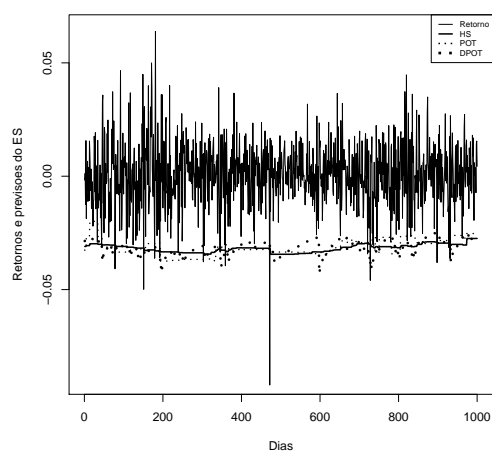


Tabela 10
Avaliação das previsões ($ws = 500$).

	$p = 0,01$			$p = 0,05$		
	HS	POT	DPOT	HS	POT	DPOT
FV	0,011	0,009	0,011	0,051	0,054	0,056
MMG ₃	0,799	0,874	0,313	0,674	0,324	0,301
CAViaR	0,394	0,296	0,023	0,317	0,001	0,470
GVaR ₁	0,999	0,756	0,998	0,481	0,294	0,306
Teste 1	0,612	0,562	0,842	0,469	0,294	0,401
Teste 2	0,646	0,398	0,721	0,554	0,648	0,765
Teste 3	0,609	0,594	0,688	0,517	0,435	0,473

8. Conclusões

O objetivo deste trabalho foi comparar o poder empírico de novas classe *backtestings* do VaR e do ES, e além disso mostrar a sua utilidade na avaliação das previsões. As Tabelas 1–6 mostram como a nova classe GVAr tende a ter um poder empírico maior a medida que o tamanho de amostra cresce na maioria dos casos. As três alternativas de testes do tipo GVAr, com função de risco em (2)–(3), apresentam comportamentos similares em termo do poder empírico. O teste CAViaR também evidencia bons resultados em comparação com os testes do tipo GMM. Isso mostra que a inclusão na previsão do VaR as covariáveis no modelo de fato acrescenta o poder dos testes. O poder empírico dos testes para o ES mostra uma forte dependência do parâmetro de distorção Δ . Logo, o poder dos testes cresce junto com Δ e n , mas, o Teste 2 mostrou um baixo poder em quase todos os cenários.

Em aplicações à série financeira do Ibovespa, as previsões do VaR e do ES obtidas pelo método HS e POT frequentemente geram agrupamento das violações. O método DPOT foi introduzido com o objetivo de superar esta deficiência. Os testes de avaliação foram ferramentas úteis para verificar a flexibilidade do método DPOT em relação ao POT e verificar que um método simples como o do tipo HS também pode ser bastante útil.

Por outro lado, na aplicação notamos que os testes podem levar a diferentes decisões acerca da hipótese de eficiência. Porém, o estudo de simulação para comparar o poder empírico pode ser uma opção para escolher o melhor teste a ser baseado na conclusão da análise. Por exemplo, para o tamanho amostral igual a 1000 poderíamos escolher o teste GVAr₀, no caso do VaR, ou o Testes 1 e 2, no caso do ES.

9. Apêndice

Nesta seção apresentamos os procedimentos de cálculo do p -valor de Monte Carlo e do poder empírico para os testes do VaR e do ES apresentados nas seções 4 e 5.

9.1 p -valor de Monte Carlo para os *backtestings* do VaR

A técnica de Monte Carlo para calcular o p -valor apresentada em Dufour (2006) começa simulando a distribuição empírica das estatísticas L_1, L_2 e L_3 sob H_{CC} . Para a i -ésima estatística do teste, geramos uma sequência L_{i1}, \dots, L_{iR} de R estatísticas, para $i \in \{1, 2, 3\}$, em que cada L_{ij} é calculada como segue:

1. Geramos uma sequência i.i.d. de variáveis indicadoras I_1, \dots, I_n tais que $I_t \sim \text{Bernoulli}(p)$ para $t = 1, \dots, n$.
2. Calculamos as durações associadas, isto é d_{t_1}, \dots, d_{t_N} .
3. Para o caso dos testes CAViaR e GVAr geramos de forma independente uma sequência $\text{VaR}_1(p), \dots, \text{VaR}_n(p)$ por meio de algum processo auto-regressivo (tipo GARCH por exemplo).
4. Finalmente, calculamos uma estatística L_{ij} . Repetindo R vezes este processo obtemos a sequência desejada.

