

CAPÍTULO V

Predicción de la volatilidad: una comparación entre métodos paramétricos y semiparamétricos

Isabel Casas*
J. Miguel Marín
Helena Veiga

En este capítulo se estudia la eficacia de distintos modelos (paramétricos, semiparamétricos y semiparamétricos con aprendizaje automático) en la predicción de la varianza diaria realizada, utilizando datos intradiarios de Bitcoin, NASDAQ y S&P500 que representan distintos grupos de mercados: criptomonedas, tecnología y el mercado de acciones estadounidense, respectivamente. La disponibilidad de datos a frecuencias muy altas, característica del *big data*, constituye una gran ventaja, ya que posibilita la estimación de la volatilidad de forma consistente y, por ende, su predicción con mayor precisión.

Se ha llevado a cabo una comparación de los modelos en términos de predicción de varianza realizada, mediante test de habilidad predictiva, tanto incondicionales como condicionales, así como en la predicción del valor en riesgo y se ha observado que el modelo autorregresivo heterogéneo de cuarticidad (o variación cuártica) sobresale en la predicción de la varianza realizada.

En cuanto al valor en riesgo, se han empleado trece contrastes condicionales para evaluar el rendimiento de los diversos modelos. Se obtiene que, en general, no presentan un buen desempeño durante el periodo correspondiente a la pandemia global del coronavirus. Los mejores resultados se observan en Bitcoin y NASDAQ, usando los modelos heterogéneos autorregresivos con cuarticidad realizada, *Random Forests* y el modelo heterogéneo autorregresivo con parámetros variables en el tiempo.

Palabras clave: modelos de aprendizaje automático, modelos HAR, varianza realizada, valor en riesgo.

* Los autores desean expresar su profundo agradecimiento al profesor Daniel Peña, así como a las profesoras Pilar Poncela y Eva Senra, por la generosa invitación a participar en estas jornadas. Además, queremos destacar nuestro agradecimiento a la Fundación Funcas por su apoyo financiero. También agradecemos la ayuda financiera de los proyectos de investigación PID2021-122919NB-I00 y PID2022-139614NB-C22, y a la Fundação para a Ciência e a Tecnologia, proyecto UIDB/00315/2020.

1. INTRODUCCIÓN

La predicción de la varianza realizada en los rendimientos de activos financieros importantes, como el Bitcoin (BTC), NASDAQ y S&P 500 (SP500), es esencial debido a su influencia significativa en los mercados financieros globales. Estos activos no solo son indicativos de la salud económica y financiera, sino que también su volatilidad es un área de interés crítico para un amplio espectro de participantes del mercado, incluidos inversores, analistas y gestores de carteras.

Entender y predecir esta volatilidad no solo facilita la toma de decisiones informadas en inversión y gestión de carteras, sino que también es vital en la asignación de activos y la gestión de riesgos. Esto es particularmente importante, ya que permite a los participantes del mercado anticipar y mitigar los efectos adversos de eventos inesperados y fluctuaciones bruscas en los precios.

Además, la capacidad de predecir con precisión la varianza realizada es un hecho muy valioso en la gestión de riesgos. Los modelos predictivos eficaces permiten identificar períodos de alta volatilidad, lo cual es crucial para la implementación de medidas preventivas, como el ajuste de las carteras de inversión y la aplicación de estrategias de cobertura. Estas acciones son fundamentales para mitigar los riesgos asociados a la volatilidad en los mercados financieros.

En la era de automatización, el seguimiento detallado de los precios de los activos financieros de forma continua facilita el cálculo de la varianza realizada (RV) diaria a partir de los rendimientos financieros intradiarios. Esta medida ofrece una representación más precisa de la volatilidad y, por ende, del riesgo financiero. La predicción de esta varianza y, en consecuencia, del riesgo del activo es crucial, por ejemplo, para tomar decisiones con conocimiento de causa en la gestión del riesgo, tanto para inversores como para instituciones financieras.

En este contexto, sugerimos la adopción de diversos métodos para predecir la varianza realizada. Estos incluyen enfoques paramétricos clásicos, así como métodos semiparamétricos y de aprendizaje automático, que se aplican a series financieras como Bitcoin (BTC), NASDAQ y S&P 500 (SP500), que representan diferentes grupos de mercados: criptomonedas, tecnología y el mercado de acciones estadounidense, respectivamente. Nuestro objetivo es realizar una comparativa exhaustiva entre las metodologías empleadas en el ámbito de predicción de la volatilidad y en el cálculo del valor en riesgo (VaR).

Dentro de los enfoques paramétricos, estimamos modelos heterogéneos autorregresivos (HAR) que se utilizan para predecir la volatilidad de los mercados financieros. Los modelos HAR utilizan información histórica de los precios para estimar cómo la volatilidad de un activo financiero cambia a lo largo del tiempo en diferentes intervalos, desde momentos cercanos en el tiempo hasta momentos más distantes en el futuro. Aunque los modelos HAR no se clasifican formalmente como modelos de memoria larga, se ha identificado que muestran

una alta persistencia y reflejan las características de memoria larga de ciertas series financieras. Por ello, son apropiados para modelar la varianza realizada.

Para poder predecir la volatilidad con mayor flexibilidad, añadimos coeficientes variantes en el tiempo a los modelos HAR para crear los TVHAR que estimaremos semiparamétricamente, lo cual permite capturar las no linealidades potenciales presentes en la varianza realizada.

Es importante destacar que los *shocks* negativos pueden influir más en la volatilidad que los positivos de igual magnitud. La razón es que los inversores suelen ser más sensibles a la incertidumbre que generan los *shocks* negativos, dada la mayor probabilidad de incurrir en pérdidas significativas. Por lo tanto, incorporar la asimetría de la volatilidad en los modelos de varianza realizada podría mejorar la precisión de las predicciones y el cálculo del valor en riesgo.

Los métodos de aprendizaje automático que utilizaremos son el *Deep Learning* (DL) y *Random Forests* (RF). Los primeros permiten ajustar redes neuronales de arquitectura *feedforward* y con capas ocultas, que posibilitan predecir observaciones futuras basadas en la estructura de los datos previos. Por otro lado, en los RF se emplean múltiples árboles de decisión en su entrenamiento, generando resultados basados en el consenso de estos árboles, o en su promedio para la predicción de nuevas observaciones.

Para analizar los resultados de la predicción de la varianza realizada, hemos utilizado contrastes de habilidad predictiva incondicionales y condicionales, llegando a la conclusión de que, para el BTC condicional a sus rendimientos, los mejores modelos son los modelos HAR que incluyen la cuartilidad realizada¹, el modelo HAR cuyos coeficientes cambian en el tiempo y RF.

Usamos como variable condicional los rendimientos, dada la ley financiera de compensación de “a mayor riesgo, mayor rendimiento”. Para el NASDAQ y SP500, utilizamos como variable condicional la volatilidad de mercado medida por el VIX, que es un índice de volatilidad en tiempo real creado por el Chicago Board Options Exchange (CBOE). Este índice fue el primero en cuantificar las expectativas del mercado respecto a la volatilidad. Los resultados indican que los mejores modelos para el NASDAQ son los modelos HAR simétrico y asimétrico que incluyen la cuartilidad realizada, y los RF. Lo mismo para el SP500, aunque el RF no ha sido calificado como un modelo dominante para la predicción de la varianza realizada. Estos resultados se basan en una función de pérdida robusta.

¹ La cuartilidad realizada es un concepto utilizado en econometría financiera, especialmente en la modelización de datos financieros de alta frecuencia. Forma parte de una familia más amplia de medidas conocidas como momentos realizados, que se utilizan para estimar la volatilidad y otros momentos superiores de los rendimientos de los activos durante un periodo de tiempo específico. Mientras que la varianza realizada (segundo momento) y la asimetría realizada (tercer momento) se discuten a menudo, la cuartilidad realizada (cuarto momento) es crucial por varias razones, en particular en la estimación de la curtosis.

En cuanto al cálculo del valor en riesgo, los mejores modelos, considerando 13 pruebas condicionales de *backtesting*, son para el Bitcoin, los modelos HAR simétrico y asimétrico que incluyen la cuartilidad realizada, y para el NASDAQ, el modelo HAR con coeficientes que varían en el tiempo. En relación al SP500, parece no haber ningún modelo que sea validado por las pruebas de *backtesting*. La razón puede ser que el periodo de análisis corresponde al de la pandemia mundial del COVID-19, y las empresas tradicionales incluidas en el SP500 fueron las que registraron más pérdidas y tardaron en recuperar sus cotizaciones de antes de la pandemia. En cuanto al BTC, se ha verificado que durante la pandemia ha tenido un buen desempeño en términos de rentabilidad. A su vez, el NASDAQ, que incluye empresas tecnológicas, como Netflix, ha registrado menos pérdidas y se ha recuperado más rápidamente que el SP500.

El trabajo se organiza de la siguiente manera: en la sección segunda se presentan los modelos que se utilizarán para predecir la varianza realizada y para el cálculo del valor en riesgo. En la sección tercera se presentan los datos y los resultados empíricos, y en la sección cuarta se presentan las principales conclusiones del estudio.

2. MODELACIÓN DE LA VARIANZA REALIZADA

Si bien los modelos heterogéneos autorregresivos (HAR) no se clasifican formalmente como modelos de memoria larga, se ha observado que exhiben una alta persistencia y capturan las características de memoria larga de ciertas series temporales financieras.

El modelo HAR es un tipo de modelo utilizado en análisis financiero para predecir la volatilidad en los mercados. Este modelo toma en cuenta diferentes horizontes de tiempo, como corto plazo, medio plazo y largo plazo, para capturar mejor la complejidad de los movimientos de los precios de los activos financieros.

La varianza realizada (RV, en inglés) del día t se define mediante

$$RV_t = \sum_{i=1}^M r_{t,i}^2, t = 1, \dots, N, \quad [1]$$

donde $r_{t,i}$ es el rendimiento del precio en el momento i del día t y M es el número total de valores de tiempo intradiario.

El modelo HAR introducido por Corsi (2009) se utiliza ampliamente para la previsión de la RV.

Sin embargo, este modelo no tiene en cuenta la respuesta asimétrica de la volatilidad a los *shocks* positivos y negativos, como se analiza en la literatura sobre los efectos del apalancamiento Christie (1982); Campbell y Hentschel (1992); Bollerslev *et al.* (2006)². De hecho,

² Aunque la asimetría y el apalancamiento no son lo mismo, los usamos indistintamente a continuación. El apalancamiento se considera un caso especial de asimetría; ver McAleer (2014).

los *shocks* negativos suelen tener un mayor impacto en la volatilidad que los *shocks* positivos de la misma magnitud. En otras palabras, los inversores suelen ser reacios a la incertidumbre asociada con los *shocks* negativos porque aumentan la probabilidad de pérdidas significativas. En consecuencia, incorporar la asimetría de la volatilidad en los modelos HAR puede mejorar la predicción de la RV.

El modelo HAR tampoco tiene en cuenta las no-linealidades en el comportamiento de la RV. Para cubrir esta necesidad, algunos autores han incluido un término extra en los modelos de la familia HAR que es una interacción entre la RV pasada y su varianza. También es posible modelar la no-linealidad usando modelos semiparamétricos que se adaptan a las condiciones del mercado. Otra forma es usar modelos de aprendizaje supervisado como las redes neuronales y los *Random Forests*. A continuación, presentamos en detalle los modelos mencionados.

