

Intervalos de confianza para VaR y ES, y su aplicación al mercado colombiano

Confidence Intervals for VaR and ES and their Application to the Colombian Market

Jorge Rosales Contreras*

Fecha de recepción: 8 enero 2016

Fecha de aceptación: 10 febrero 2016

Gerencia de Riesgo Financiero
LarrainVial
Santiago, Chile

RESUMEN

Las métricas usuales de riesgo de mercado, tales como Valor en Riesgo (VaR) o Déficit Esperado (ES), se calculan usando estimadores puntuales. Desde un punto de vista estadístico, el VaR es un cuantil y el ES una esperanza condicional de la distribución de pérdidas, la cual puede ser modelada de forma paramétrica o no paramétrica. Sin embargo, un estimador puntual es tan bueno como su precisión, por lo que cualquier estimación de riesgo debería complementarse con alguna indicación de su precisión. En este trabajo construimos intervalos de confianza para los estimadores de ambas métricas bajo las distribuciones más comúnmente usadas: normal y empírica. La utilidad de los intervalos radica en que es posible obtener conclusiones equivalentes a la prueba de backtesting desde la primera estimación de riesgos que se realice, sin necesidad de esperar a tener una muestra de estimaciones de métricas de riesgo.

Clasificación JEL: C12, C13, G17.

Palabras clave: Valor en riesgo, déficit esperado, backtesting, intervalos de confianza.

ABSTRACT

The usual market risk metrics, such as Value at Risk (VaR) or Expected Shortfall (ES), are estimated pointwise. From a statistical viewpoint, VaR is a quantile and ES is a conditional expectation of the loss distribution, which can be modeled parametrically or non-parametrically. However, a point estimator is only as good as its precision; therefore any risk estimation should be accompanied with some indication of its precision. In this paper confidence intervals for the estimators of both metrics, under the most commonly used distributions: Normal and empirical, were calculated. The usefulness of the intervals lies in the possibility of drawing a decision equivalent to backtesting from the very first risk estimation, without having to wait to gather a sample of risk estimates.

JEL Classification: C12, C13, G17.

Keywords: Value at risk, expected shortfall, backtesting, confidence intervals.

Introducción

Una vez que se han estimado las métricas de riesgo por algún tiempo, es importante que la adecuación del modelo de estimación de riesgos sea evaluada periódicamente. Ello supone tanto verificar los supuestos del modelo como cuantificar su desempeño. Esto último se refiere al uso de métodos estadísticos para contrastar las métricas de riesgo pronosticadas con los resultados —reales o hipotéticos— del período posterior.

En términos prácticos, esto significa realizar una prueba de hipótesis. Para ello, se requiere establecer una hipótesis nula que probar, una hipótesis alternativa consistente con el tipo de desviación que se desea validar; así como un nivel de significancia (la máxima probabilidad aceptable de rechazar un modelo válido) independiente del nivel de confianza de la métrica de riesgo. Seleccionar un nivel demasiado pequeño no tiene sentido porque se incrementaría la probabilidad de no rechazar un modelo inválido. Así, generalmente se elige un nivel de significancia arbitrario y se procura construir una prueba con la máxima potencia (habilidad de rechazar un modelo inválido).

Para medir dicha habilidad se construye un estadístico de prueba y se estima su distribución de probabilidades bajo la hipótesis nula. Si el valor de esta probabilidad (valor-p) es menor que el nivel de significancia, la hipótesis nula se rechaza, entendiéndose que la muestra presenta evidencia razonable en su contra.

Si la estimación de métricas de riesgo es una disciplina relativamente reciente, la validación de modelos (*backtesting*) lo es aún más. Una de las primeras y todavía más ampliamente usadas de validación para el VaR es la propuesta por Kupiec (1995), basada en el cociente de verosimilitudes de la distribución binomial. La potencia de esta prueba no es muy alta (salvo para muestras muy grandes) porque no usa ni valida la aleatoriedad ni la magnitud de los excesos.

La prueba de Christoffersen (1998) incorpora la aleatoriedad de forma explícita en el estadístico de prueba, mientras en McNeil *et al.*, (2005) propone una prueba independiente para la aleatoriedad. Engle y Manganelli

(2004) proveen un ejemplo de cómo probar la independencia usando una regresión binaria y lo llaman la prueba dinámica de cuantiles. Algunas pruebas recientes que usan el tiempo transcurrido entre excesos incluyen Christoffersen y Pelletier (2004), Candelon *et al.*, (2011) y Pelletier y Wei (2015). Dowd (2005) presenta algunas pruebas alternativas, mientras Campbell (2006) hace una revisión extensa de modelos de validación.

La popularidad del Déficit Esperado (ES por sus siglas en inglés) como métrica de riesgo creció a raíz de que Artzner *et al.*, (1999) probaron que el VaR no es una medida coherente, Acerbi y Tasche (2002), entre otros, mostraron que el ES sí lo es. Posteriormente, El Comité de Basilea (BIS 2013) decidió reemplazar al VaR a 99% por el ES al 97.5% como la métrica estándar de riesgo de mercado, pero mantuvo el VaR para efectos de validación de modelos.

Validar la estimación de VaR es relativamente sencillo porque basta verificar la fracción de tiempo que la pérdida posterior excedió el VaR y esto se puede hacer con independencia del modelo que produjo el VaR. El caso del ES no es tan directo por dos razones vinculadas entre sí: 1) el ES involucra en general el cálculo de un promedio, pero la pérdida del día siguiente es la realización de un único escenario y 2) el ES depende de la distribución subyacente. Así para validar un modelo de estimación de ES se requieren necesariamente técnicas tipo MonteCarlo o Bootstrap.

Algunos autores proponen métodos paramétricos para la estimación de ES y su *backtesting* (ver McNeil y Frey (2000), por ejemplo). Entre los métodos no paramétricos podemos destacar los trabajos de Kerkhof y Melenberg (2004), quienes usan la funcional delta para construir una prueba que se compara favorablemente con el *backtesting* del VaR de Basilea, así como Acerbi y Szekely (2014), quienes proveen tres estadísticos de prueba que mediante MonteCarlo muestran en general mejor potencia que la prueba de Basilea para VaR a 99%.

Una limitación de los procedimientos de validación de modelos de riesgos más populares es que se basan en aproximaciones asintóticas de la distribución del estadístico de prueba, lo que se traduce en la necesidad de contar con muestras grandes antes de poder tomar una decisión acerca de la adecuación del modelo de estimación de riesgos.

En la práctica, si se usan cuatro años de retornos diarios para estimar las métricas de riesgo y se realiza una prueba de adecuación (*backtest*) contrastando, según recomienda el Comité de Basilea, al menos un año de estimaciones diarias de riesgos con la pérdida del día siguiente, se tendría que

disponer de alrededor de 1 250 datos diarios antes de poder tomar una decisión sobre la bondad del modelo. Desde la primera estimación de riesgos, se debe esperar al menos un año para evaluar el modelo.

Otra limitante del procedimiento usual de *backtesting* es que existen subconjuntos no observables de la región de rechazo de la hipótesis nula. En particular, supóngase que se intenta probar la adecuación de un modelo de VaR estimado a 99% con base en una muestra de un año (250 observaciones). La región de no rechazo de la prueba de Kupiec para esta situación es $[0,6]$, esto es, se considera que el modelo subestima el riesgo si se observan más de seis excesos. Sin embargo, no será posible detectar si el modelo sobre-estima el riesgo, ya que es imposible observar un número negativo de excesos. Este será el caso siempre que el riesgo se estime a un nivel de confianza alto.

Para superar estas limitaciones es importante conocer la distribución de las métricas de riesgo al considerarlas como variables aleatorias. Para el VaR, por ejemplo, esto equivale a encontrar la distribución de un estadístico de orden, tal como lo hace Dowd (2005). Sin embargo, la distribución muestral es siempre función de la distribución de los factores de riesgo, lo que puede complicar su instrumentación práctica.

