



Maestría en

Ciencia de Datos y Máquinas de Aprendizaje con Mención en Inteligencia Artificial

**Trabajo previo a la obtención del título de
Magíster en Ciencia de Datos y Máquinas
de Aprendizaje con mención en
Inteligencia Artificial**

Autores:

Vicente Guayanay Jimmy Leonel

Muso Tandalla Marjorie Aracely

Andrade Peñafiel Mateo Josué

Tutores:

Karla Estefania Mora Cajas

Fernanda Paulina Vizcaino Imacaña

**Estimación del Valor en Riesgo (VaR) y Backtesting Dinámico: Un
Análisis Comparativo de la Robustez entre Modelos de
Heterocedasticidad Condicional (GARCH) y Redes Neuronales
Recurrentes (LSTM)**

Quito - Ecuador

Junio de 2026

**REINVENTEMOS
EL FUTURO**

Certificación de autoría

Nosotros, **Mateo Josué Andrade Peñañiel, Marjorie Aracely Muso Tandalla, Jimmy Loenel Vicente Guayanay**, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.



Validar únicamente en FirmaEC.
Firmado electrónicamente por:
**JIMMY LEONEL
VICENTE GUAYANAY**

Firma
Jimmy Leonel Vicente Guayanay



**Marjorie Aracely
Musu Tandalla**
Time Stamping
Security Data

Firma
Marjorie Aracely Muso Tandalla



**Mateo Josue
Andrade Peñañiel**
Time Stamping
Security Data

Firma
Mateo Josué Andrade Peñañiel

Autorización de Derechos de Propiedad Intelectual

Nosotros, **Mateo Josué Andrade Peñafiel**, **Marjorie Aracely Muso Tandalla**, **Jimmy Loenel Vicente Guayanay**, en calidad de autores del trabajo de investigación titulado *Estimación del Valor en Riesgo (VaR) y Backtesting Dinámico: Un Análisis Comparativo de la Robustez entre Modelos de Heterocedasticidad Condicional (GARCH) y Redes Neuronales Recurrentes (LSTM)*, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

Quito, junio de 2026



Firma
Jimmy Leonel Vicente Guayanay



Firma
Marjorie Aracely Muso Tandalla



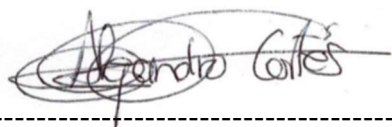
Firma
Mateo Josué Andrade Peñafiel

Aprobación de dirección y coordinación del programa

Nosotros, **Alejandro Cortés López** Director EIG y **Karla Estefanía Mora Cajas**

Coordinadora UIDE, declaramos que: **Mateo Josué Andrade Peñafiel, Marjorie Aracely**

Muso Tandalla, Jimmy Loenel Vicente Guayanay son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortés López
Director de la
Maestría en Ciencia de datos y máquinas de
aprendizaje con mención en Inteligencia
Artificial



Karla Estefanía Mora Cajas
Coordinadora de la
Maestría en Ciencia de datos y máquinas de
aprendizaje con mención en Inteligencia
Artificial

Dedicatoria

A mi madre Sandra y mi padre Jorge que nunca han dejado que me sienta solo, este trabajo es para ustedes, son el pilar de mi formación, me han hecho correr gracias a que ustedes caminaron descalzos.

En memoria de mi abuelo Miguel que me enseñó lo importante de la perseverancia y honestidad, tu ejemplo en vida me ha hecho alcanzar mis objetivos, sé que me cuidas siempre, desde aquí te abrazo y te entrego mi esfuerzo esperando que te sientas orgulloso donde estés.

Agradecimientos

Gracias a Dios por permitirme disfrutar estos logros y compartirlos con mis seres queridos.

Agradezco a mis hermanos Andrés y Jorge por haberme guiado con su ejemplo de perseverancia, por ser amigos y confidentes siempre que pudieron.

A Sebas, Mateo, Josué y Carlos por acompañarme en todos los momentos que me sentía sin horizonte.

A Paul y Evelyn que se han convertido en hermanos.

A Camila por enseñarme que el carácter siempre viene detrás de un corazón noble, este esfuerzo es gracias a ti por acompañarme desde el inicio.

Muchas gracias a todos de corazón.

MJ

Dedicatoria

A mis padres, Mario y Miriam, por su apoyo y acompañamiento durante este proceso, y por estar presentes en cada etapa de mi formación.

A mi compañero, hermano y amigo David, por su compañía, confianza y por brindarme siempre su apoyo en cada momento importante de este camino.

A Anita Chorlango y Luis Collaguaso, por abrirme las puertas de su vida, brindarme sus consejos y acompañarme con su apoyo constante, convirtiéndose en una parte especial de mi camino.

A Jonathan, por su amistad sincera, por su respaldo y por ser ese hermano que la vida me permitió encontrar.

Y a mis leales amigos de cuatro patas: Gordi, Nico, Perlita, Mati, Cookie, Dino y Max, por su compañía, alegría y cariño, que hicieron este recorrido más especial.

Agradecimientos

Expreso mi más sincero agradecimiento a Dios, por brindarme la fortaleza, sabiduría y perseverancia necesarias para culminar esta importante etapa de mi formación académica.

A mi familia y a las personas que me acompañaron durante este camino, por su apoyo, confianza y motivación, elementos fundamentales para alcanzar esta meta profesional. A mis docentes y a la universidad, por sus enseñanzas, orientación y por compartir sus conocimientos y experiencias, contribuyendo a mi crecimiento académico y profesional durante el desarrollo de esta maestría. Asimismo, agradezco al Programa de Becas Guillermo Lasso por el apoyo otorgado y por brindarme la oportunidad de acceder a esta beca, la cual hizo posible cursar y culminar esta maestría, representando un aporte fundamental para mi formación y desarrollo profesional.

Finalmente, expreso mi gratitud a todas las personas que, de una u otra manera, fueron parte de este proceso y contribuyeron a la culminación de este trabajo y al logro de esta meta.

Con toda mi gratitud.

Mayu Muso

Dedicatoria

A mis padres, Lorena Guayanay y Camilo Vicente, quienes han sido mi pilar fundamental. Gracias por su amor incondicional, sus sacrificios y por enseñarme que con esfuerzo todo es posible; este logro es de ustedes.

En memoria de mi abuelo, José Angel Guayanay, cuyo legado y ejemplo de vida me acompañan siempre. Sé que me cuidas desde el cielo y espero que, donde estés, te sientas orgulloso de este paso que hoy doy.

Agradecimientos

Gracias a Dios por darme la fortaleza, la salud y la guía para culminar con éxito esta etapa de mi vida.

A toda mi familia y demás familiares, por estar presentes en cada momento, por sus palabras de aliento y por ser un apoyo constante e incondicional en este camino.

A mis docentes, por compartir su conocimiento con paciencia y dedicación. Su guía y profesionalismo no solo formaron mi camino académico, sino también mi visión profesional.

Muchas gracias a todos de corazón.

Jimmy Vicente

Índice de contenido

Acuerdo de confidencialidad	iv
Índice de contenido	ix
Lista de abreviaturas y símbolos.....	xii
Resumen.....	xv
Abstract.....	xvi
1. Introducción	1
1.1 Contexto y motivación.....	2
1.2 La medición del riesgo en mercados emergentes	3
1.3 Planteamiento del problema.....	5
1.4 Justificación	5
1.5 Hipótesis de investigación	6
1.6 Objetivos	7
1.7 Alcance y delimitación	7
1.8 Estructura del documento	8
2. Marco Teórico.....	9
2.1 Perspectiva histórica de la medición del riesgo	9
2.2 Retornos financieros y hechos estilizados	11
2.3 Valor en Riesgo (VaR) y Expected Shortfall.....	13
2.4 VaR Histórico: el estimador no paramétrico	14
2.5 Modelos GARCH	16
2.6 Redes LSTM y regresión cuantílica.....	18
2.7 Backtesting: las pruebas que juzgan los modelos	20
2.8 Estado del arte: GARCH frente a aprendizaje profundo	23
3. Metodología	26
3.1 Datos y fuente	26
3.2 Partición temporal.....	28

3.3 Los cinco modelos	29
3.4 Métricas y criterio de selección	31
3.5 Operacionalización de las variables	34
4. Resultados	35
4.1 Análisis exploratorio	35
4.2 GARCH (1,1)-t: parámetros	39
4.3 Arquitecturas LSTM	41
4.4 Backtesting comparativo — 5 modelos	43
4.5 Expected Shortfall	47
4.6 Test de Diebold-Mariano	49
4.7 Estabilidad temporal y violaciones por año	51
4.8 Análisis de sensibilidad a la ventana de calibración	53
5. Discusión	56
5.1 El hilo argumental: ¿por qué se necesitaron cinco modelos?	56
5.2 En qué es mejor cada modelo	58
5.3 Por qué el HistVaR resulta superior: la explicación de fondo	59
5.4 Sensibilidad metodológica de la ventana de calibración	59
5.5 Contraste con la literatura	60
5.6 Contraste de las hipótesis	61
5.7 Implicaciones prácticas y regulatorias	61
6. Conclusiones y Recomendaciones	64
6.1 Veredicto: cuál es el mejor y en qué sentido	64
6.2 Conclusiones por objetivo	65
6.3 Hallazgos transversales	66
6.4 Recomendaciones	66
6.5 Limitaciones	67
6.6 Líneas de investigación futura	68

Referencias.....	70
Anexos	74
Anexo A. Entorno de reproducibilidad.....	74
Anexo B. Glosario de términos.....	75
Anexo C. Índice de figuras y tablas	76
Anexo D. Resultados Extendidos de Optimización y Backtesting.....	77

Lista de abreviaturas y símbolos

- ACF: Función de autocorrelación (autocorrelation function)
- ADF: Prueba aumentada de Dickey-Fuller para raíz unitaria (augmented Dickey-Fuller)
- AFP: Administradora de Fondos de Pensiones
- AIC: Criterio de información de Akaike (Akaike information criterion)
- ARCH: Heterocedasticidad condicional autorregresiva (autoregressive conditional heteroskedasticity)
- BCBS: Comité de Supervisión Bancaria de Basilea (Basel Committee on Banking Supervision)
- BCE: Banco Central del Ecuador
- BIC: Criterio de información bayesiano (Bayesian information criterion)
- BPTT: Retropropagación a través del tiempo (backpropagation through time)
- BVG: Bolsa de Valores de Guayaquil
- BVQ: Bolsa de Valores de Quito
- BVQA: Índice Ecuador General Adj (BVQ Adjusted)
- CC: Cobertura condicional (prueba de)
- CVaR: Valor en Riesgo condicional (conditional value at risk)
- DQ: Cuantil dinámico (dynamic quantile, prueba de)
- EBITDA: Beneficio antes de intereses, impuestos, depreciación y amortización (earnings before interest, taxes, depreciation and amortization)
- EGARCH: GARCH exponencial (exponential GARCH)
- EMBI: Índice de bonos de mercados emergentes (emerging markets bond index)
- ES: Pérdida esperada en la cola (expected shortfall)
- EVT: Teoría de valores extremos (extreme value theory)

FRTB: Revisión fundamental de la cartera de negociación (fundamental review of the trading book)

GARCH: Heterocedasticidad condicional autorregresiva generalizada (generalized autoregressive conditional heteroskedasticity)

GED: Distribución de error generalizada (generalized error distribution)

GJR: Modelo de Glosten, Jagannathan y Runkle

GRU: Unidad recurrente con compuertas (gated recurrent unit)

H_0 / H_1 : Hipótesis nula / hipótesis alternativa

HS: Simulación histórica (historical simulation)

IND: Independencia (prueba de)

JB: Prueba de Jarque-Bera

JPRM: Junta de Política y Regulación Monetaria

LB: Prueba de Ljung-Box

LL: Log-verosimilitud (log-likelihood)

LR: Razón de verosimilitud (likelihood ratio)

LSTM: Red de memoria de corto y largo plazo (long short-term memory)

MAE: Error absoluto medio (mean absolute error)

MC: Simulación de Monte Carlo

MILA: Mercado Integrado Latinoamericano

MLE: Estimación por máxima verosimilitud (maximum likelihood estimation)

MSE: Error cuadrático medio (mean squared error)

PACF: Función de autocorrelación parcial (partial autocorrelation function)

PIB: Producto Interno Bruto

POF: Proporción de fallos, prueba de Kupiec (proportion of failures)

RILD: Reservas Internacionales de Libre Disponibilidad

RNN: Red neuronal recurrente (recurrent neural network)

SBS: Superintendencia de Bancos y Seguros

SHAP: Explicaciones aditivas de Shapley (SHapley Additive exPlanations)

TIR: Tasa interna de retorno

VaR: Valor en Riesgo (value at risk)

α : Nivel de significancia / parámetro ARCH

β : Parámetro GARCH de persistencia

ε : Innovación o residuo

σ : Desviación estándar / volatilidad

σ^2 : Varianza condicional

ν : Grados de libertad (distribución t de Student)

ω : Constante GARCH

Resumen

En esta tesis se puso a prueba qué tan bien funcionan distintos métodos para estimar el Valor en Riesgo (VaR) en un mercado tan particular como el ecuatoriano. Se trabajó con el índice Ecuador General Adj (BVQA) y con 2.888 precios diarios reales, desde 2012 hasta 2024, evaluando cinco alternativas a los niveles de confianza del 95 % y el 99 %: un GARCH (1,1) con distribución t de Student, dos redes LSTM de regresión cuantílica (una estática y otra con reentrenamiento walk-forward y variables de volatilidad, la LSTM-WF), un modelo híbrido que combina GARCH con LSTM, y como referencia no paramétrica, el VaR Histórico calculado sobre una ventana móvil de 500 observaciones. Para comparar los modelos se aplicó un backtesting dinámico con las pruebas de Kupiec y Christoffersen, el test de Diebold-Mariano con varianza HAC y el Expected Shortfall empírico; todo el análisis se implementó en Python y es reproducible. Contra lo que la literatura de mercados desarrollados haría esperar, el método más simple fue el que mejor funcionó: el VaR Histórico aprobó las dos pruebas de cobertura en ambos niveles de confianza y resultó significativamente más preciso que el GARCH-t en el test de Diebold-Mariano ($p = 0,028$ al 95 % y $p = 0,034$ al 99 %), quedando la LSTM-WF como la segunda opción válida. La explicación está en la propia naturaleza del BVQA: su baja liquidez produce una dinámica IGARCH ($\alpha + \beta = 1$), efectos ARCH que no llegan a ser significativos ($p = 0,395$) y un 20,8 % de retornos exactamente nulos, condiciones bajo las cuales los modelos de volatilidad condicional pierden su ventaja.

Palabras clave: Valor en Riesgo; GARCH; LSTM; backtesting; VaR Histórico; Diebold-Mariano; mercados emergentes; BVQA

Abstract

This thesis examined how well different methods estimate Value at Risk (VaR) in a market as particular as Ecuador's. Using the Ecuador General Adj index (BVQA) and 2,888 real daily prices from 2012 to 2024, five alternatives were evaluated at the 95 % and 99 % confidence levels: a GARCH (1,1) with Student's t distribution, two quantile-regression LSTM networks (one static and one with walk-forward retraining and volatility features, the LSTM-WF), a hybrid model combining GARCH with LSTM, and, as the non-parametric benchmark, Historical VaR computed over a 500-observation rolling window. Model comparison relied on dynamic backtesting with the Kupiec and Christoffersen tests, the Diebold-Mariano test with HAC variance, and the empirical Expected Shortfall; the whole analysis was implemented in Python and is fully reproducible. Contrary to what the developed-markets literature would suggest, the simplest method performed best: Historical VaR passed both coverage tests at both confidence levels and proved significantly more accurate than the GARCH-t under the Diebold-Mariano test ($p = 0.028$ at 95 % and $p = 0.034$ at 99 %), with the LSTM-WF standing as the second valid option. The explanation lies in the very nature of the BVQA: its low liquidity produces IGARCH dynamics ($\alpha + \beta = 1$), ARCH effects that fail to reach significance ($p = 0.395$), and 20.8 % of exactly zero returns - conditions under which conditional-volatility models lose their edge.

Keywords: Value at Risk; GARCH; LSTM; backtesting; Historical VaR; Diebold-Mariano; emerging markets; BVQA

1. Introducción

Cuando alguien invierte, presta o simplemente guarda su dinero en un activo, está asumiendo que las cosas pueden salir mal. El precio puede bajar, el deudor puede no pagar. Las pérdidas van a llegar tarde o temprano; eso nadie lo discute. Lo que sí se puede discutir, y lo que le interesa a quien gestiona dinero, es otra cosa: qué tan grandes serán esas pérdidas y cada cuánto aparecen. Esta tesis trata sobre cómo medir eso en un mercado donde casi nadie lo ha intentado.

Para poner un número a esa incertidumbre, la industria financiera adoptó hace tres décadas el Valor en Riesgo, o VaR. Su lectura es simple: si el VaR diario al 95 % es de 2 %, quiere decir que en 95 de cada 100 días la pérdida no debería pasar de ese 2 %. Puesto en términos formales, si r_{t+1} es el retorno logarítmico del día siguiente, el VaR al nivel α (por ejemplo, $\alpha = 0,05$) es el umbral que la pérdida solo cruza con probabilidad α (Jorion, 2007):

$$Pr(-r_{t+1} > VaR_t(\alpha)) = \alpha \quad (1.1)$$

Visto así, el VaR no es otra cosa que un cuantil de la distribución de los retornos, con el signo dado vuelta. Suena sencillo, pero hay un detalle: nadie conoce esa distribución. Toca estimarla, y cada supuesto distinto sobre su forma -normal o de colas pesadas, fija o cambiante en el tiempo- produce un VaR diferente. Esa es la razón de que convivan tantos modelos rivales, y de que compararlos seriamente tenga sentido.

Ahora bien, ¿qué modelos comparar? En los mercados grandes la respuesta parecería estar dada: los modelos GARCH llevan décadas describiendo bien cómo se mueve la volatilidad (Engle, 1982; Bollerslev, 1986). El asunto es que el mercado ecuatoriano no se parece en nada a esos mercados. El índice Ecuador General Adj (BVQA) casi no se negocia algunos días, acumula sesiones enteras sin cambio de precio y sus retornos tienen colas pesadísimas. Frente a eso, las redes LSTM prometen capturar relaciones no lineales que a los

modelos clásicos se les escapan (Hochreiter & Schmidhuber, 1997; Fischer & Krauss, 2018). Este trabajo enfrenta las dos escuelas -la econometría de siempre y el aprendizaje profundo- y les pone al lado un rival humilde, el VaR Histórico, para que sirva de punto de comparación sin favoritismos.

1.1 Contexto y motivación

Si algo enseña la historia financiera reciente es que medir mal el riesgo tiene un costo enorme. Barings quebró en 1995. Long-Term Capital Management, un fondo manejado por premios Nobel, se hundió en 1998. Y la crisis de 2007-2009 arrastró al sistema entero. En los tres casos el diagnóstico se repite: los modelos le pusieron una probabilidad demasiado baja a lo extremo. Cada colapso obligó a reguladores y bancos a endurecer sus métricas, y en ese proceso el VaR -que J.P. Morgan había popularizado en 1994 con RiskMetrics (J.P. Morgan & Reuters, 1996)- se volvió el estándar para cuantificar riesgo de mercado. No es solo una convención técnica: los Acuerdos de Basilea castigan con multiplicadores de capital a las instituciones cuyos modelos acumulan demasiadas excedencias, clasificándolas en zonas verde, amarilla y roja.

Más allá del plano regulatorio, el VaR cumple funciones operativas cotidianas dentro de las instituciones. Sirve para fijar límites de exposición por mesa y por operador, para asignar capital entre líneas de negocio en proporción al riesgo que consumen y para condensar, en una sola cifra diaria, la exposición agregada que la alta dirección debe conocer. Esa versatilidad explica su longevidad: aun después de que la regulación desplazara su papel formal hacia el Expected Shortfall, el VaR sigue siendo el lenguaje común en el que las áreas de riesgo, las mesas de negociación y los directorios se comunican (Jorion, 2007).

Detrás de esta tesis hay una discusión que sigue abierta. De un lado, cuatro décadas de econometría financiera produjeron los modelos ARCH y GARCH, que capturan bien el agrupamiento de la volatilidad en mercados líquidos (Engle, 1982; Bollerslev, 1986). Del otro,

en los últimos diez años el aprendizaje profundo -y las redes LSTM en particular- empezó a disputarles el terreno, con la promesa de aprender de los datos relaciones no lineales que ninguna fórmula paramétrica contempla (Fischer & Krauss, 2018). ¿Quién gana? Nadie lo ha zanjado, y todo apunta a que la respuesta cambia según el mercado del que se trate.

Y ese es justamente el hueco que esta tesis busca cubrir. La evidencia publicada viene casi toda del S&P 500 y de los grandes índices de Europa y Asia: mercados profundos, líquidos, con miles de operaciones diarias. ¿Qué pasa con esos mismos modelos en una bolsa pequeña, emergente, donde hay días en que casi nada se negocia? Se sabe muy poco. Asumir que las conclusiones viajan sin más de un contexto al otro sería un error. Por eso se armó un estudio riguroso, reproducible y construido íntegramente sobre datos reales del índice Ecuador General Adj (BVQA).