2.1. Modelos de la familia HAR

Antes de presentar los modelos paramétricos sugeridos en este análisis, necesitamos definir las variables predictoras. Primero, las semivarianzas realizadas positiva y negativa como las presenta Barndorff-Nielsen *et al.* (2010):

$$RV_t^+ = \sum_{i=1}^M r_{t,i}^2, I(r_{t,i} \geq 0),$$

$$RV_t^- = \sum_{i=1}^M r_{t,i}^2, I(r_{t,i} \leq 0),$$

Como en la definición de la RV, M es el número de observaciones intradiarias, $r_{t,i}$ es el rendimiento del día t en el momento i . Aquí $I(\cdot)$ es una función indicadora que toma el valor 1 si el argumento es verdadero y cero en caso contrario.

Además, consideramos los siguientes retardos de la RV:

$$RV_{t-j|t-h} = \frac{1}{h+1-j} \sum_{i=j}^h RV_{t-i},$$

con $j \leq h$.

Por lo tanto, RV_{t-1} , $RV_{t-1|t-5}$ y $RV_{t-1|t-22}$ corresponden a los retardos diarios, semanales y mensuales, respectivamente.

Utilizando estas variables, los modelos de la familia HAR que proponemos están resumidos en el [cuadro 1](#). El modelo original propuesto por Corsi (2009) y su extensión propuesta por Patton y Sheppard (2015) que incluye las semivarianzas realizadas para modelar la asimetría. Además, Bollerslev *et al.* (2016) muestran que la RV, si bien es un estimador consistente de la volatilidad integrada bajo ciertas condiciones, es propenso a errores de medición en

muestras finitas. Estos errores pueden provocar sesgos en la estimación del modelo HAR. Para disminuir estos sesgos, proponen una nueva familia de modelos conocida como HARQ. Esta familia depende de la cuartilidad realizada, dada por:

$$RQ_t \equiv \frac{M}{3} \sum_{i=1}^M r_{t,i}^4.$$

Cuadro 1.

Modelos HAR paramétricos

Modelo	Especificación	Autores
HAR	$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t- t -5} + \beta_3 RV_{t- t -22} + u_t$	Corsi (2009)
SHAR	$RV_t = \beta_0 + \beta_1^+ RV_{t-1}^+ + \beta_1^- RV_{t-1}^- + \beta_2 RV_{t- t -5} + \beta_3 RV_{t- t -22} + u_t$	Patton y Sheppard (2015)
HARQ	$RV_t = \beta_0 + (\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2}) RV_{t-1} + \beta_2 RV_{t- t -5} + \beta_3 RV_{t- t -22} + u_t$	Bollerslev <i>et al.</i> (2016)
SHARQ	$RV_t = \beta_0 + (\beta_1^+ \beta_{1Q}^+ RQ_{t-1}^{1/2}) RV_{t-1}^+ + (\beta_1^- + \beta_{1Q}^- RQ_{t-1}^{1/2}) RV_{t-1}^- + \beta_2 RV_{t- t -5} + \beta_3 RV_{t- t -22} + u_t$	Bollerslev <i>et al.</i> (2016)

Fuente: Candel *et al.* (2020).

En nuestro estudio no consideramos la presencia de saltos en los datos. Según Patton y Sheppard (2015), los modelos que incluyen saltos, como el modelo HAR-J, tienden a comportarse de manera deficiente en comparación con el modelo SHAR. El modelo HAR-J incorpora un componente de salto al modelo básico HAR para considerar movimientos bruscos y significativos en los precios de los activos que no son capturados por las medidas estándar de volatilidad realizada. Por lo tanto, este capítulo se enfoca en modelos que excluyen la consideración de saltos.

2.2. Modelo HAR con coeficientes variables

Es necesario extender el enfoque de los modelos HAR, mediante la incorporación de coeficientes variables en el tiempo, para mejorar la predicción de la varianza realizada en los mercados financieros porque nos permite predecir cómo evoluciona el riesgo con el tiempo.

Al añadir coeficientes que cambian dinámicamente, podemos capturar la naturaleza no lineal de la volatilidad en los mercados, adaptándose mejor a los cambios en las condiciones económicas y financieras.

El modelo TVHAR ha sido propuesto por Chen *et al.* (2018) donde se explica el algoritmo para predecir futuros valores de la RV. Su expresión es:

$$RV_t = \gamma_0(\tau) + \gamma_1(\tau) RV_{t-1} + \gamma_2(\tau) RV_{t-|t|-5} + \gamma_3(\tau) RV_{t-|t|-22} + ut \quad [2]$$

para $t = 1, 2, \dots, N$ y $\tau = t/N$.

Los valores estimados de los coeficientes $\gamma(\tau) = (\gamma_0(\tau), \gamma_1(\tau), \gamma_2(\tau), \gamma_3(\tau))^t$ se obtienen mediante técnicas no paramétricas detalladas en las referencias Robinson (1989), Fan y Gijbels (1996) y Cai (2007). En particular, empleamos el enfoque del estimador local lineal.

En esquema, para cada valor de τ , calculamos los coeficientes de [2] mediante un proceso de regresión local que otorga un mayor peso a los datos cercanos a un punto temporal t . Este peso se basa en un *kernel* (núcleo).

En la siguiente ecuación [3] se muestra la expresión del estimador local lineal para γ y su primera derivada, donde el vector de predictores es $X_i = (1, RV_{t-1}, RV_{t-1|t-5}, RV_{t-1|t-22})^t$.

$$\begin{pmatrix} \hat{\gamma}_i(\tau) \\ \hat{\gamma}_i^{(1)}(\tau) \end{pmatrix} = \begin{pmatrix} S_{N,0}(\tau) & S_{N,1}^\top(\tau) \\ S_{N,1}(\tau) & S_{N,2}(\tau) \end{pmatrix}^{-1} \begin{pmatrix} T_{N,0}(\tau) \\ T_{N,1}(\tau) \end{pmatrix}, \quad [3]$$

donde las funciones $S_{N,s}(\tau)$ y $T_{N,s}(\tau)$ para $s = 0, 1, 2$ se definen como:

$$S_{N,s}(\tau) = \frac{1}{N} \sum_{i=1}^N X_i^\top X_i (\tau_i - \tau)^s K\left(\frac{\tau_i - \tau}{b}\right),$$

$$T_{N,s}(\tau) = \frac{1}{N} \sum_{i=1}^N X_i^\top (\tau_i - \tau)^s K\left(\frac{\tau_i - \tau}{b}\right) RV_i.$$

En estos estimadores, el *bandwidth* o ancho de ventana del *kernel* denotado como b , es un parámetro crítico. La elección de una ventana demasiado amplia no lograría capturar con precisión las variaciones del coeficiente, mientras que una ventana muy estrecha produciría un estimador excesivamente inestable.

Dado que los datos que analizamos tienen dependencia temporal, utilizamos el método de *leave-k-out cross-validation* con un tamaño de bloque igual a $k = \lfloor N/3 \rfloor$ para seleccionar b . De esta manera, garantizamos la independencia de las submuestras y su selección robusta.

2.2.1. Deep Learning

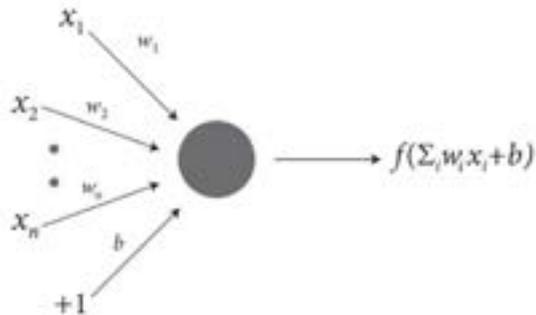
Existen varios marcos teóricos para Deep Learning (Hastie *et al.*, 2009), y aquí resumimos la arquitectura *feedforward* utilizada por librería H2O de R (ver *e.g.* Landry, 2016; LeDell *et al.*, 2018; Candel *et al.*, 2020).

La unidad básica en el modelo es la *neurona* (figura 1), un modelo inspirado en la neurona humana biológica.

En los humanos, las señales de salida de las neuronas de diferente intensidad viajan a lo largo de las uniones sinápticas y luego se agregan como entrada para la activación de una neurona conectada.

En el modelo, se agrega la combinación ponderada de señales de entrada $\alpha = \sum_{i=1}^n w_i x_i + b$ y luego se transmite una señal de salida $f(\alpha)$ por la neurona conectada. La función f repre-

Figura 1.

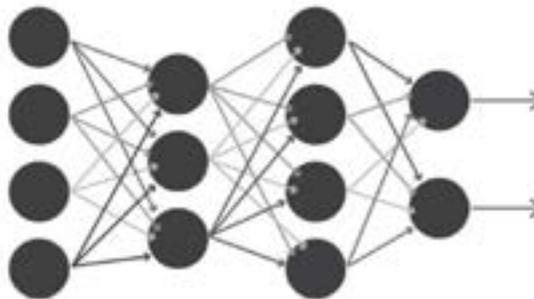
Esquema neurona básica

Fuente: Candel *et al.* (2020).

senta la función de activación no lineal utilizada en toda la red, y el sesgo b representa el umbral de activación de la neurona.

Las redes neuronales de múltiples capas predictivas (figura 2) constan de muchas capas de unidades neuronales interconectadas: comenzando con una capa de entrada para que coincida con el espacio de entrada; seguido de múltiples capas no lineales; y terminan con una regresión lineal o capa de clasificación para que coincida con el espacio de salida. Las entradas y salidas de las unidades del modelo siguen la lógica básica de la neurona única descrita anteriormente. Las constantes se incluyen en cada capa de la red que no es de salida. Los pesos que vinculan las neuronas con otras neuronas determinan completamente la salida de toda la red, y el aprendizaje ocurre cuando estos pesos se adaptan para minimizar el error en los datos de entrenamiento.

Figura 2.

Esquema de red neuronal multicapa

Fuente: Candel *et al.* (2020).

Este marco básico de redes neuronales multicapa se puede utilizar para realizar tareas de *Deep Learning*. Las arquitecturas de *Deep Learning* son modelos de extracción de características jerárquicas, que normalmente implican múltiples capas con niveles de no linealidad. Dichos modelos pueden modelizar con alto rendimiento datos complejos.

Específicamente, usamos el procedimiento `h2o.deeplearning` de la librería H2O de R, que utiliza un protocolo de entrenamiento puramente supervisado. El esquema de inicialización predeterminado es la opción de adaptación uniforme, que es una inicialización optimizada basada en el tamaño de la red.

Usamos las opciones por defecto del comando, así, la función de activación *Rectifier*: $f(\alpha) = \max(0, \alpha)$, y como función de pérdida el error cuadrático medio (ECM). A su vez, usamos dos capas ocultas con 200 neuronas en cada una.

2.2.2. *Random Forests*

Los *Random Forests* (RF) presentados por primera vez en Breiman (2001), son una modificación sustancial de las técnicas de *bagging* que generan una numerosa colección de árboles no correlacionados y luego se promedian. La idea esencial de *bagging* es promediar muchos modelos ruidosos, pero aproximadamente insesgados y, por lo tanto, reducir la varianza. Los árboles son candidatos ideales para el *bagging*, ya que pueden capturar estructuras de interacción complejas en los datos y, si crecen lo suficientemente en profundidad, tienen un sesgo relativamente bajo. Dado que los árboles son habitualmente ruidosos, se benefician enormemente de hacer promedios.