Respecto a intervalos de confianza, Dowd (2006) se basa en estadísticos de orden para estimar intervalos para VaR que funcionan bajo modelos paramétricos y no paramétricos. Moraux (2011) construye el intervalo para el VaR Normal, como en la sección 2.2.2, mejorando la estimación de Jorion (1996), quien ignoraba la media y, por tanto, su dispersión. Cotter y Dowd (2007) usan simulación MonteCarlo para construir intervalos para VaR, ES y medidas espectrales, partiendo de la Normal y luego alejándose en cuando a asimetría y kurtosis de la distribución de retornos. Gao y Song (2008) y Spierdijk (2014) construyen intervalos para estimadores basados en modelos GARCH.

En este trabajo se propone un método completamente general para el cálculo de intervalos de confianza que se basa en el estimador máximo verosímil y sus propiedades distribucionales asintóticas. Al usar propiedades asintóticas, no se pretende resolver el problema del tamaño de la muestra para la toma de decisiones. Sin embargo, en la medida que se suponga una distribución específica para los factores de riesgo, es posible construir intervalos más precisos con la misma muestra que se usó para estimar los parámetros.

En estricto sentido, las pruebas de validación tienen el componente de la aleatoriedad (McNeil *et al.*, (2005)) e independencia (Christoffersen (1998)) de los excesos, pero las métricas de riesgo heredan esa propiedad de los retornos de los factores de riesgo y consideramos, por tanto, que la independencia debe validarse antes de estimar el riesgo. Campbell (2006) recomienda efectuar las pruebas de independencia o aleatoriedad de forma separada de la prueba de cobertura no condicional (frecuencia de excesos). En este sentido, la metodología de intervalos de confianza va más allá de los modelos tradicionales de validación al incluir la magnitud de los excesos y no sólo la frecuencia de éstos.

El resto del artículo está dividido en 4 secciones. En la primera Sección se hace una revisión muy general de los procedimientos usuales de validación de modelos. En la Sección 2 se construyen los intervalos de confianza en el caso general, cuando los estimadores son máximo verosímiles, así como los intervalos de confianza para VaR y ES en los casos de Simulación Histórica (SH) y Normal. En la Sección 3 se demuestra la equivalencia entre las pruebas de validación e intervalos de confianza para ambas métricas. La Sección 4 presenta una aplicación al caso del mercado colombiano y se discuten los resultados. Finalmente, se redondean los resultados y enuncia algunas conclusiones.

1. Pruebas de validación (Backtesting)

Cuando las métricas de riesgo se han estimado sistemáticamente por algún período de tiempo, es posible evaluar el desempeño del modelo de estimación riesgo. A este proceso se le conoce como *backtesting*.¹

Sea L la variable aleatoria (v.a.) de pérdida de la cartera y supóngase que para cada período t se tienen la pérdida aleatoria L_t y estimaciones para VaR y ES a un nivel de confianza α , denotados $VaR(\alpha, t)$ y $ES(\alpha, t)$, respectivamente. Defínanse las v.a. indicadora de exceso y pérdida en exceso como

$$\begin{aligned}\hat{1}_\alpha(L_{t+1}) &:= 1_{(VaR(\alpha, t), +\infty)}(L_{t+1}) \text{ y} \\ \hat{M}_{\alpha, t+1}(L_{t+1}) &:= (L_{t+1} - ES(\alpha, t))\hat{1}_\alpha(L_{t+1}).\end{aligned}\tag{1.1}$$

¹ Aunque es posible traducir *backtesting* como prueba de validación, prueba de adecuación o prueba retrospectiva, se mantendrá el anglicismo dado su uso extendido.

Para el caso del VaR, se esperaría que el número de excesos después de n estimaciones siga una distribución binomial con parámetros n y $1-\alpha$, donde $1-\alpha$ es la probabilidad de que la pérdida del período $t+1$ exceda el VaR estimado en el período t , para cualquier valor de t . Haciendo $p=1-\alpha$, Kupiec (1995) construye una prueba para

$$H_0: p = p_0 \text{ vs } H_A: p \neq p_0 \quad (1.2)$$

basada en el cociente de verosimilitudes como estadístico de prueba. Asintóticamente, este estadístico sigue una distribución χ_1^2 , pero para un tamaño de muestra fijo Casella y Berger (2002) calculan la región de rechazo exacta para la distribución binomial.

Con respecto al ES, se puede esperar que las pérdidas en exceso del VaR se comporten como realizaciones de variables aleatorias independientes e idénticamente distribuidas que siguen una distribución con media cero y tienen masa de probabilidad de tamaño α en cero. Realizar una prueba de esta segunda propiedad es equivalente al *backtesting* del VaR, por lo que nos concentraremos en probar el comportamiento de media cero:

$$H_0: E[\widehat{M}_{\alpha,t+1}(L_{t+1})] = 0 \text{ vs } H_A: E[\widehat{M}_{\alpha,t+1}(L_{t+1})] \neq 0 \quad (1.3)$$

Dado que la prueba se realiza sobre la media de una distribución, el estadístico de prueba por excelencia es de la forma

$$t(\mathbf{m}) = \frac{\bar{m} - m_0}{\hat{\sigma}/\sqrt{n}}, \quad (1.4)$$

donde \bar{m} es la media de las pérdidas en exceso observadas $\{\widehat{m}_{\alpha,t}(l_t)\}_{t=1}^n$, n es el tamaño de la muestra, $m_0=0$ y $\hat{\sigma}$ es la desviación estándar de la muestra de excesos.

La hipótesis alternativa en (1.3) se puede plantear tanto para validar igualdad (dos colas) como la falta de simetría (una cola). En este caso se usará la hipótesis alternativa de antisimetría positiva $H_A: E[\widehat{M}_{\alpha,t+1}(L_{t+1})] > 0$.

Dowd (2005) propone hacer una prueba t "de libro de texto", pero ello no es evidente porque la distribución del estadístico depende de la distribu-

ción de pérdidas. Por ello, Efron y Tibshirani (1994) proponen una prueba de hipótesis *bootstrap* no paramétrica basada en la distribución empírica del estadístico de prueba bajo la hipótesis nula.

Si la distribución de la v.a. de pérdida L es la empírica, se puede usar el *bootstrap* no paramétrico, muestreando con reemplazo de $\{\hat{m}_{\alpha,t}(l_t)\}_{t=1}^n$; mientras que si se ha supuesto una distribución específica para L , se puede usar una versión paramétrica del *bootstrap*, muestreando de la distribución ajustada. Sólo en el caso que se suponga la distribución normal para L es que se puede realizar una prueba t de libro de texto.

En cualquier caso, si sólo se tiene un exceso sobre el VaR, no es posible estimar σ y, por tanto, tampoco es posible probar H_0 usando el estadístico de la ecuación (1.4).

El *bootstrap* no produce una región de rechazo o —su equivalente— un valor crítico, sino un nivel de significancia asintótico (ASL, por sus siglas en inglés) para la hipótesis alternativa, el cual se estima como

$$\widehat{ASL}_{boot} = \#\{t(\mathbf{m}) > t_{obs}\}/N, \quad (1.5)$$

donde $t_{obs} = t(\hat{\mathbf{m}}_{\alpha})$ es el valor del estadístico (1.4) observado en la muestra original y N es el número de repeticiones *bootstrap*.

Como lo hacen notar Efron y Tibshirani, el estimador \widehat{ASL}_{boot} no se interpreta como una probabilidad exacta, sino que, como todos los estimadores *bootstrap*, su precisión crece con el tamaño de la muestra.

2. Intervalos de confianza

2.1 Notación y fundamentos

Igual que antes, sea L la v.a. —absolutamente continua— de pérdida de la cartera, con función de distribución acumulada F_L . L_1, \dots, L_n denota una muestra aleatoria de L y $L_{(1)} < \dots < L_{(n)}$ los estadísticos de orden.