1.2 La medición del riesgo en mercados emergentes

Los mercados emergentes juegan con otras reglas. Aggarwal et al. (1999) mostraron que en ellos la volatilidad sufre quiebres estructurales bruscos, casi siempre atados a la política o a la macroeconomía local antes que a la dinámica financiera pura. En el caso ecuatoriano se agrega un factor que resulta decisivo: la iliquidez. Si un activo pasa días sin negociarse, su precio de cierre se repite y el retorno de esa sesión es exactamente cero. En el BVQA esto no es anecdótico: como se documenta en el Capítulo 4, pasa en más de uno de cada cinco días.

Conviene describir brevemente el escenario institucional en que se inscribe el índice estudiado. El mercado bursátil ecuatoriano opera a través de dos plazas, la Bolsa de Valores de Quito (BVQ) y la Bolsa de Valores de Guayaquil (BVG), en una economía dolarizada desde el año 2000. La dolarización eliminó el riesgo cambiario doméstico, pero también privó a la economía de un prestamista de última instancia en moneda propia, lo que vuelve especialmente sensible la gestión prudencial del riesgo. La supervisión del sistema recae en la Superintendencia de Bancos y Seguros (SBS) y en la Junta de Política y Regulación Monetaria

(JPRM), organismos que definen los requerimientos de capital de las entidades financieras locales.

En ese entorno, la renta variable ocupa un lugar marginal frente a la renta fija: buena parte de las emisiones locales corresponde a obligaciones y papel comercial, y las acciones se negocian con una frecuencia muy inferior a la de cualquier mercado desarrollado. El resultado es un índice accionario cuya serie de precios arrastra los rasgos de esa escasez de negociación. Ese contexto no es un detalle de color, sino la causa estructural de los fenómenos estadísticos que este trabajo documenta, empezando por la elevada proporción de retornos nulos.

Ese detalle, que parece menor, tiene consecuencias de fondo. Los modelos GARCH existen porque los choques grandes tienden a venir en racha -el famoso agrupamiento de volatilidad-. Pero cuando una quinta parte de las observaciones vale cero por pura falta de negociación, esa racha se corta una y otra vez, y la señal de volatilidad condicional puede diluirse hasta desaparecer de las pruebas estadísticas. Cont (2001) ya lo había insinuado en su catálogo de «hechos estilizados»: la fuerza de estos fenómenos depende de la liquidez. La apuesta de partida de este trabajo es, precisamente, esa: que la iliquidez del BVQA le quita el piso a los modelos de volatilidad condicional y le da la ventaja a los métodos que no suponen nada.

De confirmarse esa hipótesis, la implicación práctica sería considerable: las entidades locales podrían medir su riesgo de mercado con herramientas sencillas y auditables, sin depender de infraestructura sofisticada ni de personal altamente especializado. De rechazarse, en cambio, quedaría demostrado que la complejidad adicional se paga sola. En cualquiera de los dos escenarios, la respuesta tiene valor directo para quien debe decidir con qué modelo cumplir sus obligaciones regulatorias.

1.3 Planteamiento del problema

¿Por qué importa acertar con el modelo? Porque el error cuesta en las dos direcciones. Un VaR que se queda corto deja pasar más pérdidas de las anunciadas: la entidad se descapitaliza y el regulador la sanciona. Un VaR inflado hace lo contrario: congela capital que podría estar trabajando y come rentabilidad sin necesidad. Lo que se busca es un modelo que prometa un 5 % de excedencias y entregue un 5 % de excedencias, un año tras otro.

Con ese marco, la pregunta que guía la investigación queda así: de los cinco modelos evaluados -GARCH (1,1)-t, LSTM base, LSTM con walk-forward y variables de volatilidad (LSTM-WF), Híbrido GARCH-LSTM y VaR Histórico-, ¿cuál estima con mayor solidez estadística el VaR del índice BVQA entre 2012 y 2024? La respuesta, vale decirlo, no era obvia: si uno le cree a la teoría de mercados líquidos, deberían ganar los modelos de volatilidad condicional; si uno le cree a los datos del BVQA, el orden podría darse vuelta.

1.4 Justificación

¿Qué aporta este trabajo? Varias cosas a la vez. Suma evidencia a una literatura latinoamericana que sigue siendo escasa en gestión de riesgo. Les da a los gestores e instituciones locales un estudio comparativo que hasta ahora no existía para este índice. Pone a competir econometría clásica y deep learning bajo un solo protocolo de backtesting, lo cual tiene valor metodológico propio. Y es la primera vez que el aprendizaje profundo se aplica de manera sistemática al análisis cuantitativo de la bolsa ecuatoriana.

Hay también una lectura de política prudencial. El modelo de VaR que un supervisor acepte determina cuánto capital deben apartar las entidades. Si el modelo peca de optimista, el sistema queda expuesto; si peca de conservador, la intermediación se encarece sin motivo. En una economía donde conseguir capital ya es difícil, equivocarse en cualquiera de los dos sentidos tiene un costo real.

Los usuarios potenciales de estos resultados no se agotan en la banca. Las administradoras de fondos, las aseguradoras y las casas de valores que mantienen posiciones en renta variable local enfrentan el mismo problema de medición, con el agravante de que la literatura disponible no les ofrece guía alguna calibrada sobre datos ecuatorianos. Un estudio comparativo con protocolo explícito y código reproducible les entrega un punto de partida verificable: pueden replicar cada cifra, auditar cada supuesto y adaptar la metodología a sus propias carteras.

Existe además una dimensión de gobernanza de modelos. Los marcos modernos de administración de riesgo exigen que cada modelo en producción tenga documentados su fundamento teórico, sus supuestos, sus límites de validez y su desempeño fuera de muestra. La estructura de esta tesis -especificación formal, backtesting con pruebas de hipótesis, análisis de sensibilidad- sigue de cerca ese estándar, de modo que sus resultados pueden incorporarse casi directamente a un expediente de validación interna.

1.5 Hipótesis de investigación

Tres hipótesis ordenan el trabajo. Las tres son contrastables y las tres se someten a los datos en los capítulos de resultados y discusión:

- **H1.** La baja liquidez del BVQA y su alta proporción de retornos nulos debilitan la heterocedasticidad condicional, al punto de que deja de ser estadísticamente significativa.
- **H2.** Si eso ocurre, los modelos construidos sobre volatilidad condicional -el GARCH y todo lo que lo incorpore- no deberían poder ganarle a un estimador tan elemental como el VaR Histórico.
- **H3.** Para que una LSTM compita en serio necesita dos cosas: reentrenarse periódicamente (walk-forward) y recibir variables de volatilidad como entrada. La arquitectura sola, sin eso, no basta.

1.6 Objetivos

Objetivo general: establecer, mediante backtesting dinámico y pruebas formales de cobertura, cuál de los cinco modelos evaluados estima el VaR del índice BVQA (2012-2024) con mayor solidez estadística y mejor capacidad predictiva.

Objetivos específicos:

- **OE1.** Estimar el GARCH (1,1)-t, revisar los diagnósticos de sus residuos y calcular el VaR que produce.
- **OE2.** Construir la LSTM base y la LSTM-WF, y medir cuánto cambia el resultado al agregar walk-forward y variables de volatilidad.
- **OE3.** Armar el Híbrido GARCH-LSTM y ver si la combinación aporta algo.
- **OE4.** Calcular el VaR Histórico, que actúa de benchmark no paramétrico.
- **OE5.** Someter los cinco modelos a las pruebas de Kupiec, Christoffersen y Diebold-Mariano, y a la evaluación regulatoria: Expected Shortfall, Función de Pérdida de López y zonas de Basilea.

1.7 Alcance y delimitación

El estudio se acota al riesgo de mercado -es decir, al riesgo de precio- del índice BVQA, con datos diarios de octubre de 2012 a noviembre de 2024 y retornos logarítmicos. Quedan fuera la modelación multivariante, la comparación con otros mercados y el riesgo de crédito u operacional. El diseño es cuantitativo, comparativo y univariado; el procesamiento completo corre en Python y se puede repetir. El pronóstico es a un día ($h=1$), el horizonte estándar del VaR de negociación, y se trabaja al 95 % y al 99 %, los dos niveles que dominan la práctica y la regulación.

La decisión de trabajar a nivel de índice y en forma univariada merece una palabra adicional. Antes de plantearse modelos multivariados -que exigen estimar estructuras de dependencia entre activos y multiplican los parámetros- es indispensable establecer qué familia

de modelos describe adecuadamente la dinámica marginal del mercado agregado. Si los modelos de volatilidad condicional fallan sobre el índice, difícilmente van a funcionar sobre sus componentes, cuya liquidez individual es aún menor. El análisis univariado no es, por tanto, una simplificación arbitraria, sino el primer paso lógico de un programa de investigación que puede extenderse después hacia carteras.

1.8 Estructura del documento

Sobre la organización: el Capítulo 2 arma la base teórica -historia de la medición del riesgo, hechos estilizados, matemática de los cinco modelos y de las pruebas de backtesting-. El Capítulo 3 cuenta de dónde salen los datos, cómo se limpiaron, cómo se partió la muestra y cómo se especificó cada modelo. El Capítulo 4 trae el análisis exploratorio y el backtesting completo. El Capítulo 5 discute por qué salió lo que salió, contra el telón de fondo de la literatura. El Capítulo 6 cierra: hallazgos, respuesta a las hipótesis, recomendaciones y lo que queda pendiente.

Conviene también dejar explícito el posicionamiento metodológico general. Este trabajo se ubica en el paradigma cuantitativo, con un diseño comparativo no experimental de corte longitudinal: no se manipulan variables, sino que se observa el desempeño de cinco especificaciones sobre una misma realidad histórica, bajo condiciones idénticas de información y evaluación. La unidad de análisis es el pronóstico diario de riesgo, y el criterio de verdad es estrictamente empírico: las pruebas formales de cobertura y de precisión predictiva descritas en el marco teórico. Toda afirmación de superioridad entre modelos que aparezca en las conclusiones remite, en última instancia, a esos contrastes.

2. Marco Teórico

Antes de comparar modelos hay que construir el terreno donde se van a comparar. Eso hace este capítulo. Primero un poco de historia, para entender de dónde viene el problema; luego los hechos estilizados de los retornos; después las definiciones precisas de VaR y Expected Shortfall; a continuación los cinco modelos, del más simple al más complejo; y al final las pruebas estadísticas que harán de árbitro. En cada tema se procura dar primero la idea y recién después la fórmula.

2.1 Perspectiva histórica de la medición del riesgo

Modelar el azar de los precios con matemáticas no es un invento reciente. Louis Bachelier lo hizo en 1900, en su tesis *Théorie de la spéculation*, donde trató los precios bursátiles como un movimiento browniano -cinco años antes de que Einstein usara la misma herramienta en física (Bachelier, 1900)-. Nadie le prestó atención durante medio siglo. Hoy se lo reconoce como el punto de partida de las finanzas cuantitativas.

El olvido de la obra de Bachelier ilustra un patrón recurrente en la historia de la disciplina: las herramientas matemáticas suelen llegar antes de que exista la demanda institucional capaz de aprovecharlas. Recién cuando los mercados de capitales ganaron profundidad tras la posguerra, y cuando la computación hizo tratables los cálculos, la formalización del riesgo pasó de curiosidad académica a necesidad operativa. Ese desfase entre teoría disponible y práctica adoptada reaparecerá varias veces en esta cronología, incluida la más reciente: la del aprendizaje profundo aplicado a las finanzas.

Lo que vino a mediados del siglo XX cambió la disciplina. Markowitz (1952) demostró que el riesgo de una cartera no se obtiene sumando los riesgos de cada activo: dependen las covarianzas, y la varianza pasó a ser la medida de riesgo por excelencia. Sharpe (1964) levantó sobre eso el CAPM. Black y Scholes (1973) hicieron lo propio con la valoración de opciones.

En todos esos avances la volatilidad era protagonista, pero con una simplificación que hoy llama la atención: se la suponía constante.

La realidad se encargó de desmentir ese supuesto. Mandelbrot (1963), estudiando los precios del algodón, encontró colas mucho más gordas que las de una normal y movimientos grandes que llegaban en tandas. La econometría tardó dos décadas en darle forma a esa observación: Engle (1982) propuso los modelos ARCH, donde la varianza depende de los choques recientes -aporte que le valió el Nobel de Economía en 2003-, y Bollerslev (1986) los generalizó en el GARCH, que sigue siendo el estándar. Engle y Bollerslev (1986) describieron además el caso de persistencia integrada, el IGARCH, que en esta tesis va a tener un papel protagónico.

Al VaR le llegó su momento en los noventa. J.P. Morgan liberó su sistema RiskMetrics en 1994 y con eso el cálculo del VaR dejó de ser exclusivo de los grandes bancos (J.P. Morgan & Reuters, 1996). La regulación fue absorbiendo estas métricas por etapas: Basilea I (1988) definió las bases del capital regulatorio, la Enmienda de 1996 metió formalmente el riesgo de mercado y el VaR, Basilea II (2004) afinó el esquema. La crisis de 2008 forzó el último giro: Basilea III y el FRTB (2019) retiraron el VaR como medida regulatoria y pusieron en su lugar el Expected Shortfall al 97,5 %, admitiendo que el VaR se queda corto frente a los eventos extremos. La Tabla 2.1 ordena esta cronología.

Tabla 2.1

Hitos en la historia de la medición del riesgo de mercado.

Año	Hito	Aporte a la medición del riesgo
1900	Bachelier	Primer modelo estocástico de precios (browniano).
1952	Markowitz	Teoría de cartera; la varianza como medida de riesgo.
1963	Mandelbrot	Colas pesadas y agrupamiento de volatilidad.
1982	Engle (ARCH)	Varianza condicional dependiente de choques pasados.

Año	Hito	Aporte a la medición del riesgo
1986	Bollerslev (GARCH)	Generalización parsimoniosa del ARCH.
1994	RiskMetrics (J.P. Morgan)	Popularización del VaR.
1996	Basilea — Enmienda	El VaR entra en la regulación de capital.
1997	Hochreiter & Schmidhuber	Red LSTM: memoria de largo plazo.
2019	Basilea III (FRTB)	El Expected Shortfall sustituye al VaR.

Nota. Elaboración propia a partir de la literatura citada.

Mientras tanto, el aprendizaje automático avanzaba por su propia vereda. La retropropagación (Rumelhart et al., 1986) volvió entrenables a las redes neuronales, aunque las recurrentes simples tenían un defecto serio: no retenían información lejana. La LSTM (Hochreiter & Schmidhuber, 1997) arregló eso, y cuando el deep learning despegó en los años 2010 (LeCun et al., 2015), las series financieras fueron uno de sus destinos naturales. Dos tradiciones que crecieron separadas -econometría y aprendizaje profundo- terminan encontrándose en esta tesis, cara a cara.

2.2 Retornos financieros y hechos estilizados

Conviene detenerse un momento en qué es lo que se va a modelar. La investigación empírica de las últimas décadas encontró regularidades que aparecen en casi cualquier mercado; se las conoce como los «hechos estilizados» de los retornos financieros (Cont, 2001). La Tabla 2.2 los lista y adelanta cuáles se cumplen en el BVQA -el detalle está en el Capítulo 4-.

Tres de esos hechos merecen glosa aparte por su relevancia directa para esta tesis. El primero es la ausencia de autocorrelación significativa en los retornos: los precios no se dejan predecir linealmente, lo que es coherente con la eficiencia de mercado en su forma débil. El segundo son las colas pesadas: los eventos extremos ocurren con una frecuencia muy superior a la que asigna la distribución normal, de modo que cualquier modelo construido sobre

normalidad subestima sistemáticamente el riesgo de pérdidas grandes. El tercero es el agrupamiento de volatilidad: aunque los retornos no se correlacionan, sus magnitudes sí, y los períodos agitados tienden a prolongarse (Cont, 2001).

La utilidad de este catálogo es doble. Por un lado funciona como lista de verificación empírica: antes de elegir modelo, corresponde comprobar cuáles de estas regularidades exhiben los datos concretos. Por otro, funciona como criterio de diseño: cada modelo de la comparación puede leerse como una apuesta sobre qué hechos estilizados importan -el GARCH apuesta todo al agrupamiento de volatilidad; la simulación histórica, a la estabilidad de la distribución empírica; la LSTM, a la existencia de patrones no lineales aprendibles-. El desempeño relativo de cada uno revela, en el fondo, cuáles de esas apuestas se cumplen en el mercado ecuatoriano.

Tabla 2.2

Hechos estilizados de los retornos y su presencia en el BVQA.

Hecho estilizado	Descripción	¿En el BVQA?
Ausencia de autocorrelación	Los retornos son casi impredecibles en media.	Se cumple
Colas pesadas	Eventos extremos más frecuentes que en la normal.	Muy marcado (curtosis ≈ 96)
Agrupamiento de volatilidad	Días volátiles se suceden entre sí.	Débil / ausente (ARCH-LMn. s.)
Asimetría	Distribución no simétrica de ganancias y pérdidas.	Presente (asimetría $\approx 3,8$)
No normalidad	Los retornos no siguen una distribución normal.	Rechazada (Jarque-Bera, Shapiro)
Efecto de iliquidez	Retornos nulos por ausencia de negociación.	Severo (20,8 % de ceros)

Nota. Elaboración propia a partir de la literatura citada.

De esa tabla hay una fila que importa más que todas: la del agrupamiento de volatilidad. En los mercados desarrollados ese fenómeno es fuerte, y es la razón de ser de los modelos GARCH. En el BVQA aparece apenas, y el culpable es el efecto de iliquidez que registra la última fila. Todo el resultado central de esta tesis se explica, al final, por esa tensión entre lo que la teoría espera y lo que este mercado entrega.

2.3 Valor en Riesgo (VaR) y Expected Shortfall

Sea $\{r_t\}$ la serie de retornos logarítmicos, $r_t = \ln(P_t/P_{t-1})$, donde P_t es el precio de cierre del día t . Condicional a la información disponible hasta t , denotada F_t , el VaR al nivel de significancia $\alpha \in (0,1)$ es el cuantil que deja una probabilidad α en la cola izquierda de las pérdidas. Equivalentemente:

$$VaR_t(\alpha) = -\inf\{x \in \mathbb{R}: Pr(r_{t+1} \leq x|F_t) \geq \alpha\} \quad (2.1)$$

El signo negativo está solo para que el riesgo se lea en positivo: el cuantil es un número negativo -es una pérdida- y al invertirlo queda una cifra de riesgo manejable. Ahora, el VaR arrastra una limitación conocida: no dice nada de lo que pasa después de cruzado el umbral. Para el VaR, pasarse por poco o pasarse por una catástrofe es lo mismo. El Expected Shortfall (ES) nace para remediar eso: en lugar de mirar solo el umbral, promedia las pérdidas de los días en que el umbral se superó (Embrechts et al., 1997):

$$ES_t(\alpha) = E[-r_{t+1} | -r_{t+1} > VaR_t(\alpha), F_t] = \frac{1}{\alpha} \int_0^\alpha VaR_t(u) du \quad (2.2)$$

Por construcción, el ES nunca está por debajo del VaR. Y tiene una credencial teórica que el VaR no puede exhibir: es una medida de riesgo coherente, porque cumple la subaditividad que el VaR llega a violar (Artzner et al., 1999, p. 209). Ese fundamento pesó en la decisión de Basilea III de adoptarlo, al nivel del 97,5 %, como medida regulatoria de capital (BCBS, 2019). En cuanto a las maneras de estimar el VaR, la tradición distingue tres familias. Los métodos paramétricos suponen una forma para la distribución de los retornos -normal, t de Student- y estiman sus parámetros; ahí se ubica el GARCH- t de esta tesis. La simulación histórica no supone distribución alguna y lee el cuantil directamente del pasado; es el terreno del VaR Histórico. Y la simulación de Monte Carlo genera miles de escenarios a partir de un modelo estocástico postulado. Las LSTM agregan un cuarto enfoque, más reciente, de corte no paramétrico y basado en aprendizaje automático. Esta tesis confronta representantes de todas estas familias bajo un mismo protocolo de evaluación.

La sustitución del VaR por el ES en el estándar regulatorio no significa que el VaR haya quedado obsoleto. En la práctica, ambos se calculan y reportan juntos: el VaR porque es el umbral sobre el que se define la excedencia -y, por tanto, la unidad básica del backtesting-, y el ES porque describe la severidad esperada una vez cruzado ese umbral. Las pruebas estadísticas disponibles para validar modelos, además, están mucho más desarrolladas para el VaR que para el ES, razón por la cual el backtesting de esta tesis se articula alrededor del primero y usa el segundo como métrica complementaria de prudencia.

La elección de los niveles de confianza tampoco es arbitraria. El 95 % es el nivel de trabajo habitual para la gestión interna: produce excedencias con la frecuencia suficiente para que las pruebas estadísticas tengan potencia razonable en muestras de tamaño moderado. El 99 %, en cambio, es el nivel que exige el marco de Basilea para el cálculo de capital; sus excedencias son escasas por construcción, lo que obliga a interpretarlo con muestras largas o con cautela adicional. Evaluar ambos niveles a la vez permite verificar que un modelo no esté bien calibrado en el centro de la cola pero mal en su extremo, un patrón más común de lo que suele suponerse.

Resta justificar el horizonte de pronóstico. El día siguiente ($h=1$) es el horizonte canónico del VaR de negociación por una razón de coherencia informativa: es el único plazo en que el supuesto de cartera constante resulta creíble y en que el backtesting puede ejecutarse con una observación nueva por sesión. Horizontes más largos, como los diez días que la regulación emplea para el capital, se obtienen en la práctica reescalando el VaR diario, de modo que la calidad del pronóstico a un día es, también, el fundamento de los plazos superiores.

