

TRAVIDENCE

D1 — Technical Methodology Paper

Pre-validación con grado de auditoría de estrategias de trading: cinco contribuciones a la metodología López de Prado

Ignacio Arias · Travidence, LLC

ORCID 0009-0001-7503-3830

Versión 1.0 — agosto de 2026

DOI: 10.2139/ssrn.7308022

Abstract

La validación previa al desafío de prop firm es un punto ciego estructural del trading retail contemporáneo. FPF Tech documenta sobre 300,000+ cuentas que el 86% de los challengers no supera el desafío y solo el 7% recibe un pago — evidencia sistemática de sobreajuste que las herramientas existentes no resuelven con rigor de auditoría. Las librerías de implementación que implementan el stack López de Prado (mlfinlab, RiskLabAI) operan como infraestructura técnica sin emitir certificación; la oferta retail consumer de agregación pura entrega un score agregado sin gates ni disciplina procesal. Las firmas de prop trading, simultáneamente, mantienen un conflicto de interés estructural que les impide ofrecer pre-validación honesta. Este vacío institucional motiva una metodología independiente de tercera parte con grado de auditoría.

El presente documento presenta cinco contribuciones novel al stack López de Prado: (i) reporte dual de n_{trials} en el Sharpe Deflactado para taxonomizar el sesgo de selección; (ii) Calmar normalizado por capital que elimina la inflación por apalancamiento marginal; (iii) esquema sub-período con detección de decaimiento monotónico; (iv) test de aislamiento de componentes para estrategias multi-componente; (v) taxonomía MUST-PASS de regímenes vol-on / vol-off. Las cinco se integran en el Score de Robustez — un score híbrido continuo-puerta sobre $[0, 100]$, inspirado en la disciplina procesal de la industria auditora profesional.

La metodología se aplica a tres estrategias textbook sobre data pública: Golden Cross en crudo CL (2000–2025), RSI-14 en BTC perpetuo (2020–2025) y Asian Range Breakout en EUR/USD (2015–2025). Ninguna alcanza banda Robusta — consistente con la tasa documentada del 86%. Bajo el engine de auditoría de producción (retornos a nivel de trade, corte de régimen p50 auto-derivado, floors de admisibilidad), CL y BTC cierran con Abstención de Opinión por muestra insuficiente (18 y 31 trades), mientras EUR/USD emite el score público end-to-end del framework: 56 / 100, Robustez Limitada, sin gate vinculante. El comportamiento diferencial gate-vs-continuous frente a la agregación pura, que define el diferenciador metodológico defendible de la pre-validación con grado de auditoría, lo exhibe el caso NQ Opening Range Breakout del repositorio — admitido con 2,424 trades y capado por los gates de régimen y sub-período; a través de los cinco case studies del repositorio, el portafolio cubre los tres desenlaces públicos del framework: score emitido, gated e insufficient sample.

Palabras clave: sobreajuste de backtest; Sharpe Deflactado; Validación Cruzada Combinatoria Purgada; Probabilidad de Sobreajuste; pre-validación prop firm; Calmar normalizado por capital; asimetría de régimen; aislamiento de componentes; metodología con grado de auditoría; Score de Robustez.

JEL: G11, G14, G17, C12, C52. **MSC:** 62F03, 62P05, 91G70.

§1. Introducción

§1.1 El problema del 86%

La industria de prop firm challengers opera sobre un perfil agregado de outcomes que la literatura académica no ha incorporado plenamente al cuerpo de teoría sobre sobreajuste de estrategias.

Datos consolidados sobre más de 300,000 cuentas administradas por FPF Tech, reportados por Finance Magnates (2024), documentan que el 86% de los participantes del challenge no logra superar la evaluación inicial y solo el 7% recibe efectivamente un pago al final del ciclo de funding. La diferencia entre ambos porcentajes corresponde a participantes que pasan el challenge pero nunca alcanzan un retiro.

Este patrón es inconsistente con la hipótesis de selección adversa benigna y consistente, en cambio, con la hipótesis de selección por sobreajuste: una fracción mayoritaria de los challengers entra al desafío con estrategias cuyas métricas in-sample son producto del backtest overfitting documentado en Bailey, Borwein, López de Prado y Zhu (2017). El cliente paga la entry fee, fracasa, y la firma absorbe el ingreso sin contraparte de payout.

El problema no es la existencia del challenge ni la asimetría aritmética entre fees y payouts. El problema es la ausencia de un mecanismo independiente que permita al challenger evaluar, antes de pagar la fee, si su estrategia tiene edge genuino o si sus métricas in-sample son artefactos estadísticos. Las herramientas con las que cuenta hoy el challenger retail son insuficientes en dimensiones específicas que la siguiente sección detalla.

§1.2 El vacío de validación

Tres clases de actor ocupan parcialmente el espacio de validación cuantitativa, pero ninguno provee un servicio independiente con grado de auditoría.

Las librerías de implementación — principalmente `mlfinlab` (López de Prado et al.) y `RiskLabAI` — implementan en código la metodología del Deflated Sharpe Ratio, la validación cruzada combinatoria purgada (CPCV) y la probabilidad de backtest overfitting (PBO). Son infraestructura técnica de alta calidad, pero por construcción no emiten certificación: producen métricas que el propio analista interpreta y reporta. No hay tercera parte que firme el dictamen.

La oferta retail consumer ofrece análisis algorítmicos de bajo costo, emitiendo un score agregado en escala $[0, 100]$ sin gates duros ni disciplina procesal. La agregación pura es la decisión metodológica central de esa oferta y al mismo tiempo su debilidad estructural para uso en pre-validación de challenge: una estrategia con asimetría severa de régimen puede acumular un score numérico aceptable porque las dimensiones favorables compensan las desfavorables, sin que el sistema detenga la certificación. No existe el equivalente de un cliff-edge audit failure.

Las firmas de prop trading mismas están estructuralmente impedidas de ofrecer pre-validación honesta. Su modelo de negocio depende del flujo de fees de evaluación; una pre-validación efectiva reduciría el universo de challengers dispuestos a intentar el desafío — y con ello los ingresos de la firma. El conflicto de interés es indistinguible al cliente y no resoluble vía disclosure: una firma que vende challenge slots no puede simultáneamente avisar al cliente que su estrategia no es viable.

El vacío resultante es preciso. Existe una franja de servicio entre las librerías de implementación (que requieren competencia técnica del analista y no emiten dictamen) y la oferta consumer (que emite dictamen agregado pero sin disciplina de audit-grade). Esa franja, ocupada en otros mercados regulados por firmas de auditoría independiente, no tiene incumbente equivalente en el mercado de pre-validación de estrategias cuantitativas.

§1.3 Contribución del presente documento

TRAVIDENCE se posiciona deliberadamente en la franja descrita. La propuesta metodológica del presente documento se estructura en tres componentes integrados.

Primero, cinco contribuciones novel al stack López de Prado, cada una orientada a un modo de falla específico de la pre-validación retail: reporte dual de `n_trials` en el Sharpe Deflactado para taxonomizar el sesgo de selección; Calmar normalizado por capital que elimina la inflación por apalancamiento marginal; esquema sub-período con detección de decaimiento monotónico; test de aislamiento de componentes para estrategias multi-componente; y taxonomía MUST-PASS de regímenes vol-on / vol-off. Cada una se detalla en §4.

Segundo, un score híbrido continuo-puerta — el Score de Robustez — que integra las cinco contribuciones en una métrica $[0, 100]$, inspirado en la disciplina procesal de la industria auditora profesional. La banda pública del score (Robusta, Robustez Condicional, Robustez Limitada, No Robusta, Abstención de Opinión) traduce el resultado a una clasificación accionable para el challenger retail.

Tercero, tres estudios de caso empíricos sobre estrategias textbook y data pública — Golden Cross sobre crudo CL, RSI-14 mean reversion sobre BTC perpetuo y Asian Range Breakout sobre EUR/USD — que demuestran la aplicación de la metodología. El comportamiento diferencial gate-vs-continuous frente a la agregación pura, que define el moat metodológico defendible de TRAVIDENCE, lo exhibe el caso NQ Opening Range Breakout del repositorio.