En muchos problemas, el rendimiento de los *Random Forests* es muy similar al *boosting* y son más sencillos de entrenar y ajustar. Como consecuencia, los *Random Forests* son bastante populares y se implementan en una gran variedad de paquetes estadísticos.

El algoritmo general de *Random Forests* para usar con variables continuas es:

1. Desde $b = 1$ a B :
 - (a) Obtener una muestra *bootstrap* \mathbf{Z}^* de tamaño N a partir de los datos de entrenamiento.
 - (b) Ajustar un árbol del *Random Forest* T_b con los datos remuestreados, repitiendo de modo recursivo los siguientes pasos para cada nodo terminal del árbol, hasta alcanzar el tamaño mínimo de nodo n_{\min} .
 - i. Seleccionar m variables al azar de las p variables.
 - ii. Elegir la mejor variable o punto de división entre las m .
 - iii. Dividir el nodo en dos nodos *hijos*.

2. Generar el conjunto de árboles $\{T_b\}_1^B$.

Para hacer una predicción en un nuevo punto x :

$$\hat{f}_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x).$$

En este estudio usamos la librería H2O de R. Específicamente, usamos el procedimiento

`h2o.randomForest`.

Después de realizar un estudio previo para optimizar los valores de los parámetros del modelo, fijamos algunos valores del procedimiento para optimizar tanto el rendimiento, como la velocidad de computación.

Usamos una profundidad máxima de los árboles (`max_depth`) igual a 10, dado que valores más altos pueden llevar a sobreajustes.

Para simplificar la interpretabilidad, fijamos en 5 el número mínimo de observaciones para especificar una hoja del árbol (`min_rows`).

Finalmente, el número de árboles por defecto (`ntrees`) se fija en 30.

3. ANÁLISIS EMPÍRICO

Vamos a predecir la varianza realizada de los rendimientos del BTC, NASDAQ y SP500 que son muy relevantes en los mercados financieros globales.

3.1. Datos

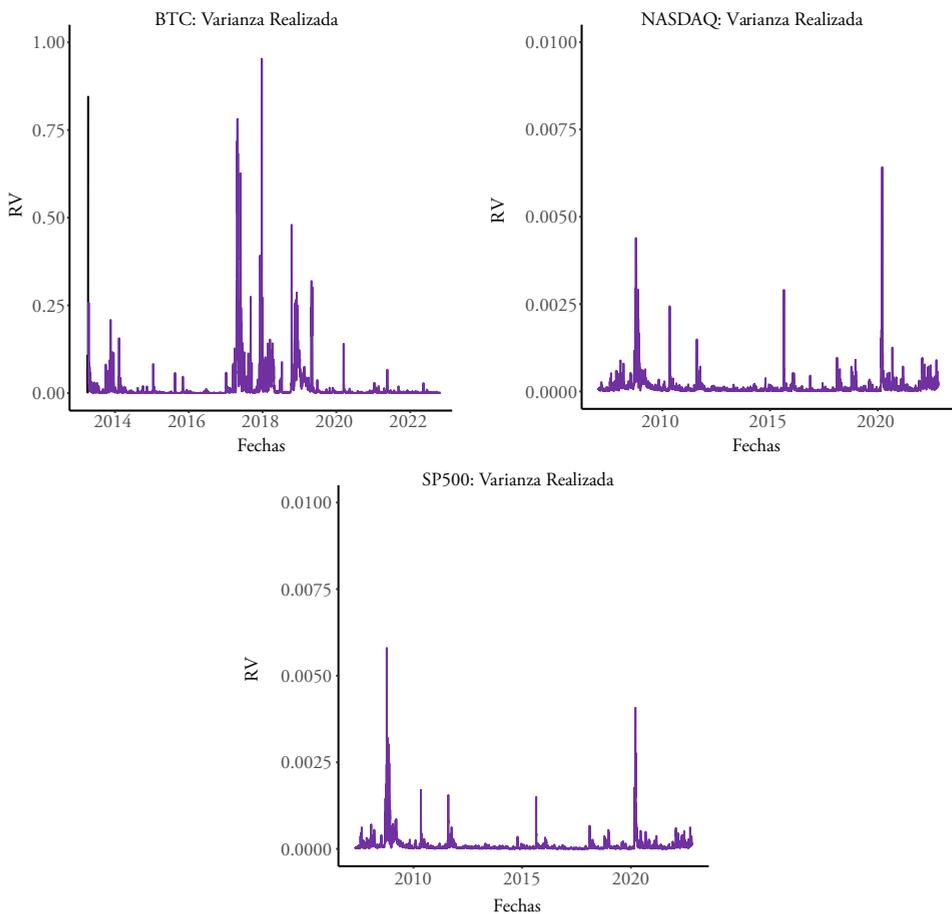
En predicción, dividimos los datos en un conjunto de entrenamiento en el que ajustamos el modelo y un conjunto de prueba donde evaluamos su poder de predicción. En nuestro análisis, empleamos los últimos 1.000 valores de cada serie, como conjunto de prueba para determinar la efectividad de cada modelo y para comparar su rendimiento prediciendo la varianza realizada. El **cuadro 2** muestra las fechas exactas de las series financieras en nuestro análisis.

Una vez ajustado el modelo con el conjunto de entrenamiento, cuyos predictores tienen índices $1, \dots, T$, se predice el valor de la RV para $T + 1$ que denotamos con $RV_{T+1|T}^*$.

Cuadro 2.**Datos**

	<i>Muestra completa</i>	<i>Muestra de predicción (1.000 valores)</i>	<i>Frecuencia</i>
BTC	23/04/2013-28/10/2022	Desde 02/02/2020	Diaria Continuo
NASDAQ	16/02/2007-21/10/2022	Desde 01/11/2018	Diaria Lunes a viernes
SP500	31/05/2007-21/10/2022	Desde 01/11/2018	Diaria Lunes a viernes

Fuentes:

Figura 3.**Varianzas realizadas**

Fuente: Candel *et al.* (2020).

Movemos la ventana de entrenamiento un día, es decir los predictores tienen índices $2, \dots, T+1$, y calculamos $RV_{T+2|T+1}^*$. Este proceso se repite 1.000 veces, siempre entrenando el modelo con conjuntos de tamaño T y moviendo la ventana un dato cada vez, para obtener la predicción diaria de la RV para las tres series en nuestro análisis.

Calculamos el error de predicción en cada paso como:

$$E_h = RV_{T+h} - RV_{T+h|T+h-1}^*$$

para $h = 1, \dots, 1000$ y RV_{T+h} calculada con la ecuación [1].

La **figura 3** muestra las varianzas realizadas, y se puede apreciar que, durante el periodo evaluado para las predicciones de volatilidad, la volatilidad del BTC es inferior en comparación con la volatilidad observada en el periodo de entrenamiento.

3.2. Evaluación del funcionamiento de los modelos de predicción

La evaluación del desempeño de los modelos de predicción es esencial en cualquier proceso de modelización de datos. No solo se trata de comprender cómo es de efectivo un modelo con datos conocidos, sino también de su capacidad para predecir datos desconocidos. La medición del desempeño en la predicción se convierte en el indicador principal para elegir un modelo que explique los patrones ocultos en los datos que nos sea útil en aplicaciones prácticas.

La evaluación comparativa de varios modelos nos desvela cuál de ellos es el mejor en nuestra tarea de predicción. Cuando decimos mejor, no nos referimos al modelo con el menor error de predicción promedio en nuestra muestra. Buscamos una respuesta estadística sólida que nos permita extrapolar los resultados a cualquier conjunto de datos con características o distribución similares.

Para ello, introducimos el procedimiento del “Model Confidence Set” y los contrastes de habilidad predictiva superior condicional y utilizamos 3 funciones de pérdida:

- El error cuadrático (EC),
- El error absoluto (EA), y
- La “Quasi-Likelihood Exponential” (QLIKE).

Patton (2011) propuso la QLIKE como medida de error de predicción de la volatilidad realizada en el contexto de modelos de volatilidad financiera. Esta medida se deriva de la función de verosimilitud, que compara la probabilidad de que los datos observados provengan de

un modelo de volatilidad específico en comparación con la volatilidad realizada. Un modelo cuyas predicciones de volatilidad son muy cercanas a la volatilidad realizada observada, el valor de QLIKE será bajo.

La **figura 4** muestra las QLIKE medias de ventana deslizante para varios modelos de predicción de la RV y para las tres series de rendimientos. Esta función de pérdida de ventana móvil se calcula para una ventana fija de 90 días. Se observa que el modelo TVHAR presenta valores elevados de QLIKE durante 2020, coincidiendo con el confinamiento y el periodo más restrictivo de la pandemia, seguido de un aumento en los errores de predicción hacia finales de 2021 que se prolonga durante 2023. Los modelos HAR, HARQ, SHARQ y RF muestran los valores más bajos de QLIKE.

En relación a la RV del NASDAQ, el patrón es similar, siendo los modelos mencionados previamente los que presentan valores más bajos de QLIKE. En cambio, para el SP500, la evidencia difiere: casi todos los modelos presentan un mal desempeño durante el confinamiento. En ese periodo, el modelo con mayores QLIKE es el SHARQ, seguido por DL, HARQ y SHAR. Sin embargo, tras ese periodo, las QLIKE de la mayoría de los modelos son similares, excepto las del DL, que son relativamente mayores que las demás.

