

Prueba de hipótesis para evaluar las previsiones del *Value-at-Risk*,

Expected Shortfall y los *Expectiles*

Jaime Enrique Lincovil Curivil *

Resumen

La evaluación de pronósticos de medidas de riesgo, tales como *Value-at-Risk* (VaR) y *Expected Shortfall* (ES), es una actividad muy relevante para las instituciones financieras. Las pruebas de hipótesis que realizan esta evaluación se introdujeron con el objetivo de verificar la eficiencia de estas predicciones. En este trabajo comparamos el poder empírico de nuevas clases de pruebas de evaluación presentes en la literatura. Además, usamos estos procedimientos para evaluar la eficiencia de las predicciones generadas por el método Simulación Histórica y métodos basados en la distribución generalizada de Pareto. Para evaluar las predicciones del VaR, la clase de pruebas Geométrico-VaR se destacó por tener un mayor poder empírico que las demás pruebas en los escenarios simulados. Esto destaca las ventajas de trabajar con duraciones discretas e incluir covariables en los modelos de prueba. Para evaluar las predicciones de ES, aquellas basadas en distribuciones condicionales de la rentabilidad del VaR mostraron una alta potencia para tamaños de muestra grandes. Adicionalmente, mostramos que el método basado en la distribución de Pareto generalizada utilizando duraciones y covariables resulta en un excelente desempeño en las predicciones VaR y ES de acuerdo a la evaluación realizada por las pruebas consideradas.

Palabras-clases: Cobertura Condicional, Expected Shortfall, Poder empírico, Prueba de evaluación, Value-at-Risk.

*E-mail: outliersj@gmail.com

1. Introducción

En el contexto del mercado financiero, las decisiones de compra o venta de activos financieros deben tomarse de acuerdo con el riesgo que implican. Por tanto, medir el riesgo de estas resoluciones es una actividad muy relevante. El *Value-at-Risk* (VaR) y *Expected Shortfall* (ES) son medidas ampliamente utilizadas por las instituciones financieras para medir el riesgo de mercado, ya que resumen el riesgo en una sola medida numérica, los cuales son fácilmente comunicables a las entidades financieras.

Tanto el VaR como el ES para una rentabilidad futura se pronostican utilizando un modelo ajustado que utiliza rentabilidades pasadas. Una buena estimación del modelo de pronóstico presenta grandes dificultades debido a que, en general, el comportamiento de los retornos es difícil de modelar. Complicaciones del modelado son los hechos estilizados de la serie de retornos tales como: *clustering* de volatilidad, varianza dinámica dependiente del tiempo, colas pesadas de la distribución, entre otros. Dadas estas dificultades, es necesario evaluar el desempeño de los pronósticos del VaR y ES bajo una determinada definición de eficiencia.

En el contexto regulatorio, el Comité de Supervisión Bancaria de Basilea (BCBS) es el principal organismo regulador de las estrategias de control del riesgo bancario. En 1996, el BCBS promulgó la regulación que recomienda el uso generalizado del VaR (BCBS, 1996). Bajo esta regulación, los bancos pueden gestionar sus propias provisiones de VaR. Recientemente, BCBS sugirió reemplazar el VaR por ES, con el fin de mitigar las debilidades encontradas en la implementación del VaR en situaciones extremas tales como: falta de información sobre el volumen de pérdida potencial y infracción de la propiedad de consistencia, para determinar los requerimientos de capital en el contexto de gestión de riesgo de mercado (BCBS, 2013). Debido a la necesidad de conocer las propiedades teóricas y empíricas de las medidas de riesgo utilizadas, surge la necesidad de evaluar los pronósticos.

La evaluación se puede realizar a través de procedimientos llamados *backtestings*, que prueban la hipótesis de eficiencia de una secuencia de predicciones. En el caso del VaR, el criterio de eficiencia se deriva del criterio de intervalos de pronóstico eficientes introducido por Christoffersen (1998). En el caso de ES, presentaremos el criterio sugerido por Acerbi e Szekely (2014). Es importante señalar que dichas pruebas evalúan el conjunto de pronósticos realizados y no el ajuste del modelo de pronóstico.

Una de las opciones presentes en la literatura para determinar el desempeño de *backtestings* es comparar su potencia empírica mediante simulación Monte Carlo. En este trabajo comparamos el poder empírico (en muestras finitas) de nuevas pruebas en diferentes escenarios simulados, ya que no existen muchos estudios en la literatura que comparen el poder empírico de dichas pruebas en un mismo escenario. En este trabajo, los procedimientos de evaluación del VaR considerados son: (i) la prueba *Generalized Moment Method* (GMM) propuesta por Candelon *et al.* (2010), (ii) la prueba *Conditional Autoregressive VaR* (CAViaR) en una versión del modelo lineal generalizado implementado por Berkowitz *et al.* (2011), (iii) la prueba *Geometric-VaR* (GVaR) presentada por Pelletier e Wei (2015). Para la evaluación de ES, consideramos dos pruebas

propuestas por Acerbi e Szekely (2014) y también la prueba basada en la distribución truncada en Righi e Ceretta (2013). En la aplicación a datos reales, usamos los métodos *Simulación histórica* (HS), *Peak-Over-Threshold* (POT) y *Duration Peak Over Threshold* (DPOT) en los pronósticos de VaR y ES para una serie de Ibovespa, dichos pronósticos fueron evaluados utilizando las pruebas mencionadas anteriormente. Usamos el lenguaje (?) para desarrollar las aplicaciones computacionales.

Como principales resultados obtenidos en este trabajo, el estudio de simulación del poder empírico del VaR mostró que la inclusión de predicciones pasadas del VaR como covariables dentro del modelo de regresión discreta para las duraciones aumenta el poder de las pruebas. Esto se evidenció en la alta potencia de las pruebas CAViaR y GVaR. Por otra parte, la potencia de las pruebas de ES mostró una fuerte dependencia de la diferencia entre la predicción de ES bajo la hipótesis nula y la predicción de ES bajo la hipótesis alternativa. En la aplicación, mostramos que el método DPOT (la extensión de POT) generó predicciones eficientes, según las pruebas, en los casos en que POT falló. Además, el método HS, aunque sencillo, también generó buenos resultados.