Siempre que la función de distribución F_L se suponga conocida, el vector $\boldsymbol{\theta}$ representará los parámetros y $\hat{\boldsymbol{\theta}}_{MV}$ el estimador de Máxima Verosimilitud (MV). Por otra parte, Φ denotará la función de distribución de la Normal estándar, ϕ su función de densidad y z_{α} su cuantil α , i.e., $\Phi(z_{\alpha}) = \alpha$, para cualquier $\alpha \in [0,1]$.

Bajo el supuesto de normalidad de la pérdida L , el único estimador insesgado de varianza mínima (UMVUE por sus siglas en inglés) de la varianza σ_L^2 es la varianza muestral $S^2 = \sum_{j=1}^n (L_j - \bar{L})^2 / (n - 1)$ y la variable aleatoria $(n - 1)S^2 / \sigma_L^2$ se distribuye χ_{n-1}^2 . En particular, $E[(n - 1)S^2 / \sigma_L^2] = n - 1$, esto es, $ES^2 = \sigma_L^2$ y además (ver Lehmann y Casella (1998) sección 1.2):

$$E\left(\sqrt{n-1} \frac{S}{\sigma_L}\right) = E[(\chi_{n-1}^2)^{1/2}] = \frac{\sqrt{2}\Gamma(n/2)}{\Gamma[(n-1)/2]} = \frac{\sqrt{n-1}}{K_{n-1}}, \quad (2.1)$$

lo que implica que $ES = \sigma_L / K_{n-1}$, con $K_{n-1} = \frac{\sqrt{n-1} \Gamma[(n-1)/2]}{\sqrt{2} \Gamma(n/2)}$. Dicho de otra forma, $\widehat{\sigma}_L = K_{n-1} S$ (y no S) es el UMVUE de la desviación estándar. Luego calculamos

$$\begin{aligned} \text{Var}(S) &= E(S^2) - (ES)^2 = \sigma_L^2 - (\sigma_L / K_{n-1})^2 \\ &= \sigma_L^2 (1 - K_{n-1}^{-2}) = \sigma_L^2 V_{n-1} / (n - 1) \\ &\approx \sigma_L^2 / [2(n - 1)], \end{aligned} \quad (2.2)$$

donde $V_{n-1} = (n - 1)(1 - K_{n-1}^{-2})$ es la varianza de una variable aleatoria con distribución χ_{n-1} y la última aproximación se obtiene de $\text{Var}[g(Y)] \approx \sigma_Y^2 [g'(\mu_Y)]^2$ con $g(y) = \sqrt{y}$. Finalmente, se sigue que

$$\text{Var}(\widehat{\sigma}_L) = K_{n-1}^2 \text{Var}(S) = K_{n-1}^2 V_{n-1} \frac{\sigma_L^2}{n - 1}. \quad (2.3)$$

Cuando n es suficientemente grande, se puede usar la expresión

$$\frac{\Gamma(n/2)}{\Gamma[(n-1)/2]} = \frac{\sqrt{n-1}}{\sqrt{2}} \left(1 - \frac{1}{4(n-1) + O(n^{-2})}\right) \quad (2.4)$$

para obtener la aproximación $K_{n-1}^2 V_{n-1} \rightarrow 1/2$ cuando $n \rightarrow \infty$. Por tanto,

$$\text{Var}(\widehat{\sigma}_L) \approx \frac{\sigma_L^2}{2(n-1)}. \quad (2.5)$$

Sea ahora $h(\boldsymbol{\theta})$ la métrica de riesgo (VaR o ES) expresada como función de los parámetros de la distribución de pérdida de la cartera. En el caso del $\text{VaR}_\alpha(\boldsymbol{\theta})$, McNeil, *et al.*, (2005) define $h(\boldsymbol{\theta})$ mediante la función cuantil de la pérdida:

$$h(\boldsymbol{\theta}) := Q(\alpha, \boldsymbol{\theta}) = F_L^{-1}(\alpha; \boldsymbol{\theta}) \quad (2.6)$$

mientras que para $\text{ES}_\alpha(\boldsymbol{\theta})$ los mismos autores definen

$$h(\boldsymbol{\theta}) := \frac{1}{1-\alpha} \int_\alpha^1 Q(\beta, \boldsymbol{\theta}) d\beta. \quad (2.7)$$

En ambas instancias $h(\boldsymbol{\theta})$ es una función continua de $(\boldsymbol{\theta})$ y también es diferenciable.

La propiedad de invarianza de los estimadores MV garantiza que $h(\widehat{\boldsymbol{\theta}}_{MV})$ es también un estimador MV (ver por ejemplo Casella y Berger (2002)) y, bajo las condiciones usuales de regularidad, se tiene:

$$[h(\widehat{\boldsymbol{\theta}}_{MV}) - h(\boldsymbol{\theta})] \xrightarrow{d} N\left(0, \text{Var}\left(h(\widehat{\boldsymbol{\theta}}_{MV})\right)\right), \text{ cuando } n \rightarrow \infty. \quad (2.8)$$

Se usará en forma reiterada este resultado, que puede ser consultado, junto con el significado de *condiciones usuales de regularidad* en Casella y Berger (2002), Sección 9.4.

Aun cuando este es un resultado asintótico, es importante aclarar que aquí n se refiere al tamaño de muestra usado para estimar $(\boldsymbol{\theta})$ (conocida como la *ventana temporal*), a diferencia del número de períodos con métricas de riesgo calculadas, como se requiere en los procedimientos usuales de *backtesting*. Así, (2.8) es válida desde la primera estimación de las métricas de riesgo, siempre que la ventana temporal sea suficientemente larga.

Basta con calcular o aproximar $\text{Var}\left(h(\widehat{\boldsymbol{\theta}}_{MV})\right)$, la varianza de la métrica de riesgo, para construir un intervalo de confianza para $h(\boldsymbol{\theta})$. Dado que $h(\cdot)$

en general no es una función sencilla de sus argumentos (θ), se puede usar la aproximación lineal dada por la expansión en series de Taylor de primer orden:

$$h(\hat{\theta}) = h(\theta) + \nabla h(\theta)'(\hat{\theta} - \theta) + Res, \quad (2.9)$$

donde

$$(\nabla h(\theta))_j = \frac{\partial h(\theta)}{\partial \theta_j}, j = 1, \dots, m. \quad (2.10)$$

Nuevamente, estas derivadas parciales pueden ser calculadas en forma cerrada o bien numéricamente, dependiendo de la complejidad de la función $h(\cdot)$. Al ignorar el residuo se obtiene la aproximación

$$h(\hat{\theta}) - h(\theta) \approx \nabla h(\theta)'(\hat{\theta} - \theta). \quad (2.11)$$

Es claro que la varianza del estimador de la métrica dependerá de la covarianza de los parámetros, $cov(\hat{\theta}_{MV})$, estimados por máxima verosimilitud.