3.2.1. Model confidence set

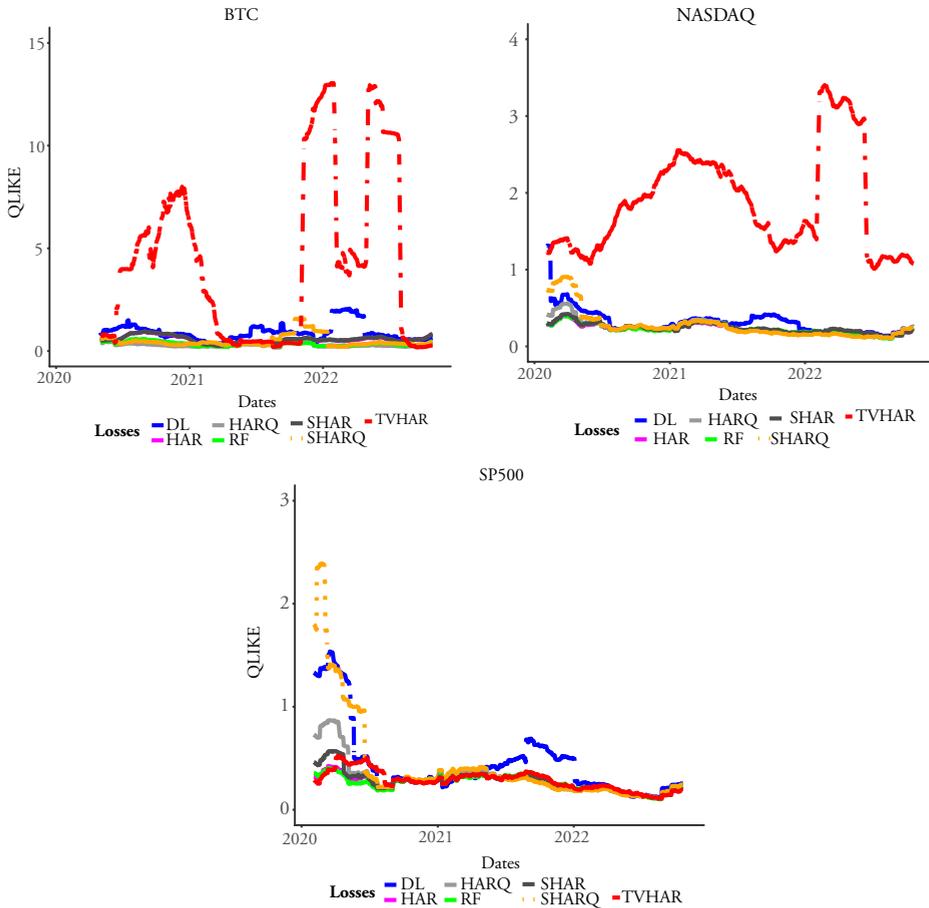
El procedimiento del *Model Confidence Set (MCS)* propuesto en Hansen *et al.* (2011) es una herramienta estadística que se utiliza para comparar la confianza en las predicciones de varios modelos. El MCS es más preciso que las métricas comunes como el *error cuadrático medio (ECM)* y el *error absoluto medio (EAM)*, ya que no solo se fija en cuánto se equivoca cada modelo en promedio, sino que también considera cuán variables son los *errores cuadráticos (EC)* y los *errores absolutos (EA)* de las predicciones. Además de con los *EC* y *EA*, este procedimiento se puede usar con otras funciones de pérdida como el QLIKE.

De manera más detallada, el MCS es un procedimiento que consiste en una serie de pruebas de hipótesis basadas en una función de pérdida. En la primera prueba estadística, el MCS selecciona el modelo con el mayor error promedio y compara el rango que abarca sus errores de predicción con los de los otros modelos. Si este rango no se superpone con ninguno de los otros modelos, se descarta y se pasa a hacer lo mismo con el siguiente modelo. Si hay superposición, se agrupan los modelos que tienen errores estadísticamente similares y se les asigna un *p*-valor. Al final de este proceso, el procedimiento MCS ha asignado un *p*-valor a cada modelo.

Los modelos cuyos *p*-valores superen cierto valor de significación (α) se consideran parte del llamado “Conjunto Superior de Modelos” (*Superior Set of Models* o SSM) y son los modelos que se eligen para realizar las predicciones.

Figura 4.

Predicción de la volatilidad: ventana deslizante de QLIKE



Fuente: Candel *et al.* (2020).

El **cuadro 3** muestra el *ECM*, *EAM* y *QLIKE* medio de los errores de predicción de los modelos propuestos usados para predecir los últimos valores de la varianza realizada de nuestras muestras del BTC, NASDAQ y SP500. También incluye los *p*-valores obtenidos con el procedimiento MCS para cada modelo.

La interpretación de los resultados del **cuadro 3** es:

- *ECM*: En la serie NASDAQ, todos los modelos tienen *p*-valores relativamente altos (por encima de 0.05). Esto significa que todos pertenecen al 95 % SSM y no hay diferencia a nivel de significación del 5 % para considerar que un modelo es mejor que los demás en términos de errores de predicción cuadráticos. Sin embargo, si se reduce el

umbral al 90 % SSM, solo el modelo SHARQ y DL pertenecen al conjunto superior de modelos del 90 %. Esto sugiere que para un nivel de significación del 10 %, SHARQ y DL son estadísticamente superiores al resto para predecir la RV. En las series SP500 y BTC, ningún modelo tiene un p -valor menor que 0.05, lo que significa que todos los modelos pertenecen al 95 % SSM.

- **EAM:** en la serie BTC, los modelos HARQ y TVHAR pertenecen al 95 % SSM, mientras que el resto de los modelos tienen errores de predicción absolutos significativamente más grandes, por lo que el HARQ y TVHAR son estadísticamente superiores en términos de EAM para predecir la RV de BTC. En las series NASDAQ y SP500, todos los modelos pertenecen al 95 % SSM para predecir la RV ya que sus p -valores son mucho más grandes que 0.05. El modelo con el menor EAM es el HAR.
- **QLIKE medio:** muchos de los modelos son estadísticamente igualmente buenos para predecir la RV de las series NASDAQ y SP500. Sin embargo, el HARQ es el único modelo que pertenece al 95 % SSM y el preferido para predecir la RV del BTC.

Cuadro 3.

Errores de predicción medios y p -valores del procedimiento MCS de los diferentes modelos

	HAR	SHAR	HARQ	SHARQ	TVHAR	DL	RF
<i>ECM</i> × 10 ⁶							
BTC	35.505	46.399	39.063	36.616	38.695	47.838	39.181
p -valor	1.000	0.569	0.705	0.780	0.832	0.549	0.666
NASDAQ	0.053	0.060	0.076	0.131	0.067	0.196	0.051
p -valor	0.508	0.316	0.562	0.585	0.518	0.289	1.000
SP500	0.027	0.031	0.026	0.032	0.038	0.039	0.027
p -valor	0.615	0.525	1.000	0.522	0.336	0.319	0.245
<i>EAM</i> × 10 ³							
BTC	2.768	2.887	1.691	1.929	1.686	3.202	1.836
p -valor	< 0.001	< 0.001	0.954	0.042	1.000	< 0.001	0.003
NASDAQ	0.079	0.082	0.084	0.087	0.083	0.102	0.080
p -valor	1.000	0.469	0.580	0.641	0.573	0.104	0.798
SP500	0.054	0.055	0.054	0.057	0.059	0.064	0.054
p -valor	1.000	0.782	0.965	0.284	0.439	0.030	0.951
QLIKE medio							
BTC	0.571	0.587	0.291	0.481	4.114	0.889	0.366
p -valor	< 0.001	< 0.001	1.000	0.039	0.006	0.002	< 0.001
NASDAQ	0.223	0.232	0.243	0.282	1.769	0.595	0.226
p -valor	1.000	0.054	0.196	0.153	< 0.001	0.060	0.529
SP500	0.256	0.269	0.308	0.958	0.265	0.547	0.256
p -valor	1.000	0.778	0.178	0.067	0.677	0.028	0.947

Fuente: Candel *et al.* (2020).

3.2.2. *Habilidad predictiva superior condicional*

En esta sección vamos a analizar la capacidad predictiva superior condicional de los modelos para predecir la volatilidad con el contraste propuesto por Li *et al.* (2022).

Sea $\{RV_t\}_{t \leq T}$ la serie temporal de volatilidad que se va a predecir; $\{RV_{0,t}\}_{T < t \leq N}$ es una serie de pronóstico de referencia y $\{\widehat{RV}_{i,t}\}_{T < t \leq N}$, $1 \leq i \leq S$, series competidoras.

El desempeño del modelo de referencia en comparación con la i -ésima alternativa competidora se evalúa mediante la diferencia de las funciones de pérdida, tal que:

$$e_{i,t} \equiv L(\widehat{RV}_{i,t}, RV_{i,t}) - L(RV_b, \widehat{RV}_{0,t}).$$

La hipótesis nula del test HPSC establece que:

$$H_0 : E(e_{i,t} | X_t = x) \geq 0, x \in \chi, \forall 1 \leq i \leq S.$$

Aquí, X_t representa una variable de estado condicional elegida por el evaluador, y χ denota la región condicional como un subconjunto del dominio de X . En este caso particular, las variables de estado condicional son el VIX que es una proxy de la volatilidad del mercado para NASDAQ y SP500, y los rendimientos de BTC para la variable BTC. En este último caso buscamos validar la teoría financiera de compensación del riesgo. El objetivo es descubrir para qué valores (bajos o altos) de la variable condicional el modelo de referencia falla o tiene éxito en predecir volatilidad futura en comparación con el modelo alternativo. La hipótesis nula establece que el modelo de referencia debe dominar débilmente a todos los modelos competidores en la región condicional χ , porque su valor de función de pérdida es estadísticamente el más pequeño. Si la hipótesis nula no puede ser rechazada, esto indica que el modelo de referencia tiene propiedades predictivas deseables.

Se pueden realizar dos tipos de test condicionales. El primero se llama “Uno-contra-Uno”, en el que cada referencia se compara con una alternativa para todos los pares de modelos. El segundo tipo se llama “Uno-contra-Todos”, que compara cada modelo de referencia con los seis modelos competidores juntos³.

El **cuadro 4** muestra los resultados de los contrastes para tres funciones de pérdida: errores cuadráticos, errores absolutos y QLIKE. Los contrastes se realizan para un nivel de confianza elegido del 95 %.

Basándonos en los contrastes (Uno-contra-Todos) para la RV de la BTC, se observa que los modelos que dominan débilmente a sus competidores en la predicción de la RV son el HAR, HARQ, SHARQ, TVHAR y RF en términos del EC. Para el EA solo el HARQ parece dominar débilmente a sus competidores. Sin embargo, para la QLIKE, parece que ningún modelo domina a sus competidores, aunque en los contrastes Uno-contra-Uno los modelos menos rechazados son el HARQ, TVHAR, DL y RF.

³ Usamos el código proporcionado por los autores disponible en <https://zenodo.org/record/4884813>

Cuadro 4.

Habilidad predictiva superior condicional para predicción de volatilidad usando tres funciones de pérdida

	EC			EA			QLIKE											
	HAR	SHARQ	TVHAR	DL	RF	HAR	SHAR	SHARQ	TVHAR	DL	RF	HAR	SHAR	SHARQ	TVHAR	DL	RF	
BTC																		
<i>Panel A: Contrastes HPSC Uno-contrá-Uno</i>																		
HAR	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0
SHAR	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
HARQ	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	0	0	0
SHARQ	0	0	0	1	0	1	0	0	1	1	0	1	1	0	0	0	0	0
TVHAR	1	1	0	1	0	1	1	0	1	1	1	1	1	0	0	0	0	0
DeepL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RF	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Panel B: Contrastes HPSC Uno-contrá-Todos</i>																		
0	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
NASDAQ																		
<i>Panel A: Contrastes HPSC Uno-contrá-Uno</i>																		
HAR	1	1	0	1	0	1	0	0	0	1	1	0	0	0	0	0	1	0
SHAR	1	1	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	0
HARQ	1	1	0	1	0	1	0	1	1	1	1	1	0	0	0	1	0	0
SHARQ	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0
TVHAR	1	1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
DeepL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RF	1	1	0	1	0	0	0	0	0	1	1	0	0	0	0	1	1	0
<i>Panel B: Contrastes HPSC Uno-contrá-Todos</i>																		
1	1	1	1	1	1	0	1	0	0	1	1	0	0	1	0	0	1	0
SP500																		
<i>Panel A: Contrastes HPSC Uno-contrá-Uno</i>																		
HAR	1	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0
SHAR	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
HARQ	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
SHARQ	1	1	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0
TVHAR	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0
DeepL	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0
RF	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0
<i>Panel B: Contrastes HPSC Uno-contrá-Todos</i>																		
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Fuente: Candel et al. (2020).