El presente trabajo está organizado de la siguiente manera. En la Sección 2 presentamos el concepto de VaR, infracciones de VaR, duraciones entre infracciones de VaR y la definición de ES. En la Sección ?? describimos brevemente los métodos de predicción HS, POT y DPOT. En las Secciones 4 y 5 exponemos el tema principal de este trabajo, esto es, las pruebas de evaluación de VaR y ES. En las Secciones ?? y ?? presentamos los resultados de las simulaciones, una aplicación y las evaluaciones de los pronósticos VaR y ES, respectivamente. Finalmente, en la Sección ?? discutimos las principales conclusiones del trabajo.

2. VaR, infracciones del VaR, duraciones y ES

Sea P_t el precio de un activo en el instante t y $r_t = \log(P_t) - \log(P_{t-1})$ el correspondiente log-retorno (o simplemente retorno) al instante t . Para un $R_t = (P_t - P_{t-1})/P_t$ pequeño, se sabe que $r_t \approx R_t$, es decir, el retorno r_t representa aproximadamente la variación de precios del activo desde el tiempo $t - 1$ a t . El VaR en el tiempo t con probabilidad $p \in (0, 1)$ y el p -cuantil $\inf\{l : F_t(l) \geq p\}$, donde F_t es la función de distribución acumulada del retorno r_t . En particular, a continuación se da la definición formal de la previsión de VaR para una posición larga para distribuciones continuas.

Definición 1 Definimos la predicción $\text{VaR}_{t|t-1}(p)$ realizado en el tiempo $t - 1$ con un horizonte de una unidad de tiempo y con probabilidad de cobertura p , como el p -cuantil que satisface

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

donde Ω_{t-1} es un conjunto que contiene informaciones hasta el tiempo $t - 1$.

Esta previsión del VaR puede considerarse como una medida de variación máxima negativa de las rentabilidades. En este caso, cuando la rentabilidad tiene una variación negativa mayor a la prevista por el VaR ,

se entiende que la variación de la rentabilidad violó el límite establecido por el VaR. Este hecho es conocido en la literatura como *infracción* de la predicción del VaR. Definimos la variable indicadora de infracción del VaR como:

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

Además, definimos la *duración* de la i -ésima infracción del VaR por:

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

donde t_i denota el instante de la i -ésima infracción del VaR. Para $i = 1$ asumimos que $t_0 = 0$.

Otra medida de riesgo muy utilizada hoy en día es el ES. El ES fue introducido por Rappoport (1993) y luego Artzner *et al.* (1997, 1999) estudió formalmente sus propiedades. El $\text{ES}(p)$ asociado al $\text{VaR}(p)$ es definido por $\text{ES}(p) = 1/(1-p) \int_p^1 \text{VaR}(u) du$. De forma simplificada, el ES es la pérdida esperada dado que la rentabilidad viola el VaR, es decir $\text{ES}(p) = \mathbb{E}[r_t | r_t < \text{VaR}(p)]$ (ver Lema 2,16 en McNeil *et al.* (2005)). La definición de predicción ES de un paso adelante se da a continuación.

Definición 2 Definimos el pronóstico $\text{ES}_{t|t-1}(p)$ realizado en el tiempo $t-1$ con un horizonte de una unidad de tiempo asociado al pronóstico $\text{VaR}_{t|t-1}(p)$, que satisface $p = \Pr(r_t < \text{VaR}_{t|t-1}(p) | \Omega_{t-1})$ para una información pasada Ω_{t-1} , como la esperanza condicional:

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

Como es descrito en la literatura (?), las medidas VaR y ES no cumplen la propiedad de coherencia. Una medida que generaliza los cuantiles y son a su vez coherentes son los espectiles propuestos ? definidos a continuación.

Definición 3 La previsión de un espectil a partir de la serie de retornos hasta r_{t-1} es dada por la siguiente solución de minimización:

$$\text{EX}_{t|t-1}(p) = \arg \min_{r \in \mathbb{R}} \left\{ \left| p - I_{\{r_t < r\}} \right| (r_t - r)^2 \mid \Omega_{t-1} \right\}.$$

El EX es una extensión del VaR que cumple la propiedad de coherencia. Análogamente, al ES, el CARES (Conditional Autoregressive Expected Shortfall) introducido por ? es dado por

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

A continuación, presentamos los métodos de predicción VaR y ES utilizados en este trabajo.

3. Métodos HS, POT e DPOT para a previsión del VaR y ES

El método HS es un procedimiento de predicción simple que asume que los procesos financieros históricos se repiten (ver Perignon e Smith (2008)). Aplicando este supuesto, el método HS estima el VaR para el instante t a través del percentil de la serie $\{r_s\}_{s=0}^{t-1} = \{r_0, \dots, r_{t-1}\}$, es decir:

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

Además, el método HS predice el ES a través de la siguiente media condicional:

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

donde I_s^* es la indicadora de la infracción de $\widehat{\text{VaR}}_{s|s-1}^{\text{HS}}(p)$ y $N^* = \sum_{s \leq t-1} I_s^*$. Dada la velocidad de cálculo y la facilidad de implementación, este método es ampliamente utilizado.

Con respecto al la estimación del EX de forma empírica es dada por:

$$\widehat{\text{EX}}_{t|t-1}^{\text{HS}}(p) = \arg \min_{r \in \mathbb{R}} \left\{ \left| p - I_{\{\bar{x}_{t-1} < r\}} \right| (\bar{x}_{t-1} - r)^2 \mid \Omega_{t-1} \right\},$$

en que $\bar{x}_{t-1} = \frac{1}{N_{ws}} \sum_{t-ws \leq t-1} r_s$ y $N_{ws} = \sum_{t-ws \leq t-1} I_{t-ws \leq t-1}$. Análogamente, $\widehat{\text{CARE}}_{t|t-1}^{\text{HS}}(p)$ es definido como el promedio condicional dado $\widehat{\text{EX}}_{t|t-1}^{\text{HS}}(p)$.