§1.4 Hoja de ruta

§2 revisa el trabajo relacionado: el stack López de Prado canónico, las implementaciones de referencia y la oferta retail consumer. §3 presenta el resumen metodológico de TRAVIDENCE. §4 detalla las cinco contribuciones novel con sus ecuaciones clave. §5 describe el diseño del estudio empírico. §6 reporta los tres estudios de caso en orden creciente de centralidad para la tesis (CL — BTC — EUR/USD; éste último emite el score público end-to-end). §7 contrasta TRAVIDENCE con la oferta consumer y las librerías de implementación (mlfinlab, RiskLabAI). §8 discute implicaciones. §9 enumera limitaciones. §10 concluye. Las referencias formales y tres apéndices (derivaciones matemáticas, certificados de auditoría como figuras, hash de reproducibilidad) cierran el documento.

§2. Trabajo relacionado

§2.1 El stack López de Prado

El cuerpo de trabajo de López de Prado y colaboradores constituye el marco de referencia académico para la validación honesta de estrategias cuantitativas. Tres componentes son particularmente relevantes para el problema de pre-validación retail.

El Deflated Sharpe Ratio (DSR) de Bailey y López de Prado (2014) extiende el Probabilistic Sharpe

Ratio (Bailey y López de Prado, 2012) para corregir simultáneamente por (a) la no-normalidad de los retornos vía momentos superiores (skew, kurtosis) y (b) el sesgo de selección inducido por evaluación múltiple sobre el mismo dataset. La fórmula del DSR estima la probabilidad de que el Sharpe verdadero exceda un umbral de referencia condicional al Sharpe observado, al número de hipótesis exploradas y al perfil distribucional de los retornos. El paper original provee la forma exacta del valor esperado del máximo Sharpe bajo el null para N trials independientes.

La validación cruzada combinatoria purgada (CPCV), detallada en López de Prado (2018, capítulo 12), aborda el problema dual del leakage temporal y el sesgo de partición única. CPCV genera múltiples paths combinando train/test splits con purging y embargo, produciendo una distribución de Sharpes en lugar de un valor único. La distribución resultante es robusta frente a la práctica común de “elegí los hiperparámetros que dieron el mejor Sharpe en la última ventana” — una forma específica de overfitting que el walk-forward simple no detecta. La superioridad empírica de CPCV sobre walk-forward y k-fold purgado en entornos sintéticos controlados, medida vía DSR y PBO, ha sido documentada por Arian, Norouzi Mobarekeh y Seco (2024).

La probabilidad de backtest overfitting (PBO), introducida en Bailey, Borwein, López de Prado y Zhu (2017), provee una métrica adicional: la probabilidad de que la mejor estrategia in-sample tenga performance OOS inferior a la mediana. PBO es complementaria al DSR y captura un modo de falla distinto — no la inflación del Sharpe sino la fragilidad del ranking entre candidatos.

Estos tres componentes, complementados por el tratamiento riguroso de feature engineering y label leakage de López de Prado (2018), forman el “stack” sobre el cual TRAVIDENCE construye sus cinco contribuciones novel detalladas en §4.

Una precisión de alcance: el remuestreo Monte Carlo de una secuencia de operaciones — práctica común en la validación retail — cuantifica el riesgo de secuencia condicional a que el edge sea genuino. Reordenar o remuestrear las mismas operaciones no puede detectar si provienen de un proceso de selección sobre N variantes: cada camino simulado hereda íntegramente el sesgo de selección del historial de entrada. Detectar sobreajuste por selección exige correcciones de la clase DSR/PBO aplicadas al proceso de búsqueda, no simulaciones sobre su resultado.

§2.2 Implementaciones de referencia: mlfinlab y RiskLabAI

La librería `mlfinlab` (Hudson & Thames) implementa CPCV, DSR, PBO y los métodos de meta-labeling de López de Prado (2018) sobre un stack Python estándar (NumPy, SciPy, pandas). `RiskLabAI` provee implementaciones independientes de un subconjunto similar con cobertura adicional de extreme value theory y métricas de cola. `mlfinlab` es hoy una librería propietaria de Hudson & Thames (código visible bajo licencia comercial desde 2021); `RiskLabAI` es open-source. Ambas librerías son herramientas técnicas de alta calidad: la implementación canónica de TRAVIDENCE del DSR fue cross-validada contra `RiskLabAI` v1.0.8 con agreement bit-identical (máxima diferencia absoluta 0.0e+00 sobre doce casos de prueba); la implementación del Calmar margin-frame fue cross-validada contra `empyrical-reloaded` v0.5.12 con agreement bit-identical hasta precisión de máquina.

La función operacional de estas librerías, sin embargo, es distinta de la de un servicio de validación independiente con grado de auditoría. Una librería entrega métricas; el analista las interpreta y reporta. No hay tercera parte que firme el dictamen, no hay umbral pre-especificado que dispare una decisión binaria PASS/FAIL para uso comercial, no hay sistema de versiones del dictamen que

permita auditar retrospectivamente la base sobre la cual se emitió. El gap es de capa de servicio, no de capa técnica.

§2.3 Oferta retail consumer: agregación pura

El servicio más visible de la franja retail consumer acepta backtests del cliente (formatos CSV / equity curve), aplica un panel de decenas de chequeos de inspiración institucional, y emite un score agregado en escala [0, 100]. La pricing micro-transaccional permite economía de scope sobre volumen alto y captura el mercado retail puro.

La debilidad estructural del servicio para uso en pre-validación de challenge prop firm es la agregación pura. El score se compone como combinación ponderada de las dimensiones evaluadas, sin gates duros que detengan la certificación cuando una de las dimensiones presenta una falla cliff-edge. Una estrategia con asimetría severa de régimen — que falla todos los criterios mínimos en condiciones vol-on, por ejemplo — puede acumular un score numéricamente aceptable porque el promedio agregado compensa con dimensiones favorables. El sistema no distingue entre debilidad continua y falla estructural binaria.

Esta debilidad es funcional al modelo de negocio consumer pero no es admisible para un dictamen que se ofrece como base de decisión sobre desembolso de capital al riesgo (fees de evaluación, configuración de cuentas financiadas). La metodología de auditoría profesional — heredada de las clasificaciones AICPA y IAASB — requiere precisamente el mecanismo opuesto: el dictamen incorpora mecanismos que modifican la opinión con independencia del resto de la evaluación — una incorrección material generalizada, por ejemplo, produce una opinión adversa.

§2.4 Statement del vacío

Recapitulando: el stack López de Prado provee el aparato técnico; las librerías de implementación lo implementan; la oferta retail consumer agrega y comercializa, pero sin gates duros; las firmas de prop trading están impedidas estructuralmente; las firmas de auditoría tradicional (Big Four) no han entrado al mercado de validación de estrategias cuantitativas retail. El vacío específico que TRAVIDENCE ocupa es la combinación de (i) implementación técnica del stack al nivel de las librerías de implementación, (ii) capa de servicio con dictamen firmado por tercera parte, (iii) sistema de gates duros calibrado para defender el cliff-edge institucional, y (iv) una banda pública de resultado que comunica simultáneamente a la audiencia retail y a los stakeholders profesionales (revisores, contrapartes B2B, auditorías futuras).

§3. Resumen metodológico TRAVIDENCE

§3.1 La arquitectura de los cinco validadores

La metodología TRAVIDENCE descansa sobre cinco validadores implementados como módulos independientes del motor de validación, cada uno orientado a un modo de falla específico documentado en la literatura y enriquecido con contribuciones novel descritas en §4.

El validador `deflated_sharpe_ratio` ataca el sesgo de selección por evaluación múltiple. Sobre los inputs `observed_sharpe`, la serie de retornos y el conteo dual (`n_trials_strict`, `n_trials_session`), devuelve la probabilidad de que el Sharpe verdadero exceda el máximo esperado bajo el null para N trials. La implementación sigue la forma exacta derivada en Bailey y López de Prado (2014).