Se sabe que la matriz de covarianza de estos estimadores es asintóticamente igual a la inversa de la matriz de información esperada. Efron y Hinkley (1978) proveen una justificación para usar la matriz de información observada en lugar de la esperada, ya que no involucra el cálculo de la esperanza. Sin embargo, calcular el Hessiano de la log-verosimilitud puede no ser sencillo, por lo que Meilijson (1989) propone usar lo que llama la matriz de información observada *empírica*, como aproximación (I_e^{-1} , ver McLachlan y Krishnan (2008)). Así, se obtiene

$$\begin{aligned} Var(h(\hat{\theta}_{MV})) &\approx [\nabla h(\theta)' I_e^{-1}(\theta) \nabla h(\theta)]|_{\theta=\hat{\theta}_{MV}} \\ &= \widehat{\sigma_{h(\hat{\theta})}^2}. \end{aligned} \quad (2.12)$$

Finalmente, el intervalo de confianza asintótico con un nivel de confianza δ para el estimador de $h(\theta)$ está dado por

$$\left[h(\hat{\theta}_{MV}) - z_{(1+\delta)/2} \sqrt{\widehat{\sigma}_{h(\theta)}^2}, h(\hat{\theta}_{MV}) + z_{(1+\delta)/2} \sqrt{\widehat{\sigma}_{h(\theta)}^2} \right]. \quad (2.13)$$

2.2. Intervalos de confianza para VaR

2.2.1 Simulación histórica

El estimador no paramétrico usual para VaR_α es $L_{\lfloor n\alpha \rfloor}$, donde $\lfloor x \rfloor$ representa el mayor entero que es menor o igual que x . Se puede construir un intervalo de confianza que no dependa de la distribución de L , a partir de la distribución del k -ésimo estadístico de orden:

$$F_{L_{(k)}}(u) = \sum_{j=k}^n \binom{n}{j} [F_L(u)]^j [1 - F_L(u)]^{n-j} \quad (2.14)$$

y de la siguiente propiedad (con $i < j$):

$$\begin{aligned} & \Pr(L_{(i)} \leq F_L^{-1}(\alpha) \leq L_{(j)}) \\ &= \Pr(L_{(i)} \leq F_L^{-1}(\alpha)) - \Pr(L_{(j)} < F_L^{-1}(\alpha)) \quad (2.15) \\ &= \sum_{k=i}^{j-1} \binom{n}{k} \alpha^k (1 - \alpha)^{n-k} = \delta(i, j). \end{aligned}$$

donde la continuidad absoluta de F_L se ha usado en la segunda igualdad. Dados α y el nivel de confianza deseado para el intervalo (δ), escójense i y j tales que $\delta(i, j) \geq \delta$. Como $\delta(i, j)$ es una función de escalón, el intervalo de confianza obtenido tiende a ser conservador.

Para n y α dados, el mapeo $k \mapsto \Pr(L_{(k)} \leq F_L^{-1}(\alpha))$ es creciente si $k < [n\alpha]$ y decreciente cuando $k > [n\alpha]$. Así, se inicia con un intervalo pequeño escogiendo valores de i y j cercanos a $[n\alpha]$ y se continúa alejándose de este valor hasta que se alcanza el nivel de confianza deseado para el intervalo.

Por lo tanto, el intervalo de confianza del 100 $\delta\%$ para VaR_α es

$$[L_{(i)}, L_{(j)}]. \quad (2.16)$$

donde $i < [n\alpha] < j$ se escogen según el proceso anterior tales que $\delta(i, j) \geq \delta$.

Obsérvese que no existe una manera única de escoger el intervalo y que éste no tiene por qué ser simétrico respecto del estimador de VaR. Por ejemplo, Dowd (2005) propone un proceso *bootstrap* equivalente a escoger i y j de modo que $\delta(1, i) \approx \delta(j, n) \approx \delta/2$.

2.2.2 Delta-Normal

Supóngase ahora que L tiene una distribución Normal con parámetros $\boldsymbol{\theta} = (\mu_L, \sigma_L^2)'$. Sean $\hat{\mu}_{MV} = \bar{L}$ y $\hat{\sigma}_{MV}^2 = \frac{1}{n} \sum_{j=1}^n (L_j - \bar{L})^2 = \frac{n-1}{n} S^2$ los estimadores MV usuales.

En este caso $\text{VaR}_\alpha = h(\boldsymbol{\theta}) = \mu_L + z_\alpha \sigma_L$. De la teoría de inferencia estadística (ver por ejemplo, Lehmann y Casella (1998) sección 2.2) sabemos que

$$\widehat{\text{VaR}}_\alpha = h(\hat{\boldsymbol{\theta}}) = \hat{\mu}_{MV} + z_\alpha \hat{\sigma}_L \quad (2.17)$$

es el UMVUE para VaR_α . Este estimador se puede escribir en términos del estimador MV como $\widehat{\text{VaR}}_\alpha = \hat{\mu}_{MV} + z_\alpha K_{n-1} \sqrt{\hat{\sigma}_{MV}^2} \sqrt{\frac{n}{n-1}}$ y por la propiedad de invarianza es él mismo un estimador MV de VaR_α .

Se puede entonces usar la normalidad asintótica de los estimadores MV (2.8) para obtener: $[\widehat{\text{VaR}}_\alpha - \text{VaR}_\alpha] \xrightarrow{d} N(0, \text{Var}(\widehat{\text{VaR}}_\alpha))$ cuando $n \rightarrow \infty$. Así, basta con calcular la varianza del estimador MV, $\text{Var}(\widehat{\text{VaR}}_\alpha)$ ya sea mediante (2.12) o directamente —como en este caso—, para construir un intervalo de confianza para VaR_α . En efecto, a partir de

$$\text{Var}(\hat{\mu}_{MV}) = \frac{\sigma_L^2}{n} \text{ y } \text{Var}(\hat{\sigma}_L) = \frac{\sigma_L^2}{n-1} K_{n-1}^2 V_{n-1} \approx \frac{\sigma_L^2}{2(n-1)}, \quad (2.18)$$

se tiene

$$\begin{aligned} \text{Var}(\widehat{\text{VaR}}_\alpha) &= \text{Var}(\hat{\mu}_{MV} + z_\alpha \hat{\sigma}_L) \\ &= \text{Var}(\hat{\mu}_{MV}) + z_\alpha^2 \text{Var}(\hat{\sigma}_L), \end{aligned} \quad (2.19)$$

ya que \bar{L} y S son independientes y, por lo tanto, $\text{cov}(\hat{\mu}_{MV}, \hat{\sigma}_L) = 0$. Luego, de las igualdades (2.18) se sigue

$$\text{Var}(\widehat{\text{VaR}}_\alpha) = \frac{\sigma_L^2}{n} + z_\alpha^2 \frac{\sigma_L^2}{n-1} K_{n-1}^2 V_{n-1} \approx \frac{\sigma_L^2}{n} + z_\alpha^2 \frac{\sigma_L^2}{2(n-1)}. \quad (2.20)$$

Como $\text{Var}(\widehat{\text{VaR}}_\alpha)$ depende de σ_L^2 , es necesario estimarla mediante

$$\widehat{\text{Var}}(\widehat{\text{VaR}}_\alpha) = \frac{\widehat{\sigma}_L^2}{n} + z_\alpha^2 \frac{\widehat{\sigma}_L^2}{2(n-1)} = \hat{\sigma}_{\widehat{\text{VaR}}_\alpha}^2 \quad (2.21)$$

o su simplificación $\hat{\sigma}_{\widehat{\text{VaR}}_\alpha}^2 = \frac{\widehat{\sigma}_L^2}{n} \left(1 + \frac{z_\alpha^2}{2}\right)$. En cualquier caso, el intervalo de confianza al $100\delta\%$ para VaR_α está dado por

$$\left[\widehat{\text{VaR}}_\alpha - \frac{z_{1+\delta}}{2} \hat{\sigma}_{\widehat{\text{VaR}}_\alpha}, \widehat{\text{VaR}}_\alpha + \frac{z_{1+\delta}}{2} \hat{\sigma}_{\widehat{\text{VaR}}_\alpha} \right], \quad (2.22)$$

donde el estimador $\widehat{\text{VaR}}_\alpha$ se obtiene según la ecuación (2.17), $\hat{\sigma}_{\widehat{\text{VaR}}_\alpha}^2$ se calcula mediante (2.21) o su simplificación y z_δ es el cuantil δ de la distribución Normal estándar.