El método POT para la predicción del VaR considera el ajuste de una Distribución Generalizada de Pareto (DGP) en base a un exceso de rentabilidad respecto de un umbral $\mu > 0$. Sea $y_t = r_t - \mu > 0$ (para $r_t > 0$) el exceso de la rentabilidad r_t con respecto a μ y $F_\mu(y_t)$ la función de distribución acumulativa de y_t . Para μ suficientemente grande, $F_\mu(y_t)$ se puede aproximar mediante la función de distribución acumulativa del DGP (ver Embrechts *et al.* (2013)), con parámetros de escala $\sigma > 0$ y de la forma $\gamma \in \mathbb{R}$, dada por:

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

El soporte de la DGP es $y \geq 0$ cuando $\gamma \geq 0$ y $0 \leq y \leq -\sigma/\gamma$ cuando $\gamma < 0$. Para un umbral fijo $\mu > 0$, el pronóstico de VaR con probabilidad de cobertura p y horizonte de una unidad de tiempo obtenido por el método POT es el siguiente:

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

donde M es el número de excesos por encima de μ en una muestra con L retornos y $\hat{\sigma}$ y $\hat{\gamma}$ son los estimadores de máximos verosímilitud respectivos (ver Tsay (2005)). Una interpretación simple de $\widehat{\text{VaR}}_{t|t-1}^{\text{POT}}(p)$ en términos de μ y $\hat{\sigma}$ de la siguiente manera: μ actúa como una intersección fija, $\hat{\sigma}$ estima el impacto de $1/\hat{\gamma}((M/Lp)^{\hat{\gamma}} - 1)$ en el pronóstico $\widehat{\text{VaR}}_{t|t-1}^{\text{POT}}(p)$, mientras que $\hat{\gamma}$ regula la escala de impacto de la proporción

de rendimientos en exceso con respecto a μ . Siguiendo a Coles *et al.* (2001), la predicción de ES basada en el enfoque POT está dada por:

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

En este caso, $\widehat{\text{ES}}_{t|t-1}^{\text{POT}}(p)$ aumenta (o disminuye) junto con $\hat{\sigma}$. Esta metodología funciona bien cuando se considera una submuestra extrema según μ . Sin embargo, Santos e Alves (2013) evidencia empíricamente que los pronósticos generados por el método POT provocan *clusters de infracciones del VaR* en períodos de alta volatilidad. Para sortear este problema, los autores proponen incorporar como covariables las duraciones entre los tiempos de exceso de rentabilidad en el método DPOT (Duration based POT). DPOT propone modelar el parámetro de escala (variando en el tiempo) por $\sigma_t = \alpha/(d_{t,\lambda})^c > 0$, donde $\alpha > 0$, $c > 0$, $\lambda \in \mathbb{N}_+$ y $d_{t,\lambda}$ es la duración entre el tiempo t y el instante antes del cual se excede el primero de los λ pasados. En este modelo, el impacto de σ_t ocurre en la dirección opuesta de la duración, es decir, σ_t disminuye al aumentar la duración de los λ excesos pasados. El impacto de $d_{t,\lambda}$ está regulado por el parámetro c . Como resultado, VaR y ES varían de acuerdo con $\hat{\sigma}_t$ en ambos enfoques. El pronóstico de VaR con probabilidad de cobertura p y horizonte de una unidad de tiempo obtenido por el método DPOT viene dado por:

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

Los parámetros λ y c son constantes determinadas según criterios empíricos. Además, la predicción de ES basada en el método DPOT viene dada por:

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

El enfoque de este trabajo es comparar el poder empírico de las pruebas para evaluar la eficiencia de los pronósticos y no en los métodos de pronóstico VaR y ES. Para más detalles sobre pronósticos ver Mancini e Trojani (2011), Abad *et al.* (2014), Nadarajah *et al.* (2014), Krause e Paoletta (2014), Nieto e Ruiz (2016), ?, Ferraty e Quintela-Del-Río (2019) y Taylor (2019).

Siguiendo a ? la estimación de la previsión de EX sobre el abordaje POT es dado por

$$\widehat{\text{EX}}_{t|t-1}^{\text{DPOT}}(p) : \frac{\hat{\sigma}}{1-\hat{\gamma}} \left[1 + \frac{2p-1}{p} \left(1 + \hat{\gamma} \frac{\widehat{\text{EX}}_{t|t-1}^{\text{DPOT}}(p)}{\hat{\sigma}} \right) \right] = \widehat{\text{EX}}_{t|t-1}^{\text{DPOT}}(p) \quad (1)$$

Finalmente, el CARES via POT es

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

4. *Backtesting* del VaR

En este trabajo, consideraremos la definición de eficiencia utilizada en trabajos como Berkowitz *et al.* (2011) y Pelletier e Wei (2015), entre muchos otros. La definición de la hipótesis de eficiencia se elabora de forma diferente según el modelo de cada prueba, aunque el criterio de eficiencia puede ser el mismo. El concepto de eficiencia es básicamente una adaptación de la definición de intervalos de predicción eficientes introducida por Christoffersen (1998).

Definición 4 Sea $r_1, \dots, r_{t_0-1}, r_{t_0}, \dots, r_T$ una serie de retornos y r_{t_0}, \dots, r_T una pós-muestra. Decimos que la secuencia $\text{VaR}_{t_0|t_0-1}(p), \dots, \text{VaR}_{T|T-1}(p)$ es eficiente si

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

Por el Lema 1 en Christoffersen (1998), la secuencia $\text{VaR}_{t_0|t_0-1}(p), \dots, \text{VaR}_{T|T-1}(p)$ es eficiente si y solo si $I_s \mid \Omega_{s-1} \stackrel{i.i.d}{\sim} \text{Bernoulli}(p)$ para $s \in \{t_0, \dots, T\}$, o equivalentemente, si $D_i \mid \Omega_{t_i-1} \stackrel{i.i.d}{\sim} \text{Geométrica}(p)$ for $i \in \{1, \dots, N\}$, en que t_i es el momento de la i -ésima infracción. Es decir, una secuencia de predicciones de VaR es eficiente si las infracciones respectivas ocurren de forma independiente, con una probabilidad de cobertura p constante durante todo el proceso de predicción. Se dice que una secuencia eficiente según la Definición de 4 tiene *cobertura condicional correcta* (CC correcta). En consecuencia, la hipótesis (nula) de CC correcta está definida por H_{CC} : “ $I_s \mid \Omega_{s-1} \stackrel{i.i.d}{\sim} \text{Bernoulli}(p)$ para $s \in \{t_0, \dots, T\}$ ” o su equivalente en cuanto a duraciones. En este trabajo consideramos la información pasada hasta el tiempo $t - 1$ dada por $\Omega_{1,t-1} = \{I_{t-1}, \dots, I_1\}$ y $\Omega_{2,t-1} = \{I_{t-1}, \dots, I_1, r_{t-1}, \dots, r_1\}$.