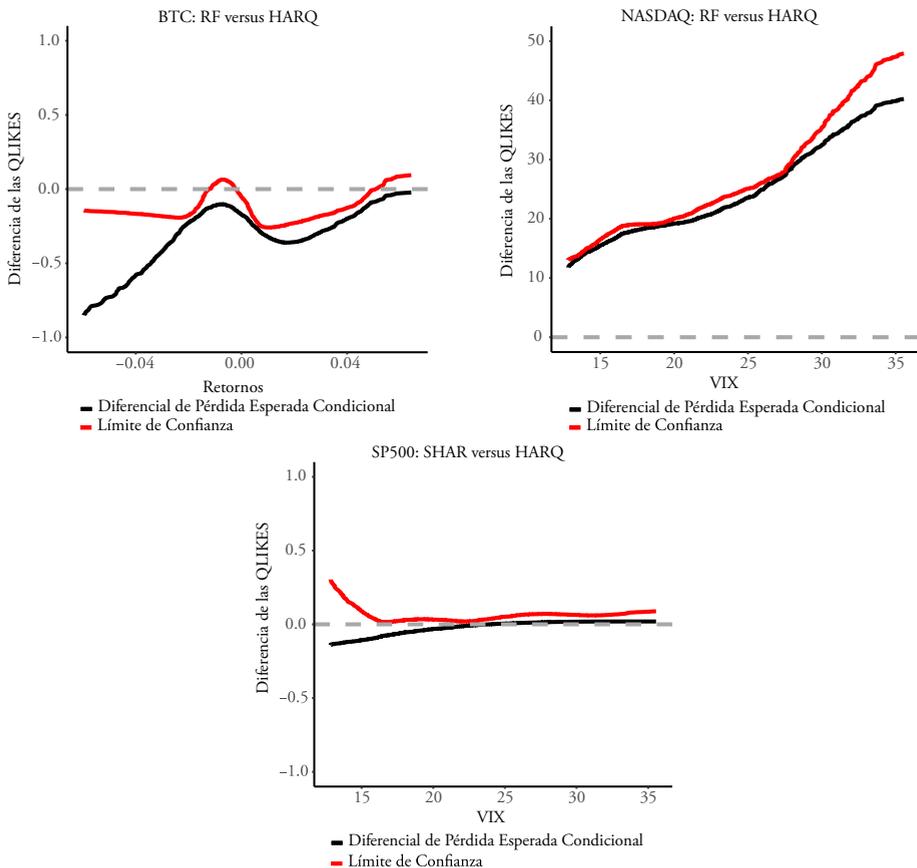
Para el NASDAQ, observamos que cuando usamos el EA o la QLIKE, los modelos HAR, HARQ, SHARQ y RF dominan débilmente a sus competidores, mientras que para la función de pérdida EC ningún modelo domina débilmente a los demás.

Finalmente, para el SP500, y también para el EC, parece no haber ningún modelo que domine débilmente en la predicción de la RV dada la volatilidad del mercado. Aunque, usando el EA, hay dos modelos que dominan, que son el HARQ y el TVHAR, y según la QLIKE, los modelos que dominan débilmente a las alternativas son el HAR, el HARQ y el SHARQ.

En resumen, los resultados de los contrastes de habilidad predictiva superior condicional parecen apuntar al HARQ como el modelo que más frecuentemente domina a sus competidores en la predicción de la RV.

Figura 5.

Predicción de la volatilidad: contrastes HPSC uno-versus-uno. Función de pérdida QLIKE



Fuente: Candel *et al.* (2020).

La **figura 5** muestra la diferencia estimada entre las funciones de pérdida esperadas condicionales, junto con los límites de confianza superior del 95 % (ver Li *et al.*, 2022) para obtener más detalles sobre los límites de confianza superior).

La función de pérdida utilizada para representar la **figura 5** es la QLIKE. Por ejemplo, la “Diferencia de las QLIKE” en el panel izquierdo se define como la diferencia entre las QLIKE del modelo RF versus HARQ. La referencia aquí es el modelo HARQ.

La prueba HPSC rechaza la hipótesis nula (es decir, HARQ es débilmente superior) si el límite de confianza está por debajo de cero en alguna región del espacio de estados condicionantes.

Comparando el RF con el QLIKE en la predicción de la RV de la BTC, no habiendo ningún modelo dominante, observamos que el RF parece tener un desempeño superior al desempeño del HARQ condicionando a los rendimientos de la BTC. También se observa que el desempeño del RF es mejor para rendimientos más negativos y empeora para valores más positivos de los rendimientos, aunque este no es estadísticamente superior para rendimientos alrededor de 0 y superiores a 0.04.

En cuanto al NASDAQ, ambos modelos son débilmente dominantes, pero el desempeño del HARQ es estadísticamente superior al del RF.

Finalmente, para el SP500 observamos que, aunque los modelos SHAR y HARQ son ambos débilmente dominantes, para volatilidades de mercado bajas, el modelo SHAR parece tener un mejor desempeño que el modelo HARQ, aunque nunca rechazamos que el modelo HARQ sea débilmente dominante.

3.2.3. Valor en riesgo

Una posibilidad es usar las predicciones del RV de nuestras series para calcular su Valor en Riesgo (VaR). Debemos recordar que en las recomendaciones del Acuerdo de Basilea III, que establece normas internacionales para regular la cantidad de capital que los bancos deben tener en reserva para protegerse contra riesgos financieros y operativos, es calcular el VaR para el nivel de confianza del 99 % ($\alpha = 0.01$).

Sea los rendimientos en el tiempo t dados por

$$y_t = \mu + \sqrt{RV_t} \varepsilon_t, \quad t = 1, \dots, T$$

donde RV_t viene dado por las especificaciones en el **cuadro 1** y ε_t es un proceso i.i.d.

La expresión matemática del VaR se define como:

$$VaR_{T+h} = -\mu_{T+h-1} - z\alpha \sqrt{RV_{T+h|T+h-1}^*}$$

donde μ_{T+h-1} es el promedio de los rendimientos de la muestra de entrenamiento, z_α es el valor crítico de una distribución normal estándar, que corresponde al nivel de confianza α , y el resto es la raíz cuadrada de la predicción de la varianza realizada.

Andersen *et al.* (2000, 2001b, 2003), Andersen *et al.* (2001a) y Andersen *et al.* (2007) encuentran que las distribuciones de los rendimientos diarios de las acciones tipificados por la correspondiente RV diaria son aproximadamente Gaussianos, especialmente considerando saltos y asimetría de la volatilidad.

Decimos que tenemos una violación en el momento $T+h$, si el VaR obtenido es menor que el rendimiento cambiado de signo:

$$\text{violación}_{T+h} = I(\text{VaR}_{T+h} < -r_{T+h}).$$

Un modelo que predice adecuadamente, tendrá un promedio de violaciones cercano a α .

El **cuadro 5** muestra el promedio de violaciones en las 1.000 predicciones de RV para cada modelo y para $\alpha = 0,01; 0,05$.

En líneas generales vemos que la tasa de fallo no está próxima a la tasa de fallo teórica. Esto puede ocurrir porque nuestro periodo de predicción del VaR corresponde a la época de la pandemia mundial de coronavirus (COVID-19), la cual provocó que todos los principales índices del mercado de valores experimentaran una caída abrupta en marzo de 2020. Sin embargo, tanto la magnitud de esa caída como la forma de la recuperación subsiguiente variaron enormemente.

A modo de ejemplo, el 15 de marzo de 2020, los principales mercados europeos y las acciones tradicionales en Estados Unidos habían perdido aproximadamente el 40 % de su valor en comparación con el 5 de enero de 2020. En cambio, los mercados asiáticos y el índice compuesto NASDAQ solo perdieron alrededor del 20 al 25 % de su valor. Una situación similar se observa en la recuperación posterior al coronavirus. A partir del 14 de noviembre de 2021, el valor del índice compuesto NASDAQ era aproximadamente un 65 % más alto que en enero de 2020, mientras que la mayoría de los otros mercados solo habían recuperado entre un 20 y un 40 %. Esto sugiere que ha habido menos turbulencias en el NASDAQ en comparación con el SP500, que incluye empresas más tradicionales.

En cuanto al BTC, hay estudios que indican que la pandemia de COVID-19 no influyó en la volatilidad de las criptomonedas, Sifat (2021), mientras que otros concluyeron que los precios de BTC aumentaron en el período de la pandemia. Así, Goodell y Goutte (2021) y Corbet *et al.* (2020) sugirieron que las grandes criptomonedas actuaron como una reserva de valor durante este tiempo. Lo que sugiere que este periodo fue favorable a las principales criptomonedas.

Además, en el marco de nuestro análisis, realizamos 13 contrastes condicionales. Los contrastes condicionales evalúan si la proporción de violaciones es consistente con el

nivel de confianza del VaR, teniendo en cuenta la dependencia temporal de las violaciones. Las pruebas condicionales son más generales que las pruebas incondicionales y pueden incluir pruebas de independencia como parte de su evaluación. Los contrastes incondicionales contrastan si la tasa de fallo es similar a la tasa de fallo esperada, mientras que los contrastes de independencia chequean si los fallos son independientes entre sí.

La independencia de los fallos es un requisito importante para asegurar que el modelo de VaR es válido. Si los fallos están correlacionados, podría indicar que el modelo no está capturando adecuadamente el riesgo en diferentes momentos del tiempo.

Cuando se hace *backtesting* el principal desafío y problema asociado con la validación del pronóstico del VaR es la baja potencia de las pruebas de *backtesting*. Se hace necesario, así, una metodología de *backtesting* libre de modelos y tener en cuenta el error de estimación.

Cuadro 5.

Promedio de violaciones del VaR (en %) de los diferentes modelos de predicción de la RV

	<i>HAR</i>	<i>SHAR</i>	<i>HARQ</i>	<i>SHARQ</i>	<i>TVHAR</i>	<i>DL</i>	<i>RF</i>
$\alpha = 1\%$							
BTC	0,4	0,4	1,1	1,5	6,9	2,5	1,2
NASDAQ	3,9	4,2	3,7	3,3	1,6	4,1	3,5
SP500	4,1	3,3	3,8	3,2	3,7	4,4	3,6
$\alpha = 5\%$							
BTC	1,4	1,5	3,5	3,8	10,6	4,7	2,4
NASDAQ	10,0	10,3	9,6	8,6	2,8	10,2	9,6
SP500	9,3	8,5	8,9	8,2	9,0	9,3	8,8

Fuente: Candel *et al.* (2020).

Se han propuesto varias pruebas de *backtesting* para abordar estos problemas, incluida la prueba de Cuantil Dinámico (DQ) propuesta por Engle y Manganelli (2004), Escanciano y Olmo (2010, 2011) y, más recientemente, la propuesta de Barendse *et al.* (2021).

Barendse *et al.* (2021) también proponen versiones robustas para las pruebas de *Expected shortfall* producidas por error de estimación y muestran vía simulaciones de Monte Carlo que las pruebas estándar pueden sufrir distorsiones de tamaño debido al error de estimación, con frecuencias de rechazo empíricas que exceden los niveles de significación nominal. La robustificación de las pruebas de *backtesting* generalmente corrige este problema con bastante éxito, aunque puede llevar a alguna pérdida de potencia y no son de fácil implementación.