2.3. Intervalos de confianza para ES

2.3.1. Simulación histórica

El estimador no paramétrico de ES_α está dado por

$$\widehat{ES}_\alpha = \frac{1}{n - [n\alpha]} \sum_{j=[n\alpha]+1}^n L_{(j)} \quad (2.23)$$

Para aproximar la distribución de \widehat{ES}_α se usará el siguiente resultado asintótico de David y Nagaraja (2003):

$$\sqrt{n}[\widehat{ES}_\alpha - ES_\alpha] \xrightarrow{d} N(0, \sigma_\alpha^2 + \alpha(ES_\alpha - VaR_\alpha)^2) \quad (2.24)$$

donde $\sigma_\alpha^2 = Var(L|L > VaR_\alpha)$. De esta manera, el intervalo de confianza al nivel $100\delta\%$ para ES_α es

$$\left[\widehat{ES}_\alpha - \frac{z_{1+\delta}}{2} \hat{\sigma}_{\widehat{ES}_\alpha}, \widehat{ES}_\alpha + \frac{z_{1+\delta}}{2} \hat{\sigma}_{\widehat{ES}_\alpha} \right] \quad (2.25)$$

donde \widehat{ES}_α está dado por la ecuación (2.23), $\hat{\sigma}_{\widehat{ES}_\alpha}^2$ es la varianza de la distribución Normal en (2.24) y z_δ es el cuantil δ de la distribución Normal estándar.

2.3.2. Delta-Normal

Sean $\widehat{\mu}_L = \bar{L}$ y $\widehat{\sigma}_L = K_{n-1}S$ y los estimadores descritos anteriormente, basados en los estimadores MV, con varianzas dadas por (2.18). McNeil *et al.*, (2005) muestran que, bajo el supuesto de normalidad, $ES_\alpha = h(\boldsymbol{\theta}) = \mu_L + c_\alpha \sigma_L$, donde $c_\alpha = \frac{1}{1-\alpha} \phi(\Phi^{-1}(\alpha))$. Su estimador es

$$\widehat{ES}_\alpha = h(\widehat{\boldsymbol{\theta}}) = \widehat{\mu}_L + c_\alpha \widehat{\sigma}_L \quad (2.26)$$

el cual se puede escribir en términos del estimador MV tal como en la Sección 2.2.2 y es por tanto un estimador MV en sí mismo para ES_α debido a la propiedad de invarianza de los estimadores MV.

De la linealidad de la ecuación (2.26) y el hecho de que $\widehat{\mu}_L$ y $\widehat{\sigma}_L$ son insesgados, se sigue que \widehat{ES}_α es un estimador insesgado de ES_α . Entonces, por (2.8) basta con encontrar la varianza de \widehat{ES}_α para construir un intervalo de confianza para ES_α :

$$\begin{aligned} Var(\widehat{ES}_\alpha) &= Var(\widehat{\mu}_L + c_\alpha \widehat{\sigma}_L) \\ &= Var(\widehat{\mu}_L) + c_\alpha^2 Var(\widehat{\sigma}_L), \end{aligned} \quad (2.27)$$

dado que \bar{L} y S son independientes y, por lo tanto, $cov(\widehat{\mu}_L, \widehat{\sigma}_L) = 0$. Siguiendo un desarrollo análogo a la Sección 2.2.2 se obtiene

$$\begin{aligned} \widehat{Var}(\widehat{ES}_\alpha) &= \frac{\sigma_L^2}{n} + c_\alpha^2 \frac{\sigma_L^2}{n-1} K_{n-1}^2 V_{n-1} \Big|_{\theta=\widehat{\theta}} \\ &= \frac{\widehat{\sigma}_L^2}{n} + c_\alpha^2 \frac{\widehat{\sigma}_L^2}{n-1} K_{n-1}^2 V_{n-1} \\ &\approx \frac{\widehat{\sigma}_L^2}{n} \left[1 + \frac{n}{n-1} \frac{c_\alpha^2}{2} \right] = \widehat{\sigma}_{ES_\alpha}^2, \end{aligned} \quad (2.28)$$

lo cual se puede simplificar a $\widehat{\sigma}_{ES_\alpha}^2 = \frac{\widehat{\sigma}_L^2}{n} \left[1 + \frac{c_\alpha^2}{2} \right]$ para n grande. El intervalo de confianza al nivel $100\delta\%$ para ES_α es entonces

$$\left[\widehat{ES}_\alpha - \frac{z_{1+\delta}}{2} \widehat{\sigma}_{ES_\alpha}, \widehat{ES}_\alpha + \frac{z_{1+\delta}}{2} \widehat{\sigma}_{ES_\alpha} \right], \quad (2.29)$$

donde \widehat{ES}_α está dado por la ecuación (2.26), $\widehat{\sigma}_{ES_\alpha}^2$ está dado por (2.28) y Z_δ es el cuantil δ de la distribución Normal estándar.

3. Equivalencia entre Backtesting e Intervalos de Confianza

En la Sección 1 se establecieron las hipótesis de los procedimientos de *backtesting* para VaR y ES. Por otro lado, de la teoría de inferencia estadística sabemos que existe una correspondencia entre un intervalo de confianza y una prueba de hipótesis (ver por ejemplo Casella y Berger (2002), Teorema 9.2.1). Así, la prueba de hipótesis que está asociada al intervalo (2.13) es

$$H_0: h(\boldsymbol{\theta}) = h(\widehat{\boldsymbol{\theta}}) \text{ vs } H_A: h(\boldsymbol{\theta}) \neq h(\widehat{\boldsymbol{\theta}}). \quad (3.1)$$

En el resto de esta sección enunciamos y probamos, con base en la propiedad anterior, los resultados que permiten establecer una equivalencia entre las pruebas de hipótesis de la Sección 1 y los intervalos de confianza de la Sección 2.

Proposición 1. Sea $h(\cdot)$ como en la ecuación (2.6). Entonces, la región de no rechazo de la prueba de hipótesis (1.2) es igual al intervalo de confianza (2.13).

Demostración. Basta probar que la prueba de hipótesis (1.2) corresponde a (3.1). En efecto, partiendo de la primera ecuación se tiene:

$$\begin{aligned} p &= p_0 \\ \Leftrightarrow E \left[1_{\{L > \widehat{\text{VaR}}_\alpha\}} \right] &= p_0 \\ \Leftrightarrow P(L > \widehat{\text{VaR}}_\alpha) &= p_0 \\ \Leftrightarrow P(L \leq \widehat{\text{VaR}}_\alpha) &= 1 - p_0 = \alpha \\ \Leftrightarrow \widehat{\text{VaR}}_\alpha &= Q(\alpha, \boldsymbol{\theta}) \\ \Leftrightarrow h(\widehat{\boldsymbol{\theta}}) &= h(\boldsymbol{\theta}). \end{aligned} \quad (3.2)$$

Así, por ejemplo, sean $n = 250$, $p = 1\%$ y $\delta = 99\%$, el intervalo para np usando (2.16) arriba, es $[L_{(1)}, L_{(6)}]$. Dicho de otra manera, el modelo se considera adecuado si se observan entre 0 y 6 excesos en 250 días hábiles, tal como ocurre con la prueba de Kupiec.

Por lo tanto, se tiene una alerta temprana cuando una observación excede el límite superior del intervalo de confianza; mientras que el modelo está siendo demasiado conservador cuando no existen observaciones por encima del límite inferior de dicho intervalo.

Obsérvese que no ocurre la limitante de la prueba de Kupiec discutida en la Sección 1, ya que es factible observar pérdidas tanto por abajo de $L_{(1)}$ como por arriba de $L_{(6)}$.

La siguiente Proposición establece la equivalencia entre el proceso de *backtesting* y la construcción del intervalo de confianza para ES.

Proposición 2. Sea $h(\cdot)$ como en la ecuación (2.7). Entonces, la región de no rechazo de la prueba de hipótesis (1.3) es igual al intervalo de confianza (2.13).