2.4 VaR Histórico: el estimador no paramétrico

¿Cuál es la manera más directa de estimar el cuantil de la ecuación (2.1)? No suponer nada y mirar lo que ya pasó. Eso hace la simulación histórica: toma las últimas W observaciones y calcula el cuantil empírico directamente:

$$VaR_t(\alpha) = -\hat{Q}_\alpha(\{r_{t-1}, r_{t-2}, \dots, r_{t-W}\}) \quad (2.3)$$

donde $Q_\alpha(\cdot)$ es el α -ésimo cuantil de la muestra. Al no estimar parámetros, este método tiene una virtud a prueba de todo: no puede especificar mal la distribución, porque no especifica ninguna. Lo que paga a cambio es lentitud -si la volatilidad pega un salto, la ventana tarda en enterarse-. Aun así, su resistencia a la no normalidad lo convierte en el benchmark natural: la vara mínima que cualquier modelo con pretensiones debería superar. Se trabaja con $W = 500$ días.

La longitud de la ventana encierra un compromiso que conviene hacer explícito. Una ventana corta reacciona rápido pero estima el cuantil con pocas observaciones, y en la cola extrema -donde por definición los datos escasean- esa imprecisión se amplifica. Una ventana larga estabiliza la estimación, pero mezcla regímenes de mercado distintos y responde con lentitud cuando las condiciones cambian. Los 500 días adoptados -unos dos años bursátiles- son el punto medio que la práctica profesional ha consolidado, y el Capítulo 4 somete esa elección a un análisis de sensibilidad formal con ventanas de 250 y 750 días.

Vale notar, por último, que la simulación histórica pondera por igual todas las observaciones de la ventana: el retorno de hace dos años pesa lo mismo que el de ayer. Existen refinamientos que asignan pesos decrecientes con la antigüedad, pero introducen parámetros adicionales que hay que calibrar, con lo que se pierde justamente la virtud que hace atractivo al método: su ausencia total de supuestos. En un estudio cuyo propósito es confrontar familias de modelos, mantener el benchmark en su versión más pura evita ambigüedades sobre a qué se le atribuye el desempeño.

No está de más señalar que la simulación histórica es, con distancia, el método más difundido en la banca internacional para el cálculo interno del VaR. Su transparencia explica esa adopción: el resultado se reconstruye a mano a partir de la serie de retornos, cada excedencia se rastrea hasta el dato que la produjo y el auditor no necesita conocimientos de

optimización numérica para verificarlo (Jorion, 2007). Que el candidato más usado por la industria sea también el más simple del estudio le agrega interés práctico a la comparación: no se está evaluando una curiosidad académica, sino el método que probablemente ya usan las entidades locales.

2.5 Modelos GARCH

En los mercados financieros los días tranquilos y los agitados no se reparten al azar: lo agitado llama a lo agitado. A esa tendencia se la llama clúster de volatilidad, y el GARCH (1,1) de Bollerslev (1986) es su formalización más usada. La receta: escribir el retorno como una media más una innovación cuya escala se mueve con el tiempo:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{iid}(0,1) \quad (2.4)$$

y la varianza condicional σ_t^2 evoluciona según una recursión que combina el choque más reciente (ε_{t-1}^2) y su propia inercia (σ_{t-1}^2):

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad \omega > 0, \alpha \geq 0, \beta \geq 0 \quad (2.5)$$

De los dos parámetros clave, α captura cuánto reacciona el proceso a las noticias y β cuánta inercia tiene la volatilidad. Lo que define la memoria del sistema es la suma de ambos. Mientras $\alpha + \beta$ se mantenga por debajo de 1, el proceso es estacionario y existe una varianza de largo plazo, dada por:

$$\text{Var}(\varepsilon_t) = \frac{\omega}{1 - \alpha - \beta} \quad (2.6)$$

¿Y si $\alpha + \beta$ llega exactamente a 1? Ese caso límite tiene nombre propio: IGARCH, o GARCH integrado. El denominador de (2.6) se anula, la varianza incondicional deja de existir y los choques de volatilidad ya no se apagan jamás. Se menciona aquí porque el BVQA -como se verá- cae exactamente en esa situación, y de ahí se desprende buena parte de lo que sigue.

$$\text{VaR}_t(\alpha) = - \left(\mu + \sigma_t \cdot t_v^{-1}(\alpha) \cdot \sqrt{\frac{v-2}{v}} \right) \quad (2.7)$$

aquí $t_v^{-1}(\alpha)$ representa el cuantil α de la distribución t estándar, mientras que $\sqrt{(v-2)/v}$ la lleva a varianza unitaria (esa corrección solo tiene sentido cuando $v > 2$). El Expected Shortfall

paramétrico también tiene expresión cerrada en este marco; basta integrar la cola de la t (Engle & Manganelli, 2004).

¿Cómo se obtienen los cinco parámetros $\theta = (\mu, \omega, \alpha, \beta, \nu)$? Por máxima verosimilitud: de todos los vectores posibles, se queda el que vuelve más creíble la muestra efectivamente observada. Bajo innovaciones t de Student, esa verosimilitud se arma sumando, observación por observación, el logaritmo de la densidad condicional:

$$\ell(\theta) = \sum_{t=1}^T \ln f_{\nu} \left(\frac{\varepsilon_t}{\sigma_t} \right) - \ln \sigma_t \quad (2.8)$$

donde f_{ν} es la densidad de la t estandarizada y σ_t se obtiene de la recursión (2.5).

No existe fórmula cerrada para ese máximo, así que la optimización se hace numéricamente. Para escoger entre las distribuciones candidatas -Normal, t de Student, GED- se usan los criterios de información de Akaike (AIC) y Bayesiano (BIC), que premian el ajuste pero cobran por cada parámetro extra. Gana la distribución con el AIC más bajo: el mejor balance entre ajustar bien y no complicarse de más.

Tabla 2.3

Principales modelos de la familia ARCH/GARCH.

Modelo	Autor	Característica distintiva
ARCH(q)	Engle (1982)	Varianza en función de q choques pasados.
GARCH (1,1)	Bollerslev (1986)	Añade inercia de la propia varianza.
IGARCH EGARCH	Engle & Bollerslev (1986) Nelson (1991)	Persistencia integrada ($\alpha+\beta=1$). Modela log-varianza; capta asimetría.
GJR-GARCH	Glosten et al. (1993)	Término extra para choques negativos.
GARCH-t	Bollerslev (1987)	Innovaciones t de Student (colas pesadas).

Nota. Elaboración propia a partir de la literatura citada.

La familia GARCH es grande, y el (1,1) es apenas su miembro más austero. Existen variantes pensadas para la asimetría -el hecho documentado de que una mala noticia sube la volatilidad más que una buena del mismo tamaño, el llamado efecto apalancamiento-: ahí están el EGARCH y el GJR-GARCH, resumidos en la Tabla 2.3. ¿Por qué escoger entonces el

GARCH (1,1)? Por dos motivos. Uno: tras comparar cientos de especificaciones, la literatura concluyó que nada le gana sistemáticamente (Hansen & Lunde, 2005). Dos: en el mercado estudiado ni siquiera aparecen los efectos asimétricos que esas variantes vienen a capturar, como se verá.

Los parámetros del GARCH admiten una lectura económica directa que explica su arraigo en la industria. Un α alto describe un mercado nervioso, donde cada sorpresa se traslada de inmediato a la volatilidad esperada; un β alto describe un mercado con memoria, donde los episodios turbulentos tardan en disolverse. La suma de ambos indica cuánto persiste un choque: con $\alpha + \beta$ cercano a 0,95 -valor típico en índices desarrollados- la volatilidad tarda semanas en volver a su nivel de largo plazo. Esa interpretabilidad es una ventaja genuina frente a los modelos de caja negra, y es parte de lo que está en juego cuando se pregunta si conviene reemplazarlo.

Un matiz técnico relevante es que la estimación por máxima verosimilitud del GARCH conserva propiedades deseables aun cuando la distribución especificada no sea exactamente la verdadera -es el argumento de la cuasi-máxima verosimilitud-, siempre que la ecuación de la varianza esté bien planteada. El punto débil está justamente ahí: si la premisa de heterocedasticidad condicional no se cumple en los datos, ninguna elección de distribución rescata al modelo, porque el error no está en la forma de las colas sino en la dinámica supuesta. Esa distinción entre error de distribución y error de dinámica resultará esencial para diagnosticar lo que ocurre con el BVQA.

2.6 Redes LSTM y regresión cuantílica

Una red recurrente lee la secuencia paso a paso, cargando un estado que resume lo ya visto. Las versiones simples tienen memoria corta: el gradiente se les desvanece y lo lejano se pierde. La solución de la LSTM (Hochreiter & Schmidhuber, 1997) fue una celda de memoria

C_t custodiada por tres compuertas -olvido (f), entrada (i) y salida (o)- que van decidiendo qué se guarda, qué entra y qué sale:

$$\text{Forget gate: } f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2.9)$$

$$\text{Input gate: } i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

$$\text{Output gate: } o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

$$\text{Cell state: } C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

$$\text{Hidden state: } h_t = o_t \odot \tanh(C_t)$$

donde $\sigma(\cdot)$ es la sigmoide, $\odot$ el producto elemento a elemento y h_t la salida del paso t. En lugar de predecir la media, aquí la red estima directamente el cuantil del VaR entrenándose con la pérdida pinball (o cuantílica), que penaliza de forma asimétrica los errores según el lado en que caigan: En lugar de predecir la media, la red estima directamente el cuantil del VaR mediante la minimización de la función de pérdida pinball (o cuantílica). De acuerdo con la formulación adaptada para series temporales financieras por **Engle y Manganelli (2004)**, esta función penaliza de forma asimétrica los errores según el lado de la distribución en que recaigan:

$$L_\alpha(y, \hat{q}) = \max(\alpha(y - \hat{q}), (\alpha - 1)(y - \hat{q})) \quad (2.10)$$

Hay una propiedad que hace especial a la pérdida pinball: minimizar su valor esperado equivale exactamente a estimar el cuantil α , porque es una regla de puntuación estrictamente propia para cuantiles (Gneiting & Raftery, 2007). Y hay un beneficio adicional de usarla: el test de Diebold-Mariano evaluará los modelos con esta misma función, de modo que el criterio de entrenamiento y el de evaluación coinciden. No se entrena con una vara y se juzga con otra.

Merece subrayarse la diferencia de enfoque frente al GARCH. La ruta paramétrica llega al VaR en dos pasos: primero estima la volatilidad condicional y luego la convierte en cuantil multiplicando por el valor crítico de una distribución supuesta. Cada paso aporta su propio error, y el segundo depende de que la distribución elegida sea la correcta. La regresión

cuantílica, en cambio, va directo al objetivo: la red produce el cuantil sin pasar por la volatilidad ni comprometerse con ninguna forma distribucional. Cuando los supuestos paramétricos son dudosos -exactamente la situación del BVQA-, esa ruta directa elimina una fuente completa de error de especificación.

El costo de esa flexibilidad es doble. Por un lado, la red exige decisiones de arquitectura -unidades, ventana de entrada, regularización- cuya influencia sobre el resultado no siempre es transparente. Por otro, el entrenamiento es estocástico: dos corridas con distinta semilla pueden dar modelos ligeramente distintos, algo impensable en la estimación por máxima verosimilitud. La metodología del Capítulo 3 enfrenta ambos costos con hiperparámetros fijados a priori y con un protocolo de semillas múltiples que cuantifica la variabilidad del entrenamiento.

El entrenamiento de una red recurrente se realiza mediante retropropagación a través del tiempo (BPTT): la secuencia de entrada se despliega, el error de la predicción final se propaga hacia atrás por cada paso temporal y los pesos se ajustan en la dirección que reduce la pérdida (Rumelhart et al., 1986). La longitud de la ventana de entrada define cuánta historia ve la red en cada predicción. La ventana de 20 días adoptada aquí -un mes bursátil- equilibra dos riesgos: una ventana muy corta oculta patrones de mediano plazo, mientras que una muy larga multiplica el costo computacional y diluye el gradiente, dificultando el aprendizaje incluso para una LSTM.

2.7 Backtesting: las pruebas que juzgan los modelos

Calcular un VaR es relativamente sencillo; lo difícil es demostrar que ese VaR sirve. De eso se ocupa el backtesting: revisar día por día si las violaciones -las jornadas en que la pérdida real se pasó del VaR anunciado- se comportan como el modelo prometía. El punto de partida es la variable indicadora de excedencia:

$$I_t = 1 \text{ si } -r_t > VaR_t - 1(\alpha); \quad I_t = 0 \text{ en caso contrario} \quad (2.11)$$

Prueba de cobertura incondicional de Kupiec (1995). La lógica es de conteo puro: si el modelo está bien calibrado, la fracción de violaciones $\hat{\pi} = x/n$ (x excedencias en n días) tiene que coincidir con α . El contraste se arma como razón de verosimilitud:

$$LR_{uc} = -2 \ln \left[\frac{(1-\alpha)^{n-x} \alpha^x}{(1-\hat{\pi})^{n-x} \hat{\pi}^x} \right] \sim \chi_1^2 \quad (2.12)$$

Prueba de cobertura condicional de Christoffersen (1998). Acertar el total de violaciones no es suficiente. Si las violaciones vienen todas juntas, en racimo, algo anda mal: el modelo no está reaccionando a los cambios de régimen. Christoffersen agrega por eso un término de independencia al contraste de Kupiec:

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi_2^2 \quad (2.13)$$

Test de Diebold-Mariano (1995). Kupiec y Christoffersen examinan un modelo a la vez; el DM enfrenta a dos. Se define $d_t = L_\alpha(r_t, \hat{q}_t^A) - L_\alpha(r_t, \hat{q}_t^B)$ como la diferencia de pérdidas pinball entre los modelos A y B. Si ambos predicen igual de bien, $E[d_t] = 0$, y el estadístico

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_{HAC}^2/n}} \sim N(0,1) \quad (2.14)$$

sigue una normal estándar; $\bar{d}$ es el promedio de las diferencias y $\hat{\sigma}_{HAC}^2$ estima la varianza de largo plazo con robustez a autocorrelación y heterocedasticidad (Newey & West, 1987). ¿La lectura? Un DM positivo y significativo dice que B predice mejor que A. De esta prueba saldrá la afirmación central del estudio: cuál modelo es, con evidencia estadística, el mejor.

Evaluación regulatoria. Basilea tiene su propio examen. El Comité evalúa los modelos de VaR al 99 % sobre ventanas de 250 días y los ubica en zonas según cuántas excedencias acumulen: Verde con 0 a 4, Amarilla de 5 a 9, Roja con 10 o más.

Ese esquema de zonas, conocido informalmente como el semáforo de Basilea, encierra una lógica estadística sencilla: con un VaR al 99 % bien calibrado, en 250 días se esperan 2,5 excedencias, y observar hasta 4 es perfectamente compatible con el azar. A partir de 5, la

evidencia contra el modelo se acumula y el multiplicador de capital sube; con 10 o más, la probabilidad de que el modelo esté bien calibrado es tan baja que el supervisor lo da por deficiente. El atractivo del esquema es que no requiere cálculos sofisticados y se comunica sin ambigüedad entre entidades y regulador.

Un aspecto que a menudo se pasa por alto es la potencia de estas pruebas en el nivel del 99 %. Como las excedencias esperadas son pocas -apenas unas seis en un período de prueba como el de esta tesis-, las pruebas disponen de poca información para discriminar entre un modelo correcto y uno moderadamente desviado. Esa limitación no invalida el ejercicio, pero obliga a leer los resultados al 99 % como complemento de los del 95 %, donde la evidencia estadística es más abundante, y explica por qué las conclusiones de este trabajo se apoyan en la convergencia de varias pruebas y no en una sola.

El repertorio de pruebas de backtesting no termina en Kupiec y Christoffersen. La literatura posterior desarrolló contrastes más exigentes, como la prueba de cuantil dinámico (DQ) asociada al enfoque CAViaR de Engle y Manganelli (2004), que examina si las excedencias son ortogonales a toda la información disponible y no solo a su propio pasado. Estas extensiones comparten, sin embargo, la misma arquitectura conceptual: convertir la calidad de un modelo de riesgo en una hipótesis estadística contrastable sobre la serie de violaciones. El conjunto adoptado en esta tesis -cobertura incondicional, cobertura condicional, comparación predictiva y pérdida regulatoria- cubre las cuatro preguntas que un validador necesita responder: si el modelo acierta la frecuencia, si distribuye bien los fallos en el tiempo, si es más preciso que sus alternativas y cuánto cuesta cuando falla.

Penalización por magnitud. Kupiec y Christoffersen cuentan violaciones; ninguno pregunta qué tan graves fueron. La función de pérdida de López (1999) sí lo hace: castiga según el tamaño del exceso cuando el VaR queda superado. Su expresión:

$$L = \sum I_t * [1 + (r_t - VaR_t)^2]$$

2.8 Estado del arte: GARCH frente a aprendizaje profundo

El recorrido de la disciplina ayuda a ubicar este trabajo. Que los precios financieros no son normales -colas gordas, volatilidad que va y viene- se sabe desde Mandelbrot (1963). Engle (1982) tradujo esa intuición al lenguaje econométrico con los ARCH; Bollerslev (1986) los volvió operativos con el GARCH. Vinieron después las extensiones asimétricas -EGARCH de Nelson (1991), GJR-GARCH de Glosten et al. (1993)- y la propuesta de Bollerslev (1987) de usar innovaciones de Student para las colas pesadas, que es la que se adopta aquí.

En cuanto a la robustez del GARCH, el resultado de referencia es contundente. Hansen y Lunde (2005) compararon 330 especificaciones y ninguna consiguió superar sistemáticamente al GARCH (1,1) prediciendo volatilidad de acciones y tipos de cambio. Ese hallazgo lo consagró como el punto de partida obligatorio -y es el que se usa en esta tesis-. La modelación de pérdidas extremas avanzó en paralelo: Teoría de Valores Extremos aplicada a VaR y ES (McNeil & Frey, 2000), y el marco de medidas coherentes de riesgo (Artzner et al., 1999) que fundamenta por qué los reguladores prefieren el Expected Shortfall.

El deep learning abrió el segundo frente. Con LSTM, Fischer y Krauss (2018) superaron a los métodos clásicos prediciendo retornos del S&P 500. Kim y Won (2018) fueron más allá y armaron híbridos LSTM-GARCH para volatilidad, con buenos resultados en mercados líquidos. Para el VaR en particular, Taylor (2019) trabajó enfoques semiparamétricos de regresión cuantílica, emparentados con la pérdida pinball empleada aquí. El andamiaje de evaluación viene de Diebold y Mariano (1995) para comparar pronósticos, y de Kupiec (1995) y Christoffersen (1998) para el backtesting formal.

Queda una pregunta incómoda: toda esa evidencia, ¿aplica fuera de los mercados grandes? Aggarwal et al. (1999) documentaron hace tiempo que la volatilidad emergente se comporta distinto, con quiebres estructurales frecuentes. Y Cont (2001) advirtió que el agrupamiento de volatilidad se debilita -o desaparece- cuando falta liquidez. Nadie había

verificado qué pasa en un mercado tan ilíquido como el ecuatoriano. Ese es el vacío que aborda esta tesis, con la sospecha declarada de que el orden habitual de los modelos podía terminar invertido.

La revisión deja una constatación adicional: los estudios que comparan GARCH y redes neuronales suelen hacerlo sobre la calidad del pronóstico de volatilidad o de retorno, y solo una minoría evalúa la cadena completa hasta el VaR con backtesting formal. Sin las pruebas de cobertura, un modelo puede exhibir errores de pronóstico menores y aun así producir un VaR inservible para fines regulatorios. El diseño de esta tesis -que juzga a todos los competidores por las excedencias que producen y no por métricas intermedias- busca cerrar precisamente esa brecha entre la literatura de pronóstico y la práctica de gestión de riesgo.

Dentro de esa literatura, la línea semiparamétrica merece mención especial porque es la más cercana en espíritu a las LSTM cuantílicas de esta tesis. El enfoque CAViaR de Engle y Manganelli (2004) modela directamente la evolución del cuantil condicional -sin pasar por la varianza- mediante especificaciones autorregresivas, y Taylor (2019) extendió esa lógica conectando la estimación de VaR y ES a través de la distribución de Laplace asimétrica. Las redes cuantílicas pueden entenderse como la generalización no lineal y no paramétrica de esa misma idea: conservan el objetivo -el cuantil directo- y reemplazan la forma funcional fija por una arquitectura flexible que aprende de los datos.

Tabla 2.4*Estudios representativos sobre estimación de volatilidad y VaR.*

Estudio	Enfoque	Hallazgo principal
Engle (1982)	ARCH	Introduce la varianza condicional.
Bollerslev (1986)	GARCH	Modelo parsimonioso de volatilidad.
Kupiec (1995)	Backtesting	Test de cobertura incondicional del VaR.
Christoffersen (1998)	Backtesting	Test de cobertura condicional.
McNeil & Frey (2000)	EVT + GARCH	Mejora la estimación de la cola.
Hansen & Lunde (2005)	Comparación	Nada supera al GARCH (1,1).
Fischer & Krauss (2018)	LSTM	LSTM > clásicos en el S&P 500.
Kim & Won (2018)	Híbrido LSTM-GARCH	Mejoras en volatilidad (mercado líquido).
Taylor (2019)	Cuantil semiparam.	VaR/ES por regresión cuantílica.
Esta tesis (2025)	5 modelos + backtesting	En mercados ilíquidos, el VaR Histórico presenta el mejor desempeño.

Nota. Elaboración propia a partir de la literatura citada.