Para evaluar los modelos en el contexto del VaR además de los contrastes propuestos en Christoffersen (1998) y Engle y Manganelli (2004) usaremos una herramienta para el *backtesting* de pronósticos del VaR que se basa en un modelo no lineal, llamado modelo de regresión Binaria Dinámica (DB), propuesto por Dumitrescu *et al.* (2012). Esta propuesta

mejora la propuesta de Engle y Manganelli (2004) que implementa un contraste basado en una regresión simple de las violaciones en las violaciones pasadas. Se sabe que este modelo no es adecuado para variables dependientes binarias dado que los residuos son heterocedásticos por construcción y su distribución es discreta, Gouriéroux y Jasiak (2001).

El modelo de regresión binaria dinámica, además de ser adecuado para variables dependientes binarias, tiene en cuenta la posible correlación entre violaciones (*clusters*) en la estimación y exhibe una potencia más alta que las pruebas de *backtesting* usuales en la literatura.

Usaremos siete especificaciones diferentes para el modelo DB, denotadas por DB_1 a DB_7 , que están inspiradas en las especificaciones CAViaR propuestas por Engle y Manganelli (2004). Estos contrastes son más robustos frente al error de estimación y tienen buenas propiedades en muestras finitas. Además, tienen en cuenta la posible correlación entre violaciones, que no es abordada por algunas de las pruebas existentes. Por otro lado, pueden implementarse fácilmente y permiten pruebas separadas de las hipótesis de cobertura incondicional, independencia y cobertura condicional.

El modelo de respuesta binaria dinámica considera que la probabilidad condicional de violación en el tiempo t viene dada por:

$$Pr(I_t(\alpha) = 1 | F_{t-1}) = E[I_t(\alpha) | F_{t-1}] = F(\pi_t),$$

donde $F(\cdot)$ denota una función de distribución acumulada y F_{t-1} es el conjunto de información disponible en $t - 1$.

Se asume que el índice π_t satisface la siguiente representación autorregresiva:

$$\pi_t = c + \sum_{j=1}^{q_1} \beta_j \pi_{t-j} + \sum_{j=1}^{q_2} \delta_j I_{t-j}(\alpha) + \sum_{j=1}^{q_3} \psi_j l(x_{t-j}, \phi) + \sum_{j=1}^{q_4} \gamma_j l(x_{t-j}, \phi) I_{t-j}, \quad [4]$$

donde $l(\cdot)$ es una función de un número finito de valores retardados observables, y x_t es un vector de variables explicativas. El papel de $l(\cdot)$ es vincular el índice t con las variables observables que pertenecen al conjunto de información. Una elección natural para x_{t-j} podría ser los rendimientos retardados o los retardos del VaR.

Dada la ecuación [4], Dumitrescu *et al.* (2012) sugieren usar siete especificaciones. Las primeras cuatro especificaciones (DB_1 a DB_4) corresponden a un modelo de respuesta binaria dinámica que incluye el índice π_t retardado como variable explicativa y alguna información adicional a través de los valores pasados observados del proceso de violación.

Los modelos quinto y sexto (DB_5 y DB_6) se derivan de las especificaciones autorregresivas de cuantiles CAViaR utilizadas por Engle y Manganelli (2004).

La séptima especificación (DB_7) incluye una interacción entre el VaR retardado y la variable binaria de violación para incluir una respuesta asimétrica del índice a valores pasados del VaR en caso de una violación.

$$DB_1: \quad \pi_t = c + \beta_1 \pi_{t-1}, \quad [5]$$

$$DB_2: \quad \pi_t = c + \beta_1 \pi_{t-1} + \delta_1 I_{t-1}(\alpha) \quad [6]$$

$$DB_3: \quad \pi_t = c + \beta_1 \pi_{t-1} + \delta_1 I_{t-1}(\alpha) + \delta_2 I_{t-2}(\alpha), \quad [7]$$

$$DB_4: \quad \pi_t = c + \beta_1 \pi_{t-1} + \delta_1 I_{t-1}(\alpha) + \delta_2 I_{t-2}(\alpha) + \delta_3 I_{t-3}(\alpha), \quad [8]$$

$$DB_5: \quad \pi_t = c + \beta_1 \pi_{t-1} + \psi_1 \text{VaR}_{t-1}, \quad [9]$$

$$DB_6: \quad \pi_t = c + \beta_1 \pi_{t-1} + \delta_1 I_{t-1}(\alpha) + \psi_1 \text{VaR}_{t-1} \quad [10]$$

$$DB_7: \quad \pi_t = c + \beta_1 \pi_{t-1} + \delta_1 I_{t-1}(\alpha) + \psi_1 \text{VaR}_{t-1} + \gamma_1 \text{VaR}_{t-1} I_{t-1}. \quad [11]$$

Cada una de estas especificaciones tiene diferentes variables explicativas y, por lo tanto, captura diferentes aspectos del proceso de violación.

Para la estimación de cada uno de los modelos mencionados anteriormente, Dumitrescu *et al.* (2012) han usado el método propuesto por Kauppi y Saikkonen (2008) que se basa en la estimación de máxima verosimilitud con restricciones. Este método requiere que las variables explicativas sean estacionarias y que las variables aleatorias sigan una distribución normal para cumplir con las condiciones de regularidad necesarias.

De manera formal, sea π_t proveniente de cualquiera de las especificaciones [5] a [11], el logaritmo de la función de verosimilitud viene dado por:

$$\ln L(\theta, c; I(\alpha), Z) = \sum_{t=1}^T \left[I_t(\alpha) \ln F(\pi_t(\theta, c, Z_t)) + (1 - I_t(\alpha)) \ln (1 - F(\pi_t(\theta, c, Z_t))) \right],$$

donde θ es el vector de parámetros de las especificaciones anteriores excepto la constante, *i.e.*, $\theta = [\beta_1, \delta_1, \delta_2, \delta_3, \psi_1, \gamma_1]$ y Z_t es el vector de variables explicativas en el tiempo t . La estimación por máxima verosimilitud lleva a estimadores con buenas propiedades estadísticas en muestras grandes.

Contrastar la hipótesis nula de cobertura condicional ($E[I_t(\alpha)|F_{t-1}] = \alpha$) utilizando el modelo binario dinámico es similar al caso de la prueba DQ de Engle y Manganelli (2004), ya que se basa en el contraste estándar sobre los coeficientes de la regresión:

$$H_0: \beta_1 = 0, \delta_1 = 0, \delta_2 = 0, \delta_3 = 0, \psi_1 = 0, \gamma_1 = 0 \text{ y } c = F^{-1}(\alpha).$$

Bajo la hipótesis nula, la variable de violaciones es ortogonal a cualquier variable explicativa que pertenezca al conjunto de información F_{t-1} y la probabilidad condicional de una violación es igual a α . Se pueden usar contrastes de razón de verosimilitudes (LR), *Wald* y multiplicadores de *Lagrange* (LM).

De manera similar, la prueba de independencia se basa en $H_0: \beta_1 = 0, \delta_1 = 0, \delta_2 = 0, \delta_3 = 0, \psi_1 = 0, \gamma_1 = 0$, es decir, la nulidad de los coeficientes, pero esta vez en un modelo donde la constante ya no es una función de α . Bajo la hipótesis nula de independencia, no hay dependencia temporal de las violaciones, pero la probabilidad de violación de VaR podría ser diferente de la tasa de cobertura α .

Cuadro 6.

Valor en riesgo: nivel de confianza del 99 %

Variable	Modelo	Promedio VaR	Contrastes de "backtesting" de cobertura condicional												
			DB1	DB2	DB3	DB4	DB5	DB6	DB7	LR	DQ1	DQ2	DQ3	DQV aR _c	DQV aR _c
BTC															
HAR	0.140	15.98 (0.00)	15.98 (0.00)	15.98 (0.00)	15.98 (0.01)	16.89 (0.00)	16.90 (0.00)	16.71 (0.00)	6.79 (0.03)	3.62 (0.16)	3.62 (0.31)	3.63 (0.46)	3.68 (0.30)	3.75 (0.59)	3.80 (0.80)
SHAR	0.141	15.98 (0.00)	15.98 (0.00)	15.98 (0.00)	15.98 (0.01)	16.96 (0.00)	16.96 (0.00)	16.86 (0.00)	6.79 (0.03)	3.62 (0.16)	3.62 (0.31)	3.63 (0.46)	3.72 (0.59)	3.74 (0.81)	3.77 (1.10)
HARQ	0.101	9.31 (0.01)	9.95 (0.02)	12.71 (0.01)	14.02 (0.02)	12.62 (0.01)	12.62 (0.01)	12.34 (0.01)	0.28 (0.87)	7.35 (0.03)	13.60 (0.05)	14.16 (0.01)	7.67 (0.01)	14.48 (0.01)	15.80 (0.03)
SHARQ	0.106	10.68 (0.00)	10.70 (0.01)	12.76 (0.01)	13.23 (0.02)	15.23 (0.00)	15.24 (0.01)	15.22 (0.00)	1.90 (0.39)	2.91 (0.23)	6.98 (0.07)	7.26 (0.12)	5.16 (0.16)	10.03 (0.07)	10.76 (0.15)
TVHAR	0.076	153.92 (0.00)	192.92 (0.00)	193.20 (0.00)	194.86 (0.00)	218.15 (0.00)	218.15 (0.00)	217.92 (0.00)	160.25 (0.00)	507.83 (0.00)	628.71 (0.00)	648.47 (0.00)	649.37 (0.00)	735.61 (0.00)	768.87 (0.00)
DL	0.119	20.15 (0.00)	24.66 (0.00)	23.14 (0.00)	23.16 (0.00)	25.92 (0.00)	26.16 (0.00)	23.78 (0.00)	11.33 (0.00)	26.22 (0.00)	34.21 (0.00)	40.28 (0.00)	27.28 (0.00)	37.13 (0.00)	45.66 (0.00)
RF	0.111	9.21 (0.01)	12.16 (0.01)	12.80 (0.01)	13.02 (0.02)	13.76 (0.01)	13.78 (0.02)	11.31 (0.01)	0.22 (0.89)	6.52 (0.04)	12.07 (0.01)	12.61 (0.01)	7.01 (0.07)	12.74 (0.03)	13.46 (0.06)
NASDAQ															
HAR	0.024	49.27 (0.00)	50.84 (0.00)	50.91 (0.00)	53.49 (0.00)	51.48 (0.00)	51.64 (0.00)	51.02 (0.00)	49.40 (0.00)	85.99 (0.00)	86.75 (0.00)	103.44 (0.00)	95.75 (0.00)	102.06 (0.00)	116.42 (0.00)
SHAR	0.024	57.88 (0.00)	65.05 (0.00)	65.14 (0.00)	65.31 (0.00)	65.74 (0.00)	68.98 (0.00)	59.61 (0.00)	58.64 (0.00)	107.74 (0.00)	107.84 (0.00)	120.20 (0.00)	116.63 (0.00)	116.66 (0.00)	126.76 (0.00)
HARQ	0.024	43.80 (0.00)	45.92 (0.00)	46.23 (0.00)	48.08 (0.00)	50.54 (0.00)	50.93 (0.00)	50.41 (0.00)	44.05 (0.00)	75.14 (0.00)	76.21 (0.00)	95.47 (0.00)	105.39 (0.00)	108.63 (0.00)	128.74 (0.00)
SHARQ	0.024	33.55 (0.00)	36.50 (0.00)	36.61 (0.00)	37.47 (0.00)	40.70 (0.00)	40.91 (0.00)	40.12 (0.00)	34.18 (0.00)	56.34 (0.00)	58.81 (0.00)	69.75 (0.00)	86.30 (0.00)	87.30 (0.00)	94.92 (0.00)
TVHAR	0.078	3.15 (0.21)	12.71 (0.01)	12.78 (0.01)	13.04 (0.02)	33.27 (0.00)	33.26 (0.00)	33.27 (0.00)	4.43 (0.11)	7.23 (0.03)	10.45 (0.02)	13.40 (0.01)	27.14 (0.00)	30.80 (0.00)	34.06 (0.00)
DL	0.024	54.96 (0.00)	56.84 (0.00)	57.59 (0.00)	57.81 (0.00)	58.77 (0.00)	58.79 (0.00)	57.87 (0.00)	54.99 (0.00)	97.80 (0.00)	102.22 (0.00)	106.53 (0.00)	105.25 (0.00)	116.37 (0.00)	119.11 (0.00)
RF	0.025	38.56 (0.00)	40.34 (0.00)	40.64 (0.00)	41.52 (0.00)	44.56 (0.00)	44.73 (0.00)	44.49 (0.00)	38.98 (0.00)	65.25 (0.00)	66.95 (0.00)	75.98 (0.00)	89.99 (0.00)	93.28 (0.00)	116.48 (0.00)
SP500															
HAR	0.019	54.96 (0.00)	58.13 (0.00)	58.88 (0.00)	58.94 (0.00)	59.34 (0.00)	59.42 (0.00)	58.03 (0.00)	55.85 (0.00)	101.99 (0.00)	115.24 (0.00)	118.75 (0.00)	116.07 (0.00)	125.29 (0.00)	140.27 (0.00)
SHAR	0.019	33.55 (0.00)	34.88 (0.00)	35.08 (0.00)	39.03 (0.00)	35.80 (0.00)	35.82 (0.00)	34.81 (0.00)	34.18 (0.00)	56.34 (0.00)	58.81 (0.00)	61.14 (0.00)	59.94 (0.00)	64.98 (0.00)	66.74 (0.00)
HARQ	0.019	46.51 (0.00)	48.99 (0.00)	49.23 (0.00)	49.24 (0.00)	51.60 (0.00)	51.66 (0.00)	50.48 (0.00)	47.89 (0.00)	86.25 (0.00)	92.37 (0.00)	92.87 (0.00)	103.15 (0.00)	106.40 (0.00)	118.35 (0.00)
SHARQ	0.019	31.13 (0.00)	31.13 (0.00)	31.14 (0.00)	31.14 (0.00)	36.42 (0.00)	36.52 (0.00)	35.62 (0.00)	31.12 (0.00)	49.17 (0.00)	49.17 (0.00)	49.18 (0.00)	63.91 (0.00)	71.35 (0.00)	73.07 (0.00)
TVHAR	0.019	43.80 (0.00)	44.19 (0.00)	44.25 (0.00)	44.39 (0.00)	44.68 (0.00)	44.68 (0.00)	47.20 (0.00)	44.05 (0.00)	75.14 (0.00)	76.21 (0.00)	76.65 (0.00)	77.21 (0.00)	80.69 (0.00)	82.94 (0.00)
DL	0.019	63.86 (0.00)	66.78 (0.00)	67.35 (0.00)	67.37 (0.00)	67.43 (0.00)	67.54 (0.00)	66.41 (0.00)	64.39 (0.00)	120.02 (0.00)	129.93 (0.00)	132.14 (0.00)	130.87 (0.00)	139.66 (0.00)	142.46 (0.00)
RF	0.020	41.15 (0.00)	41.22 (0.00)	41.23 (0.00)	42.13 (0.00)	45.43 (0.00)	45.36 (0.00)	45.25 (0.00)	41.21 (0.00)	68.91 (0.00)	69.17 (0.00)	70.58 (0.00)	88.38 (0.00)	94.56 (0.00)	104.81 (0.00)