4.1. Tests basados en el método de los momentos generalizados

La prueba GMM para probar H_{CC} con respecto a $\Omega_{1,t}$, presentada en Candelon *et al.* (2010), usa el polinomio ortogonal de la distribución Geométrica de cierto orden. A continuación, definimos el $(j + 1)$ -ésimo término de este polinomio.

Definición 5 El $(j + 1)$ -ésimo término del polinomio ortogonal asociado a la distribución Geométrica con probabilidad de éxito p evaluada en la duración observada $D = d$ está definido por la siguiente relación 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 $j \in \mathbb{N}$, con $M_{-1}(d, p) = 0$ y $M_0(d, p) = 1$. Así, el polinomio ortogonal de orden k asociado a la distribución Geométrica está dado por $\sum_{j=1}^k M_j(d, p)$.

Sea D_1, \dots, D_N una secuencia de duraciones entre infracciones del VaR y sea $\mathbb{E}[M_{j+1}(D_i, p)]$ la esperanza del polinomio aleatorio $M_{j+1}(D_i, p)$. El enfoque GMM utiliza la siguiente propiedad de los términos del

polinomio $\sum_{j=1}^k M_j(d, p)$ de la siguiente manera: si la secuencia $D_1|\Omega_{t_1-1}, \dots, D_N|\Omega_{t_N-1}$ es i.i.d. con distribución geométrica (p), entonces $\mathbb{E}[M_{j+1}(D_i, p)] = 0, \forall j \in \mathbb{N}, \forall i = 1, \dots, N$ (Candelon *et al.* (2010) página 320). Bajo estas condiciones de momento, la hipótesis H_{CC} con respecto al polinomio ortogonal de orden k se formula como:

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

Si asumimos que el proceso de duraciones $\{D_i : i \geq 1\}$ es estacionario y ergódico, entonces bajo H_{CC} el estadístico de prueba $J_{CC}(k)$ tiene la siguiente propiedad asintótica:

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

donde $\mathbf{M}(D_i, p) = (M_1(D_i; p), \dots, M_k(D_i; p))^\top$ y $\xrightarrow{\mathcal{D}}$ denota la convergencia en la distribución. En este enfoque, la hipótesis alternativa no se especifica. Ver Hansen (1982) para más detalles.

4.2. Prueba de hipótesis del tipo CAViaR

Siguiendo ?, Berkowitz *et al.* (2011) consideramos un modelo logístico para la probabilidad de cobertura, donde $I_t | \Omega_{2,t-1} \sim \text{Bernoulli}(\pi_t)$, $\pi_t = \Pr(I_t = 1 | \Omega_{2,t-1}) = (1 + e^{-\eta_t})^{-1}$, $\eta_t = \delta_0 + \delta_1 I_{t-1} + \delta_2 \text{VaR}_{t|t-1}(p)$ es el predictor lineal y $\delta = (\delta_0, \delta_1, \delta_2) \in \mathbb{R}^3$.

Bajo H_{CC} , π_t no depende de I_{t-1} ni de $\text{VaR}_{t|t-1}(p)$. O sea, $H_{CC} : \pi_t = p, \forall t \in \{t_0, \dots, T\}$. Las hipótesis de prueba que utilizan los parámetros en δ están dadas por:

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

Se utiliza el test de razón de verosimilitud generalizada, cuyo estadístico L_2 viene dado 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 $\tilde{I}_T = (I_{t_0}, \dots, I_T)$, $\delta_0 = \log(p/(1-p))$, $\delta^0 = (\delta_0, 0, 0)$, $\hat{\delta} = (\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2)$ es el estimador de máxima verosimilitud de δ e \mathcal{L}_1 es su respectiva función de log-verosimilitud.

4.3. Prueba de hipótesis del tipo Geométrico-VaR (GVaR)

Hass (2005) propone utilizar la distribución Weibull discreta tipo II caracterizada en Rinne (2008) para modelar la duración de las infracciones, cuya función de riesgo viene dada por $\lambda_d = \pi d^{b-1}$. Posteriormente, Pelletier e Wei (2015) propuso agregar el pronóstico VaR como covariable en la función de riesgo a través

del siguiente modelo:

$$\begin{aligned}\lambda_d^i &= \Pr\left(I_{t_{i-1}+d} = 1 \mid I_{t_{i-1}+d-1} = 0, \dots, I_{t_{i-1}+1} = 0, I_{t_{i-1}} = 1, \mathbf{r}_i(d)\right) \\ &= \pi d^{b-1} \exp\{\beta \text{VaR}_{t_{i-1}+d}(p)\}\end{aligned}\quad (2)$$

donde $\beta \geq 0$, $b > 1$, $\pi \in (0, 1)$ y t_{i-1} es el instante de la $(i-1)$ -ésima infracción, $t_i = t_{i-1} + d$ es el instante en que se produciría la i -ésima infracción del VaR con una duración igual a d y $\mathbf{r}_i(d) = \{r_{t_{i-1}+d-1}, r_{t_{i-1}+d-2}, \dots\}$. Por lo tanto, λ_d^i mide la probabilidad de i -ésima infracción del pronóstico $\text{VaR}_{t_{i-1}+d}(p)$ después de d períodos.