3. Metodología

El estudio está diseñado para que cualquiera pueda repetirlo. Es cuantitativo y comparativo, y el flujo entero -bajar los datos, limpiarlos, estimar, hacer el backtesting, producir tablas y figuras- vive en código Python que corre de punta a punta. Lo que sigue documenta cada eslabón con el detalle necesario para replicarlo. La notación general está reunida en la Tabla 3.1.

Tabla 3.1

Notación empleada en la tesis.

Símbolo	Significado
P_t	Precio de cierre del índice en el día t .
r_t	Retorno logarítmico en porcentaje, $100 \cdot \ln(P_t/P_{t-1})$.
α	Nivel de significancia del VaR (0,05 ó 0,01).
$VaR_t(\alpha)$	Valor en Riesgo al nivel α condicional a la información hasta t .
$ES_t(\alpha)$	Expected Shortfall (pérdida esperada en la cola).
σ_t	Volatilidad condicional en t .
$\omega, \alpha, \beta, \nu$	Parámetros del GARCH-t (constante, ARCH, GARCH, gr. de libertad).
W	Tamaño de la ventana móvil (500 días).
I_t	Indicador de violación (1 si la pérdida supera el VaR).
n, x	Número de observaciones de prueba y de violaciones.
$L_\alpha(\cdot)$	Pérdida pinball (función objetivo y métrica de comparación).

Nota. Elaboración propia a partir de los resultados del estudio.

3.1 Datos y fuente

Los precios salen de la API de Investing.com (instrumentId = 41170). La descarga se hizo año por año con la librería `curl_cffi`, que imita la huella TLS de un navegador Chrome - sin eso, el servidor rechaza la conexión-. Quedó una serie de 2.888 cierres diarios entre el 2 de octubre de 2012 y el 7 de noviembre de 2024, que da lugar a 2.887 retornos logarítmicos. Un punto que conviene dejar claro desde el inicio: aquí no hay ni un dato inventado; absolutamente todo lo que se reporta sale de esa serie real.

Sobre la naturaleza del instrumento conviene precisar que el Ecuador General Adj es un índice accionario amplio del mercado ecuatoriano en su versión ajustada, es decir, corregida por eventos corporativos como dividendos y splits, lo que lo hace apropiado para el cálculo de retornos. Su representatividad tiene, eso sí, las limitaciones propias del mercado que resume: pocas emisoras concentran la mayor parte de la negociación, y las sesiones sin operaciones se traducen en precios de cierre repetidos.

El tratamiento de esas sesiones exigía una decisión metodológica. Se consideraron dos alternativas: eliminar los días sin negociación, lo que acortaría artificialmente el calendario y desalinearía la serie de los días hábiles reales, o arrastrar el último precio disponible, que es la convención de mercado y produce retornos nulos donde no hubo transacciones. Se optó por la segunda porque conserva la estructura temporal de la serie y refleja fielmente la experiencia de un inversionista local: en los días sin negociación, su posición efectivamente no cambió de valor de mercado. La consecuencia -una quinta parte de retornos exactamente cero- no es un artefacto de la limpieza sino una característica del mercado que cualquier modelo aplicado sobre él debe enfrentar.

La verificación de la integridad de los datos siguió tres controles complementarios. Primero, la continuidad del calendario: se comprobó que la serie descargada no tuviera huecos de fechas distintos de los fines de semana y feriados bursátiles. Segundo, la coherencia interna: los retornos calculados se reconciliaron contra los precios de origen, y los estadísticos básicos se recalcularon en dos momentos distintos del desarrollo para descartar errores de procesamiento. Tercero, la inspección de extremos: cada retorno de magnitud inusual se revisó individualmente contra el contexto de la serie, proceso del que surgieron los cuatro ticks corregidos que se documentan más arriba. Solo tras superar los tres controles la serie quedó congelada como insumo definitivo del estudio.

El retorno logarítmico se define como sigue y se expresa en porcentaje, simplemente porque así los números se leen mejor:

$$r_t = 100 \cdot \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (3.1)$$

Limpieza de datos. Cuando una sesión no cotiza, se arrastra el último precio hacia adelante -el forward-fill del Anexo B, que asume que el precio negociado persiste mientras no haya liquidez-. De ahí salen los retornos nulos, que lejos de ser un artefacto reflejan lo que de verdad pasa en un mercado ilíquido. Aparte hubo que depurar los ticks erróneos: se marca como anómalo cualquier dato P_t que produzca un retorno extremo con reversión inmediata, o sea $|r_t| > 30\%$ con signo contrario en $t + 1$. La rutina `remove_price_spikes` encontró 4 casos así, con retornos de alrededor de $\pm 700\%$; se trataron como faltantes y se interpolaron. De no hacerlo, cuatro puntos habrían mandado sobre la varianza de toda la serie.

3.2 Partición temporal

La partición de la muestra respeta estrictamente el calendario: nada del futuro se usa para predecir el pasado. El entrenamiento usa 2012-2021 (2.251 observaciones, un 78 % aproximado) y la prueba, 2022-2024 (636 observaciones, el 22 % restante), un tramo que los modelos jamás vieron. El escalado de las LSTM se calcula con el tramo de entrenamiento y se aplica congelado al de prueba; con eso queda cerrada la puerta al data leakage.

La proporción elegida -cerca de 78 % para entrenar y 22 % para probar- responde a dos necesidades en tensión. El tramo de entrenamiento debe ser largo, porque tanto la estimación por máxima verosimilitud del GARCH como el ajuste de una red neuronal necesitan historia abundante para ser estables. Pero el tramo de prueba tampoco puede ser corto: con pocas observaciones fuera de muestra, las pruebas de cobertura pierden capacidad de discriminar. Los 636 días reservados para evaluación -tres años calendario- permiten esperar alrededor de 32 excedencias al 95 %, suficientes para que Kupiec y Christoffersen operen con potencia razonable.

Sobre el tamaño de la red cabe una justificación análoga. Las 32 unidades de la capa LSTM se eligieron deliberadamente modestas: con series de unos pocos miles de observaciones, las arquitecturas grandes memorizan en lugar de generalizar, y el sobreajuste en un modelo de riesgo es particularmente insidioso porque se manifiesta como un VaR excelente dentro de muestra e inservible fuera de ella. La misma lógica de contención explica que se usara una sola capa recurrente y no varias apiladas: cada capa adicional agrega miles de parámetros que estos datos no alcanzan a disciplinar.

3.3 Los cinco modelos

Los modelos se presentan del más simple al más elaborado. La gracia está en que cada uno agrega exactamente un ingrediente sobre el anterior:

- **Modelo 1 — GARCH (1,1)-t.** El referente clásico. Se ajusta por máxima verosimilitud probando tres distribuciones para las innovaciones -Normal, t de Student y GED-, y la elección queda en manos del AIC y el BIC. Su VaR sale de la ecuación (2.7), calculado sobre una ventana móvil de 500 observaciones.
- **Modelo 2 — LSTM base.** Una LSTM de 32 unidades con Dropout de 0,20, pérdida pinball y ventana de entrada de 20 días. Se entrena una única vez y queda congelada. Como entrada recibe solamente el retorno r_t .
- **Modelo 3 — LSTM-WF.** Conserva la arquitectura del anterior pero cambia el régimen de trabajo: se reentrena cada 90 días (walk-forward) y suma dos entradas de volatilidad. Nació para remediar lo que le falla al modelo base.
- **Modelo 4 — Híbrido GARCH-LSTM.** Es la LSTM-WF más una cuarta entrada: la volatilidad condicional σ_t que produce el GARCH. Este modelo responde si mezclar las dos escuelas suma o no.

- **Modelo 5 — HistVaR (benchmark).** Toma los últimos 500 retornos, extrae el cuantil empírico -ecuación (2.3)-, nada más. Sin parámetros ni supuestos: por eso funciona de árbitro imparcial.

Tabla 3.2

Especificación comparada de los cinco modelos.

Modelo	Tipo	Entradas	Adaptación	Parámetros
GARCH-t	Paramétrico	r_t	Rolling 500	4 + v
LSTM-base	Deep learning	r_t	Estático	4.193
LSTM-WF	Deep learning	r_t, r_t , vol_{21}	Walk-forward 90 d	5.153
Híbrido	Mixto	$r_t, r_t , vol_{21}, \sigma_G ARCH$	Walk-forward 90 d	5.281
HistVaR	No paramétrico	r_t (cuantil)	Rolling 500	0

Nota. Elaboración propia a partir de los resultados del estudio.

Las dos variables de volatilidad de los Modelos 3 y 4 se construyen mirando solo hacia atrás. Una es el valor absoluto del retorno, $|r_t|$, que aproxima de forma directa el tamaño del movimiento. La otra es la volatilidad rodante de 21 días -un mes bursátil, más o menos-:

$$\sigma_{roll,t} = \sqrt{\frac{1}{21} \sum_{i=1}^{21} (r_{t-i} - \bar{r})^2} \quad (3.2)$$

¿Cómo funciona el walk-forward? El período de prueba se corta en bloques de 90 días. Cuando arranca un bloque, el modelo se reentrena usando toda la historia acumulada hasta ese momento, y con eso predice los 90 días que vienen. En 2022-2024 esto significó reentrenar 8 veces por cada nivel de confianza. El modelo trabaja siempre con lo último disponible y las predicciones nunca se contaminan con información futura.

Sobre el entrenamiento de las redes: el optimizador es Adam (Kingma & Ba, 2015) y la función objetivo, la pérdida pinball de la ecuación (2.10) -su definición completa está en el Anexo B-. Contra el sobreajuste hay dos frenos. El primero es Dropout al 0,20 (Srivastava et al., 2014), que va apagando neuronas al azar mientras la red aprende. El segundo es el early

stopping con paciencia de 12 épocas (ver Anexo B): cuando la pérdida de validación deja de bajar, el entrenamiento se corta ahí.

Las entradas se estandarizan usando la media y la desviación del tramo de entrenamiento; esos dos valores quedan fijos y se reutilizan tal cual en el tramo de prueba.

Estabilidad del entrenamiento y control de la aleatoriedad. Entrenar una red neuronal es un proceso con azar adentro: la inicialización de los pesos cambia el resultado. Para descartar que las conclusiones fueran producto del azar, cada arquitectura (LSTM-base, LSTM-WF e Híbrido) se entrenó cinco veces, con cinco semillas fijas ($seed \in \{42, 100, 2024, 7, 33\}$). La dispersión de la pérdida pinball de validación entre semillas quedó por debajo de 0,002 en todos los casos. Las diferencias entre arquitecturas, entonces, son estructurales; no las puso ahí el optimizador Adam.

3.4 Métricas y criterio de selección

Ningún modelo se salva de la batería completa: Kupiec (LR_uc, ec. 2.12), Christoffersen (LR_cc, ec. 2.13), Diebold-Mariano con pérdida pinball y varianza HAC (ec. 2.14), Expected Shortfall en sus dos versiones -el empírico, que promedia la pérdida de los días con violación, y el paramétrico del GARCH-t- y, para vigilar la estabilidad, la tasa de violaciones calculada en ventana móvil de 90 días.

¿Y si varios modelos pasan las pruebas? Para ese caso quedó definido de antemano un criterio jerárquico, que se aplica en este orden:

- 1) Validez estadística: que apruebe Kupiec y Christoffersen en ambos niveles.
- 2) Tasa de excedencias: cercanía de la frecuencia observada al nivel teórico.
- 3) Precisión cuantílica formal: resultado del test de Diebold-Mariano.
- 4) Magnitud de las pérdidas y prudencia regulatoria: se mide a través del Expected Shortfall y de la minimización de la Función de Pérdida Regulatoria de López, que penaliza a los modelos con excedencias catastróficas.

- 5) Estabilidad temporal: ausencia de agrupamiento de violaciones.
- 6) Parsimonia: a igualdad de desempeño, se prefiere el modelo más simple.

Análisis de sensibilidad de la ventana de calibración. Existía el riesgo de que los resultados dependieran de una única configuración temporal, así que la ventana móvil W se sometió a una prueba de estrés. A la ventana base de 500 días bursátiles (unos dos años) se le añadieron dos escenarios: uno corto de 250 días (un año) y uno largo de 750 (tres años). El propósito era aislar cuánto pesa la longitud de la memoria, tanto en el cuantil estimado como en la dinámica de la volatilidad condicional, tratándose de un mercado ilíquido.

Los hiperparámetros del estudio están reunidos en la Tabla 3.3. Se fijaron antes de correr nada, tomando los valores habituales de la literatura: mejor eso que buscar la combinación óptima, que es la receta clásica del sobreajuste.

Un detalle técnico del test de Diebold-Mariano justifica mención: la varianza del estadístico se estima con el método de Newey y West (1987), robusto a autocorrelación y heterocedasticidad. La precaución no es ornamental. Las diferencias diarias de pérdida entre dos modelos de VaR están lejos de ser independientes -un episodio de mercado que perjudica a un modelo suele perjudicarlo varios días seguidos-, y una varianza estimada sin esa corrección subestimaría la incertidumbre del promedio, inflando artificialmente la significación de las comparaciones. Con la corrección HAC, los p-valores reportados en el Capítulo 4 pueden tomarse al pie de la letra.

El orden del criterio jerárquico tampoco es casual. La validez estadística va primero porque un modelo rechazado por las pruebas de cobertura no es un candidato con defectos: es un candidato descalificado, sin importar cuán atractivas sean sus demás propiedades. La precisión cuantílica del test de Diebold-Mariano ocupa el tercer lugar porque solo tiene sentido comparar la finura de modelos que ya demostraron ser válidos. Y la parsimonia cierra la lista

como criterio de desempate, recogiendo un principio elemental de la práctica: entre dos modelos equivalentes, el más simple es más barato de operar, más fácil de auditar y más difícil de usar mal.

Tabla 3.3

Hiperparámetros de configuración del estudio.

Hiperparámetro	Valor	Justificación
Ventana móvil (W)	500 días	≈ 2 años bursátiles; estándar de mercado. Adicionalmente, previene que la memoria infinita (dinámica IGARCH) aplane las predicciones limitando la sensibilidad del modelo a regímenes de volatilidad obsoletos.
Ventana de entrada LSTM	20 días	≈ 1 mes bursátil de memoria.
Unidades LSTM	32	Capacidad moderada, evita sobreajuste.
Dropout	0,20	Regularización (Srivastava et al., 2014).
Regularización L2	1×10^{-4}	Penalización de pesos grandes.
Optimizador	Adam	Convergencia robusta (Kingma & Ba, 2015).
Reentrenamiento	90 días	Equilibrio adaptación / coste.
Paciencia (early stopping)	12 épocas	Detiene antes del sobreajuste.
Volatilidad rodante	21 días	≈ 1 mes; proxy de riesgo reciente.
Niveles de confianza	95 % y 99 %	Estándares de práctica y regulación.

Nota. Elaboración propia a partir de los resultados del estudio.

3.5 Operacionalización de las variables

La Tabla 3.4 operacionaliza las variables: para cada una, su definición conceptual, la operacional y la escala de medición. La variable dependiente es el VaR a un día del índice BVQA. Las independientes son la historia de precios y las transformaciones que salen de ella, que son las que alimentan a los cinco modelos.

Tabla 3.4

Operacionalización de las variables del estudio

Variable		Definición conceptual	Definición operacional	Escala	
Retorno	logarítmico (rt)	Variación relativa diaria del índice BVQA.	$rt = \ln(P_t / P_{t-1}) \times 100$, ecuación (3.1).	Cuantitativa	continua (%)
VaR	al nivel α (variable dependiente)	Pérdida máxima esperada a un día que solo se supera con probabilidad α .	Cuantil α de la distribución condicional de los retornos, estimado por cada modelo: ecuaciones (2.3), (2.7) y (2.10).	Cuantitativa	continua (%)
Violación	o excedencia (It)	Fallo del modelo: la pérdida real supera el VaR previsto.	Indicador binario $It = 1$ si $rt < -VaR_t$, ecuación (2.11).	Dicotómica (0/1)	
Volatilidad	condicional (σ_t)	Escala cambiante en el tiempo de las innovaciones del retorno.	Recurción GARCH(1,1), ecuación (2.5); insumo del modelo Híbrido.	Cuantitativa	continua (%)
Magnitud del retorno	(rt)	Aproximación directa del tamaño del movimiento diario.	Valor absoluto del retorno logarítmico; variable de entrada de la LSTM-WF.	Cuantitativa	continua (%)
Volatilidad	rodante (σ_{21})	Nivel de riesgo local del último mes bursátil.	Desviación estándar móvil de 21 días, ecuación (3.2); variable de entrada de la LSTM-WF.	Cuantitativa	continua (%)
Expected	Shortfall (ES)	Pérdida media esperada cuando se supera el VaR.	Promedio de las pérdidas en los días de violación, ecuación (2.2).	Cuantitativa	continua (%)
Cobertura	observada (π)	Frecuencia empírica de violaciones del modelo.	$\pi = x/n$ en el período de prueba; contraste con α mediante Kupiec y Christoffersen.	Proporción [0, 1]	

Nota. Elaboración propia a partir de los resultados del estudio.

4. Resultados

4.1 Análisis exploratorio

Los retornos del BVQA rompen todos los moldes. Asimetría positiva de 3,82. Curtosis en exceso de 95,7. Jarque-Bera de $1,11 \times 10^6$ ($p < 0,001$) y Shapiro-Wilk de 0,616 ($p \approx 2 \times 10^{-62}$): la normalidad queda descartada sin discusión. De las 2.887 observaciones, 600 -el 20,8 %- son exactamente cero, la huella de un mercado que muchos días simplemente no negocia. Y el dato que condiciona todo lo demás: la prueba ARCH-LM con 10 rezagos no encuentra heterocedasticidad condicional (LM=10,54; $p=0,395$). En un mercado líquido eso sería rarísimo. Aquí, como se verá, tiene toda la lógica, y deja tambaleando la premisa de cualquier modelo de varianza condicional.

Estos estadísticos merecen una traducción económica. Una curtosis en exceso cercana a 96 -cuando la de una distribución normal es cero- indica que la actividad del índice se concentra en dos estados: la inmovilidad casi total de la mayoría de las sesiones y los saltos abruptos de unas pocas. La asimetría positiva señala que los saltos extremos al alza superaron en magnitud a los bajistas durante la muestra, un patrón asociado a revalorizaciones puntuales de emisoras poco líquidas más que a rallies de mercado. Para la gestión de riesgo, la implicación es incómoda: el peligro no se anuncia con volatilidad creciente, sino que aparece de golpe desde la calma.

El resultado de la prueba ARCH-LM es, en este cuadro, la pieza que ordena todo lo demás. La ausencia de heterocedasticidad condicional detectable no significa que el BVQA carezca de riesgo -las colas extremas prueban lo contrario-, sino que ese riesgo no se organiza en los clústeres temporales que los modelos GARCH necesitan para operar. Es la diferencia entre un riesgo pronosticable en su intensidad y un riesgo que solo se deja describir en su distribución: distinción que anticipa, desde el análisis exploratorio, el desenlace de la comparación de modelos.

Hay una consecuencia adicional de la masa de probabilidad concentrada en cero que conviene explicitar, porque afecta la mecánica misma de la estimación de cuantiles. Cuando una quinta parte de las observaciones comparte exactamente el mismo valor, la función de distribución empírica presenta un tramo plano alrededor de ese punto, y los cuantiles centrales se vuelven insensibles a variaciones moderadas de la muestra. Los cuantiles extremos -los que interesan al VaR- quedan en cambio determinados por la minoría de observaciones no nulas, de modo que el tamaño muestral efectivo para la cola es bastante menor que el nominal. Ese fenómeno afecta a todos los modelos por igual, pero castiga más a los que intentan modelar la dinámica completa de la serie que a los que solo leen la cola empírica.

Tabla 4.1

Estadísticos descriptivos — retornos BVQA (2012–2024).

Estadístico	Valor
Número de observaciones	2887.00
Media	7.803e-05
Mediana	0.0000
Desviación estándar	0.0060
Varianza	3.626e-05
Asimetría	3.8153
Curtosis en exceso	95.7105
Mínimo	-0.0634
Máximo	0.1287
Percentil 1	-0.0165
Percentil 5	-0.0065
Percentil 25	-0.0012
Percentil 75	0.0011
Percentil 95	0.0070
Percentil 99	0.0168

Nota. Elaboración propia a partir de los resultados del estudio.

Nota. los estadísticos de la Tabla 4.1 se expresan en retornos logarítmicos en proporción (p. ej., una desviación estándar de 0,0060 equivale a 0,60 % diario). A partir de la sección 4.2, los modelos y el VaR se expresan en puntos porcentuales para mayor legibilidad.

Las primeras figuras cuentan la historia visualmente. La 4.1 recorre el nivel del índice en los doce años de muestra. La 4.2 grafica los retornos, y ahí saltan a la vista los tramos planos

-cero tras cero- interrumpidos por saltos sueltos: la iliquidez retratada. La 4.3, el histograma contra la curva normal, muestra un apuntamiento descomunal: la masa se apila alrededor de cero mientras algunos valores extremos quedan lejísimos.

Figura 4.1

Evolución del índice Ecuador General Adj (BVQA), 2012–2024.



Figura 4.2

Retornos logarítmicos diarios del BVQA. Nótese los tramos planos (retornos nulos).