Demostración. Claramente, la prueba de hipótesis (1.3) corresponde exactamente a (3.1), debido a que:

$$\begin{aligned} E[\widehat{M}_{\alpha,t+1}(L)] &= E\left[(L_{t+1} - \widehat{ES}_{\alpha})1_{\{L > \widehat{VaR}_{\alpha}\}}\right] \\ &= E(L|L > \widehat{VaR}_{\alpha}) - E\left[\widehat{ES}_{\alpha} 1_{\{L > \widehat{VaR}_{\alpha}\}}\right] \quad (3.3) \\ &= ES_{\alpha} - E(\widehat{ES}_{\alpha}) \\ \therefore E[\widehat{M}_{\alpha,t+1}] = 0 &\Leftrightarrow ES_{\alpha} = E(\widehat{ES}_{\alpha}). \end{aligned}$$

Se concluye entonces que es equivalente hacer el *backtesting* para ES (respecto de la propiedad de media cero) y construir el intervalo de confianza, con la ventaja de que no se requiere esperar para acumular la muestra que permita probar H_0 .

4. Resultados empíricos

En esta sección se propone una cartera de activos, a la cual se le calculan las métricas de riesgo, se aplican las pruebas de validación y se construyen los intervalos de confianza descritos, comparando las conclusiones obtenidas con cada estrategia.

La cartera propuesta (ver Cuadro 1) se compone de posiciones largas en seis acciones del mercado colombiano y una posición corta en dólares.

El período de vigencia de la cartera incluye desde julio de 2001 hasta junio de 2015. Cuando no existe precio para una acción, se ha usado su beta respecto del índice de la bolsa de valores de Colombia (IGBC) y el rendimiento de éste.

Cuadro 1. Descripción de la cartera

Instrumento	PFBCOLO	ECOPETL	ISAGEN	PREC	CNEC	ISA	USDCOP
Acciones o Nominal	1 800	30 000	18 000	5 000	8 500	7 000	-20 000

Fuente: Elaboración propia.

Todos los cálculos de VaR se han realizado al 99% con una ventana histórica de 1 000 días hábiles y un horizonte de tenencia de un día hábil. Los intervalos de confianza para el VaR también se han calculado al 99% de confianza. Para el ES, los cálculos y los intervalos se han hecho al 97.5% de confianza, según lo propone el BIS (2013), con la misma ventana histórica y horizonte. Así, tanto VaR como ES se han calculado desde agosto 2005 hasta junio 2015. El Cuadro 2 muestra un resumen de las métricas de riesgo individuales y para la cartera durante los 10 años considerados por ambos métodos de cálculo.

Cuadro 2. VaR(99%) para la Cartera Considerada (MXN mln)

Métrica	VaR(99%)		ES(97.5%)	
	SH	Delta-Normal	SH	Delta-Normal
PFBCOLO	1.955	1.654	2.006	1.662
ECOPETL	4.317	3.552	4.364	3.570
ISAGEN	1.389	1.149	1.426	1.155
PREC	16.439	12.821	18.084	12.885
CNEC	8.115	6.116	8.498	6.146
ISA	3.205	2.502	3.19	2.515
USDCOP	0.839	0.661	0.853	0.665
Cartera	32.424	24.09	34.217	24.211

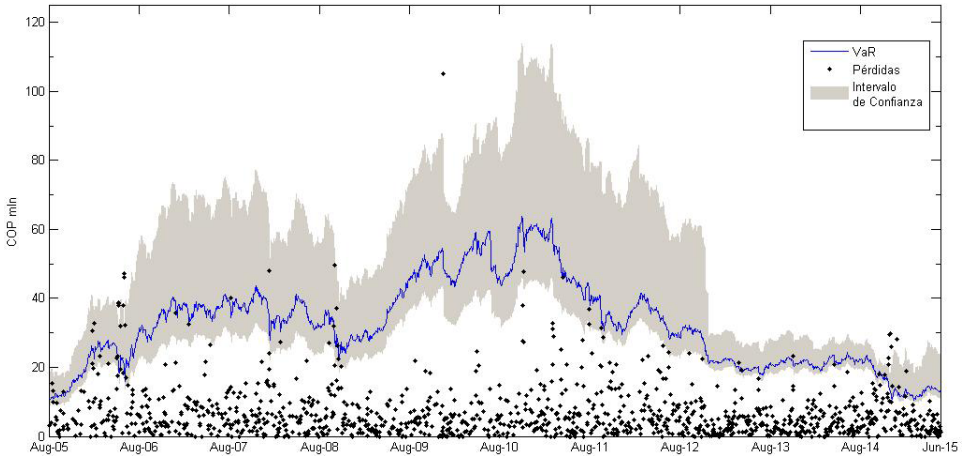
Fuente: Elaboración propia.

La Figura 1 muestra el VaR al 99% estimado por SH con los intervalos de confianza y las pérdidas del día siguiente, mientras que el Cuadro 3 despliega los datos agrupados por año y magnitud: la columna [Var, LimSup) muestra los excesos sobre el VaR pero no sobre el límite superior del intervalo, la

columna $[\text{LimSup}, +\infty)$ muestra los excesos sobre el límite superior del intervalo y la columna *Excesos* es igual a la suma de las dos columnas previas.

Es clara la consistencia de metodologías para los años 2006 y 2014 en los que ambas rechazan la adecuación del modelo, así como para los años 2005, 2007 y 2013, en los que ninguna rechaza la hipótesis nula.

Figura 1. VaR(99%) SH e Intervalo de confianza



Fuente: Elaboración propia.

Cuadro 3. Backtesting e Intervalos de Confianza para VaR(99%) SH

Año	Datos	(LimInf, VaR]	(VaR, LimSup]	(LimSup, +∞)	Excesos
2005	99	2	3	0	3
2006	242	9	3	7	10
2007	243	2	1	0	1
2008	244	6	2	1	3
2009	242	0	0	1	1
2010	245	1	0	0	0
2011	246	5	0	0	0
2012	244	2	0	0	0
2013	244	1	2	0	2
2014	244	2	2	5	7
2015	118	0	2	1	3
2014-15	242	1	4	6	10

Fuente: Elaboración propia.

Del 2010 al 2012 la prueba de Kupiec no es capaz de indicar si el modelo es demasiado conservador. Existen, sin embargo, 1, 5 y 2 observaciones sobre el límite inferior del intervalo, lo que apunta a que el modelo no está siendo conservador. En estos casos, la metodología de intervalos de confianza provee información adicional que la prueba de hipótesis no puede.

Los años 2008, 2009 y 2015 presentan discrepancias. En 2008 y 2009 se tienen pocos excesos (3 y 1) pero de tal magnitud que uno de ellos en cada año supera el límite superior del intervalo de confianza. Ambos constituyen observaciones atípicas: Lehmann Brothers en 2008 y una pérdida de 30% en Pacific Rubiales en 2009.

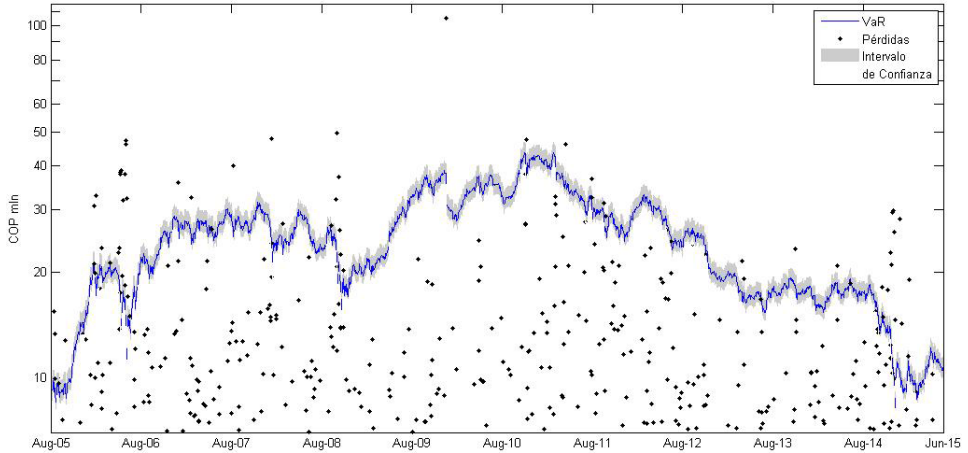
Finalmente, en 2015 se han presentado 3 excesos sobre el VaR, que es el límite superior de la región de aceptación, pero uno de ellos excede el intervalo. Esto hace suponer que el modelo es demasiado agresivo (como en 2006 y 2014) pero falta un exceso más para confirmarlo. Sin embargo, si se completa una muestra de un año desde julio 2014, desaparece la inconsistencia y ambas metodologías rechazan la bondad del modelo. En situaciones como ésta, los intervalos de confianza proveen una alerta temprana.