Adicionalmente, en este trabajo consideramos los modelos logísticos y probito para la función de riesgo de duraciones. Esto es

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

$$\lambda_d^i = \Phi\{\Phi^{-1}(\lambda(d)) + \beta \text{VaR}_{t_{i-1}+d}(p)\}, \quad (4)$$

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

Considerando los modelos (2), (3) y (4) la hipótesis CC correcta se puede expresar mediante $H_{CC} : \lambda_d^i = p, \forall i, \forall d$, es decir, la probabilidad de que ocurra la i -ésima infracción del VaR después de d periodos es igual a p , no depende de su duración d y de $\text{VaR}_{t_{i-1}+d}(p)$. Las hipótesis de prueba son:

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

En este caso también se utiliza la prueba de razón de verosimilitud generalizada, cuyo estadístico L_3 viene dado por:

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

donde $\tilde{D}_N = (D_1, \dots, D_N)$, $\theta^0 = (p, 1, 0)$, $\hat{\theta} = (\hat{p}, \hat{b}, \hat{\beta})$ es el estimador de máxima verosimilitud de $\theta = (\pi, b, \beta)$ y \mathcal{L}_2 es la función log- -probabilidad de la distribución discreta de Weibull tipo II considerando una censura no informativa para la primera y última duración. El procedimiento se denomina prueba Geometric-VaR (GVaR).

Para los procedimientos presentados en esta sección, la distribución asintótica de las estadísticas de prueba se puede usar para calcular el valor p y tomar una decisión sobre H_{CC} . Otra alternativa es estimar el valor p mediante métodos de Monte Carlo.

5. *Backtesting* del ES

Los métodos para la evaluación de predicciones de ES que consideramos en este trabajo son del tipo no paramétrico, es decir, prueban hipótesis que asumen condiciones generales sobre una posible función de

distribución de probabilidad que sigue a los retornos.

Sean $K_t(x) = \Pr(r_t < x)$ y $K_t^{[p]}(x) = \Pr(r_t < x | r_t < \text{VaR}(p)) = \min(1, K_t(\cdot)/p)$ la distribución acumulada y la distribución acumulada en la cola superior utilizadas para realizar las predicciones de VaR y ES, respectivamente. Por ejemplo, K_t podría ser la distribución acumulada Generalizada de Pareto y $K_t^{[p]}$ su respectiva condicional. En términos simples, Acerbi e Szekely (2014) propone que una distribución adecuada de $K_t^{[p]}$ utilizada para hacer la predicción de ES debería ser aproximada, o equivalente, la verdadera distribución de las colas de los retornos $F_t^{[p]}(\cdot) = \min(1, F_t(\cdot)/p)$. Esto se establece en la siguiente definición.

Definición 6 Decimos que la secuencia $\text{ES}_{t_0|t_0-1}(p), \dots, \text{ES}_{T|T-1}(p)$ es eficiente si

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

Contrariamente a la definición 2, esta definición solo considera las características de la población de la distribución de retornos y no una evaluación de los pronósticos de ES como el promedio condicional de los rendimientos observados de la pós-muestra. De lo contrario, las predicciones de ES realizadas por el modelo ajustado ($\text{ES}_{t|t-1}^K(p)$) podrían subestimar (o sobrestimar) las predicciones generadas a partir del modelo real ($\text{ES}_{t|t-1}^F(p)$) aunque las predicciones de VaR ($\text{VaR}_{t|t-1}^K(p)$) fueron iguales a las generadas a través de F_t ($\text{VaR}_{t|t-1}^F(p)$). Como se describió anteriormente y la definición de 2, se establecen las siguientes hipótesis para ser probadas.

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

donde $\text{ES}_{t|t-1}^F(p)$ y $\text{VaR}_{t|t-1}^F(p)$ son las predicciones generadas por $F_t^{[p]}$ y $\text{ES}_{t|t-1}^K(p)$ y $\text{VaR}_{t|t-1}^K(p)$ son los pronósticos generados por $K_t^{[p]}$, respectivamente. Se han propuesto los siguientes procedimientos para probar estas hipótesis. En este caso, el par de hipótesis son iguales para las diferentes pruebas que presentaremos a continuación.

5.1. Prueba de hipótesis 1

Luego de procedimientos algebraicos simples basados en la definición 2, podemos verificar que una secuencia de predicciones eficiente $\widehat{\text{ES}}_{1|0}(p), \dots, \widehat{\text{ES}}_{T|T-1}(p)$ debe satisfacer $\mathbb{E} \left[r_t / \widehat{\text{ES}}_{t|t-1}(p) - 1 \mid r_t < \widehat{\text{VaR}}_{t|t-1}(p) \right] = 0$ para cada $t \in \{1, \dots, T\}$. Sea I_1, \dots, I_T una muestra de infracciones del VaR correspondientes a una serie de rendimientos r_1, \dots, r_T y $N_T = \sum_{t=1}^T I_t$. Las estadísticas¹ propuesta por Acerbi e Szekely (2014) para

¹En Acerbi e Szekely (2014), $-\text{VaR}(p) < 0$ y $-\text{ES}(p) < 0$, obteniendo una versiónãl Z_1 estadístico dado por $(1/N_T) \sum_{t=1}^T r_t I_t / (\widehat{\text{ES}}_{t|t-1}(p)) + 1$. Lo mismo se aplica a las estadísticas Z_2 y Z_3 . Nuestra versión és equivalente.

la Prueba 1 viene dada por:

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

Tengamos en cuenta que bajo H_0 , $\mathbb{E}[Z_1] = 0$, pero bajo H_A , $\mathbb{E}[Z_1] < 0$. En consecuencia, valores de Z_1 cercanos a cero indican evidencia a favor de H_0 y evidencia en contra en el caso contrario.

5.2. Prueba de hipótesis 2

Reescribiendo el ES dado en la definición 2 como una esperanza no condicional, una predicción correcta del ES debe satisfacer $\widehat{\text{ES}}_{t|t-1}(p) = \mathbb{E}[r_t I_t / p]$ para todo $t \in \{1, \dots, T\}$. Análogamente al caso anterior, Acerbi e Szekely (2014) propone una segunda opción de estadísticos de contraste para las hipótesis anteriores 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 en la prueba de hipótesis 1, bajo H_0 , $\mathbb{E}[Z_2] = p^{-1} - 1$, pero bajo H_A , $\mathbb{E}[Z_2] < p^{-1} - 1$. Análogamente al caso anterior, los valores de Z_2 cercanos a cero indican evidencia a favor de H_0 y evidencia en contra de lo contrario.