El validador `capital_normalized_calmar` ataca la inflación por apalancamiento en el Calmar histórico. Sobre los inputs serie de retornos, `capital_baseline` declarado y factor de anualización, devuelve el par (`calmar_margin_frame`, `calmar_capital_frame`) y la `inflation_ratio` que diagnostica la divergencia. El frame de capital es la métrica canónica para comparación cross-asset; el frame de margen se preserva por compatibilidad histórica.

El validador `subperiod_monotonic_decay` ataca la inestabilidad temporal del edge. Sobre los inputs serie de retornos, número de chunks K (default cinco), métrica configurable y umbral de loss endpoint-to-endpoint, devuelve el booleano canónico `passes_no_decay` con semántica AND: la estrategia pasa solo si la secuencia de la métrica no es monotónicamente no-creciente y la pérdida endpoint controlada está bajo el umbral.

El validador `component_isolation` ataca el portfolio-level overfitting de estrategias multi-componente. Sobre los inputs retornos agregados, retornos por componente y tolerancia de interacción, descompone el Sharpe agregado en la suma de los Sharpes standalone más un residual de interacción, y aplica atribución leave-one-out para identificar el componente dominante. La tolerancia recomendada escala con $|1 - \sqrt{K}|$ por la no-aditividad estructural del Sharpe (ver §4.4).

El validador `regime_must_pass` ataca la asimetría de régimen. Sobre los inputs serie de retornos, etiquetas de régimen (por defecto vol-on / vol-off según `threshold` de volatilidad rolling), criterios mínimos por régimen y mínimo de observaciones por régimen, devuelve el booleano `passes_all_regimes` con agregación binaria AND across regimes. La asimetría observacional entre regímenes es esperada y no anómala; lo que el validador penaliza es la asimetría de desempeño, no la asimetría de cobertura.

Las cinco contribuciones operacionalizan el principio de exigir evidencia de robustez en cada dimensión, en línea con el stack López de Prado.

§3.2 El Score de Robustez — estructura del híbrido continuo-puerta

El Score de Robustez integra las salidas de los cinco validadores en una métrica sobre $[0, 100]$, traducida a una banda pública de resultado (ver §3.3). La estructura es deliberadamente híbrida: combina una agregación continua ponderada de tres dimensiones de calidad con un sistema de gates duros que actúan como caps cuando una dimensión presenta falla cliff-edge.

Las tres dimensiones de la agregación continua son:

- **Edge Quality (EQ)** — derivada del Deflated Sharpe Ratio en su variante conservadora `dsr_strict`, mediante una función de mapeo lineal por tramos con knots calibrados por bucket de clase de activo. EQ captura la fortaleza estadística del edge corregida por sesgo de selección.
- **Risk Profile (RP)** — derivada del Calmar capital-frame, eliminando la inflación margen-frame que diferencia frente a la convención histórica de tear sheets. RP captura la calidad del perfil retorno/drawdown comparable cross-asset.

- **Sample & Coverage (SC)** — composite de tres componentes: adecuación del número de trades, cobertura calendario y mínimo de observaciones por régimen. SC captura la credibilidad muestral de las dos dimensiones anteriores.

La agregación produce un `base_score` continuo. Sobre ese base score actúan tres gates duros:

- **Gate de régimen** — disparado cuando `regime_must_pass` reporta `passes_all_regimes = False`. Cap conservador (el más bajo de los tres). Justificación: la asimetría de régimen es el modo de falla con mayor potencia predictiva sobre el fracaso en cuenta real, dado que el mercado cambiará de régimen y la estrategia, por construcción, encontrará su lado débil.
- **Gate de sub-período decay** — disparado cuando `subperiod_monotonic_decay` reporta `passes_no_decay = False`. Cap intermedio. Justificación: el edge está muriendo *ahora*; cuando el decay aparece estadísticamente en los chunks, la estrategia está bien adentrada en su trayectoria de extinción.
- **Gate de aislamiento de componentes** — condicional a `n_components ≥ 2`, disparado cuando `component_isolation` reporta `passes_isolation = False`. Cap más alto de los tres. Justificación: el fallo es estructural pero típicamente refinable; la estrategia agregada puede recuperarse mediante re-diseño de la lógica de combinación.

Cuando dos o más gates disparan simultáneamente, el cap más conservador (el más bajo) prevalece. El score final se obtiene como `min(round(base_score), effective_cap)`. La asimetría entre el carácter probabilístico del base score (sujeto a intervalo de confianza) y el carácter determinista del cap (el gate disparó o no) se preserva en el reporte: el IC del 95% se computa sobre el base score; el evento de gate se reporta como hecho binario.

Las dimensiones, los nombres de los gates, las definiciones conceptuales y la lógica del híbrido continuo-puerta son la superficie pública de la metodología. Los componentes internos del scoring — pesos dimensionales, caps de gating, posiciones de knots de la curva de mapeo y parámetros de calibración — se mantienen como secreto comercial. La nomenclatura de tiers consumer-facing y sus cortes (publicados en el Brief Ejecutivo D2 §4) son labels públicos que permiten al cliente interpretar su score; el modelo interno que asigna esos labels es propietario y se mantiene fuera del presente documento (ver §3 de la especificación interna de TRAVIDENCE (no publicada)).

§3.3 La banda pública de resultado

El score numérico se traduce a una banda pública: Robusta (80–100), Robustez Condicional (65–79), Robustez Limitada (50–64), No Robusta (0–49) y Abstención de Opinión (N/A, cuando la muestra es insuficiente para un veredicto). La banda es la clasificación accionable en lenguaje comercial que recibe el cliente retail, y es a la vez la clasificación trazable al criterio institucional que necesita el stakeholder profesional — revisor, contraparte B2B, escrutinio mediático futuro.

La disciplina procesal del cuerpo auditor profesional — codificada en AICPA (2011) Statement on Auditing Standards No. 122 §AU-C 705 (Modifications to the Opinion in the Independent Auditor's Report) e IAASB (2015) ISA 705/706 (Modifications to the Opinion in the Independent Auditor's Report; Emphasis of Matter Paragraphs) — es la inspiración de diseño del sistema de banda: un dictamen firmado, con criterios pre-especificados, que puede acotar el resultado por debajo de lo que asignaría la agregación pura. El dictamen TRAVIDENCE no pretende equivaler funcionalmente

a una auditoría financiera (no es una opinión sobre estados financieros bajo GAAP/IFRS), pero sí adopta deliberadamente esa disciplina procesal como base de defensibilidad legal y reputacional.

§4. Las cinco contribuciones novel

Las siguientes cinco subsecciones describen las contribuciones novel de TRAVIDENCE al stack López de Prado. Cada subsección sigue la misma estructura: (i) statement explícito de la novedad respecto al baseline canónico, (ii) ecuación clave y forma operacional, (iii) implicación práctica sobre el verdict del Score de Robustez. Las derivaciones completas se difieren al Apéndice A.

§4.1 Reporte dual de n_trials en el Sharpe Deflactado

La aplicación canónica del Deflated Sharpe Ratio en Bailey y López de Prado (2014) opera sobre un único valor n_trials declarado por el analista. En la práctica, post-mortems de discovery runs revelan consistentemente que el valor reportado refleja únicamente el espacio de selección de la *sesión actual* del scan — la cardinalidad del grid efectivamente recorrido en el run que produjo la hipótesis declarada — mientras que el conteo *institucional verdadero* incluye todas las exploraciones históricas previas sobre el mismo dataset que informaron la formulación de la hipótesis. La diferencia entre ambos conteos es típicamente de un orden de magnitud o más.

La contribución de TRAVIDENCE es operacionalizar el dual reporting como invariante de implementación. El validador `deflated_sharpe_ratio` requiere dos inputs paralelos: `n_trials_strict` (universo cross-session institucional de todas las variantes que tocaron data sobre este dataset) y `n_trials_session` (subconjunto comparado en la selección final del run actual). Por construcción algebraica del problema, $n_trials_strict \geq n_trials_session$; la implementación eleva `ValueError` si el invariante se viola.