Los resultados correspondientes para el modelo Normal se encuentran en la Figura 2 y en el Cuadro 4. Al comparar las metodologías de *backtesting* e intervalos de confianza para este modelo se obtienen las mismas conclusiones que con SH para 2006, 2009, 2012 y 2014. Para 2005, 2008 y 2015, el modelo Normal sería rechazado ya sea mediante pruebas de hipótesis o intervalos de confianza.

En contraste, las metodologías discrepan para los años 2007, 2010, 2011 y 2013, ya que prácticamente cualquier exceso sobre el VaR sobrepasó también el intervalo de confianza. Es decir, la metodología de intervalos se habría rechazado la bondad del modelo para los cuatro años, mientras que no habría sido así mediante pruebas de hipótesis.

La razón de esta discrepancia es que el intervalo de confianza es muy angosto, pues la dispersión estimada del VaR es muy pequeña. Está ya bien documentado, desde el trabajo seminal de Fama (1965), que la distribución Normal no modela adecuadamente los retornos de los factores de riesgo de mercado y los activos usados en este ejercicio dejan ver claramente esa debilidad.

Figura 2. VaR(99%) Normal e Intervalo de Confianza



Fuente: Elaboración propia.

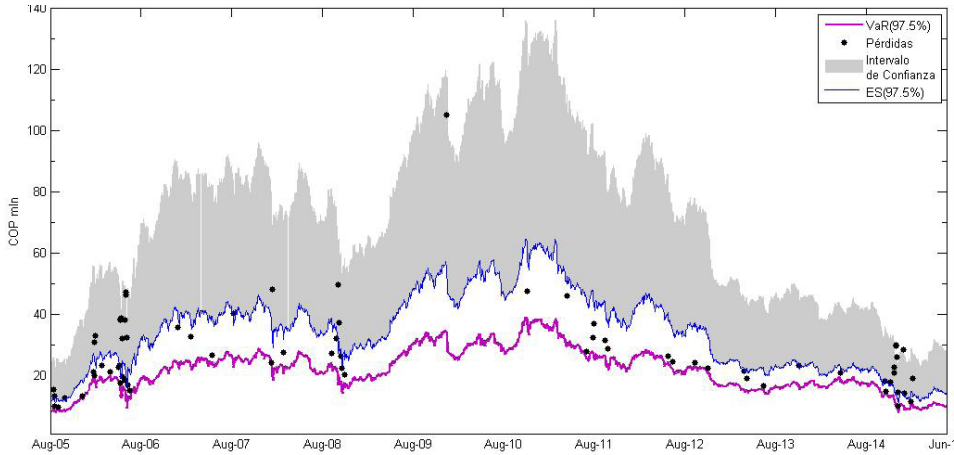
Cuadro 4. Backtesting e Intervalos de Confianza para VaR(99%) Normal

Año	Datos	(LimInf,VaR]	(VaR,LimSup]	(LimSup,+∞)	Excesos
2005	99	1	2	3	5
2006	242	1	4	16	20
2007	243	1	0	3	3
2008	244	1	1	8	9
2009	242	0	0	1	1
2010	245	0	0	1	1
2011	246	0	1	4	5
2012	244	2	0	0	0
2013	244	1	0	3	3
2014	244	2	1	10	11
2015	118	1	1	3	4

Fuente: Elaboración propia.

Respecto al ES(97.5%), el intervalo de confianza de una cola y el *backtesting* para el modelo de SH se muestran en la Figura 3 y en el Cuadro 5.

Figura 3. ES(97.5%) SH e Intervalo de Confianza



Fuente: Elaboración propia.

Cuadro 5. Backtesting e Intervalos de Confianza para ES(97.5%) SH

Año	Datos	(VaR,ES]	(ES,LimSup]	(LimSup,+∞)	ASL
2005	99	3	3	0	0.714
2006	242	11	8	2	0.017
2007	243	3	1	0	0.966
2008	244	7	3	0	0.502
2009	242	0	1	0	ND
2010	245	1	0	0	ND
2011	246	6	0	0	0.999
2012	244	4	0	0	0.936
2013	244	4	0	0	0.972
2014	244	4	4	3	0.019
2015	118	1	2	1	0.143

Fuente: Elaboración propia.

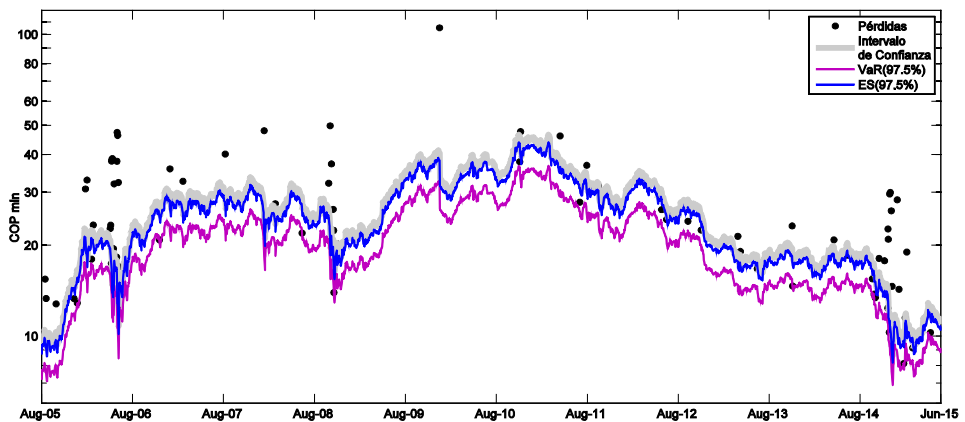
En la Figura 3 se visualizan líneas para el VaR(97.5%), el ES(97.5%) y el extremo superior del intervalo de confianza al 97.5% para este último. También se grafican como puntos los excesos sobre el VaR(97.5%).

De ellos, seis caen por arriba de la cota superior del intervalo de confianza: 2 en el año 2006, 3 en 2014 y 1 en 2015 (ver la quinta columna del Cuadro 5). El nivel de significancia asintótico, calculado mediante un proceso *bootstrap* de 10 mil repeticiones, rechaza la hipótesis nula para 2006 y 2014, no así para 2015. Adicionalmente, este método no pudo pronunciarse para los años 2009 y 2010, ya que, al existir un único exceso, no es posible estimar la dispersión de los excesos sobre el VaR. El intervalo de confianza para estos años permite concluir que el modelo es adecuado, entregando nuevamente más información que el *backtesting*.

La única discrepancia entre los métodos reside entonces en 2015. Como se indicó en la sección 2 la precisión del ASL depende del tamaño de la muestra, que en este caso es demasiado pequeña (sólo 118 datos y 4 excesos).

Para el modelo Normal, los intervalos de confianza de una cola y el *backtesting* para el ES(97.5%) se observan en la Figura 4 y en el Cuadro 6. Esta Figura tiene escala logarítmica para su mejor visualización.

Figura 4. ES(97.5%) Normal e Intervalo de Confianza



Fuente: Elaboración propia.

Cuadro 6. Backtesting e Intervalos de Confianza para ES(97.5%) Normal

Año	Datos	(VaR,ES]	(ES,LimSup]	(LimSup,+∞)	ASL
2005	99	2	2	3	0.110
2006	242	3	4	16	0.000
2007	243	1	0	3	0.091
2008	244	4	1	8	0.007
2009	242	0	0	1	ND
2010	245	1	0	1	0.252
2011	246	1	1	4	0.069
2012	244	4	0	0	0.942
2013	244	2	0	3	0.173
2014	244	7	0	10	0.003
2015	118	3	1	3	0.045

Fuente: Elaboración propia.