5.3. Prueba de hipótesis 3

En otra línea, Righi e Ceretta (2013) propone estudiar las fluctuaciones de la distribución estandarizada truncada de los retornos r_t , que es la distribución de la transformación $(r_t - \widehat{\text{ES}}_{t|t-1}(p))I_t / \text{DP}_{t|t-1}$, donde $\text{DP}_{t|t-1}(p) = \sqrt{\widehat{\text{Var}}(r_t | r_t < \widehat{\text{VaR}}_{t|t-1}(p))}$. Por lo tanto, proponemos considerar la estadística

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

para el procedimiento de prueba, donde $\widehat{\text{Var}}(r_t | r_t < \widehat{\text{VaR}}_{t|t-1}(p)) = \sum_{s=t_{a_0}}^{t-1} (r_s - \widehat{\text{ES}}_{t|t-1}(p))^2 I_s / ws$ y t_{a_0} es el momento del primer retorno de la muestra utilizada para estimar $\widehat{\text{ES}}_{t|t-1}(p)$. La interpretación de los valores de Z_3 es similar a las pruebas 1 y 2. Para las tres pruebas *backtesting* presentadas aquí, decidimos usar H_0 usando el valor p de Monte Carlo descrito en la sección 8.3.

6. Evaluaciones de las previsiones generadas por los métodos POT e DPOT

AQUI QUEDE.

Utilizaremos os *backtestings*, apresentados nas Seções 4 e 5, para realizar uma avaliação 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 histórico 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 e 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 e Alves (2012) para aplicar o método DPOT, o pacote *evir* (Pfaff e McNeil (2018)) para aplicar o método POT e o pacote *PerformanceAnalytics* (Peterson e Peter (2019)) para aplicar o método HS. Escolhemos o valor de μ e c 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 decréscimo até o início do ano 2016 e um crescimento até o presente ano. A Tabela 1 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).

Cuadro 1: Estatísticas descritivas dos retornos

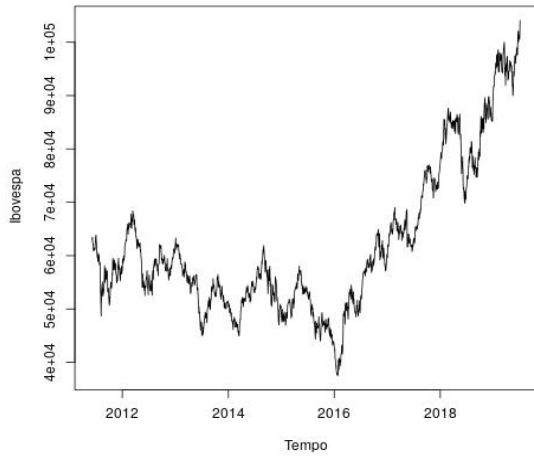
Média	0,0002	1º quartil	-0,0083
Desviación estandar	0,0146	3º quartil	0,0089
Curtosis	1,9170	Mínimo	-0,0921
Assimetria	-0,1489	Máximo	0.0638

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 $\alpha = 0,05$.

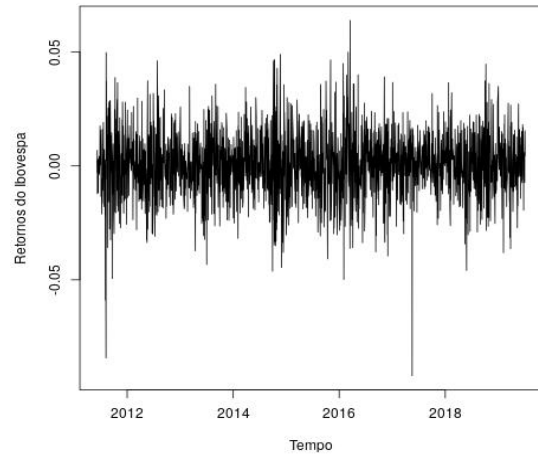
Para $ws = 250$, escolhemos $p = 0,01$ e $th = 0,0215$ para o método POT, mas escolhemos $th = 0,08$ e

²Por exemplo, para uma sequência de previsões $\text{VaR}_1(p), \dots, \text{VaR}_n(p)$, a FV é a proporção em que ocorrem os eventos $\{r_i < \text{VaR}_{i-1}(p)\}_{i=1}^n$.

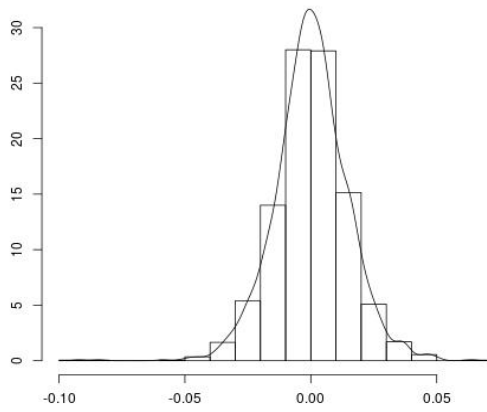
Figura 1: Gráficos de análise descritivo