Nota: Se utilizaron trece contrastes de cobertura condicional, a saber, siete especificaciones binarias dinámicas DB Dumitrescu *et al.* (2012), seis contrastes DQ Engle y Manganelli (2004) incluyendo varios retardos de la variable de violaciones y VaR, así como el contraste LR_c Christoffersen (1998). Los valores *p* correspondientes se presentan entre paréntesis.

Fuente: Candel *et al.* (2020).

Cuadro 7.

Valor en riesgo: nivel de confianza del 95 %

Variable	Modelo	Promedio VaR	Contrastes de "backtesting" de cobertura condicional												
			DB1	DB2	DB3	DB4	DB5	DB6	DB7	LR	DQ1	DQ2	DQ3	DQV aR _t	DQV aR _c
BTC															
HAR	0.098	48.93 (0.00)	53.61 (0.00)	54.34 (0.00)	54.35 (0.00)	53.75 (0.00)	53.84 (0.00)	48.96 (0.00)	43.25 (0.00)	29.69 (0.00)	30.60 (0.00)	31.41 (0.00)	29.72 (0.00)	30.63 (0.00)	31.44 (0.00)
SHAR	0.099	46.08 (0.00)	49.86 (0.00)	50.60 (0.00)	50.61 (0.00)	50.24 (0.00)	50.27 (0.00)	46.08 (0.00)	40.46 (0.00)	28.03 (0.00)	28.83 (0.00)	29.55 (0.00)	28.03 (0.00)	28.83 (0.00)	29.59 (0.00)
HARQ	0.073	12.73 (0.00)	17.09 (0.00)	18.92 (0.00)	19.28 (0.00)	17.72 (0.00)	24.49 (0.00)	15.80 (0.00)	6.75 (0.00)	5.71 (0.06)	10.56 (0.01)	10.85 (0.03)	7.00 (0.07)	11.72 (0.04)	12.00 (0.10)
SHARQ	0.074	10.43 (0.01)	13.80 (0.00)	14.87 (0.00)	14.80 (0.01)	13.05 (0.01)	13.04 (0.02)	14.62 (0.00)	4.54 (0.10)	3.69 (0.16)	5.13 (0.16)	6.50 (0.16)	5.39 (0.15)	6.78 (0.24)	8.12 (0.32)
TVHAR	0.053	53.92 (0.00)	97.48 (0.00)	97.57 (0.00)	97.84 (0.00)	142.86 (0.00)	142.89 (0.00)	142.83 (0.00)	59.37 (0.00)	95.36 (0.00)	115.18 (0.00)	117.89 (0.00)	165.46 (0.00)	174.92 (0.00)	190.28 (0.00)
DL	0.084	6.73 (0.03)	14.69 (0.00)	17.97 (0.00)	18.37 (0.00)	20.06 (0.00)	20.31 (0.00)	15.58 (0.00)	4.39 (0.11)	7.88 (0.02)	18.01 (0.00)	18.01 (0.00)	8.85 (0.03)	22.74 (0.00)	22.74 (0.00)
RF	0.078	26.45 (0.00)	29.06 (0.00)	31.50 (0.00)	31.63 (0.00)	30.34 (0.00)	30.32 (0.00)	28.25 (0.00)	21.50 (0.00)	15.40 (0.00)	17.26 (0.00)	17.40 (0.00)	15.89 (0.00)	17.70 (0.00)	17.84 (0.01)
NASDAQ															
HAR	0.017	41.85 (0.00)	48.82 (0.00)	49.70 (0.00)	49.82 (0.00)	48.97 (0.00)	54.53 (0.00)	46.94 (0.00)	41.74 (0.00)	53.11 (0.00)	56.80 (0.00)	57.71 (0.00)	67.37 (0.00)	70.61 (0.00)	71.35 (0.00)
SHAR	0.017	46.45 (0.00)	59.57 (0.00)	61.17 (0.00)	61.26 (0.00)	60.29 (0.00)	64.51 (0.00)	52.16 (0.00)	46.40 (0.00)	59.74 (0.00)	64.07 (0.00)	66.74 (0.00)	74.78 (0.00)	77.01 (0.00)	78.04 (0.00)
HARQ	0.017	36.01 (0.00)	37.34 (0.00)	41.54 (0.00)	42.02 (0.00)	41.95 (0.00)	47.03 (0.00)	41.87 (0.00)	36.13 (0.00)	45.35 (0.00)	48.74 (0.00)	48.75 (0.00)	63.75 (0.00)	69.85 (0.00)	70.57 (0.00)
SHARQ	0.017	23.07 (0.00)	24.52 (0.00)	30.73 (0.00)	30.76 (0.00)	31.77 (0.00)	31.80 (0.00)	29.35 (0.00)	24.02 (0.00)	29.16 (0.00)	29.77 (0.00)	29.89 (0.00)	44.48 (0.00)	44.70 (0.00)	44.86 (0.00)
TVHAR	0.055	11.96 (0.00)	30.01 (0.00)	31.05 (0.00)	33.80 (0.00)	55.06 (0.00)	55.10 (0.00)	54.69 (0.00)	15.83 (0.00)	13.87 (0.00)	17.09 (0.00)	28.89 (0.00)	23.40 (0.00)	27.03 (0.00)	36.93 (0.00)
DL	0.017	44.90 (0.00)	45.30 (0.00)	47.78 (0.00)	47.53 (0.00)	48.88 (0.00)	49.95 (0.00)	48.72 (0.00)	45.05 (0.00)	57.90 (0.00)	58.45 (0.00)	59.05 (0.00)	61.52 (0.00)	65.47 (0.00)	66.45 (0.00)
RF	0.017	36.01 (0.00)	37.15 (0.00)	43.86 (0.00)	43.87 (0.00)	43.47 (0.00)	43.77 (0.00)	40.99 (0.00)	36.63 (0.00)	46.19 (0.00)	46.90 (0.00)	47.07 (0.00)	60.49 (0.00)	60.50 (0.00)	63.12 (0.00)
SP500															
HAR	0.014	31.88 (0.00)	38.92 (0.00)	39.13 (0.00)	40.16 (0.00)	39.11 (0.00)	41.27 (0.00)	34.94 (0.00)	32.01 (0.00)	39.75 (0.00)	46.78 (0.00)	47.12 (0.00)	46.79 (0.00)	51.83 (0.00)	54.84 (0.00)
SHAR	0.014	21.92 (0.00)	28.35 (0.00)	29.11 (0.00)	31.07 (0.00)	28.35 (0.00)	29.43 (0.00)	25.91 (0.00)	21.82 (0.00)	26.11 (0.00)	35.07 (0.00)	35.23 (0.00)	34.02 (0.00)	42.33 (0.00)	43.09 (0.00)
HARQ	0.013	26.70 (0.00)	30.52 (0.00)	31.68 (0.00)	31.67 (0.00)	30.05 (0.00)	30.02 (0.00)	29.69 (0.00)	26.74 (0.00)	32.60 (0.00)	42.09 (0.00)	42.46 (0.00)	40.66 (0.00)	47.17 (0.00)	49.62 (0.00)
SHARQ	0.013	18.60 (0.00)	20.92 (0.00)	22.17 (0.00)	22.29 (0.00)	20.50 (0.00)	20.92 (0.00)	20.50 (0.00)	18.51 (0.00)	21.84 (0.00)	26.91 (0.00)	27.79 (0.00)	24.89 (0.00)	30.31 (0.00)	32.89 (0.00)
TVHAR	0.014	27.96 (0.00)	35.75 (0.00)	32.35 (0.00)	37.67 (0.00)	36.31 (0.00)	40.01 (0.00)	33.62 (0.00)	27.96 (0.00)	34.24 (0.00)	43.09 (0.00)	43.10 (0.00)	40.59 (0.00)	47.10 (0.00)	47.36 (0.00)
DL	0.013	31.88 (0.00)	35.03 (0.00)	35.36 (0.00)	35.35 (0.00)	40.05 (0.00)	40.21 (0.00)	40.03 (0.00)	31.79 (0.00)	39.34 (0.00)	46.42 (0.00)	46.44 (0.00)	55.95 (0.00)	59.94 (0.00)	60.42 (0.00)
RF	0.014	25.47 (0.00)	31.88 (0.00)	34.23 (0.00)	34.29 (0.00)	32.80 (0.00)	33.00 (0.00)	30.84 (0.00)	25.59 (0.00)	31.13 (0.00)	48.84 (0.00)	49.11 (0.00)	41.75 (0.00)	55.74 (0.00)	59.46 (0.00)

Nota: Se utilizaron trece contrastes de cobertura condicional, a saber, siete especificaciones binarias dinámicas DB Dumitrescu et al. (2012), seis contrastes DQ Engle y Manganelli (2004) incluyendo varios retardos de la variable de violaciones y VaR, así como el contraste LR_C Christoffersen (1998). Los valores *p* correspondientes se presentan entre paréntesis.