Los años 2006, 2008, 2012 y 2014 proveen la misma conclusión mediante pruebas de hipótesis e intervalos de confianza. Los años 2005, 2007, 2011 y 2015 generan la misma conclusión si se reduce el nivel de confianza (entre 96.5% para 2015 y hasta 88% para 2005). Igual que antes, no es posible probar la hipótesis para 2009 pero el intervalo de confianza indica un modelo no válido para ese año. Existen discrepancias entre los métodos para los años 2010 y 2013, en que el método de intervalos de confianza rechaza la validez del modelo mientras la prueba de hipótesis no encuentra evidencia suficiente para hacerlo. Igual que antes, esto se debe a la combinación de dos factores: *a*) una muestra de excesos muy pequeña y *b*) la distribución Normal no describe adecuadamente el comportamiento de los factores de riesgo.

La primera dificultad se puede subsanar realizando la prueba de hipótesis con los excesos combinados de 2, 3 o hasta 4 años, lo cual le otorgaría mayor precisión, pero con el inconveniente de tomar decisiones sobre el modelo de forma potencialmente tardía.

En el caso de SH, por ejemplo, tomando los excesos de períodos de dos años no existen discrepancias entre intervalos de confianza y pruebas de hipótesis para ningún período.

Conclusiones

El resultado más importante de este trabajo es la equivalencia entre el *backtesting* y el intervalo de confianza para cualquier métrica basada en el estimador máximo verosímil de los factores de riesgo. En particular, se demostraron resultados específicos para VaR y ES y se aplicaron empíricamente bajo las metodologías Simulación Histórica y Normal, con resultados generalmente consistentes y discrepancias que pueden depender de observaciones atípicas, de la bondad del modelo de estimación o del hecho que en diversos momentos se han usado aproximaciones.

La estrategia de intervalos de confianza tiene dos ventajas evidentes: 1) no se requiere acumular una muestra de métricas que permita probar la hipótesis nula, ya que el intervalo se construye con la misma muestra que se estima la métrica y 2) permite decidir, a diferencia de la prueba de hipótesis, si el modelo está siendo demasiado conservador, aun cuando el número de estimaciones de riesgo sea “pequeño”. Gracias a esta combinación de propiedades la metodología podría establecerse como un *sistema de alertas tempranas*.

Una tercera propiedad, sobre la cual el Comité de Basilea está poniendo énfasis, es que las metodologías tradicionales de validación de modelos no toman en cuenta la magnitud de los excesos. El BIS se basa en ello —entre otros motivos— para proponer el cambio de VaR a ES como métrica de riesgo, pero aquí se observa que la metodología de intervalos de confianza permite incorporar la magnitud del exceso de forma natural, por lo que es posible poner a revisión un modelo aún desde el primer exceso observado, dependiendo si éste superó el límite superior del intervalo de confianza.

Por último, la asimetría de los intervalos es un punto a destacar. Si bien el intervalo (2.13) es simétrico por construcción, ello se debe a que se basa en la distribución asintótica de los estimadores MV, que es simétrica. Cuando la distribución supuesta de los parámetros es tratable, como es el caso de la distribución Empírica, los intervalos no necesariamente son simétricos, como se aprecia en la Figura 1.

Referencias bibliográficas

- Acerbi, C. y Székely, B (2014). "Back-Testing Expected Shortfall". *Risk* (21), pp. 76-81.
- Acerbi, C. y Tasche, D (2002). "On the Coherence of Expected Shortfall". *Journal of Banking and Finance*, 26(7), pp. 1487-1503.
- Artzner, P., Delbaen, F., Eber, J. M., y Heath, D. (1999). "Coherent Measures of Risk". *Math Finance* 9(3), pp. 203-228.
- BIS (2013). "Fundamental Review of the Trading Book: A Revised Market Risk Framework". Disponible en www.bis.org/publ/bcbs265.pdf.
- Campbell, S. D. (2006). "A Review of Backtesting and Backtesting Procedures". *Journal of Risk*, 9(2), pp. 1-17.
- Candelon, B., Colletaz, G., Hurlin, C., y Tokpavi, S. (2011). "Backtesting Value-at-Risk: A GMM Duration-Based Test". *Journal of Financial Econometrics*, 9(2), pp. 314-343.
- Casella, G. y Berger, R. L. (2002). *Statistical Inference*. Pacific Groove, CA: Wadsworth and Brooks/Cole Advanced Books and Software.
- Christoffersen, P. y Pelletier, D. (2004). "Backtesting Value-at-Risk: A Duration-Based Approach". *Journal of Financial Econometrics*, 2(1), pp. 84-108.
- Christoffersen, P. F. (1998). "Evaluating Interval Forecasts". *International Economic Review* 39(4), pp. 841-862.
- Cotter, J. y Dowd, K. (2007). "Evaluating the Precision of Estimators of Quantile-Based Risk Measures". *Working paper 200743*, Geary Institute, University College Dublin.
- David, H. A. y Nagaraja, H. N. (2003). *Order statistics*. Hoboken, NJ: J Wiley and sons.
- Dowd, K. (2005). "Measuring Market Risk". Hoboken, NJ: John Wiley and sons.
- Dowd, K. (2006). "Using Order Statistics to Estimate Confidence Intervals for Probabilistic Risk Measures". *Journal of Derivatives*, 14(2), pp. 77-81.
- Efron, B. y Hinkley, D. V. (1978). "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information". *Biometrika*, 65, pp. 457-487.
- Efron, B. y Tibshirani, R. J. (1994). *An introduction to the bootstrap*. New York, NY: Chapman and Hall.
- Engle, R. F. y Manganelli, S. (2004). "Caviar: Conditional Autoregressive Value at Risk by Regression Quantiles". *Journal of Business and Economic Statistics*, 22(4), pp. 367-381.
- Fama, E. (1965). "The Behavior of Stock-Market Prices". *Journal of Business*, 38(1), pp. 34-105.

- Gao, F. y Song, F. (2008). "Estimation risk in GARCH VaR and ES estimates". *Econometric Theory*, 24(5), pp. 1404–1424.
- Jorion, P. (1996). "Risk2: Measuring the Risk in Value at Risk". *Financial Analysts Journal*, 52(6), pp. 47–56.
- Kerkhof, J. y Melenberg, B. (2004). "Backtesting for Risk-Based Regulatory Capital". *Journal of Banking and Finance*, 28(8), pp. 1845–1865.
- Kupiec, P. H. (1995). "Techniques for Verifying the Accuracy of Risk Measurement Models". *The Journal of Derivatives*, 3(2), pp. 73–84.
- Lehmann, E. y Casella, G. (1998). *Theory of Point Estimation*. Springer Verlag.
- McLachlan G. y Krishnan, T. (2008). *The EM Algorithm and Extensions*, 2nd edition. Hoboken, NJ: John Wiley and sons.
- McNeil, A. J. y Frey, R. (2000). "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach". *Journal of Empirical Finance*, 7(3-4), pp. 271–300.
- McNeil, A.J., Frey, R. y Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: Princeton University Press.
- Meilijson, I. (1989). "A Fast Improvement to the EM Algorithm on its Own Terms". *J. R. Stat. Soc., Ser. B*, 51(1), pp. 127–138.
- Morau, F. (2011). "How Valuable is your VaR? Large Sample Confidence Intervals for Normal VaR". *Journal of Risk Management in Financial Institutions*, 4(2), pp. 189–200.
- Pelletier, D. y Wei, W. (2015). "The Geometric-Var Backtesting Method". *Journal of Financial Econometrics*.
- Spierdijk, L. (2014). "Confidence Intervals for ARMA–GARCH Value-at-Risk: The Case of Heavy Tails and Skewness". *Computational Statistics & Data Analysis*. En prensa.