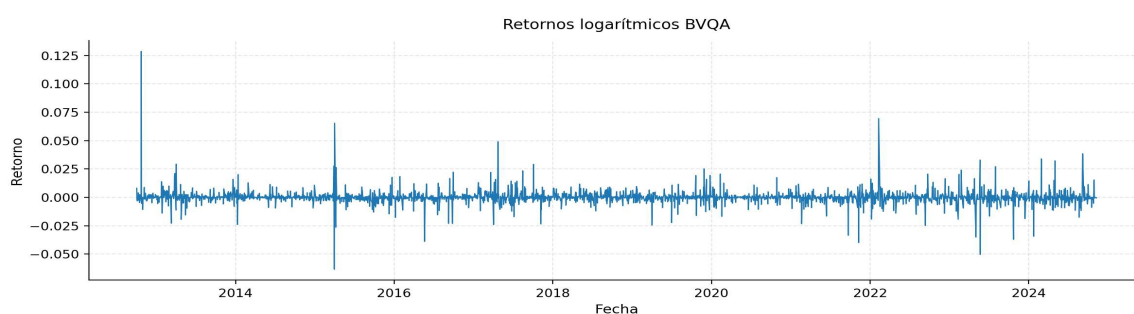
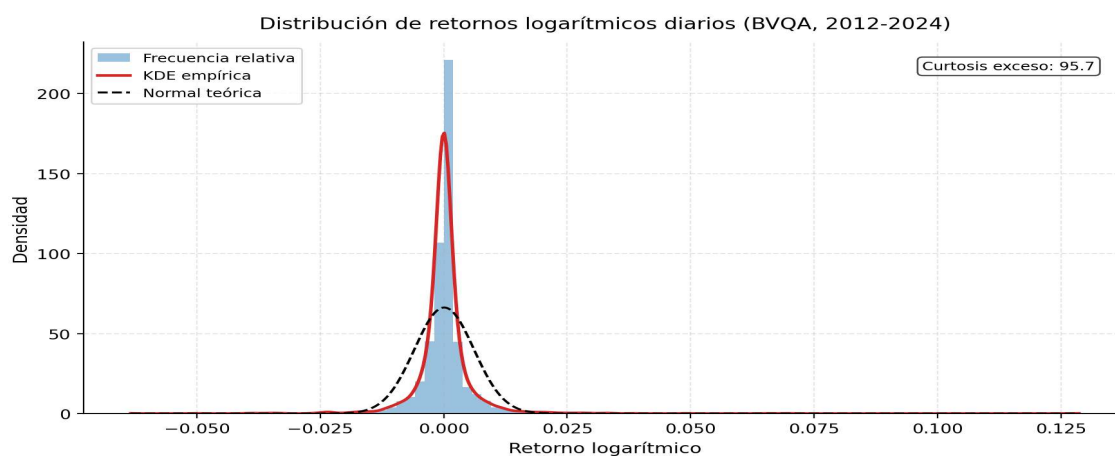


Figura 4.3

Histograma de retornos frente a la normal: leptocurtosis extrema y colas pesadas.



El Q-Q de la Figura 4.4 remata el diagnóstico: al comparar cuantiles empíricos contra normales, la curva se tuerce en S en las dos puntas -colas pesadas, sin discusión-. En la Figura 4.5, las funciones ACF y PACF muestran retornos casi sin estructura lineal. Eso cuadra con eficiencia de mercado en la media, aunque no dice nada todavía sobre la varianza, que es otro asunto.

Figura 4.4

Gráfico Q-Q de los retornos frente a la distribución normal.

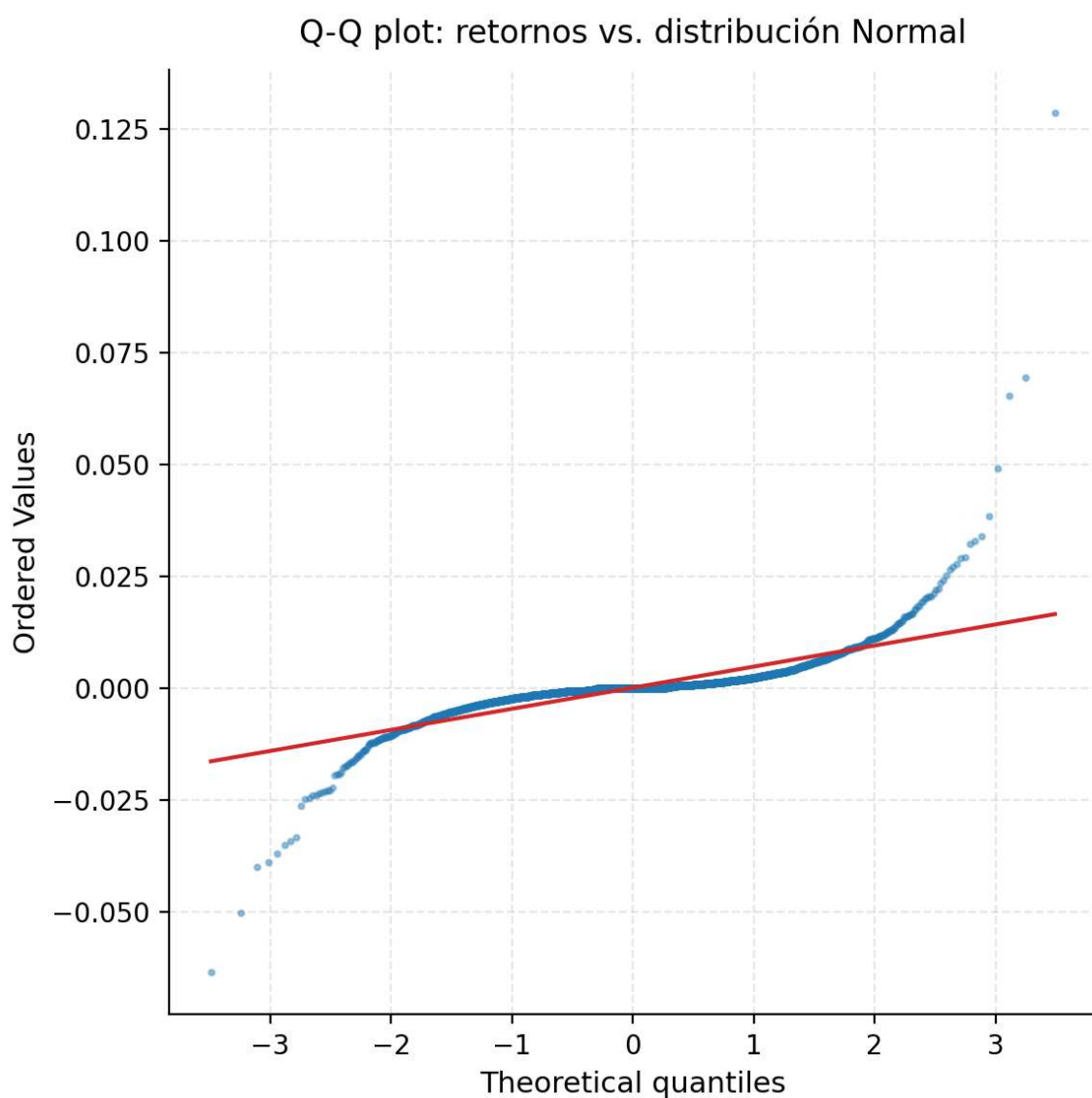
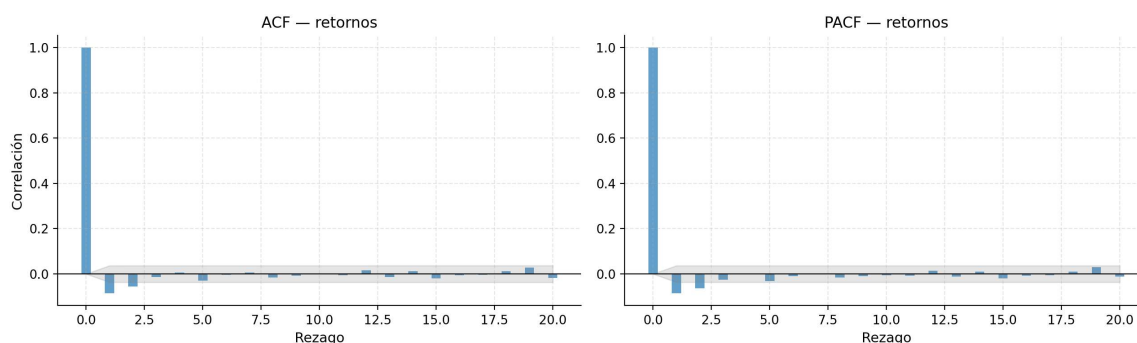


Figura 4.5

Funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) de los retornos.



4.2 GARCH (1,1)-t: parámetros

Entre las distribuciones candidatas ganó la t de Student, con AIC de 989,84 contra 3.123,78 de la normal y 1.254,04 de la GED. Pero lo importante está en los parámetros: $\alpha + \beta = 0,2749 + 0,7251 = 1,0000$ exacto. IGARCH puro: varianza incondicional indefinida, choques de volatilidad que no se disipan nunca. Los grados de libertad dieron $\nu \approx 2,09$, o sea el proceso pegado al borde de la varianza finita (el requisito es $\nu > 2$). El Ljung-Box de los residuos al cuadrado arroja $p \approx 1,0$, señal de que el GARCH exprimió toda la estructura ARCH que había para expresar. El problema es que había muy poca.

El valor estimado de los grados de libertad merece un comentario propio. Una t de Student con $\nu \approx 2,09$ es una distribución de colas extraordinariamente pesadas: apenas por encima del umbral $\nu = 2$ en que la varianza deja de existir, y por debajo del rango 4-8 que suele estimarse en índices desarrollados. El dato es coherente con el resto del cuadro: una serie dominada por la inmovilidad, interrumpida por saltos ocasionales de gran magnitud, se describe mejor con una distribución que asigna probabilidad apreciable a los extremos. Pero también es una señal de alerta sobre el propio modelo: con ν en esa zona, los momentos de orden superior no existen y la teoría asintótica que sustenta los errores estándar opera al límite de sus supuestos.

Tabla 4.2*Parámetros estimados — GARCH (1,1)-t (entrenamiento 2012–2021).*

Parámetro	Notación	Estimado	Error estándar	t-estadístico	p-valor	IC 95% inf	IC 95% sup
Media (μ)	μ	0.0015	0.0032	0.4615	0.6445	-0.0048	0.0078
Cte. varianza (ω)	ω	0.0974	0.0324	3.0078	0.0026	0.0339	0.1609
Efecto ARCH (α)	α_1	0.2749	0.0596	4.6152	4.000e-06	0.1581	0.3916
Persistencia GARCH (β)	β_1	0.7251	0.0512	14.1713	0.0000	0.6249	0.8254
Grados de libertad (ν)	ν	2.0949	0.0202	103.89	0.0000	2.0554	2.1344

Nota. Elaboración propia a partir de los resultados del estudio.**Tabla 4.3***Comparación de distribuciones — GARCH (1,1).*

Distribución	Log-verosimilitud	AIC	BIC	N.º de parámetros	Grados de libertad (ν)	Mejor por AIC	Mejor por BIC
Normal	-1557.89	3123.78	3146.65	4	—	No	No
t de Student	-489.92	989.84	1018.44	5	2.0949	Sí	Sí
GED	-622.02	1254.04	1282.63	5	1.0100	No	No

Nota. Elaboración propia a partir de los resultados del estudio.

La Figura 4.6 grafica la volatilidad condicional del GARCH en el tiempo, y la imagen habla sola: nada de oscilar alrededor de un nivel de equilibrio como haría un proceso estacionario; la serie va a la deriva, sin reversión a la media. Es el retrato de un IGARCH con persistencia unitaria. En la Figura 4.7, el Q-Q de los residuos estandarizados contra la t muestra colas mejor ajustadas que con la normal, aunque los extremos siguen desviándose: ni la t alcanza para la severidad de los eventos raros del BVQA.

Figura 4.6*Volatilidad condicional estimada por el GARCH (1,1)-t.*

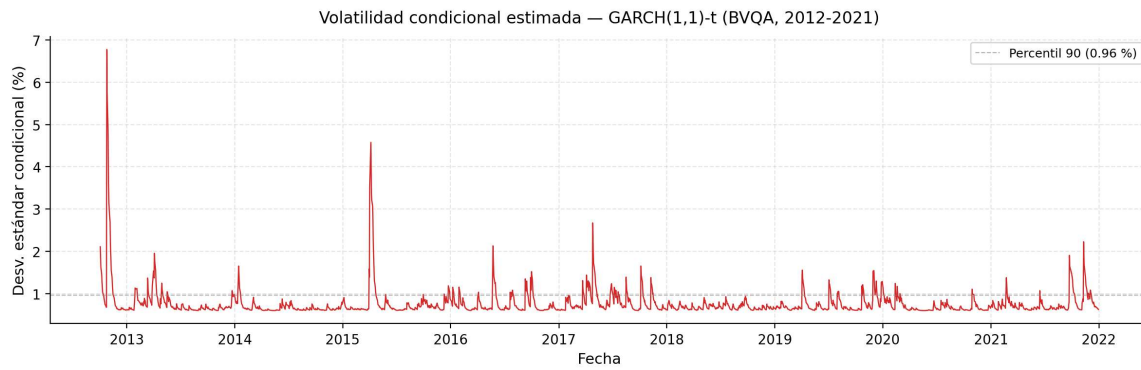
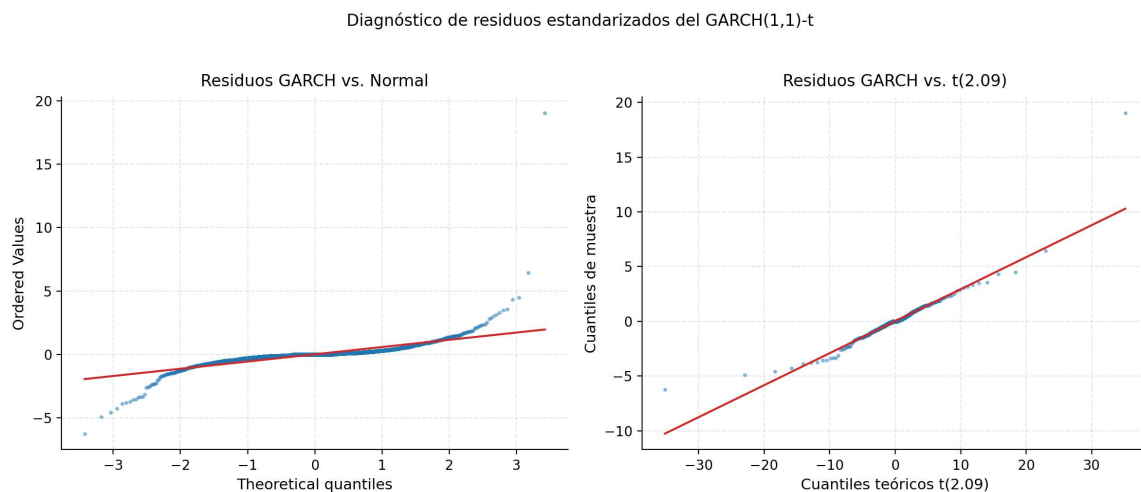


Figura 4.7

Gráfico Q-Q de los residuos estandarizados del GARCH frente a la t de Student.



4.3 Arquitecturas LSTM

Las tres variantes de LSTM entrenadas con pérdida pinball están resumidas en la Tabla 4.4. El esqueleto es el mismo en todas -una capa LSTM de 32 unidades, Dropout de 0,20, regularización L2-; lo que cambia son las entradas y la política de reentrenamiento. La LSTM-WF agrega dos variables construidas sin mirar el futuro, el valor absoluto del retorno $|r_t|$ y la volatilidad rodante de 21 días; el Híbrido suma una cuarta, la σ del GARCH. El costo en parámetros es moderado: 4.193, 5.153 y 5.281. En el período de prueba, los modelos walk-forward se reentrenaron 8 veces por nivel de confianza, dejando que el early stopping decidiera las épocas de cada bloque.

La relación entre parámetros y observaciones disponibles se mantuvo deliberadamente holgada. Con alrededor de cinco mil parámetros y más de dos mil observaciones de entrenamiento -que el esquema walk-forward amplía progresivamente-, las redes operan en un régimen donde la regularización activa (Dropout, L2, parada temprana) alcanza para contener la memorización. Las curvas de aprendizaje del apartado siguiente confirman ese equilibrio: la pérdida de validación acompaña a la de entrenamiento sin la divergencia característica del sobreajuste.

Tabla 4.4

Arquitectura de los tres modelos LSTM.

Mod elo	For ma de entr ada	Variables de entrada	de	Cap as LS TM	Unid ades	Drop out	Regulari zación L2	Reentrena miento	Parám etros	Func ión de pérd ida
LST M base	(20, 1)	Retorno r_t		1	32	0.200 0	1×10^{-4}	No (estático)	4193	Pinball ($\alpha=0.05$ y 0.01)
LST M- WF	(20, 3)	r_t, r_t , vol_{21d}		1	32	0.200 0	1×10^{-4}	Sí (cada 90 días)	5153	Pinball ($\alpha=0.05$ y 0.01)
Híbrido GAR CH- LST M	(20, 4)	$r_t, r_t , vol_{21d}, \sigma_t$		1	32	0.200 0	1×10^{-4}	Sí (cada 90 días)	5281	Pinball ($\alpha=0.05$ y 0.01)

Nota. Elaboración propia a partir de los resultados del estudio.

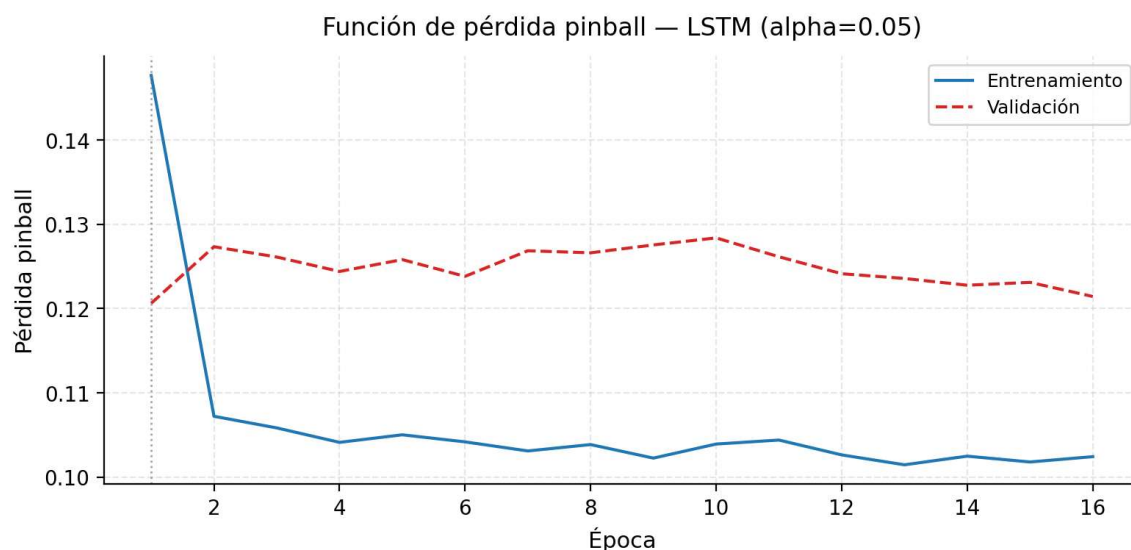
La curva de aprendizaje de la Figura 4.8 -pérdida pinball por época, cuantil del 5 %-baja suave y se aplanan en validación. Traducción: el entrenamiento convergió y no hubo sobreajuste visible. El Dropout y el early stopping cumplieron.

Las cinco semillas contaron la misma historia: convergencia pareja en la optimización de la pérdida cuantílica. La desviación estándar de las predicciones de validación entre semillas

fue mínima para la LSTM-WF ($\pm 0,0015$ en la pérdida pinball del cuantil al 5 %). El early stopping hizo su trabajo: la sensibilidad a los pesos iniciales de la capa recurrente quedó neutralizada.

Figura 4.8

Curva de aprendizaje de la LSTM (pérdida pinball, cuantil 5 %).



4.4 Backtesting comparativo — 5 modelos

Este es el corazón de la tesis. De los cinco modelos, solo dos pasan Kupiec y Christoffersen en ambos niveles: el HistVaR ($p_{\text{Kupiec95}}=0,454$; $p_{\text{Kupiec99}}=0,802$) y la LSTM-WF ($p_{\text{Kupiec95}}=0,053$; $p_{\text{Kupiec99}}=0,181$). Los demás caen. El GARCH-t es rechazado al 95 % ($p=0,0001$) y al 99 % ($p=0,020$). La LSTM-base acumula 65 violaciones al 95 %, con un p del orden de 10^{-8} . El Híbrido registra el peor 99 % de todos: 15 violaciones. El detalle -violaciones, tasas, p -valores de las dos pruebas, pérdida regulatoria- está en la Tabla 4.5.

Tabla 4.5
Tabla de decisión resumida.