Fuente: Candel et al. (2020).

El **cuadro 6** muestra los resultados de las pruebas de *backtesting* para los niveles de confianza del 99 % y 95 %, respectivamente, para las tres series de rendimientos financieros. Teniendo en cuenta el nivel de confianza del 99 % y los rendimientos de BTC, observamos que los modelos HARQ y RF son los que mejor predicen el VaR, dado que las hipótesis nulas de cobertura condicionales son menos rechazadas.

Los contrastes de Engle y Manganelli (2004) y de Christoffersen (1998) no son rechazados casi nunca, y los contrastes en los modelos dinámicos binarios no rechazan la cobertura condicional al nivel de significación del 1 %.

En cuanto a los rendimientos de NASDAQ, los modelos en general predicen muy mal el VaR, excepto el modelo TVHAR, que genera coberturas condicionales más cercanas a las teóricas. En cuanto a los rendimientos de SP500, los modelos predicen VaR que no superan las pruebas de *backtesting*.

Los resultados del *backtesting* al nivel de confianza del 5 % son un poco peores (véase **cuadro 7**). Solo para el BTC, el modelo SHARQ logra predecir el valor en riesgo de forma más o menos adecuada.

4. CONCLUSIONES

El propósito de nuestro estudio se centra en profundizar en la predicción de la volatilidad medida por la varianza realizada que es calculada a partir de datos intradiarios, particularmente de datos de cinco minutos. Se ha demostrado que esta medida es una medida consistente de la volatilidad diaria, una variable que no es observable. El objetivo es analizar qué modelos son los más adecuados para prever las fluctuaciones en el riesgo, con el fin de optimizar las estrategias de inversores y gestores de carteras.

En el capítulo que presentamos, se llevaron a cabo predicciones utilizando diversos modelos paramétricos, semiparamétricos y de aprendizaje automático. La evaluación de su desempeño se hizo mediante funciones de pérdida específicas y contrastes de habilidad predictiva superior, tanto incondicionales como condicionales. Además, se contrastaron los modelos en el contexto de la gestión del riesgo mediante el cálculo del valor en riesgo.

Los datos analizados incluyen los rendimientos de Bitcoin y dos índices bursátiles: NASDAQ y S&P 500, que representan diferentes grupos de mercados: criptomonedas, tecnología y el mercado de acciones estadounidense, respectivamente.

Se observó que algunos modelos tenían un desempeño superior en función de la serie de rendimientos financieros utilizada. Sin embargo, en términos generales, se encontró que el modelo heterogéneo autorregresivo y los modelos heterogéneos autorregresivos con cuartilidad simétrico y asimétrico y *Random Forests* presentaban los valores más bajos de QLIKE, indicando que son los más adecuados para predecir la varianza realizada (a un día) en las tres series de rendimientos analizadas.

En cuanto al cálculo del valor en riesgo, se aplicaron trece pruebas condicionales de *backtesting*, que tienen la ventaja de una mayor potencia en comparación con las pruebas convencionales y tienen en cuenta el error de estimación. Las pruebas muestran evidencia de que los modelos heterogéneos autorregresivos con cuarticidad, tanto simétrico como asimétrico, y modelos semiparamétricos como *Random Forests*, parecen adecuados para el cálculo del valor en riesgo para Bitcoin. Mientras que el modelo heterogéneo autorregresivo con parámetros que cambian con el tiempo, es el único modelo avalado por las pruebas de *backtesting* para el NASDAQ.

En relación al S&P 500, los modelos muestran un bajo desempeño para el cálculo del valor en riesgo en el periodo considerado. Es relevante señalar que, durante periodos de alta volatilidad, como el confinamiento por la pandemia, que afectó en mayor medida a las empresas convencionales incluidas en el índice S&P 500, se observó un peor desempeño de los modelos en términos del cálculo del valor en riesgo.

En conclusión, el estudio de la predicción de la varianza realizada en los mercados financieros es un campo de gran importancia, que tiene un impacto directo en la toma de decisiones de inversión y en la gestión del riesgo. Contar con modelos precisos y confiables es esencial para navegar en los dinámicos y competitivos mercados financieros actuales.

Referencias

- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. y EBENS, H. (2001a). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61, pp. 43–76. doi: [https://doi.org/10.1016/S0304-405X\(01\)00055-1](https://doi.org/10.1016/S0304-405X(01)00055-1)
- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. y LABYS, W. C. (2000). Great realizations in the exchange rate: Exchange rates and inflation in industrialized countries. *Journal of International Economics*, 50, pp. 87–108.
- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. y LABYS, W. C. (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96, pp. 42–55.
- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. y LABYS, W. C. (2003). Modeling and forecasting realized volatility. *Econometrica* 71, pp. 579–625.
- ANDERSEN, T. G., BOLLERSLEV, T. y DOBREV, D. (2007). No-arbitrage semimartingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics*, 138, pp. 125–180.
- BARENDSE, S., KOLE, E. y VAN DIJK, D. (2021). Backtesting Value-at-Risk and Expected Shortfall in the Presence of Estimation Error. *Journal of Financial Econometrics*, 21, pp. 528–568. doi:10.1093/jfinec/nbab008
- BARNDORFF-NIELSEN, O. E., KINNEBROCK, S. y SHEPHARD, N. (2010). Measuring downside risk? Realised semivariance. En T. BOLLERSLEV, J. RUSSELL y M. WATSON, M. (Eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Oxford University Press, New York.
- BOLLERSLEV, T., LITVINOVA, J. y TAUCHEN, G. (2006). Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, 4, pp. 353–384.
- BOLLERSLEV, T., PATTON, A. J. y QUAEDVLIEG, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192, pp. 1–18.

- BREIMAN, L. (2001). Random forests. *Machine learning*, 45, pp. 5–32.
- CAI, Z. (2007). Trending time-varying coefficient time series with serially correlated errors. *Journal of Econometrics*, 136, pp. 163–188.
- CAMPBELL, J. Y. y HENTSCHL, L. (1992). No news is good news: A asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31, pp. 281–318.
- CANDEL, A., PARMAR, V., LEDELL, E. y ARORA, A. (2020). Deep learning with h2o. H₂o. ai Inc , pp. 1–21.
- CHEN, X. B., GAO, J., LI, D. y SILVAPULLE, P. (2018). Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. *Journal of Business & Economic Statistics*, 36, pp. 88–100.
- CHRISTIE, A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10, pp. 407–432.
- CHRISTOFFERSEN, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39, pp. 841–862.
- CORBET, S., LARKIN, C. y LUCEY, B. (2020). The contagion effects of the COVID-19 pandemic: Evidence from gold and cryptocurrencies. *Finance Research Letters*, 35, 101554. doi: <https://doi.org/10.1016/j.frl.2020.101554>.
- CORSI, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7, pp. 174–196.
- DUMITRESCU, E. I., HURLIN, C. y PHAM, V. (2012). Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Test. *Revue de l'association française de finance*, 33, pp. 79–112.
- ENGL, R. F. y MANGANELLI, S. (2004). CAViaR. *Journal of Business and Economic Statistics*, 22, pp. 367–381. doi:10.1198/073500104000000370
- ESCANCIANO, J. y OLMO, J. (2010). Backtesting parametric Value-at-Risk with estimation risk. *Journal of Business and Economic Statistics*, 28, pp. 36–51. doi:10.1198/jbes.2009.07063
- ESCANCIANO, J. C. y OLMO, J. (2011). Robust backtesting test for value-at-risk models. *Journal of Financial Econometrics*, 9, pp. 132–161. doi:10.1093/jfinec/nbq021
- FAN, J. y GIJBELS, I. (1996). Local Polynomial Modeling and Its Applications. Hapman and Hall, London.
- GOODELL, J. W. y GOUTTE, S. (2021). Co-movement of COVID-19 and Bitcoin: Evidence from wavelet coherence analysis. *Finance Research Letters*, 38, 101625. doi: <https://doi.org/10.1016/j.frl.2020.101625>
- GOURIEROUX, C. y JASIAK, J. (2001). Financial Econometrics: Problems, Models, and Methods. Princeton University Press.
- HANSEN, P. R., LUNDE, A. y NASON, J. M. (2011). The model confidence set. *Econometrica*, 79, pp. 453–497.
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H. y FRIEDMAN, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Volume 2. Springer.
- KAUPPI, H. y SAIKKONEN, P. (2008). Predicting US recessions with dynamic binary response models. *Review of Economics and Statistics*, 90, pp. 777–791.
- LANDRY, M. (2016). Machine learning with r and h2o. H₂o. ai: Mountain View, CA, USA .
- LEDELL, E., GILL, N., AIELLO, S., FU, A., CANDEL, A., CLICK, C., KRALJEVIC, T., NYKODYM, T., ABOYOUN, P., KURKA, M., et al., (2018). *Package h2o*. R Foundation for Statistical Computing: Vienna, Austria , 17.

- LI, J., LIAO, Z. y QUAEDVLIEG, R. (2022). Conditional superior predictive ability. *The Review of Economic Studies*, 89, pp. 843–875.
- MCALIEER, M. (2014). Oil shocks and the macroeconomy: The role of price variability. *Econometrics*, 2, pp. 145–150.
- PATTON, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160, 246–256.
- PATTON, A. J. y SHEPPARD, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility full access. *Review of Economics and Statistics*, 97, pp. 683–697.
- ROBINSON, P. (1989). Nonparametric estimation of time-varying parameters. En P. HACKL (Ed.), *Statistical Analysis and Forecasting of Economic Structural Change*. Springer, Berlin.
- SIFAT, I. (2021). On cryptocurrencies as an independent asset class: Long-horizon and COVID- 19 pandemic era decoupling from global sentiments. *Finance Research Letters*, 43, 102013. doi: <https://doi.org/10.1016/j.frl.2021.102013>