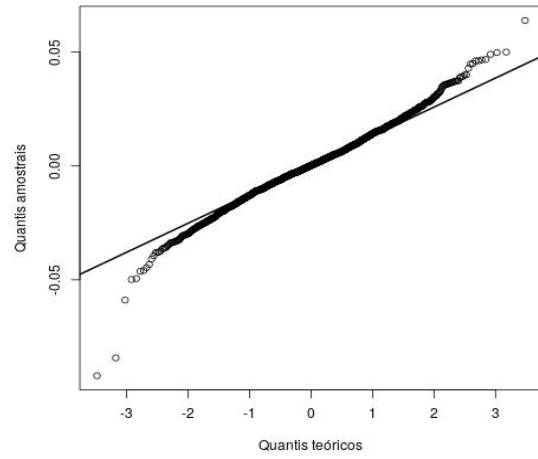
(a) Índice Bovespa



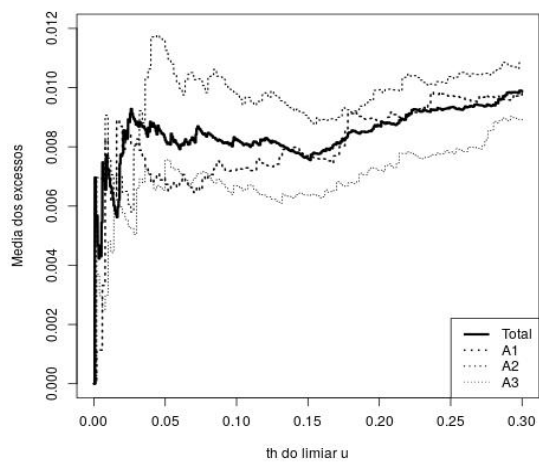
(b) Retornos do índice Bovespa



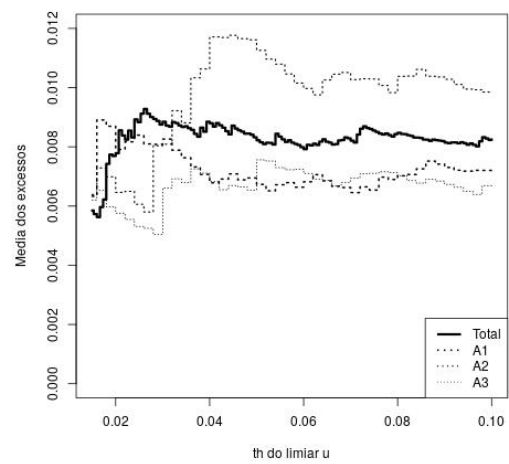
(c) Histograma dos retornos do índice Bovespa



(d) Gráfico Quantil-Quantil normal dos retornos



(e) Média dos excessos parcial e total ($0 \leq th \leq 0,3$)



(f) Média dos excessos parcial e total ($0 \leq th \leq 0,1$)

$c = 1$ 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 2. 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 3. 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 reflexivo 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.

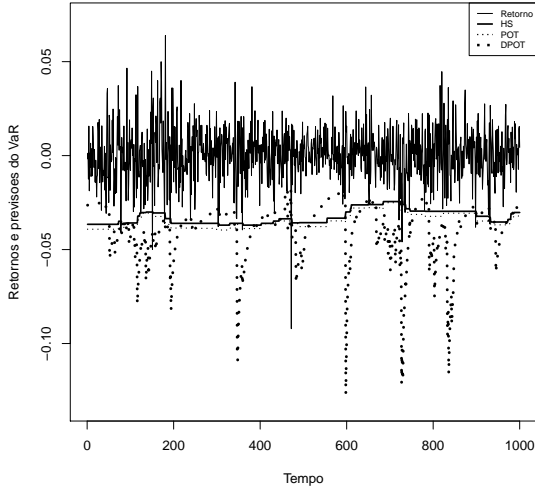
Cuadro 2: Evaluaciones de previsiones ($ws = 250$).

	$p = 0,01$			$p = 0,05$		
	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

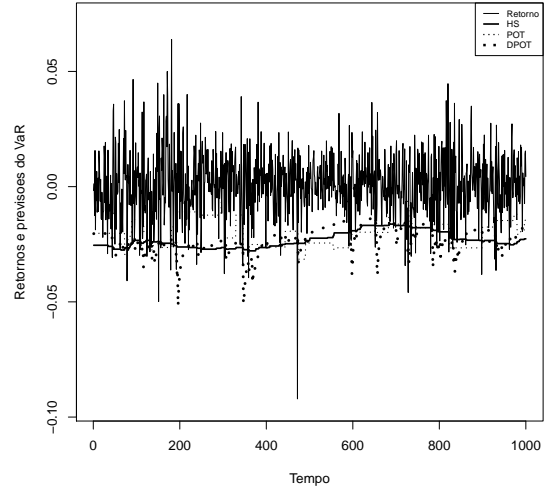
Cuadro 3: Evaluaciones de previsiones ($ws = 500$).

	$p = 0,01$			$p = 0,05$		
	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

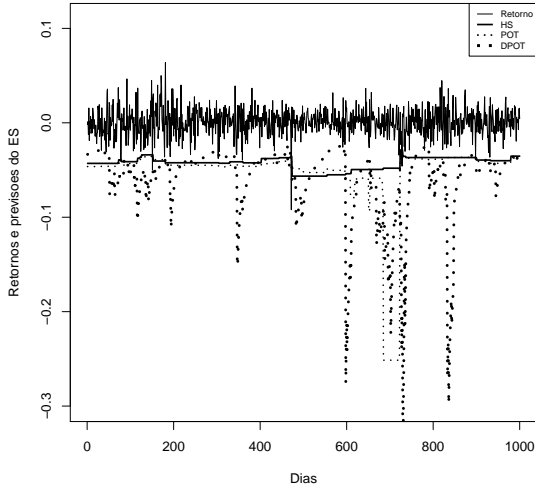
Figura 2: PÃss-amostra e previsÃes do VaR e do ES para $ws = 250$.



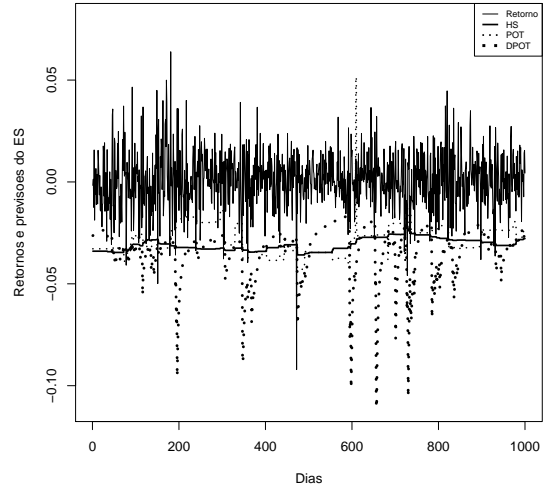
(a) PÃss-amostra e previsÃes do VaR para $p = 0,01$: $th = 0,0215$ para o POT; $th = 0,08$, $c = 1$ e $\lambda = 3$ para o DPOT.