Bailey y López de Prado (2014) derivan la forma exacta del valor esperado del máximo Sharpe bajo el null:

$$E[\max SR \mid N] = \sqrt{V[\{SR_n\}]} \cdot \left[(1 - \gamma_E) \Phi^{-1} \left(1 - \frac{1}{N} \right) + \gamma_E \Phi^{-1} \left(1 - \frac{1}{N_e} \right) \right]$$

donde $\gamma_E \approx 0.5772$ es la constante de Euler-Mascheroni, Φ^{-1} es la inversa de la CDF normal estándar, y $V[\{SR_n\}]$ es la varianza del estimador del Sharpe a lo largo de los N trials. El estadístico DSR se calcula como:

$$DSR = \Phi \left(\frac{SR_{obs} - E[\max SR \mid N]}{\hat{\sigma}_{SR}} \right)$$

con $\hat{\sigma}_{SR}$ la desviación del estimador según la corrección de Mertens (2002) por momentos superiores, sobre la base de Lo (2002):

$$\hat{\sigma}_{SR}^2 = \frac{1 - \gamma_3 \cdot SR_{obs} + \frac{\gamma_4 - 1}{4} \cdot SR_{obs}^2}{T - 1}$$

donde γ_3 es la skew muestral unbiased y γ_4 la kurtosis muestral Pearson (no excess; Normal $\rightarrow 3$). La implementación canónica usa la convención `scipy.stats bias=False, fisher=False` para alinearse exactamente con la forma de Mertens (2002).

El dual reporting produce dos verdicts paralelos del Test 5 de TRAVIDENCE:

- `dsr_strict` $\geq 0.95 \rightarrow$ PASS estricto.
- `dsr_session` $\geq 0.95 \rightarrow$ PASS sesión.

(El umbral 0.95 corresponde al nivel de confianza estándar del paper de referencia, no a una constante calibrada de TRAVIDENCE.)

Por construcción, `dsr_strict` $\leq$ `dsr_session` siempre (el cap de $E[\max \text{SR} \mid N]$ crece con N). El modo de falla diagnósticamente interesante es PASS @ session AND FAIL @ strict: la estrategia sobrevive solo bajo el conteo cómodo del analista y cae bajo el conteo institucional. Esta brecha — la distancia entre la declaración honesta-de-la-sesión y la realidad cross-session — es precisamente el patrón que el dual reporting hace auditable. El verdict primary de TRAVIDENCE usa `dsr_strict`; `dsr_session` queda como sensitivity informational.

La implementación de referencia vive en el módulo `deflated_sharpe.py` del motor de validación, función `deflated_sharpe_ratio`. La política híbrida del estimador de varianza $V[\{SR_n\}]$ (empírica si el caller provee el array de SRs explorados con longitud ≥ 2 ; asintótica Lo (2002) en caso contrario) se traza en el campo `variance_source` del resultado para auditabilidad.

§4.2 Calmar normalizado por capital — frame de capital vs frame de margen

El Calmar ratio histórico (Young, 1991), en la convención de tear sheets retail, se calcula como retorno anualizado sobre máximo drawdown porcentual, donde el drawdown se reporta como porcentaje sobre el rolling peak de la curva de equity. En contextos con apalancamiento — futuros con margin requirements, forex retail bajo regulación ESMA 30:1, crypto perpetual con cross-margin — el rolling peak crece a velocidad proporcional al notional controlado por el trader, no al capital real puesto al riesgo. El Calmar resultante está sistemáticamente inflado por el factor de apalancamiento embebido en el sizing por margen, y los Calmar reportados entre asset classes no son comparables.

La contribución de TRAVIDENCE es separar explícitamente dos frames del Calmar y declarar el frame de capital como canónico para comparación cross-asset. El validador `capital_normalized_calmar` requiere `capital_baseline` declarado y estrictamente positivo. La convención TRAVIDENCE (v1.3) define `capital_baseline` como el capital real al riesgo del trader (margin posted), no como el notional:

$$\text{capital_baseline} = \frac{\text{nominal_capital}}{\text{leverage_factor}}$$

Sobre esta definición, las dos frames del Calmar se calculan como:

$$\text{Calmar}_{\text{margin}} = \frac{r_{\text{ann}}}{|DD_{t^*}/P_{t^*}|}, \quad \text{Calmar}_{\text{capital}} = \frac{r_{\text{ann}}}{|DD_{t^*}/\text{capital_baseline}|}$$

donde P_{t^*} es el rolling peak al momento del peor drawdown, DD_{t^*} es el drawdown en dólares en ese evento, y r_{ann} es el retorno anualizado geométrico computado sobre la equity curve geométrica $E[t] = \text{capital_baseline} \cdot \prod_{i \leq t} (1 + r_i)$.

El ratio entre las dos frames provee un diagnóstico de la magnitud de la inflación:

$$\text{inflation_ratio} = \frac{\text{Calmar}_{\text{margin}}}{\text{Calmar}_{\text{capital}}} = \frac{P_{t^*}}{\text{capital_baseline}}$$

algebraicamente derivado por cancelación del numerador y del DD_{t^*} común. $\text{inflation_ratio} \neq \text{leverage_factor}$ en general: las dos magnitudes solo coincidirían si el equity creciera por un factor leverage_factor durante el período y el peor drawdown ocurriera al final. En la práctica, con position sizing unitario y holding periods limitados, los inflation ratios observados quedan típicamente entre 1.0× y 5.5× incluso bajo leverage nominal de 30×. Esta divergencia es información útil del validador, no defecto: mide la divergencia *real* entre frames al momento del peor evento, no el leverage nominal del contexto.

El umbral institucional del Calmar capital-frame es privado y calibrado por clase de activo; se aplica al Calmar_capital exclusivamente. Una estrategia con Calmar_margin alto que no acompaña con un Calmar_capital que supere ese umbral es rechazada en el bloque institucional independientemente del valor margen-frame.

La implementación canónica vive en el módulo `capital_normalized_calmar.py` del motor de validación, función `capital_normalized_calmar`. El frame de margen se preservó por compatibilidad y diagnóstico; el frame de capital es la métrica novel TRAVIDENCE que alimenta la dimensión Risk Profile del Score de Robustez (ver §3.2).

§4.3 Esquema sub-período con detección de decaimiento monótonico

El edge de una estrategia que opera sobre data financiera no es atemporal. Los regímenes cambian, los participantes mutan, la microestructura se ajusta; un edge presente en el período de descubrimiento puede haber decaído antes incluso de que el resultado del backtest se materialice. La práctica retail común — reportar métricas agregadas sobre todo el histórico — oculta esa trayectoria temporal. Una estrategia con $\text{PF} = 1.8$ agregado puede esconder una caída de $\text{PF} = 2.5$ en el período más antiguo a $\text{PF} = 1.1$ en el más reciente; el promedio aritmético no captura ese hecho.

La contribución de TRAVIDENCE es operacionalizar el test de monotonic decay como invariante binario canónico, generalizando el schema 3-period histórico de la metodología a una versión K-chunk equal-split aplicable de modo uniforme sobre cualquier período de evaluación. El validador `subperiod_monotonic_decay` parte la serie de retornos en K chunks contiguos de longitud igual (default K=5), evalúa una métrica configurable (Sharpe por defecto) sobre cada chunk, y emite el booleano canónico `passes_no_decay` con semántica AND locked:

$$\text{passes_no_decay} = \neg \text{is_weakly_monotonic_decay} \wedge (\text{overall_decay_pct} < \theta)$$

donde θ es el umbral de loss endpoint-to-endpoint (default 20%), $\text{is_weakly_monotonic_decay}$ es verdadero cuando la secuencia $m_1, m_2, \dots, m_K$ es no-creciente, y overall_decay_pct es $(m_1 - m_K)/m_1$. La semántica AND es deliberada: el operador OR dejaría pasar estrategias estrict-

tamente decayendo pero superficialmente, mientras que el AND exige simultáneamente ausencia de monotonía no-creciente y caída endpoint controlada.