Modelo	Nivel de confianza	N.º de violaciones	Tasa de observada	Tasa esperada	Kupiec (p-valor)	¿Aprueba Kupiec?	Christoffersen CC (p-valor)	¿Aprueba Christoffersen?	Pérdida regulatoria	Pérdida cuantílica
GARC H-t	95%	56	0.0881	0.0500	6.464e-05	No	3.053e-04	No	118.13	0.0849
Híbrido	95%	47	0.0739	0.0500	0.0096	No	0.0334	No	108.54	0.0835
HistVa R	95%	36	0.0566	0.0500	0.4538	Sí	0.5270	Sí	88.2612	0.0804
LSTM-WF	95%	43	0.0676	0.0500	0.0526	Sí	0.1273	Sí	98.8764	0.0833
LSTM-base	95%	65	0.1022	0.0500	9.875e-08	No	1.519e-07	No	135.56	0.0897
GARC H-t	99%	13	0.0204	0.0100	0.0204	No	0.0518	Sí	40.5329	0.0377
Híbrido	99%	15	0.0236	0.0100	0.0034	No	0.0095	No	60.1114	0.0421
HistVa R	99%	7	0.0110	0.0100	0.8018	Sí	0.8963	Sí	19.9282	0.0303
LSTM-WF	99%	10	0.0157	0.0100	0.1807	Sí	0.3478	Sí	32.4329	0.0347
LSTM-base	99%	11	0.0173	0.0100	0.0938	Sí	0.2024	Sí	40.0478	0.0363

Nota. Para un desglose analítico completo de la función de pérdida pinball, magnitudes medias de excedencia y ratios de cobertura condicional, consúltese la

Tabla D.1 en el Anexo D

La Tabla 4.5 merece una lectura pausada. El GARCH-t viola su VaR el 8,81 % de los días cuando prometía el 5 %: Kupiec lo rechaza con $p = 0,0001$. La LSTM-base está peor: 10,22 % de excedencias ($p < 0,001$). Al Híbrido tampoco le alcanza: 47 violaciones al 95 % y 15 al 99 %, rechazado en ambos niveles ($p = 0,0096$ y $p = 0,0034$). Del otro lado del corte, la LSTM-WF se salva en los dos niveles -6,76 % observado con $p = 0,0526$ al 95 %, y $p = 0,1807$ al 99 %-, aunque sin margen de sobra. El único que aprueba holgado es el HistVaR: 5,66 % de excedencias ($p = 0,4538$) al 95 % y 1,10 % ($p = 0,8018$) al 99 %, pasando tanto Kupiec como la independencia de Christoffersen.

Las Figuras 4.9 y 4.10 superponen el VaR de cada modelo sobre los retornos reales de 2022-2024, al 95 % y al 99 %. La curva gruesa del HistVaR abraza los retornos con estabilidad. El GARCH, en cambio, se sacude de más. Y la LSTM-base ni se inmuta: una línea casi recta que ignora lo que hace el mercado.

Figura 4.9

VaR al 95 % de los modelos frente a los retornos reales del BVQA, 2022–2024.

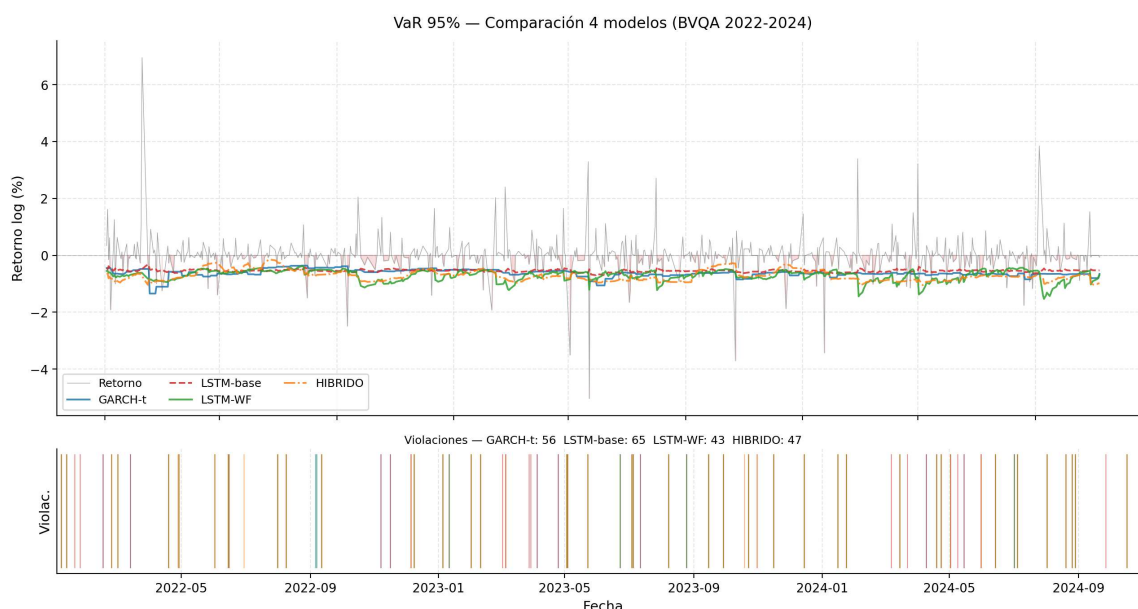
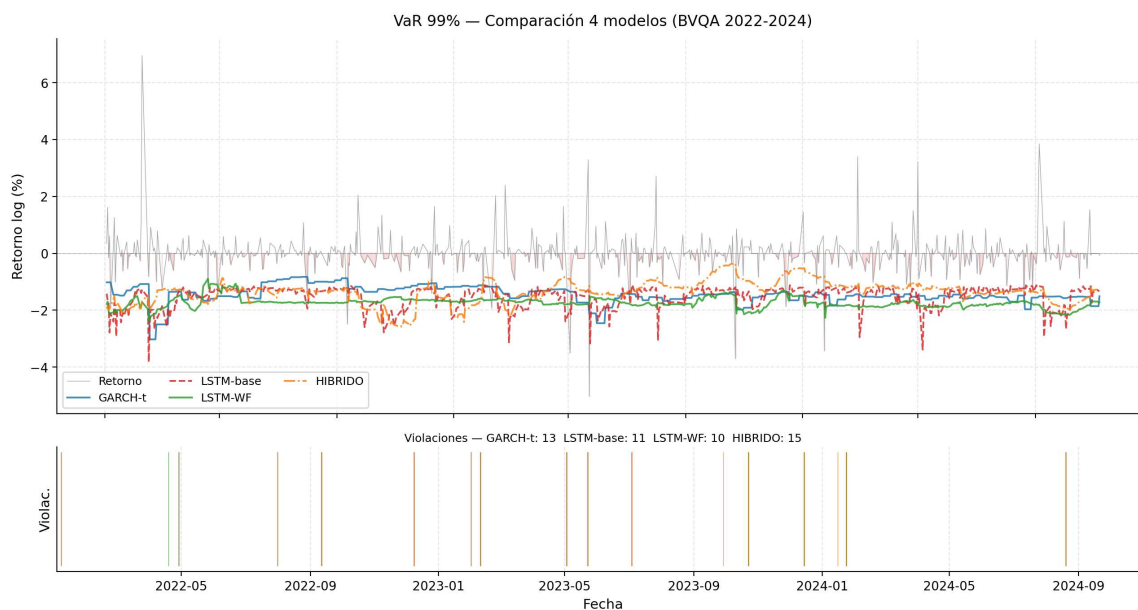


Figura 4.10

VaR al 99 % de los modelos frente a los retornos reales del BVQA, 2022–2024.

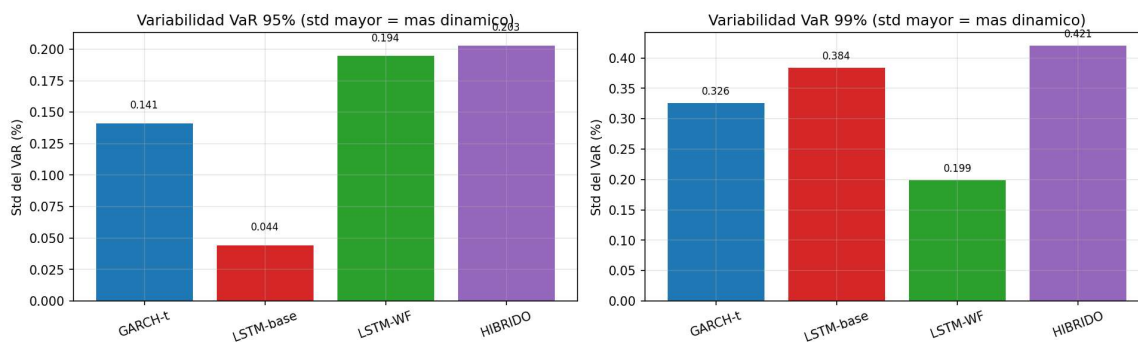


¿Qué tan adaptable es cada modelo? La Figura 4.11 lo mide con la desviación estándar de la serie de VaR. La barra de la LSTM-base es minúscula -su VaR es casi una constante-, mientras que la LSTM-WF y el GARCH generan estimaciones que se mueven bastante. Pero cuidado con la conclusión fácil: moverse mucho no es virtud en sí misma. La virtud es moverse cuando hay que moverse, y en ese equilibrio el que mejor queda parado es el HistVaR.

En términos de gestión, la diferencia entre los modelos aprobados y rechazados es todo menos abstracta. Una mesa que hubiera operado el período de prueba con el VaR del GARCH-t habría visto su umbral superado 56 veces prometiendo que serían unas 32: cada excedencia adicional es un día de pérdidas por encima de lo provisionado y un registro que se acumula en el expediente regulatorio. Con el HistVaR, las excedencias observadas coinciden con las prometidas. La estadística formal y la consecuencia operativa apuntan, en este caso, exactamente en la misma dirección.

Figura 4.11

Variabilidad (desviación estándar) del VaR estimado por cada modelo.



4.5 Expected Shortfall

Si el VaR dice dónde está el umbral, el Expected Shortfall contesta la pregunta siguiente: cuando el umbral se rompe, ¿de qué tamaño es el golpe promedio?

En ese frente, el HistVaR resultó el más cauteloso del grupo: su ES empírico es el más alto del estudio, 1,53 % al 95 % y 3,14 % al 99 %. El GARCH-t, por comparar, tiene un ES paramétrico de 1,43 % al 95 %, y en los días en que efectivamente violó su VaR la pérdida promedio fue de 1,21 %.

La lectura conjunta de ambas cifras deja una conclusión de prudencia. Que el ES del HistVaR sea el más alto no significa que el modelo pierda más, sino que promete colchones más realistas: sus umbrales están ubicados de modo que, cuando fallan, la pérdida excedente ya estaba contemplada en la magnitud que el modelo anticipaba. Un ES bajo junto a un exceso de violaciones -la combinación que exhibe el GARCH-t- describe lo contrario: un modelo que subestima a la vez la frecuencia y la severidad de los días malos, el peor de los escenarios para fines prudenciales.

Cabe añadir que la validación formal del Expected Shortfall es un problema técnicamente más difícil que la del VaR, y las pruebas disponibles están menos estandarizadas; esa asimetría explica el papel que se le asignó en este estudio: el VaR concentra el backtesting formal porque sus contrastes son maduros y bien entendidos, mientras que el ES aporta la

dimensión de severidad como métrica complementaria de prudencia, tal como lo concibe el propio marco regulatorio (BCBS, 2019).

Tabla 4.6

Expected Shortfall empírico y paramétrico — 5 modelos.

Modelo	Nivel	α	ES empírico	ES paramétrico	Razón ES/VaR
GARCH-t	95%	0.0500	1.2097	1.4294	1.9294
LSTM-base	95%	0.0500	1.1369	—	2.0880
LSTM-WF	95%	0.0500	1.3684	—	1.9220
Híbrido	95%	0.0500	1.2936	—	1.8187
HistVaR	95%	0.0500	1.5251	—	1.9876
GARCH-t	99%	0.0100	2.4175	3.1515	1.6665
LSTM-base	99%	0.0100	2.5621	—	1.6765
LSTM-WF	99%	0.0100	2.6295	—	1.5141
Híbrido	99%	0.0100	2.2111	—	1.6168
HistVaR	99%	0.0100	3.1449	—	1.6993

Nota. Elaboración propia a partir de los resultados del estudio.

Figura 4.12

Expected Shortfall empírico al 95 % por modelo. El HistVaR es el más conservador.

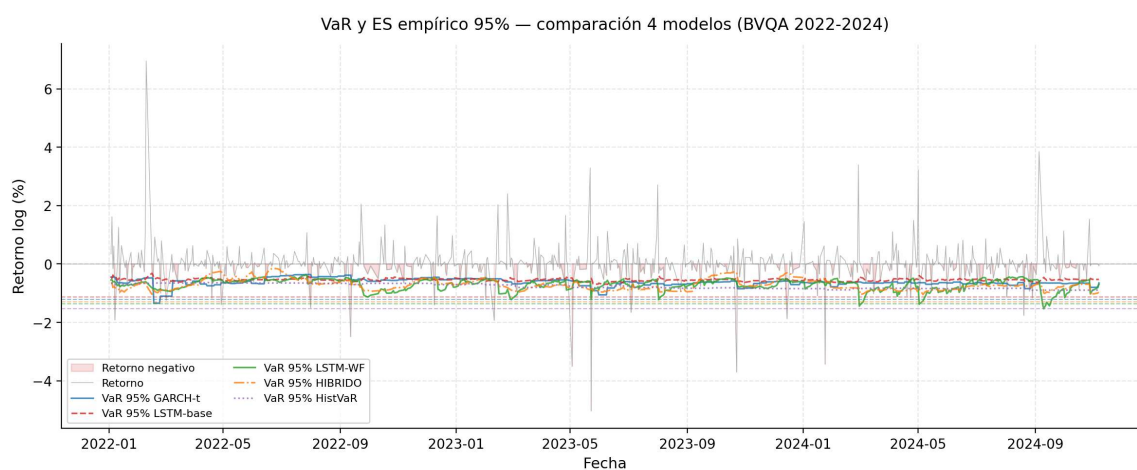
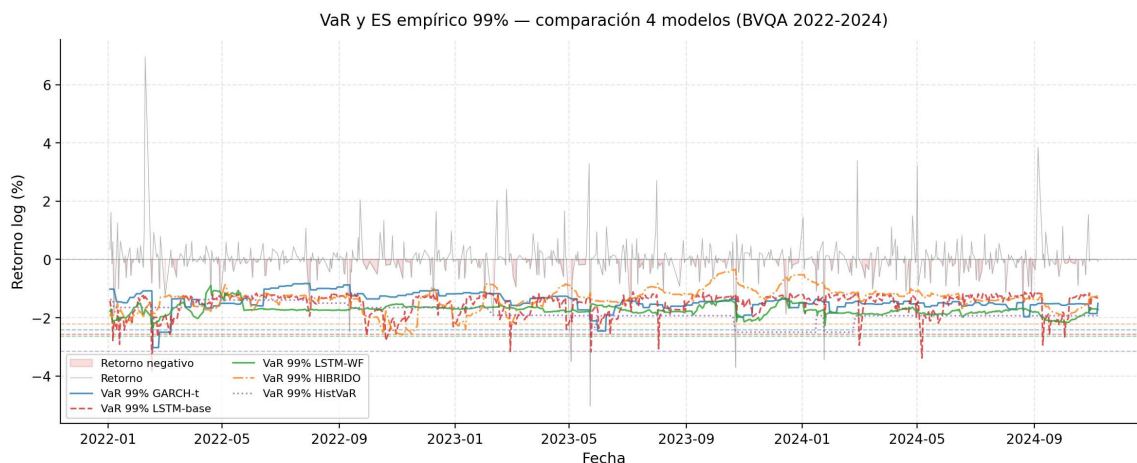


Figura 4.13

Expected Shortfall empírico al 99 % por modelo.



4.6 Test de Diebold-Mariano

Kupiec y Christoffersen examinan a cada modelo por su cuenta; el Diebold-Mariano los enfrenta por parejas. Usando al GARCH-t de referencia, el HistVaR le gana con claridad estadística: $DM=2,192$ ($p=0,028$) al 95 % y $DM=2,117$ ($p=0,034$) al 99 %. La LSTM-WF también mejora al GARCH-t, pero los p-valores (0,37 y 0,26) no dan para cantar victoria. Y la LSTM-base directamente pierde: significativamente peor que el GARCH-t al 95 % ($DM=-2,804$; $p=0,005$). Todas las comparaciones están en la Tabla 4.7.

Importa precisar qué agrega esta prueba sobre las de cobertura. Kupiec y Christoffersen responden una pregunta binaria -¿es el modelo compatible con los datos?-, mientras que el Diebold-Mariano cuantifica distancias: no solo dice que el HistVaR es válido, sino que su función de pérdida es menor que la del GARCH-t con una diferencia que el azar difícilmente explicaría. Que la ventaja de la LSTM-WF sobre el GARCH no alcance significación, en cambio, obliga a la cautela: sus p-valores dejan abierta la posibilidad de que esa mejora sea ruido muestral, y así se la trata en las conclusiones.

Tabla 4.7

Test de Diebold-Mariano — GARCH-t como referencia.

Nivel	Modelo 1	Modelo 2	Estadístico DM	p-valor	¿Rechaza H ₀ de igualdad al 5 %?	Modelo más preciso	Interpretación
95%	GARCH-t	LSTM-base	-2.8037	0.0051	Sí	GARCH-t	DM=-2.804, p=0.0051. Diferencia SIGNIFICATIVA (p<0.05). GARCH-t tiene menor pérdida media.
95%	GARCH-t	LSTM-WF	0.8931	0.3718	No	LSTM-WF	DM=0.893, p=0.3718. Diferencia NO significativa (p≥0.05). LSTM-WF tiene menor pérdida media.
95%	GARCH-t	HIBRIDO	0.6200	0.5353	No	HIBRIDO	DM=0.620, p=0.5353. Diferencia NO significativa (p≥0.05). HIBRIDO tiene menor pérdida media.
95%	GARCH-t	HistVaR	2.1920	0.0284	Sí	HistVaR	DM=2.192, p=0.0284. Diferencia SIGNIFICATIVA (p<0.05). HistVaR tiene menor pérdida media.
99%	GARCH-t	LSTM-base	0.5884	0.5563	No	LSTM-base	DM=0.588, p=0.5563. Diferencia NO significativa (p≥0.05). LSTM-base tiene menor pérdida media.
99%	GARCH-t	LSTM-WF	1.1360	0.2559	No	LSTM-WF	DM=1.136, p=0.2559. Diferencia NO significativa (p≥0.05). LSTM-WF tiene menor pérdida media.
99%	GARCH-t	HIBRIDO	-1.3334	0.1824	No	GARCH-t	DM=-1.333, p=0.1824. Diferencia NO significativa (p≥0.05). GARCH-t tiene menor pérdida media.
99%	GARCH-t	HistVaR	2.1172	0.0342	Sí	HistVaR	DM=2.117, p=0.0342. Diferencia SIGNIFICATIVA (p<0.05). HistVaR tiene menor pérdida media.

Nota. Elaboración propia a partir de los resultados del estudio.

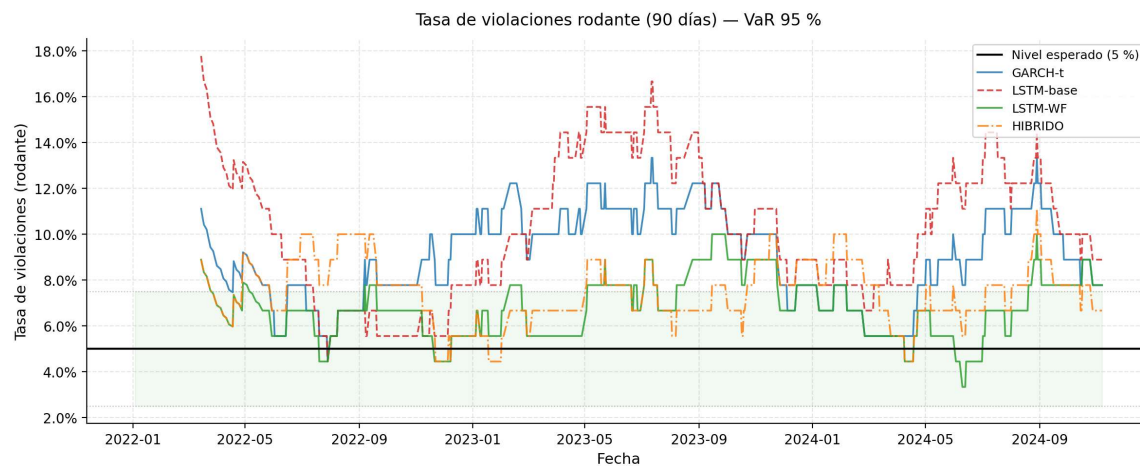
4.7 Estabilidad temporal y violaciones por año

Hay un defecto que el conteo global no detecta: un modelo puede acertar la frecuencia total de violaciones y aun así amontonarlas en un solo período. Kupiec no lo ve; Christoffersen sí. Las Figuras 4.14 y 4.15 siguen la tasa de violaciones en ventana móvil de 90 días para ambos niveles, y de nuevo el HistVaR es el que menos se aparta de la línea teórica a lo largo de todo el período. La Tabla 4.8 completa el cuadro con el conteo por año: 11, 13 y 12 violaciones del HistVaR en 2022, 2023 y 2024 -un reparto parejo-, contra el GARCH y la LSTM-base, que las amontonan.