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



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

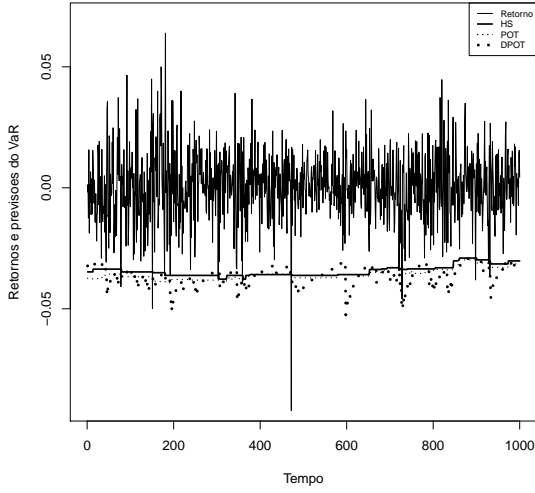


(d) PÃss-amostra e previsÃes do ES para $p = 0,05$: $th = 0,023$ para POT; $th = 0,1$ e $c = 1$ para o DPOT.

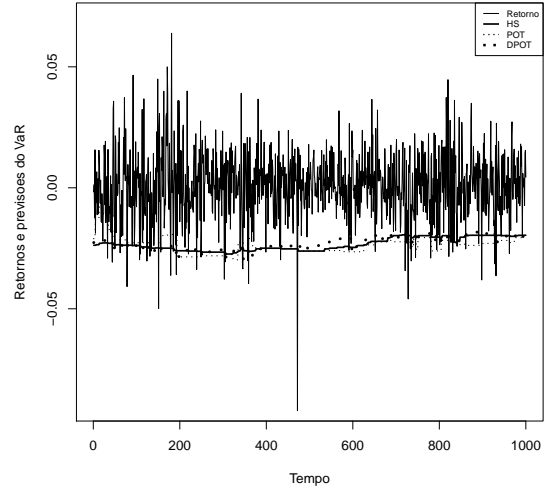
7. Conclusões

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 ??-?? 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 trÃas alternativas de testes do tipo GVAr, com funÃção de risco em (3)–(4), apresentam comportamentos similares em termo do poder empírico. O teste CAViAR tambÃm evidencia bons resultados em comparaÃção 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

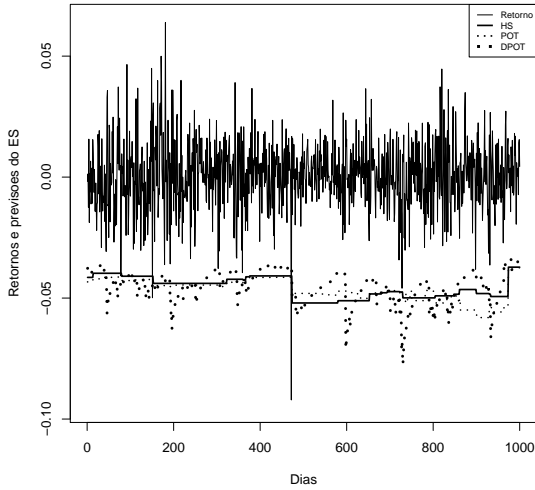
Figura 3: PÃss-amostra e previsÃtes do VaR e do ES para $ws = 500$.



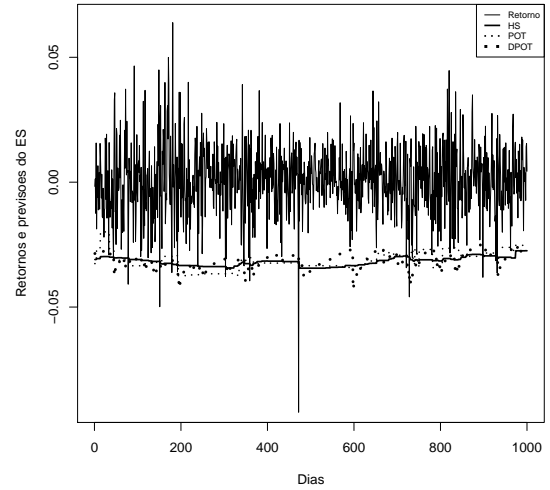
(a) PÃss-amostra e previsÃtes do VaR para $p = 0,01$: $th = 0,025$ para o POT; $th = 0,1$ e $c = 0,3$ para o DPOT.



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



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



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

o ES mostra uma forte dependÃncia do parÃmetro de distorÃÃo Δ . Logo, o poder dos testes cresce junto com Δ e n , mas, o Teste 2 mostrou um baixo poder em quase todos os cenÃrios.

Em aplicaÃÃes Ã sÃrie financeira do Ibovespa, as previsÃtes do VaR e do ES obtidas pelo mÃtodo HS e POT frequentemente geram agrupamento das violaÃÃes. O mÃtodo DPOT foi introduzido com o objetivo de superar esta deficiÃncia. Os testes de avaliaÃÃo foram ferramentas Ãteis para verificar a flexibilidade do mÃtodo DPOT em relaÃÃo ao POT e verificar que um mÃtodo simples como o do tipo HS tambÃm pode ser bastante Ãtil.

Por outro lado, na aplicaÃÃo notamos que os testes podem levar a diferentes decisÃes acerca da hipÃtese de eficiÃncia. PorÃm, o estudo de simulaÃÃo para comparar o poder empÃrico 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.

8. 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.

8.1. 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 $\alpha \in (0, 1)$.

8.2. 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:

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

8.3. 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 e Szekely (2014), o p-valor é estimado por:

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

8.4. 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 8.1.

8.5. 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 e 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

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

em que $\epsilon_{t+1} \sim N(0,1)$, $\phi_1 = -0,051$, $w = 0,00013$, $\alpha = 0,044$, $\gamma = 0,063$ 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 Chann que se transformou no presente artigo. Caso o leitor tivesse 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

Referencias

- Abad, P., Benito, S., & López, C. (2014). A comprehensive review of Value at Risk methodologies. *The 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. November, 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, jan.* 1996.
- BCBS. Fundamental review of the trading book: A revised market risk framework. *Second Consultative Paper*, 2013.
- 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*, 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., Kluppelberg, 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. *The Journal of Risk*, 8(2), 17.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 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. and 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.
- Pérlignon, 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. and 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. URL <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, 2010.
- 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.
- Truquero, 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.