Un punto técnico relevante es la convención de display del campo `overall_decay_pct`. La fórmula es matemáticamente exacta pero produce valores visualmente engañosos cuando m_1 es pequeño, cercano a cero o negativo. En el caso 02 (BTC RSI mean reversion), por ejemplo, la secuencia de Sharpes por chunk es $[0.39, 0.98, -0.12, 1.01, 1.71]$ — el edge se fortalece netamente del primer al último chunk, con una caída puntual en el tercero — y el `overall_decay_pct` calculado es -332.04% matemáticamente correcto pero visualmente confundente. El booleano canónico `passes_no_decay = True` provee la respuesta correcta independientemente de la magnitud de m_1 . (Secuencia del engine vigente re-auditado; §6.2 reporta la del dictamen original congelado — ver nota al inicio de §6.)

La convención TRAVIDENCE para reportes de cara al cliente es: mostrar siempre la serie `metric_per_subperiod` completa (legible directamente), reportar el booleano `passes_no_decay` como verdict, y suprimir el porcentaje raw cuando $|m_1| < 0.10$ (sustituyéndolo por N/A — `m_1 too small for meaningful percentage` en el certificado).

La implementación canónica vive en el módulo `subperiod_decay.py` del motor de validación, función `subperiod_monotonic_decay`. La schema 3-period histórica (pre-discovery, discovery, post-discovery con anclaje de fechas y flags multi-nivel) se preserva como instantiación recomendada para discovery runs con reporting expandido, alimentada por el verdict binario del validador.

§4.4 Aislamiento de componentes — el floor aritmético $|1 - \sqrt{K}|$

Las estrategias multi-componente — ensemble de dos o más señales independientes combinadas mediante un signal aggregator o portfolio — son particularmente vulnerables a un modo de falla específico que los tests univariados aplicados al agregado no detectan. La metáfora operacional es la siguiente: cuatro componentes individualmente débiles, ninguno con edge distinguible estadísticamente del ruido, pueden producir un agregado con Sharpe aparentemente saludable por ruido de la combinación. Profit Factor, Calmar y DSR aplicados al agregado pueden lucir saludables; pero la fragilidad es estructural y el edge no es replicable, no es robusto a cambios menores en la lógica de combinación, y no resiste despliegue institucional.

La contribución de TRAVIDENCE es la descomposición SR + atribución leave-one-out como primitiva estructural, complementada por la calibración explícita de la tolerancia de interacción contra el floor aritmético determinístico del problema.

La descomposición opera sobre los inputs `strategy_returns` (retornos del agregado) y `component_returns` (mapping nombre → retornos standalone). El procedimiento canónico es:

$$\begin{aligned} SR_{full} &= \text{annualized_sharpe}(\text{strategy_returns}) \\ SR_i^{\text{standalone}} &= \text{annualized_sharpe}(\text{component_returns}[i]) \\ \rho_{\text{int}} &= SR_{full} - \sum_i SR_i^{\text{standalone}}, \quad \rho_{\text{int,pct}} = \frac{\rho_{\text{int}}}{SR_{full}} \end{aligned}$$

La atribución leave-one-out identifica el componente dominante:

$$\Delta_i = SR_{full} - SR(\text{strategy_returns} - \text{component_returns}[i]), \quad \text{dominant} = \arg \max_i |\Delta_i|$$

El veredicto canónico es `passes_isolation = (| $\rho_{\text{int,pct}}$ |  $\leq$  interaction_tolerance)`.

La calibración de `interaction_tolerance` es donde reside la sutileza. Bajo aditividad ($\mathbf{r_full} = \sum_i \mathbf{r_i}$) e independencia (componentes IID con media μ y varianza σ^2), un cálculo directo da:

$$SR_{full} = \sqrt{K} \cdot \frac{\mu}{\sigma}, \quad \sum_i SR_i^{\text{standalone}} = K \cdot \frac{\mu}{\sigma}, \quad \frac{\rho_{\text{int}}}{SR_{full}} = 1 - \sqrt{K}$$

La baseline naive $\sum SR_i \approx SR_{full}$ **falla**, entonces, con un floor determinístico de magnitud $|1 - \sqrt{K}|$ incluso bajo aditividad e independencia perfectas. El floor no es interacción genuina; es la huella aritmética de que el Sharpe es una razón mean/std y por construcción no es aditivo.

La calibración recomendada del `interaction_tolerance` por K se estructura como el floor aritmético $|1 - \sqrt{K}|$ más un buffer muestral finito. El buffer absorbe slack muestral en los Sharpes estimados y leve no-linealidad del agregador no atribuible a overfitting estructural; una estrategia cuyo $|\rho_{\text{int,pct}}|$ excede la tolerancia recomendada para su K está consumiendo el buffer con interacción real y debe revisarse. Las tolerancias numéricas por K y la magnitud del buffer están calibradas y documentadas internamente per la Disclosure Boundary Policy.

K (componentes)	Floor aritmético $ 1 - \sqrt{K} $	<code>interaction_tolerance</code> recomendado
2	0.414	(internal — see Disclosure Boundary Policy)
3	0.732	(internal — see Disclosure Boundary Policy)
4	1.000	(internal — see Disclosure Boundary Policy)
5	1.236	(internal — see Disclosure Boundary Policy)

La implementación canónica vive en el módulo `component_isolation.py` del motor de validación, función `component_isolation_test`. Para estrategias single-component (K=1), el test se reporta como N/A — `single-component strategy` y no afecta el veredicto del Score de Robustez. Para $K \geq 2$, el gate `component_isolation` del Score de Robustez (§3.2) se activa cuando `passes_isolation = False`.

§4.5 MUST-PASS regimes con taxonomía vol-on / vol-off

El último de los cinco validadores ataca la asimetría de régimen, un modo de falla que las métricas agregadas ocultan estructuralmente. Una estrategia que opera bien en regímenes de baja volatilidad y catastróficamente en regímenes de alta volatilidad puede acumular un Sharpe agregado aceptable si la fracción de tiempo en cada régimen es funcional al sesgo direccional del backtest. La promediación silenciosa entre regímenes es invisible al cliente y al revisor sin desglose explícito.

La contribución de TRAVIDENCE es operacionalizar el test MUST-PASS con agregación binaria AND across regimes como invariante canónico, manteniendo la taxonomía 4+2 vol-on/vol-off como perfil institucional recomendado. El validador `regime_must_pass` opera sobre `returns` indexado por `datetime`, `regime_labels` aligned, un criterio mínimo por régimen y por métrica (estructura y cortes privados, no publicados) y `min_observations_per_regime` (el mínimo de observaciones por régimen por debajo del cual el régimen se trata como FAIL conservador).

La regla de pase es estricta conjunción:

$$\text{passes_all_regimes} = \bigwedge_{r \in R} \text{passes_per_regime}[r] \wedge |\{r : n_r < \text{min_obs}\}| = 0$$

Cualquier régimen que falle, cualquier régimen con observaciones insuficientes y cualquier criterio que devuelva NaN cuenta como FAIL global. No hay promedio cross-régimen; el contrato del validador es AND.

Cada criterio por régimen lleva una dirección de comparación implícita — un piso que la métrica debe igualar o superar, o un techo que no debe excederse — que la implementación preserva de forma auditable y sin ambigüedad de codificación. Los outputs preservan la clave del usuario verbatim, permitiendo round-trip exacto y desambiguación cuando un mismo nombre de métrica base aparece bajo dos direcciones simultáneas en el mismo régimen.

El registry canónico soporta siete métricas computadas internamente: `sharpe`, `sortino`, `calmar` (margen-frame; para capital-frame usar el validador especializado), `dd_pct`, `return`, `win_rate` y `profit_factor`. La composición mínima recomendada para perfil institucional es 4 regímenes vol-on + 2 regímenes vol-off cubriendo el período pre-discovery + discovery del estudio. La taxonomía concreta por clase de activo se detalla en el Apéndice A.