Seja L_0 o valor de uma estatística obtida da amostra observada, o p -valor é obtido como segue. Como estamos trabalhando com uma distribuição discreta, com probabilidade não nula, L_0 poderia ter o mesmo valor a alguma estatística da sequência L_{i1}, \dots, L_{iR} . Para quebrar os empates, geramos, independentemente, uma sequência U_0, \dots, U_R i.i.d., em que $U_1 \sim \mathcal{U}(0, 1)$. O p -valor de Monte Carlo para o i -ésimo teste é obtido por:

$$\hat{\alpha}(L_0) = \frac{R \times \hat{G}(L_0) + 1}{R + 1},$$

em que

$$\hat{G}(L_0) = 1 - \frac{1}{R} \sum_{j=1}^R \mathbb{I}_{\{L_{ij} \leq L_0\}} + \frac{1}{R} \sum_{j=1}^R \left(\mathbb{I}_{\{L_{ij} = L_0\}} \times \mathbb{I}_{\{U_j = U_0\}} \right),$$

e \mathbb{I}_A é a variável indicadora do evento A . Rejeitamos a hipótese nula se $\hat{\alpha}(L_0) \leq \alpha$, para um determinado nível de significância $\alpha \in (0, 1)$.

9.2 Poder empírico para *backtestings* do VaR

O poder de um teste de hipóteses, denotada por $\hat{\beta}$, é calculada como segue. Simulamos a distribuição empírica das estatísticas dos testes sob a hipótese alternativa (de agrupamento das violações) gerando, para a i -ésima estatística do teste, uma sequência $L_{i1}^A, \dots, L_{iR}^A$ de R estatísticas como segue:

1. Geramos uma sequência de retornos via algum PGD.
2. Para cada uma das n pós-amostras obtemos uma previsão do VaR via o método HS seguindo o procedimento descrito em Tashman (2000), obtendo assim uma sequência $\text{VaR}_1, \dots, \text{VaR}_n$ sob a hipótese de alternativa.
3. Usando a sequência dos retornos, de previsões do VaR e uma sequência de violações do VaR.
4. Calculamos a respectiva sequência de durações das violações.
5. Calculamos a estatística L_{ij}^A . Repetindo R vezes este processo obtemos a sequência desejada.
6. Utilizando o procedimento de calculo do p -valor $\hat{\alpha}(\cdot)$ e um nível de significância α , o poder do teste é estimado por:

$$\hat{\beta} = \frac{1}{R} \sum_{j=1}^R \mathbb{I}_{\{\hat{\alpha}(L_{ij}^A) \leq \alpha\}}.$$

9.3 p -valor de Monte Carlo para os *backtestings* do ES

Simulamos a distribuição empírica das estatísticas Z_1, Z_2 e Z_3 sob H_0 gerando, para a i -ésima estatística do teste, uma sequência Z_{i1}, \dots, Z_{iR} de R estatísticas, para $i \in \{1, 2, 3\}$, como segue:

1. De uma mesma distribuição F geramos uma sequência de (i) variáveis i.i.d. r_1, \dots, r_n , (ii) quantis $\text{VaR}_1^F(p), \dots, \text{VaR}_n^F(p)$ e (iii) das esperanças condicionais $\text{ES}_1^F(p), \dots, \text{ES}_n^F(p)$.
2. No caso do teste Z_3 calculamos $\text{DP}_{n+1|n}(p)$.
3. Finalmente obtemos uma estatística do teste Z_{ij} . Repetindo R vezes este processo obtemos a sequência desejada.

Dada uma estatística Z_0 baseada em uma amostra observada e seguindo a Acerbi & Szekely (2014), o p -valor é estimado por:

$$\hat{\alpha}(Z_0) = \frac{1}{R} \sum_{j=1}^R \mathbb{I}_{\{Z_{ij} \leq Z_0\}}.$$

9.4 Poder empírico para os *backtestings* do ES

Simulamos a distribuição empírica das estatísticas Z_1, Z_2 e Z_3 sob H_A gerando, para a i -ésima estatística do teste, uma sequência $Z_{i1}^A, \dots, Z_{iR}^A$ de R estatísticas como segue:

1. Geramos uma sequência de variáveis i.i.d. r_1, \dots, r_n , de quantis $\text{VaR}_1^F(p), \dots, \text{VaR}_n^F(p)$ e das esperanças condicionais $\text{ES}_1^F(p), \dots, \text{ES}_n^F(p)$ geradas de uma mesma distribuição F .
2. Geramos duas sequências de variáveis i.i.d. uniformes z_1, \dots, z_n , em que $z_j \sim \mathcal{U}(0, \Delta)$ para $i = 1, \dots, n$.
 - Obtemos uma sequência $\text{ES}_1(p), \dots, \text{ES}_n(p)$, em que $\text{ES}_i(p) = \text{ES}_i^F(p) - z_i$, tais que $\text{ES}_j^F(p) \geq \text{ES}_j(p)$.
3. Finalmente obtemos uma estatística do teste Z_{ij}^A . Repetindo R vezes este processo obtemos a sequência desejada.
4. O poder do Testes 1 e 2 é calculado por $\hat{\beta} = \frac{1}{R} \sum_{j=1}^R \mathbb{I}_{\{\hat{\alpha}(Z_{ij}^A) \leq \alpha\}}$ e no caso do Teste 3 o poder é análogo ao p -valor da Seção 9.1.

9.5 Processos Geradores dos Dados (PGDs)

O PGD1 gerado via o modelo APARCH(1,1) é dado por

$$\begin{aligned} r_{t+1} &= \sigma_{t+1} z_{t+1}, \\ \sigma_{t+1}^\delta &= w + \alpha(|r_t| - \gamma r_t)^\delta + \beta \sigma_t^\delta, \end{aligned}$$

em que z_t segue uma distribuição *t-Student* assimétrica $t(10)$, $\varphi = 0,88$, $w = 0,03$, $\alpha = 0,086$, $\gamma = 0,64$, $\beta = 0,91$ e $\delta = 1,15$. Este processo foi utilizado por Santos & Alves (2013).

O PGD2 gerado via o modelo GARCH(1,1) – $t(v)$ assimétrico é dado por

$$\begin{aligned} r_{t+1} &= \sigma_{t+1} \sqrt{\frac{v-2}{v}} z_{t+1}, \\ \sigma_{t+1}^2 &= w + \alpha \sigma_t^2 \left(\sqrt{\frac{v-2}{v}} z_t - \theta \right) + \beta \sigma_t^2, \end{aligned}$$

em que $\alpha = 0,1$, $\theta = 0,5$, $\beta = 0,85$, $w = 3,9683e^{-6}$ e $v = 8$. Os parâmetros são selecionados seguindo a Hass (2005).

O PGD3 gerado via o modelo AR(3)–TGARCH(1,1) é definido por

$$\begin{aligned} r_{t+1} &= \phi_1 r_t + a_{t+1}, \\ a_{t+1} &= \sigma_{t+1} \varepsilon_{t+1}, \\ \sigma_{t+1}^2 &= w + \alpha a_t^2 + \gamma a_t^2 d_t + \beta \sigma_t^2. \end{aligned}$$

em que $\varepsilon_{t+1} \sim N(0,1)$, $\phi_1 = -0,051$, $w = 0,00013$, $\alpha = 0,044$, $\gamma = 0,063$ e $\beta = 0,910$. Neste caso $d_t = 1$ se $a_t < 0$ e $d_t = 0$ no caso contrário. A seleção desses parâmetros foi baseada no Exemplo 5.4 em Morettin (2011).

Agradecimentos

Agradecemos a instituição CAPES pelo apoio financeiro oferecido para a elaboração da Dissertação de mestrado do aluno Jaime Lincovil sob a orientação da professora Dra. Chang Chiann que se transformou no presente artigo. Caso o leitor tenha interesse nos códigos das programações dos testes aqui apresentados, estes podem ser encontrados no seguinte link: https://github.com/JaimeLincovil/Backtesting-VaR-ES/blob/master/Tests_medidas_Riesgo_VaRES.R

Referências

- Abad, P., Benito, S., & López, C. (2014). A comprehensive review of Value-at-Risk methodologies. *Spanish Review of Financial Economics* **12**(1): 15–32.
- Acerbi, C., & Szekely, B. (2014). Back-testing expected shortfall. *Risk* **27**(11): 76–81.
- Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1997). Thinking coherently. *Risk* **10**: 68–71.
- Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* **9**(3): 203–228.
- Basel Committee on Banking Supervision (1996). *Amendment to the Capital Accord to Incorporate Market Risks*. Basle, Switzerland.
- Basel Committee on Banking Supervision (2013). *Fundamental Review of the Trading Book: A Revised Market Risk Framework*. Bank for International Settlements, <https://www.bis.org/publ/bcbs265.pdf>.