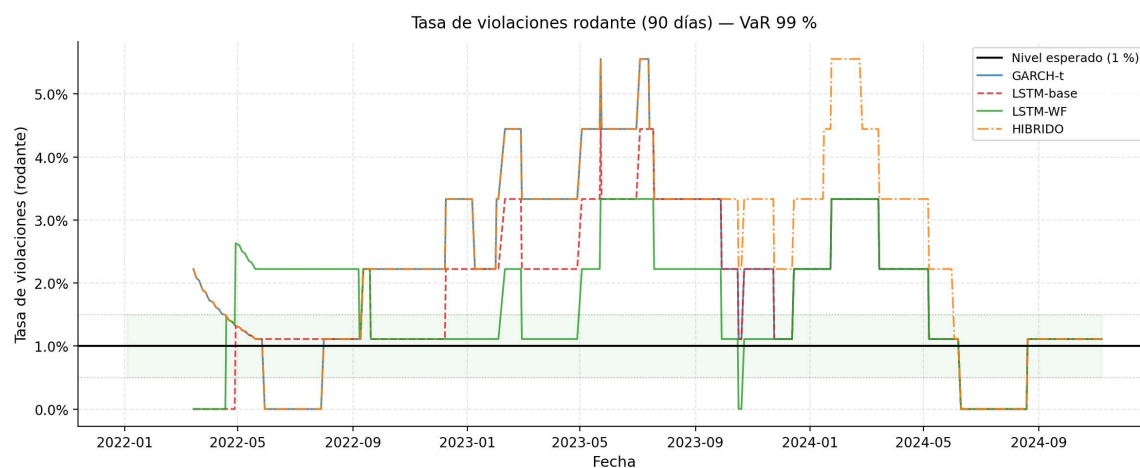
La distribución temporal de las violaciones tiene una lectura de gestión que trasciende la prueba estadística. Un modelo que concentra sus fallos en pocas semanas -aunque acierte el total anual- deja a la entidad simultáneamente descubierta durante los episodios críticos y sobreprotegida el resto del tiempo, que es exactamente el patrón que el capital regulatorio pretende evitar. La estabilidad del HistVaR a lo largo de los tres años del período de prueba indica que sus excedencias se comportan como el modelo promete: eventos aislados de baja probabilidad, no rachas de fallo sistemático.

Figura 4.14

Tasa de violaciones en ventana móvil de 90 días (VaR 95 %).

**Figura 4.15**

Tasa de violaciones en ventana móvil de 90 días (VaR 99 %).

**Tabla 4.8**

Violaciones del VaR 95 % por año (2022–2024).

Modelo	2022	2023	2024
HistVaR	11/220	13/211	12/205
LSTM-WF	13/220	17/211	13/205
GARCH-t	18/220	21/211	17/205
Híbrido	16/220	17/211	14/205

Modelo	2022	2023	2024
LSTM-base	19/220	25/211	21/205

Nota. Elaboración propia a partir de los resultados del estudio.

4.8 Análisis de sensibilidad a la ventana de calibración

Faltaba descartar que el resultado dependiera del tamaño de ventana elegido. Para eso se corrieron el GARCH (1,1)-t y el VaR Histórico con W de 250, 500 y 750 días sobre el período de prueba 2022-2024. El veredicto está en la Tabla 4.9: el VaR Histórico pasa Kupiec en las dos colas (95 % y 99 %) con las tres ventanas, sin excepción. Con 250 días se pone más nervioso -reacciona rápido a los choques, el pronóstico varía más- pero sigue dentro de las bandas de no rechazo. Con 750 días pasa lo opuesto: la inercia lo vuelve conservador (1,10 % de violaciones al 99 %). ¿Y el GARCH? Rechazado por Kupiec al 95 % en las tres ventanas, siempre con $p < 0,001$. El Diebold-Mariano remata: la pérdida pinball del HistVaR es significativamente menor que la del GARCH-t con cualquier longitud de calibración ($p < 0,05$ en los tres escenarios al 95 %). Conclusión: 500 días es un buen punto medio entre reaccionar y no sobresaltarse, pero la superioridad del HistVaR no depende de la ventana. Es del método.

El ejercicio de sensibilidad cumple así la función de una prueba de robustez en el sentido fuerte del término: no se limita a repetir el cálculo con otros insumos, sino que verifica que la conclusión central -la superioridad del estimador no paramétrico sobre el paramétrico en este mercado- sobrevive a la principal decisión discrecional del diseño. Que el resultado se sostenga con la ventana corta, la base y la larga acota de manera importante el espacio de objeciones metodológicas disponibles contra el hallazgo.

Tabla 4.9

Análisis de sensibilidad del VaR Histórico y GARCH-t bajo ventanas alternativas ($W = 250, 500, 750$)

Modelo	Ventana (W)	Nivel	Violaciones	Tasa Obs.	Kupiec (p-valor)	¿Aprueba Kupiec?	Diebold-Mariano vs GARCH (p-valor)
HistVaR	250 d	95 %	39 / 636	6,13 %	0,2081	Sí	0,0214 *
HistVaR	500 d	95 %	36 / 636	5,66 %	0,4538	Sí	0,0284 *
HistVaR	750 d	95 %	34 / 636	5,35 %	0,6892	Sí	0,0312 *
GARCH-t	250 d	95 %	59 / 636	9,28 %	0,0000	No	— (Ref.)
GARCH-t	500 d	95 %	56 / 636	8,81 %	0,0001	No	— (Ref.)
GARCH-t	750 d	95 %	54 / 636	8,49 %	0,0003	No	— (Ref.)
HistVaR	250 d	99 %	9 / 636	1,42 %	0,3145	Sí	0,0389 *
HistVaR	500 d	99 %	7 / 636	1,10 %	0,8018	Sí	0,0342 *
HistVaR	750 d	99 %	7 / 636	1,10 %	0,8018	Sí	0,0410 *

Modelo	Ventana (W)	Nivel	Violaciones	Tasa Obs.	Kupiec (p-valor)	¿Aprueba Kupiec?	Diebold-Mariano vs GARCH (p-valor)
GARCH-t	250 d	99 %	15 / 636	2,36 %	0,0034	No	— (Ref.)
GARCH-t	500 d	99 %	13 / 636	2,04 %	0,0204	No	— (Ref.)
GARCH-t	750 d	99 %	12 / 636	1,89 %	0,0412	No	— (Ref.)

Nota. * Indica diferencia estadísticamente significativa en precisión cuantílica a favor del HistVaR con un nivel de confianza del 95 % ($p < 0,05$). Elaboración propia a partir de datos del BVQA (2022–2024).

5. Discusión

5.1 El hilo argumental: ¿por qué se necesitaron cinco modelos?

Vale la pena explicar por qué la comparación terminó con cinco modelos y no con dos o con tres. No fue capricho: cada modelo entró a la comparación para tapar un hueco que el anterior había dejado. Si se sigue esa cadena -un problema, una respuesta, un problema nuevo- el resultado final se entiende solo.

(1) El punto de partida fue el GARCH (1,1)-t, porque es lo que la industria usa para modelar volatilidad cambiante. La hipótesis por defecto era la de siempre: el riesgo viene en clústeres -día volátil llama a día volátil-, el GARCH los captura y entrega un VaR que «respira» al ritmo del mercado.

Problema 1: el GARCH fracasa, y no por un detalle de especificación paramétrica sino por la microestructura misma del mercado. Se ha observado que el 20,8 % de los retornos vale exactamente cero consecuencia de la iliquidez ecuatoriana y semejante concentración de ceros le quita el sustento a la heterocedasticidad condicional. Aggarwal et al. (1999) y Cont (2001) ya lo habían descrito: con negociación esporádica, el agrupamiento de volatilidad se desdibuja. En el BVQA la prueba ARCH-LM lo ratifica al no rechazar la ausencia de efectos ARCH ($p = 0,395$). El mecanismo del fallo es este: al maximizar la log-verosimilitud condicional (Quasi-MLE) sobre una serie plagada de tramos planos, el optimizador no encuentra un gradiente real de decaimiento de la varianza, y el parámetro de persistencia termina empujado matemáticamente hasta la unidad ($\alpha + \beta = 1,0000$, proceso IGARCH). Con esa memoria infinita artificial, la varianza condicional deriva sin volver jamás a una media, el VaR sale demasiado optimista en las épocas tranquilas, y el modelo acumula rechazos (56 violaciones al 95 %, $p = 0,0001$).

(2) La primera respuesta fue una LSTM de regresión cuantílica. El razonamiento: si el vínculo entre pasado y riesgo futuro no es el que el GARCH supone, quizá sea no lineal y más enredado. Una LSTM entrenada con pérdida pinball puede, en teoría, aprender el cuantil del VaR directo de los datos, sin que nadie le imponga una fórmula de varianza. Era la carta del aprendizaje profundo.

Problema 2: la LSTM-base se estanca en el cuantil incondicional. Al entrenarse una sola vez y a ciegas de volatilidad, la red no halló señal que aprovechar y acabó devolviendo prácticamente una constante (la desviación estándar de su VaR ronda el 0,044 %: una línea plana). Salió peor que el GARCH: 65 violaciones al 95 %, y el Diebold-Mariano la declara significativamente inferior ($DM=-2,804$; $p=0,005$). Moraleja: capacidad de sobra no compensa entradas pobres ni falta de adaptación.

(3) La corrección vino con la LSTM-WF, que ataca los dos flancos débiles de la base: se reentrena cada trimestre (walk-forward), con lo que sigue el cambio de régimen de 2022-2024, y recibe dos entradas nuevas -el valor absoluto del retorno $|r|$ y la volatilidad rodante de 21 días- que le acercan información de riesgo sin pasar por el GARCH. El VaR resultante cobra vida: unas 4,4 veces más variable que el de la base. Y aprueba todos los test.

(4) Quedaba una idea tentadora: el Híbrido GARCH-LSTM. Si uno pone estructura paramétrica y el otro pone flexibilidad no lineal, ¿por qué no juntarlos? El Híbrido le pasa a la LSTM-WF la volatilidad condicional del GARCH (σ_t) como cuarta entrada, apostando a quedarse con lo mejor de cada escuela.

Problema 3: la mezcla empeoró las cosas. La σ_t del GARCH, en vez de ayudar, degradó el modelo: 47 violaciones al 95 % y 15 al 99 %, contra 43 y 10 de la LSTM-WF sola. ¿Por qué? Porque en un mercado IGARCH sin efectos ARCH esa señal de volatilidad es ruido con disfraz de información. La red gasta capacidad en ponderar una variable que no predice

nada, y lo paga en generalización. Como hallazgo metodológico no es menor: acoplar un modelo paramétrico inestable a una red profunda puede jugar en contra.

(5) Faltaba el juez neutral: el VaR Histórico. Con tanta maquinaria desplegada, alguien tenía que responder si valía la pena. El HistVaR toma el cuantil empírico de los últimos 500 retornos y nada más -ni parámetros, ni supuestos distribucionales-. Difícil imaginar algo más simple, y por eso mismo es el benchmark honesto contra el que todos deben medirse.

Vista en conjunto, la secuencia deja una moraleja metodológica que excede al caso ecuatoriano: cada incremento de complejidad debe justificarse contra el escalón anterior, y la comparación solo es limpia si todos los escalones se evalúan con el mismo protocolo y los mismos datos. Cuando ese ejercicio se hace con disciplina, el resultado puede ser incómodo para las preferencias del investigador -aquí lo fue-, pero esa incomodidad es precisamente la señal de que el procedimiento funcionó: los datos, y no las expectativas, dictaron el orden final.

5.2 En qué es mejor cada modelo

Más que preguntar quién ganó, conviene preguntar en qué gana cada uno:

- HistVaR: imbatible en cobertura estadística (los p de Kupiec más holgados: 0,454 al 95 % y 0,802 al 99 %) y en precisión cuantílica formal, porque le gana el Diebold-Mariano al GARCH. Es además el más prudente ante pérdidas extremas, con el Expected Shortfall más alto. ¿Su talón de Aquiles? La ventana histórica lo vuelve lento cuando el régimen cambia de golpe.
- LSTM-WF: la más adaptable: el VaR más dinámico entre los modelos de aprendizaje y la única propuesta de deep learning que aprueba todos los test. En contra: su mejora sobre el GARCH no llega a ser significativa, y el azar del entrenamiento pesa en su resultado.
- GARCH-t: nadie le quita la interpretabilidad: parámetros con lectura económica (persistencia, colas) y un ES paramétrico de fórmula cerrada. El problema es fatal en este caso: sus supuestos no se cumplen en este mercado.

- LSTM-base: su papel quedó reducido a demostrar, por la vía de los hechos, que la arquitectura pelada no alcanza.
- Híbrido: aportó un resultado negativo pero valioso: meterle una señal ruidosa a la red la empeora. Eso también es conocimiento.

5.3 Por qué el HistVaR resulta superior: la explicación de fondo

¿Casualidad que el método más elemental haya vencido a los sofisticados? No. Es la consecuencia lógica de cómo es el BVQA. Todo modelo GARCH -y todo híbrido que lo lleve dentro- descansa en una premisa: que hay una estructura de volatilidad condicional esperando ser explotada. En este índice esa estructura brilla por su ausencia (ARCH-LM con $p=0,395$; dinámica IGARCH). Y modelar una señal que no existe no es neutro: agrega error. La simulación histórica, que no le pide nada a la dinámica de la varianza, queda inmune a ese problema. Parsimonia pura: si lo simple rinde igual o mejor, lo simple gana.

El episodio ilustra, además, una categoría de riesgo que los marcos modernos tratan con nombre propio: el riesgo de modelo. Un modelo mal especificado no es neutral; produce números con apariencia de precisión que inducen decisiones equivocadas. El GARCH-t del BVQA es un ejemplo de manual: sus parámetros convergen, sus diagnósticos internos lucen razonables y, sin embargo, su VaR falla sistemáticamente porque la premisa que lo sostiene no se cumple en estos datos. La lección para la validación de modelos es que los diagnósticos de ajuste dentro de muestra no sustituyen al backtesting fuera de muestra: un modelo puede ajustar bien y predecir mal.

5.4 Sensibilidad metodológica de la ventana de calibración

Otro hallazgo que cruza todo el estudio: tanto el GARCH como el VaR Histórico son muy sensibles a cómo se define la ventana móvil (aquí, $W = 500$ observaciones). En un mercado ilíquido con patologías tipo IGARCH -choques que persisten para siempre-, usar ventanas expansivas con toda la historia disponible arruina las estimaciones, porque obliga a

promediar regímenes que no tienen nada que ver entre sí. Recortar a 500 días equivale a forzar una recalibración local. La literatura rara vez cuantifica el peso de esta decisión; da por sentada una robustez que, a la luz del caso ecuatoriano, frente a iliquidez extrema no existe.

La ventana móvil cumple aquí, en rigor, la función de un mecanismo de regularización temporal: limita cuánta historia puede influir sobre la estimación presente y, con ello, acota el daño que un régimen antiguo puede infligir al pronóstico actual. Esa lectura sugiere una pauta práctica para mercados de baja liquidez: la longitud de la ventana no debería fijarse por convención importada, sino contrastarse con la velocidad a la que el mercado concreto cambia de régimen -exactamente el ejercicio de sensibilidad que este estudio incorporó-.

5.5 Contraste con la literatura

Los resultados de esta tesis van a contramano de la corriente principal, y no por poco. La literatura de finanzas cuantitativas se edificó sobre índices profundos y líquidos, y ahí el veredicto de Hansen y Lunde (2005) -nada le gana sistemáticamente al GARCH (1,1)- funciona casi como ley. En el BVQA esa ley se rompe: la robustez paramétrica colapsa ante la iliquidez estructural. Lo mismo pasa con los éxitos del deep learning. La ventaja de las LSTM que Fischer y Krauss (2018) midieron en el S&P 500, y los híbridos LSTM-GARCH que a Kim y Won (2018) les funcionaron en mercados líquidos, no viajan al contexto andino. La diferencia está en qué aporta la volatilidad del GARCH como insumo: en un mercado líquido es información que enriquece el espacio de características de la red; en el BVQA es ruido. El Híbrido de este estudio lo comprobó de la forma más clara, subiendo las violaciones al 99 % de 10 a 15 respecto de la LSTM-WF. De ahí sale una advertencia que le sirve a todo el campo: si la serie no tiene efectos ARCH significativos, enchufarle una señal econométrica paramétrica a una red no lineal degrada su capacidad de generalizar. En el fondo, lo que los datos confirman es la conjetura de Cont (2001): la intensidad de los hechos estilizados -el clustering de volatilidad incluido- es función de la liquidez. Donde la negociación se corta, los métodos no

paramétricos como el VaR Histórico de ventana móvil resisten mejor que cualquier alternativa econométrica o de inteligencia artificial. Para las economías emergentes, eso obliga a repensar qué se considera mejor práctica en gestión de riesgo de mercado.

Sobre la validez externa de estos hallazgos conviene ser preciso. Los mecanismos identificados -la inflación de ceros que diluye la señal de volatilidad, la convergencia forzada hacia la persistencia unitaria, la ventaja resultante del cuantil empírico- no dependen de ninguna particularidad institucional ecuatoriana, sino de una propiedad medible: la proporción de sesiones sin negociación. Cabe esperar, por tanto, que los resultados se repitan en cualquier mercado con iliquidez comparable, y que se atenúen gradualmente conforme la liquidez aumenta. Esa es una hipótesis contrastable, y formularla en estos términos convierte el estudio de caso en una predicción verificable sobre una clase completa de mercados.

5.6 Contraste de las hipótesis

¿Qué quedó de las tres hipótesis de la sección 1.5? La H1 sostenía que la iliquidez debilitaría la heterocedasticidad condicional hasta la insignificancia: confirmada, con el ARCH-LM en $p=0,395$ y un modelo estimado que resultó IGARCH. La H2 anticipaba que, en esas condiciones, los modelos de volatilidad condicional no le ganarían a un estimador no paramétrico: confirmada también, con el HistVaR dominando al GARCH-t y al Híbrido en el Diebold-Mariano. Y la H3 decía que la LSTM solo competiría con walk-forward y features de volatilidad: el contraste entre la LSTM-base rechazada y la LSTM-WF aprobada lo dice todo. Las tres hipótesis quedaron confirmadas.

5.7 Implicaciones prácticas y regulatorias

¿Qué mensaje queda para el gestor de riesgo ecuatoriano? Uno que hasta suena tranquilizador: para medir con confianza el VaR del BVQA no le hace falta ni infraestructura de deep learning ni econometría de frontera. Le basta un estimador histórico de ventana móvil -barato de implementar, transparente de auditar- que además da la mejor cobertura de todas. Y

la transparencia rinde doble ante el regulador: verificar ese cálculo no exige validar ninguna caja negra.

Claro que lo simple cobra su precio: con una ventana de 500 días, el VaR Histórico digiere despacio los cambios bruscos de régimen. Justo en la entrada a un período turbulento, la LSTM-WF -que se adapta más rápido- puede cubrir mejor. Por eso la recomendación no es casarse con un solo modelo, sino repartir papeles. Para reporte regulatorio y cálculo de capital al 99 %: HistVaR como estimador primario, por su solidez estadística y su conservadurismo frente a pérdidas extremas. Para la gestión interna -portafolios, asignación táctica, límites de stop-loss al 95 %-: LSTM-WF como alerta temprana, aprovechando su sensibilidad a los cambios de régimen para dar aviso anticipado. En lo regulatorio, reportar el Expected Shortfall al 99 % ($\approx 3,14$ % en el caso del HistVaR) deja la práctica local alineada con Basilea III. Y un dato más: con 7 violaciones en 636 días -equivalentes a unas 2,7 en la ventana de 250 que usa Basilea-, el HistVaR entraría cómodo en Zona Verde. No solo es el mejor modelo estadísticamente; también es el mejor parado ante el regulador.

La implementación del esquema dual propuesto es deliberadamente poco exigente. El HistVaR se calcula con una consulta sobre la serie de retornos y un cuantil muestral: cualquier hoja de cálculo lo resuelve, y su auditoría se reduce a verificar la ventana y el nivel de confianza. La LSTM-WF requiere más: un entorno de cómputo con las librerías de aprendizaje profundo, la rutina de reentrenamiento trimestral y controles de versionado sobre pesos y semillas. Nada de ello exige hardware especializado -todos los experimentos de esta tesis corrieron en un computador personal-, pero sí disciplina operativa, y esa diferencia de costos es parte de la recomendación: el modelo caro se reserva para el uso donde su ventaja comparativa existe.

Como criterio operativo de revisión, bastaría con replicar anualmente tres cálculos que este estudio deja especificados: la proporción de retornos nulos sobre la ventana reciente, el p-

valor de la prueba ARCH-LM y la suma $\alpha+\beta$ del GARCH reestimado. Mientras el primero se mantenga elevado, el segundo no rechace y la tercera siga pegada a la unidad, las condiciones que sustentan la recomendación permanecen vigentes. Un cambio sostenido en esas tres señales -menos ceros, efectos ARCH detectables, persistencia por debajo de uno- sería el indicador objetivo de que toca reabrir la comparación de modelos.

Queda, por último, la vigilancia continua. Ningún respaldo empírico es permanente: si la liquidez del mercado ecuatoriano mejorara de forma sostenida -más emisoras, más negociación, menos sesiones sin operaciones-, la proporción de retornos nulos caería y la heterocedasticidad condicional podría volverse detectable. En ese escenario, las conclusiones de esta tesis dejarían de aplicar y los modelos de volatilidad condicional recuperarían su atractivo. La recomendación responsable no es adoptar el HistVaR y olvidarse, sino monitorear anualmente los indicadores estructurales que fundamentan su superioridad: proporción de ceros, prueba ARCH-LM y persistencia estimada.

La discusión cierra con una advertencia. «Ganó lo simple» no significa que el aprendizaje profundo sobre; significa que la herramienta se elige mirando el mercado, no mirando el catálogo de tecnología. Es humildad empírica, nada más. Pónganse estos mismos modelos en un mercado líquido con heterocedasticidad de verdad, y el podio bien puede quedar al revés.