Un punto técnico relevante es la asimetría observacional esperada entre regímenes. La mayoría de los assets pasan más tiempo en un régimen que en el otro, dependiendo del threshold elegido y la historia del asset; el régimen vol-on es típicamente minoritario. La asimetría no es defecto del test ni anomalía: es estructural. El validador trata regímenes con `n < min_observations_per_regime` como FAIL conservador para evitar verdict optimista basado en muestra insuficiente; por encima de ese floor, la asimetría no compromete la validez estadística del test AND-across-regimes — el régimen minoritario simplemente tiene métricas estimadas con mayor varianza, lo cual es información honesta sobre la exposición de la estrategia a ese régimen.

La implementación canónica vive en el módulo `regime_must_pass.py` del motor de validación, función `regime_must_pass`. El gate `regime` del Score de Robustez (§3.2) se activa cuando `passes_all_regimes = False` y constituye, conceptualmente, el gate de severidad más alta del sistema de tres gates: la asimetría de régimen es el modo de falla con mayor potencia predictiva sobre fracaso en cuenta real, dado que el mercado cambiará de régimen y la estrategia, por construcción, encontrará su lado débil.

§5. Diseño del estudio empírico

§5.1 Por qué tres estrategias textbook sobre data pública

El objetivo de los tres casos de estudio reportados en §6 no es descubrir alpha. Es validar la metodología TRAVIDENCE como filtro. Tres estrategias textbook, ampliamente documentadas en la literatura de trading retail (Chan, 2013; Clenow, 2013) y derivadas de ideas con décadas de visibilidad, son adversarial test cases del filtro: si la metodología TRAVIDENCE fuera laxa, una de las tres alcanzaría banda Robusta y la pretensión de rigor quedaría refutada por contraejemplo. Una precisión de honestidad metodológica: la calibración v1.0 (§9.3) se construyó en parte a partir de una población sintética derivada de estos mismos tres casos, de modo que su valor aquí es ilustrativo del mecanismo, no una prueba independiente de estrictez. Esa prueba, fuera de la población de calibración, la aporta el caso NQ Opening Range Breakout del repositorio (Apéndices B–C). Si ninguna de las tres alcanza honestamente banda Robusta — consistente con la tasa documentada del 86% — la metodología sostiene su pretensión; bajo la re-auditoría con el engine de producción, las tres se resuelven en dos Disclaimers of Opinion por muestra insuficiente y un score de Robustez Limitada público sin gate vinculante (nota de re-auditoría, §6).

La elección de estrategias textbook responde también a una restricción de transparencia. Cuando el objeto del estudio es textbook sobre data pública, el paper puede divulgar la especificación completa de la estrategia, la data exacta usada, la implementación del backtest y los resultados intermedios sin tropezar con propiedad intelectual privada. El contraste con una validación sobre estrategias propietarias o data privada es estructural: ese reporting sería necesariamente incompleto.

§5.2 Política de separación firma/fondo

TRAVIDENCE, como firma de servicio de validación independiente, mantiene una política operacional explícita de separación entre las estrategias internas que cualquier accionista o consultor pueda operar como inversionista privado y las estrategias evaluadas por los validadores publicados en este documento o utilizadas en cualquier material institucional. La política se traduce en la siguiente regla de selección de casos: ningún estudio publicado por TRAVIDENCE usa estrategias del fondo propio de ningún empleado, accionista o asesor; los tres casos del presente paper son sintéticos en el sentido de estar contruidos sobre especificaciones públicas (textbook) aplicadas a series públicas, no privadas. La política protege la independencia institucional del dictamen y evita el conflicto de interés estructural que afecta a las firmas de prop trading mismas.

§5.3 Fuentes de data, períodos y supuestos

Los tres casos usan series públicas accesibles vía librerías estándar. La tabla siguiente resume las fuentes:

Caso	Estrategia	Período	Fuente de data	Tier simulado
01	Golden Cross — CL futuros	2000–2025	yfinance, CL=F daily	T2 (Audit Verified)
02	RSI(14) mean reversion — BTC perpetuo	2020–2025	back-adjusted ccxt Binance BTC/ USDT:USDT 1d	T2
03	Asian Range Breakout — EUR/USD	2015–2025	dukascopy- python EUR/USD 1h	T2

En los tres casos, el tier T2 corresponde a “Audit Verified” en la jerarquía de evidencia TRAVIDENCE (§0.1 de la metodología canónica). En producción comercial, una validación T1 (“Verified Audit”) requeriría conexión de solo lectura vía la API del broker; las series públicas usadas en los tres casos están aceptadas para los estudios reportados aquí pero no calificarían para engagements institucionales B2B en su forma actual.

Los supuestos de costos del backtest vanilla son cero en los tres casos (cero comisión, cero spread, cero slippage). Esta elección es deliberada: el objetivo del estudio es testear la metodología TRAVIDENCE, no estimar la P&L realista. Cualquier versión con costos realistas — que estimaría un round-trip de aproximadamente \$10 USD por trade en CL, fees típicos de exchange en BTC perpetual, y spread + commission típicos retail en EUR/USD — degradaría las métricas y alejaría aún más cada score de la banda Robusta; la versión cero-costos es el lado más favorable a la estrategia y constituye, por tanto, el test más exigente para el filtro.

Los position sizings de los tres casos son unitarios (1 unidad por trade, sin compounding ni pyramiding). La interpretación de “una unidad” varía por contexto operacional (1 contrato de CL en futuros, 1 unidad BTC nominal en perpetual, 1 lote estándar de EUR/USD en forex retail), y el `capital_baseline` para el validador `capital_normalized_calmar` se ajusta consecuentemente vía la convención `capital_baseline = nominal_capital / leverage_factor` documentada en §4.2 y aplicada en §6.

§6. Tres estudios de caso

Nota de re-auditoría (2026-06-10, engine de producción). Los walkthroughs de validadores en §6.1–§6.3 son el registro congelado de la auditoría original, producida bajo la calibración de régimen per-bucket ya retirada y el conteo de observaciones a nivel de barra (nota histórica, Apéndice A §A.5). La re-auditoría con el engine de producción — retornos a nivel de trade, corte de régimen p50 auto-derivado y aplicación de los floors de admisibilidad — supersede los dictámenes finales como sigue. El Caso 01 (CL) y el Caso 02 (BTC) cierran con **Abstención de Opinión — Evidencia Insuficiente**: sus muestras a nivel de trade ($n_{\text{trades}} = 18$ y 31) quedan por debajo del floor de admisibilidad. El Caso 03 (EUR/USD) emite el primer score público end-to-end real del framework: **56 / 100, Robustez Limitada, sin gate vinculante**. El rol de demostración de gates que cargaba EUR/USD pasa al Caso 04 del repositorio (NQ Opening Range Breakout): declarado insufficient

por la auditoría original a nivel de barra, es admisible bajo conteo a nivel de trade ($n_trades = 2,424$) y queda capado por los gates [**regime, subperiod**] — comportamiento de gate vinculante sobre una muestra admitida. El Caso 05 del repositorio (XAU/USD Bollinger reversal) queda igualmente gated [**regime, subperiod**]. A través de los cinco case studies del repositorio, el portafolio cubre los tres desenlaces públicos del framework: score emitido (03), gated (04, 05), insufficient sample (01, 02). Los certificados del Apéndice B y los hashes de reproducibilidad del Apéndice C corresponden a los dictámenes re-auditados. Per la nota metodológica del spec interno §12.3: los hallazgos de auditoría actualizan la visión del auditor; los auditores no alteran auditorías para ajustarlas a expectativas previas.

§6.1 Caso 01 — Golden Cross sobre crudo CL (2000–2025)

Descripción de la estrategia

Cruce de medias móviles simples sobre el precio de cierre diario del contrato continuo de crudo WTI (CL). Entrada larga cuando la SMA(50) cruza por encima de la SMA(200); salida cuando la SMA(50) cruza por debajo. Long-only, sin stops, sin position sizing más allá de una unidad por trade. El crossover 50/200 es la convención focal de la literatura de trend-following retail desde al menos los años setenta (Brock, Lakonishok y LeBaron, 1992).