- Berkowitz, J., Christoffersen, P., & Pelletier, D. (2011). Evaluating Value-at-Risk models with desk-level data. *Management Science* **57**(12): 2213–2227.
- Candelon, B., Colletaz, G., Hurlin, C., & Tokpavi, S. (2010). Backtesting Value-at-Risk: A GMM duration-based test. *Journal of Financial Econometrics* **9**(2): 314–343.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review* **39**:4: 841–862.
- Christoffersen, P., & Pelletier, D. (2004). Backtesting Value-at-Risk: A duration-based approach. *Journal of Financial Econometrics* **2**(1): 84–108.
- Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An Introduction to Statistical Modeling of Extreme Values* (Vol. 208). London: Springer.
- Dufour, J. M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics* **133**(2): 443–477.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling Extremal Events: For Insurance and Finance* (Vol. 33). Springer Science & Business Media.
- Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive Value-at-Risk by regression quantiles. *Journal of Business & Economic Statistics* **22**(4): 367–381.
- Ferraty, F., & Quintela-Del-Río, A. (2016). Conditional VaR and expected shortfall: a new functional approach. *Econometric Reviews* **35**(2): 263–292.
- Haas, M. (2005). Improved duration-based backtesting of Value-at-Risk. *Journal of Risk* **8**(2): 17–38.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4): 1029–1054.
- Krause, J., & Paoletta, M. (2014). A fast, accurate method for Value-at-Risk and expected shortfall. *Econometrics* **2**(2): 98–122.
- Mancini, L., & Trojani, F. (2011). Robust Value-at-Risk prediction. *Journal of Financial Econometrics* **9**(2): 281–313.
- McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools* (Vol. 3). Princeton: Princeton university press.
- Morettin, P. A. (2011). *Econometria Financeira: Um Curso em Séries Temporais Financeiras*, 2ª Edição. São Paulo, Editora Edgar Blücher.
- Nadarajah, S., Zhang, B., & Chan, S. (2014). Estimation methods for expected shortfall. *Quantitative Finance* **14**(2): 271–291.
- Nieto, M. R., & Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting* **32**(2): 475–501.
- Pfaff B. & McNeil A. (2018). *evir: Extreme Values in R*. R package version 1.7–4.
- Pelletier, D., & Wei, W. (2015). The geometric-VaR backtesting method. *Journal of Financial Econometrics* **14**(4): 725–745.
- Pérignon, C., & Smith, D. R. (2010). The level and quality of Value-at-Risk disclosure by commercial banks. *Journal of Banking & Finance* **34**(2): 362–377.

- Peterson B. G. & Peter C. (2019). *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis*. R package version 1.5.3.
- R Core Team (2018). *R: A language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Rappoport, P. (1993). A new approach: Average Shortfall. *JP Morgan Fixed Income Research Technical Document*.
- Righi, M., & Ceretta, P. S. (2013). Individual and flexible expected shortfall backtesting. *Journal of Risk Model Validation* 7(3): 3–20.
- Rinne, H. (2008). *The Weibull Distribution: A Handbook*. Chapman and Hall/CRC.
- Santos, P. A. (2010). Interval forecasts evaluation: R programs for a new independence test. *Notas e Comunicações CEAUL* 17/10.
- Santos, P. A., & Alves, M. F. (2012). R program to implement the DPOT model. *Unpublished article*.
- Santos, P. A., & Alves, M. F. (2013). Forecasting Value-at-Risk with a duration-based POT method. *Mathematics and Computers in Simulation* 94: 295–309.
- Taylor, J. W. (2019). Forecasting Value-at-Risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics* 37(1): 121–133.
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting* 16(4): 437–450.
- Trucíos, C., Hotta, L. K., & Ruiz, E. (2017). Robust bootstrap forecast densities for GARCH returns and volatilities. *Journal of Statistical Computation and Simulation* 87(16): 3152–3174.
- Tsay, R. S. (2005). *Analysis of financial time series* (Vol. 543). John Wiley & Sons.