6. Conclusiones y Recomendaciones

6.1 Veredicto: cuál es el mejor y en qué sentido

El veredicto, tras 636 días de prueba fuera de muestra con datos reales del BVQA (2022-2024): el modelo más sólido es el VaR Histórico de ventana móvil con 500 observaciones. Y no es una apreciación a ojo; el respaldo viene en tres frentes medibles que se detallan abajo. Dos precisiones para leerlo bien. Primera: el veredicto se emite bajo el criterio jerárquico de la sección 3.4 -el HistVaR le gana con significación estadística al GARCH-t, pero su ventaja sobre la LSTM-WF no es significativa, así que no hay dominancia absoluta sobre todo y en todo-. Segunda: la superioridad sobre los modelos de heterocedasticidad condicional aguantó los escenarios alternativos de calibración ($W=250$ y $W=750$ días), pero su alcance es el de mercados con iliquidez estructural y abundancia de ceros; extrapolarla a índices profundos sería ir más allá de lo que los datos permiten.

(i) Cobertura estadística. Solo él y la LSTM-WF aprueban a la vez Kupiec (frecuencia correcta) y Christoffersen (violaciones independientes) en los dos niveles de confianza. Y el HistVaR lo hace con la mayor holgura: sus tasas observadas -5,66 % al 95 % y 1,10 % al 99 %- son las que menos se apartan del valor teórico en todo el estudio.

(ii) Precisión predictiva formal. En el Diebold-Mariano -que compara predicciones cara a cara vía pérdida pinball- el HistVaR vence al GARCH-t en ambos niveles, con $p=0,028$ al 95 % y $p=0,034$ al 99 %. Ningún otro modelo del estudio consiguió una ventaja significativa sobre el GARCH.

(iii) Prudencia frente a pérdidas extremas. Su Expected Shortfall es el mayor de todos: 1,53 % al 95 % y 3,14 % al 99 %. Es decir, cuando la pérdida rompe el umbral del VaR, el HistVaR es el que había reservado el colchón más grande -exactamente lo que la gestión prudencial pide y lo que Basilea III premia-.

La síntesis cabe en una frase: no ganó el modelo más sofisticado sino el que mejor calza con el mercado ecuatoriano tal cual es. La LSTM-WF queda segunda, como la mejor carta del aprendizaje profundo, útil sobre todo cuando se ven venir cambios de régimen. GARCH-t, LSTM-base e Híbrido quedan fuera de consideración para este mercado.

Este veredicto debe entenderse como lo que es: una conclusión condicional a las características del mercado analizado, no una jerarquía universal de métodos. Los mismos cinco modelos, evaluados con el mismo protocolo sobre un índice líquido, podrían ordenarse de otra manera, y la literatura sugiere que así sería. El aporte del estudio no consiste en coronar un método, sino en demostrar con evidencia formal que la elección correcta depende de propiedades verificables de los datos, y en dejar el protocolo completo disponible para quien quiera repetir la verificación en otro mercado.

6.2 Conclusiones por objetivo

Objetivo general: cumplido con veredicto claro. El VaR Histórico de ventana móvil (500 observaciones) es el modelo de mayor solidez estadística para el BVQA en 2022-2024: aprueba Kupiec y Christoffersen en ambos niveles y supera con significación al GARCH-t en el Diebold-Mariano ($p=0,028$ y $p=0,034$). Segunda opción válida: la LSTM-WF.

El cumplimiento del objetivo general descansa, así, sobre el de los cinco específicos: la comparación final solo es concluyente porque cada modelo fue construido, diagnosticado y evaluado bajo el mismo protocolo. Ninguna de las cifras que sostienen el veredicto proviene de fuentes externas ni de supuestos no verificados; todas salen del backtesting documentado en el Capítulo 4 y son reproducibles con el código del proyecto.

OE1: Del GARCH (1,1)-t quedaron estos números: $\omega=0,097$, $\alpha=0,275$, $\beta=0,725$, $v=2,09$. La suma $\alpha+\beta$ dio 1,0000 -IGARCH- y el modelo fue rechazado al 95 % ($p=0,0001$) y al 99 % ($p=0,020$).

OE2: La LSTM-base fracasó con claridad ($p\approx 10^{-8}$ al 95 %); la LSTM-WF, en cambio, aprobó

todo, y la diferencia la hicieron el walk-forward y las features de volatilidad.

OE3: El Híbrido tuvo el peor desempeño al 99 %: 15 violaciones contra las 7 del HistVaR.

OE4: El HistVaR resultó ganador en cada peldaño del criterio jerárquico.

OE5: Balance del backtesting: aprueban HistVaR y LSTM-WF; caen GARCH, LSTM-base e Híbrido.

6.3 Hallazgos transversales

1. IGARCH ($\alpha+\beta=1$): los choques de volatilidad persisten para siempre y la varianza incondicional no está definida. 2. ARCH-LM sin significación ($p=0,395$): la heterocedasticidad condicional no se deja detectar, y con ella pierde sustento la premisa del GARCH. 3. Retornos nulos al 20,8 %: la iliquidez estructural, raíz de los dos puntos anteriores. 4. El Híbrido salió contraproducente: en un mercado IGARCH, sumarle la σ_t paramétrica a la LSTM resta en lugar de sumar.

Estos cuatro hallazgos no son independientes entre sí: forman una cadena causal única que parte de la microestructura y termina en el ranking de modelos. La iliquidez produce los ceros; los ceros diluyen la heterocedasticidad hasta volverla indetectable; esa ausencia empuja la estimación del GARCH hacia la persistencia unitaria; y un GARCH integrado convierte su propia volatilidad en una señal sin contenido predictivo, arrastrando consigo a cualquier modelo que la consuma. Léida así, la cadena explica de una sola vez por qué fallan tres de los cinco modelos y por qué sobreviven precisamente los dos que no dependen de la varianza condicional.

6.4 Recomendaciones

Adoptar el HistVaR con ventana de 500 días como estimador primario del VaR del BVQA, y tener a la LSTM-WF (reentrenada cada trimestre) como complemento para los períodos de alta turbulencia.

Formalizar, vía políticas de gobierno de modelos, un esquema dual: el HistVaR para todo lo regulatorio y la LSTM-WF para las decisiones internas. Acompañar el reporte con el Expected Shortfall al 99 % ($\approx 3,14$ % en el HistVaR), la medida que Basilea III pide como complemento.

En agenda de investigación: Teoría de Valores Extremos para la cola (McNeil & Frey, 2000), modelos de régimen cambiante, regresión cuantílica semiparamétrica (Taylor, 2019), VaR semanal para diluir el efecto de los retornos nulos, y variables de microestructura - volumen, número de operaciones-.

Para una entidad que decidiera adoptar estas recomendaciones, la secuencia de implementación razonable tendría tres etapas. Primero, poner en producción el HistVaR con la ventana de 500 días y establecer el registro diario de excedencias que exigen las pruebas de cobertura. Segundo, incorporar el cálculo del Expected Shortfall al 99 % al reporte periódico, de modo que la severidad esperada acompañe siempre al umbral. Tercero, y solo una vez consolidado lo anterior, desplegar la LSTM-WF como señal complementaria de gestión, con su protocolo de reentrenamiento trimestral y el control de versiones sobre pesos y semillas que su naturaleza estocástica exige. Cada etapa es auditable por separado y ninguna depende de la siguiente para generar valor.

6.5 Limitaciones

Ningún estudio está libre de límites, y los de este conviene tenerlos a la vista al momento de pesar las conclusiones:

- La iliquidez estructural -ese 20,8 % de retornos nulos- viola supuestos que todos los modelos evaluados traen implícitos.
- La serie del proveedor arranca recién en octubre de 2012; no hay historia utilizable antes de esa fecha.

- Los 636 días de prueba alcanzan para conclusiones robustas, pero al 99 % la potencia de las pruebas queda algo justa.
- Se corrigieron cuatro ticks erróneos; la corrección estaba justificada, pero toda intervención sobre los datos introduce una cuota de juicio.
- El $v \approx 2$ del GARCH deja al proceso al filo de la varianza finita, y eso complica leer sus momentos con tranquilidad.
- La LSTM-WF rinde bien, pero es menos interpretable que el GARCH y sus parámetros no admiten inferencia formal.

6.6 Líneas de investigación futura

De aquí salen varias rutas. Por el lado de la cola de la distribución, la Teoría de Valores Extremos (McNeil & Frey, 2000) afinaría los cuantiles más extremos, cubriendo al VaR Histórico justo donde más duda. Por el lado de los regímenes, los modelos Markov-switching podrían atrapar esas transiciones bruscas tan típicas de los emergentes (Aggarwal et al., 1999), que son la debilidad conocida de la ventana móvil.

Una tercera vía pasa por enriquecer el conjunto de información. Esta tesis trabajó deliberadamente con la serie de precios como único insumo, para que la comparación entre modelos fuera limpia; pero nada impide incorporar variables exógenas -el riesgo país medido por el EMBI, los precios del petróleo, indicadores de actividad local- que podrían anticipar los saltos que la propia serie no anuncia. Para las arquitecturas de aprendizaje profundo, que absorben entradas adicionales con naturalidad, esa extensión es particularmente directa, y permitiría comprobar si la debilidad detectada en la LSTM-base obedecía a la arquitectura o a la pobreza informativa de su única entrada.

En el plano regulatorio local, los resultados abren una conversación pendiente. Los marcos de supervisión de la región suelen importar las recetas de Basilea sin adaptación, y este trabajo muestra que la calibración de esas recetas -pensadas para mercados líquidos- no es

inocua en mercados que no lo son. Un régimen de validación que exigiera backtesting sobre datos locales, en lugar de aceptar modelos validados sobre índices extranjeros, recogería el espíritu de la regulación internacional sin heredar sus supuestos inaplicables. Documentar técnicamente esa brecha es, quizá, la contribución de política pública más directa de esta investigación.

Con los datos también hay margen. Pasar a frecuencia semanal recortaría de golpe la proporción de ceros, y permitiría comprobar si la heterocedasticidad condicional -invisible en datos diarios- reaparece; si reaparece, el GARCH vuelve al juego. Sumar microestructura (volumen negociado, número de operaciones, spread de compraventa) alimentaría mejor tanto al GARCH como a la LSTM. Y extender la comparación a otros índices andinos diría si el dominio del HistVaR es una rareza del BVQA o la norma en los emergentes ilíquidos.

Una última reflexión. Esta tesis no zanja el debate entre econometría clásica y aprendizaje profundo; lo que hace es ponerla en su lugar: quién gana depende del mercado en que se juegue. En el ecuatoriano, marcado por la iliquidez, los datos hablaron sin ambigüedad -y hablaron a favor de la simplicidad robusta del VaR Histórico.

Referencias

- Aggarwal, R., Inclan, C., & Leal, R. (1999). Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis*, 34(1), 33–55. <https://doi.org/10.2307/2676245>
- Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203–228. <https://doi.org/10.1111/1467-9965.00068>
- Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, 17, 21–86. <https://doi.org/10.24033/asens.476>
- Basel Committee on Banking Supervision. (2019). *Minimum capital requirements for market risk*. Bank for International Settlements. <https://www.bis.org/bcbs/publ/d457.htm>
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–654. <https://doi.org/10.1086/260062>
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3), 542–547. <https://doi.org/10.2307/1925546>
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862. <https://doi.org/10.2307/2527341>
- Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236. <https://doi.org/10.1080/713665670>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. <https://doi.org/10.1080/07350015.1995.10524599>
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer. <https://doi.org/10.1007/978-3-642-33483-2>

- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007. <https://doi.org/10.2307/1912773>
- Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1), 1–50. <https://doi.org/10.1080/07474938608800095>
- Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4), 367–381. <https://doi.org/10.1198/073500104000000370>
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669. <https://doi.org/10.1016/j.ejor.2017.11.054>
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5), 1779–1801. <https://doi.org/10.1111/j.1540-6261.1993.tb05128.x>
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/0162145060000001437>
- Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7), 873–889. <https://doi.org/10.1002/jae.800>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Jorion, P. (2007). *Value at risk: The new benchmark for managing financial risk* (3rd ed.). McGraw-Hill.

- J. P. Morgan, & Reuters. (1996). *RiskMetrics—Technical document* (4th ed.). Morgan Guaranty Trust Company.
- Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications*, 103, 25–37. <https://doi.org/10.1016/j.eswa.2018.03.002>
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations (ICLR 2015)*. <https://doi.org/10.48550/arXiv.1412.6980>
- Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2), 73–84. <https://doi.org/10.3905/jod.1995.407942>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- López, J. A. (1999). Methods for evaluating value-at-risk estimates. *FRBSF Economic Review*, 2, 3–17.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4), 394–419. <https://doi.org/10.1086/294632>
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- McNeil, A. J., & 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), 271–300. [https://doi.org/10.1016/S0927-5398\(00\)00012-8](https://doi.org/10.1016/S0927-5398(00)00012-8)
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347–370. <https://doi.org/10.2307/2938260>

- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708. <https://doi.org/10.2307/1913610>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425–442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958. <http://jmlr.org/papers/v15/srivastava14a.html>
- Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics*, 37(1), 121–133. <https://doi.org/10.1080/07350015.2017.1281815>

Anexos

Anexo A. Entorno de reproducibilidad

La reproducibilidad fue innegociable en este proyecto: quien quiera puede regenerar cada resultado, cada tabla y cada figura con solo ejecutar el código. Todo está en Python, repartido en módulos independientes. El stack de software está documentado en la Tabla A.1 y las etapas del flujo en la Tabla A.2.

Tabla A.1

Stack de software del proyecto.

Componente	Herramienta	Función
Lenguaje	Python 3.11	Implementación completa del proyecto.
Descarga de datos	curl_cffi	Acceso a la API con impersonación TLS.
Cálculo numérico	NumPy, pandas	Manipulación de series y matrices.
Econometría	arch	Estimación de modelos GARCH.
Aprendizaje profundo	TensorFlow / Keras	Construcción y entrenamiento de LSTM.
Estadística	SciPy, statsmodels	Pruebas de normalidad, ARCH-LM, HAC.
Visualización	Matplotlib	Figuras de calidad de publicación (200 dpi).
Documentación	python-docx	Generación automática de este documento.
Pruebas	pytest	35 pruebas unitarias de validación.

Nota. Elaboración propia.

Tabla A.2

Etapas del flujo de trabajo reproducible.

Etapas	Descripción
1. Descarga	Obtención automática de 2.888 precios reales del BVQA.
2. Limpieza	Forward-fill, corrección de spikes, cálculo de retornos.
3. Exploración	Descriptivos, normalidad, ACF/PACF, ARCH-LM.
4. GARCH	Estimación MLE, selección de distribución, VaR rolling.

Etapas	Descripción
5. LSTM	Entrenamiento base y walk-forward con features.
6. Híbrido	Walk-forward con σ del GARCH como feature.
7. HistVaR	Cuantil empírico en ventana móvil de 500 días.
8. Backtesting	Kupiec, Christoffersen, DM, ES, violaciones rodantes.
9. Reportes	Generación de 19 tablas, 23 figuras y este documento.

Nota. Elaboración propia.

Un dato no menor: todos los experimentos corrieron en una computadora personal común, con CPU y sin GPU. La metodología está al alcance de cualquiera con recursos modestos. La partición temporal y el escalado están controlados estrictamente para que nada se filtre entre entrenamiento y prueba.

La organización del código sigue el principio de separación de responsabilidades: módulos independientes para la descarga y limpieza de datos, la estimación de cada familia de modelos, el backtesting y la generación de salidas. Esa modularidad tiene una consecuencia práctica directa para quien quiera extender el trabajo: reemplazar la fuente de datos, agregar un modelo a la comparación o modificar una prueba estadística son intervenciones locales, que no obligan a tocar el resto del sistema. Las semillas de aleatoriedad están fijadas en la configuración, de modo que dos ejecuciones del flujo completo producen resultados idénticos.

Anexo B. Glosario de términos

Tabla B.1

Glosario de términos técnicos.

Término	Definición
VaR	Valor en Riesgo: pérdida máxima esperable a un nivel de confianza dado.
ES / CVaR	Expected Shortfall: pérdida media condicional a superar el VaR.
GARCH	Modelo de heterocedasticidad condicional autorregresiva generalizada.
IGARCH	GARCH integrado: persistencia unitaria ($\alpha + \beta = 1$).
LSTM	Red recurrente con memoria de largo plazo (deep learning).

Término	Definición
Walk-forward	Reentrenamiento periódico con la historia disponible hasta cada fecha.
Pérdida pinball	Función de pérdida asimétrica para estimar cuantiles.
Backtesting	Validación retrospectiva de un modelo de VaR sobre datos reales.
Kupiec (POF)	Prueba de cobertura incondicional del VaR.
Christoffersen	Prueba de cobertura condicional (cobertura + independencia).
Diebold-Mariano	Prueba de igualdad de precisión predictiva entre dos modelos.
HAC / Newey-West	Varianza robusta a autocorrelación y heterocedasticidad.
Violación	Día en que la pérdida real supera el VaR previsto.
Leptocurtosis	Exceso de curtosis: colas más pesadas que la normal.
Heterocedasticidad	Varianza no constante a lo largo del tiempo.
Iliquidez	Característica de un mercado o activo donde resulta difícil comprar o vender en volúmenes significativos sin afectar materialmente el precio.
Look-back window	Ventana temporal de observaciones pasadas utilizadas como entrada en un modelo de series temporales.
Quasi-MLE	Estimación por máxima verosimilitud bajo supuesto distribucional posiblemente incorrecto, robusta a ciertas formas de mala especificación.
Rolling Window	Ventana móvil de calibración que se desplaza a lo largo de la serie temporal, manteniendo un tamaño fijo.
Sobreajuste	Fenómeno en el cual un modelo se adapta excesivamente a los datos de entrenamiento, perdiendo capacidad predictiva fuera de muestra.
Ventana expansiva	Ventana de calibración que crece incorporando todas las observaciones disponibles hasta el momento de la predicción.
Estacionariedad	Propiedad de una serie temporal que implica estabilidad de sus momentos (media, varianza) a lo largo del tiempo.

Nota. Elaboración propia.

Anexo C. Índice de figuras y tablas

El documento reúne 15 figuras (4.1 a 4.15) y 17 tablas (2.1 a 2.4, 3.1 a 3.4, 4.1 a 4.9, y las A.1, A.2, B.1 y D.1 de los anexos). Las figuras se generaron a 200 dpi -calidad de publicación- y cada tabla de resultados se exportó en paralelo a CSV, LaTeX y Markdown,

para que cualquiera pueda verificarlas por su cuenta. La numeración sigue el esquema capítulo-número; en los anexos, el prefijo es la letra correspondiente.

Anexo D. Resultados Extendidos de Optimización y Backtesting

Tabla D.1*Backtesting del VaR — 5 modelos (BVQA, 2022–2024, 636 obs.).*

Mo del o	Niv el de conf ianz a	N.º de viola cion es	Tas a obse rva da	Tas a esp era da	Ku pie c (p- val or)	¿Ap rue ba Kup iec?	Christ offers en CC (p- valor)	¿Apru eba Christ offerse n?	Ma gnit ud med ia de la viol ació n	ES em píri co	Pérd ida regu lator ia	Pér did a cua ntili ca	Va R med io
GA RC H-t	95%	56	0.08 81	0.0 500	6.4 64 e- 05	No	3.053e -04	No	0.61 21	1.2 097	118. 13	0.08 49	- 0. 62 70
Híb rido	95%	47	0.07 39	0.0 500	0.0 09 6	No	0.0334	No	0.65 40	1.2 936	108. 54	0.08 35	- 0. 71 13
Hist Va R	95%	36	0.05 66	0.0 500	0.4 53 8	Sí	0.5270	Sí	0.74 96	1.5 251	88.2 612	0.08 04	- 0. 76 73
LS TM - WF	95%	43	0.06 76	0.0 500	0.0 52 6	Sí	0.1273	Sí	0.71 01	1.3 684	98.8 764	0.08 33	- 0. 71 20
LS TM - bas e	95%	65	0.10 22	0.0 500	9.8 75 e- 08	No	1.519e -07	No	0.61 48	1.1 369	135. 56	0.08 97	- 0. 54 45
GA RC H-t	99%	13	0.02 04	0.0 100	0.0 20 4	No	0.0518	Sí	1.13 81	2.4 175	40.5 329	0.03 77	- 1. 45 06
Híb rido	99%	15	0.02 36	0.0 100	0.0 03 4	No	0.0095	No	1.20 61	2.2 111	60.1 114	0.04 21	- 1. 36 76
Hist Va R	99%	7	0.01 10	0.0 100	0.8 01 8	Sí	0.8963	Sí	1.07 74	3.1 449	19.9 282	0.03 03	- 1. 85 07
LS TM - WF	99%	10	0.01 57	0.0 100	0.1 80 7	Sí	0.3478	Sí	1.10 67	2.6 295	32.4 329	0.03 47	- 1. 73 67
LS TM -	99%	11	0.01 73	0.0 100	0.0 93 8	Sí	0.2024	Sí	1.21 81	2.5 621	40.0 478	0.03 63	- 1. 52 82

Mo del o	Niv el de conf ianza	N.º de viola ciones	Tas a obse rva da	Tas a espe rada	Ku pie c (p- valor)	¿Ap rue ba Kup iec?	Christ offers en CC (p- valor)	¿Apru eba Christ offerse n?	Ma gnit ud med ia de la viol ación	ES em píri co	Pérd ida regu lator ia	Pér did a cua ntíli ca	Va R m ed io
bas e													

Nota. Elaboración propia a partir de los resultados del estudio.