Hipótesis económica subyacente

La hipótesis es trend persistence en commodity markets impulsada por inercias estructurales de oferta/demanda (shocks geopolíticos, cuotas OPEC, decisiones de capacidad del ciclo shale) y por dinámicas macro de movimiento lento (PIB global, fortaleza del USD) (Hong y Yogo, 2012). Una media móvil de largo plazo actúa como filtro de baja frecuencia que participa en los grandes bull markets (2003–2008, 2009–2014, 2020–2022) mientras permanece flat durante regímenes de rango comprimido. La hipótesis cumple los cinco criterios PFP-MP (Previa, Falsable, Parsimoniosa, Mecanística, Preespecificada) detallados en §0.3 de la metodología canónica.

Resultados del backtest vanilla

El backtest produce 18 trades en 25 años con win rate de 44.4%, average winner de +36.4%, average loser de -8.9%, profit factor de 3.26, retorno total de 244%, retorno anualizado geométrico de 5.1%, volatilidad anualizada de 25.4%, Sharpe ratio anualizado de 0.32, Sortino ratio de 0.46, Calmar ratio clásico de 0.083 y máximo drawdown porcentual de -62.2%. Average trade duration de 288 días. La densidad de trades es típica de estrategias de trend-following con holdings multi-año.

Aplicación de los validadores TRAVIDENCE

Cifras del dictamen original congelado; no reproducibles contra metrics.json vigente — ver nota §6.

El validador deflated_sharpe_ratio con $n_trials_strict = n_trials_session = 2$ (las dos ventanas SMA, sin grid search) devuelve $dss_strict = 0.862$. El umbral institucional es 0.95; el verdict del Test 5 es FAIL marginal por insuficiencia muestral relativa al ruido del Sharpe.

El sesgo de selección es bajo por construcción (espacio de hipótesis pequeño), pero la magnitud del Sharpe observado tampoco es suficientemente alta para superar el umbral.

El validador `capital_normalized_calmar` con `nominal_capital = $50,000` y `leverage_factor = 9` (CME margin típico para CL) devuelve `capital_baseline = $5,555.56`. El Calmar margen-frame es 0.083 y el Calmar capital-frame es 0.015, con `inflation_ratio = 5.34`×. El umbral institucional del Calmar capital-frame falla en ambas frames. El `inflation_ratio` alto (el más alto de los tres casos) refleja que el peak del equity sobre el período de 25 años creció considerablemente arriba del baseline antes del peor drawdown.

El validador `subperiod_monotonic_decay` con `K=5` chunks sobre los 25 años devuelve la secuencia de Sharpes por chunk `[0.78, 0.49, -0.24, -0.10, 0.41]`. El `verdict.passes_no_decay = False` por `overall_decay_pct = 47.0% > 20%`, gate disparado: la estrategia perdió aproximadamente la mitad del Sharpe entre el primer y el último chunk.

El validador `regime_must_pass` reparte las observaciones del período entre el régimen vol-on (1,073) y el vol-off (5,135) —clasificados por la volatilidad anualizada propia de la estrategia— y confronta cada régimen, por separado, con los criterios mínimos pre-registrados del bucket commodities: un piso de Sharpe y un techo de drawdown que cada régimen debe satisfacer. El régimen vol-on registra un Sharpe de 0.3 y un drawdown máximo de 46.4%; el vol-off, un Sharpe de 0.4 y un drawdown de 71.0%. Ninguno de los dos regímenes alcanza el piso de Sharpe ni se mantiene bajo el techo de drawdown del bucket, de modo que `passes_all_regimes = False`. Como la regla es conjunción estricta (AND across regimes, sin promediación entre regímenes), basta con que uno falle —aquí fallan los dos— para que el gate de régimen dispare.

El validador `component_isolation` no aplica (estrategia single-component).

Score de Robustez y dictamen

Dos gates duros disparan simultáneamente (régimen y sub-período); el cap más conservador prevalece. La banda resultante es No Robusta — `capped_by: regime` registrado en el verdict. La estrategia, en su forma textbook actual, no es viable para despliegue retail prop firm. (Dictamen original congelado; ver la nota de re-auditoría al inicio de §6.)

La Figura 1 (Apéndice B; PDF en `../.../.../case_studies/01-golden-cross-cl/audit_certificates/01-cl-golden-cross-audit-certificate.pdf`) reproduce el certificado de auditoría completo correspondiente.

Interpretación

El Golden Cross sobre CL es un caso clásico de estrategia con mecanismo plausible — trend persistence en commodities está bien documentado — pero con métricas operacionales insuficientes para superar el filtro audit-grade. El Sharpe agregado es bajo y la asimetría sub-período revela decay material entre la primera mitad de la muestra (cuando el shale revolution aún no había alterado la estructura del mercado) y la segunda (cuando los regímenes commodity se volvieron más volátiles y menos trended). La estrategia es un ejemplo informativo del modo de falla “mecanismo correcto, calibración insuficiente para audit-grade”.

§6.2 Caso 02 — RSI(14) mean reversion sobre BTC perpetuo (2020–2025)

Descripción de la estrategia

RSI(14) con smoothing Wilder aproximado vía rolling mean sobre el cierre diario de BTC/USDT:USDT (perpetual contract). Entrada larga cuando $RSI(14) < 30$ y no hay posición abierta; salida cuando $RSI(14) > 50$. Long-only, sin stops, position sizing unitario. La ventana (14) y el umbral de entrada (30) son valores del framework original de Wilder (1978); la salida en la línea media (50) es convención posterior.

Hipótesis económica subyacente

Los mercados crypto exhiben cascadas de liquidación forzada periódicas — flush de longs apalancados en el mercado perpetuo — que llevan el RSI a lecturas extremas oversold cuando el precio se desvía sustancialmente del rango de volatilidad normal. Eventos documentados durante el período de muestra: caída COVID de marzo 2020, corrección Elon-tweet de mayo 2021, colapso Luna de junio 2022, colapso FTX de noviembre 2022. La mean reversion sigue conforme funding rates se normalizan y los dip buyers reentran. La estrategia captura la pierna de recuperación sin intentar predecir la cascada. La hipótesis cumple los cinco criterios PFP-MP.

Resultados del backtest vanilla

El backtest produce 31 trades en 5 años con win rate de 54.8%, average winner de +8.9%, average loser de -5.5%, profit factor de 1.96, retorno total de 80.3%, retorno anualizado de 7.9%, volatilidad anualizada de 30.7%, Sharpe ratio de 0.41, Sortino ratio de 0.57, Calmar ratio clásico de 0.16 y máximo drawdown porcentual de -47.9%. Average trade duration de 13 días.

Aplicación de los validadores TRAVIDENCE

Cifras del dictamen original congelado; no reproducibles contra metrics.json vigente — ver nota §6.

El validador `deflated_sharpe_ratio` con `n_trials_strict = n_trials_session = 3` devuelve `dss_strict = 0.61`. FAIL claro respecto al umbral institucional 0.95: la combinación de Sharpe modesto y espacio de selección de tres trials (los tres parámetros Wilder) produce un máximo esperado bajo el null que el Sharpe observado no supera.

El validador `capital_normalized_calmar` con `nominal_capital = $30,000` (mediana BTC durante el período) y `leverage_factor = 20×` (cross-margin típico crypto perpetual) devuelve `capital_baseline = $1,500`. El Calmar margen-frame es 0.242 y el Calmar capital-frame es 0.233, con `inflation_ratio = 1.04×`. El ratio es el más bajo de los tres casos, reflejando que el peor drawdown ocurre en early-sample (cuando $\text{peak} \approx \text{baseline}$) — el strategy decae con DD profundo cerca del inicio del período antes de recuperarse. Ambas frames fallan el umbral institucional del Calmar capital-frame.

El validador `subperiod_monotonic_decay` con `K=5` devuelve la secuencia $[0.17, 0.72, 0.29, 0.62, 1.10]$. El veredict `passes_no_decay = True`: la estrategia *fortalece* su edge a lo largo del período. El `overall_decay_pct` reportado de -540.76% es matemáticamente correcto pero engañoso por

m_1 pequeño (0.17); el booleano canónico provee la respuesta interpretable. La estrategia mejoró sustancialmente entre el primer y el quinto chunk.

El validador `regime_must_pass` clasifica la muestra en el régimen vol-on (50 observaciones, por encima del mínimo de observaciones por régimen) y el vol-off (1,914) por la volatilidad anualizada propia de la estrategia, y evalúa cada régimen contra el piso de Sharpe y el techo de drawdown pre-registrados del bucket. El régimen vol-on registra un Sharpe de 0.2 —por debajo del piso— y un drawdown máximo de 14.5%, dentro del techo: pasa el criterio de drawdown pero falla el de Sharpe. El vol-off registra un Sharpe de 0.6, por encima del piso, y un drawdown de 31.2% que excede el techo: pasa Sharpe pero falla drawdown. Cada régimen incumple al menos un criterio, así que el AND across regimes es False y el gate de régimen dispara.

Score de Robustez y dictamen

El gate de régimen disparado produce banda No Robusta, con `capped_by: regime` registrado en el verdict. (Dictamen original congelado; ver la nota de re-auditoría al inicio de §6.)

La Figura 2 (Apéndice B; PDF en `.././././././case_studies/02-rsi-mean-reversion-btc/audit_certificates/02-btc-rsi-mean-reversion-audit-certificate.pdf`) reproduce el certificado de auditoría completo.

Interpretación

El caso BTC es informativo por tres razones. Primero, ilustra que `passes_no_decay = True` (la estrategia no decae; mejora) es condición *necesaria pero no suficiente* para banda Robusta: los otros gates pueden disparar incluso cuando el sub-período test pasa. Segundo, demuestra el manejo correcto de la asimetría observacional entre regímenes — vol-on con 50 obs versus vol-off con 1,914 obs — sin que la asimetría comprometa la validez del verdict cuando ambos regímenes están por encima del floor mínimo. Tercero, el ejemplo es operacionalmente relevante para inversionistas crypto: estrategias mean-reversion sobre perpetuals con holding period multi-day tienden a exhibir el patrón “fortalecimiento aparente del edge en early sample post-launch” combinado con drawdowns severos en vol-off, dado que la mayoría del período de muestra del perpetual launching post-2020 está dominada por regímenes vol-off. La metodología TRAVIDENCE captura ese perfil correctamente.

§6.3 Caso 03 — Asian Range Breakout sobre EUR/USD (2015–2025) — caso del score público end-to-end

Descripción de la estrategia

Breakout del rango de la sesión asiática durante la apertura de la sesión de Londres en EUR/USD. Sesiones en UTC: asiática 00:00–07:00, Londres 07:00–12:00, NY 12:00–17:00. Por cada día: computar `high_asian = max(high in 00:00–07:00)` y `low_asian = min(low in 00:00–07:00)`. Durante la sesión Londres, monitorear el primer breakout del rango asiático: entrar long si el high de la barra excede `high_asian`, entrar short si el low excede `low_asian`. Salida obligatoria al cierre de NY (17:00 UTC). Sin stops en la versión vanilla. La estrategia es ampliamente documentada en literatura forex retail y material educativo de prop firms desde al menos 2010.

Hipótesis económica subyacente

El volumen EUR/USD durante la sesión asiática está dominado por flujos asiáticos de arbitraje y actividad retail pequeña; la paridad es típicamente range-bound en 00:00–07:00 UTC. Cuando los desks de Londres abren a 07:00 UTC, el posicionamiento relacionado con ECB y los flujos institucionales europeos establecen el sesgo direccional del día. El breakout del rango asiático captura la dirección dominante del flujo intra-día con follow-through medio hacia el cierre de NY. La hipótesis cumple los cinco criterios PFP-MP.

Resultados del backtest vanilla

El backtest produce 2,584 trades en 10 años con win rate de 59.6%, average winner de +0.33%, average loser de -0.28%, profit factor de 1.76, retorno total de 771%, retorno anualizado de 15.5%, volatilidad anualizada de 5.4%, Sharpe ratio de 2.67, Sortino ratio de 4.48, Calmar ratio clásico de 2.38 y máximo drawdown porcentual de -6.5%. Average trade duration de 8.6 horas. Los números son sustancialmente más fuertes que los de los dos casos anteriores y, en aislamiento, sugerirían una estrategia con perfil candidato a banda alta.

Aplicación de los validadores TRAVIDENCE

Cifras del dictamen original congelado; no reproducibles contra metrics.json vigente — ver nota §6.

El validador `deflated_sharpe_ratio` con `n_trials_strict = n_trials_session = 3` (las tres ventanas de sesión) devuelve `dss_strict = 1.0`. PASS limpio: el Sharpe observado es suficientemente alto y el espacio de selección suficientemente pequeño para superar el umbral institucional con margen.

El validador `capital_normalized_calmar` con `nominal_capital = $100,000` (1 lote estándar) y `leverage_factor = 30×` (forex retail ESMA) devuelve `capital_baseline = $3,333.33`. El Calmar margen-frame es 2.38 y el Calmar capital-frame es 1.66, con `inflation_ratio = 1.44×`. El Calmar margen-frame supera claramente el umbral institucional; el Calmar capital-frame también lo hace, con margen menor. Esta es la divergencia central que el dual frame de TRAVIDENCE hace visible: una estrategia que aparentaría perfil “strong edge” en frame de margen (2.38) clarifica su perfil real como “credible edge pero ya no claramente strong” en frame de capital (1.66) — el re-framing del 44% es información operacionalmente relevante para la decisión de fundability.

El validador `subperiod_monotonic_decay` con `K=5` devuelve la secuencia [2.59, 1.78, 3.03, 3.67, 2.32]. `passes_no_decay = True`: la estrategia no decae; oscila en torno a un Sharpe alto sin trayectoria monotónica.

El validador `regime_must_pass` es donde el caso revela su perfil crítico. Las 203 observaciones vol-on registran un Sharpe de 3.2 y un drawdown máximo de 4.1%: el régimen vol-on supera con holgura tanto el piso de Sharpe como el techo de drawdown del bucket. Las 3,589 observaciones vol-off registran un Sharpe de 2.7 —también por encima del piso— pero un drawdown de 5.4% que no satisface el criterio de drawdown del bucket: el régimen vol-off pasa el criterio de Sharpe y **falla** el de drawdown. Esa falla es suficiente para que el AND across regimes sea False y el gate de régimen dispare.

Score de Robustez y dictamen — el comportamiento gate-vs-continuous

La agregación continua de las tres dimensiones produce un perfil situado en la zona media “ambigua” del consumer band ladder. En un sistema de agregación pura — el modelo de agregación pura — ese perfil se reportaría probablemente como una banda intermedia favorable, con recomendación de revisión menor. La estrategia obtendría una clasificación favorable o tibiamente favorable.

En el sistema híbrido continuo-puerta TRAVIDENCE, el gate de régimen disparado reemplaza el perfil continuo con el cap del gate: banda No Robusta, capped_by: regime. La transición cualitativa de un perfil continuo-favorable a No Robusta es el comportamiento diferencial gate-vs-continuous que constituye el moat metodológico defendible de TRAVIDENCE. El cliente recibe una señal binaria clara: la estrategia no es viable para despliegue retail prop firm en su forma actual, independientemente de que la calidad continua de las dimensiones EQ y RP la situaría en un rango más favorable. (Dictamen original congelado; ver la nota de re-auditoría al inicio de §6.)

La Figura 3 (Apéndice B; PDF en `../.../.../case_studies/03-asian-range-breakout-eurusd/audit_certificates/03-eurusd-asian-range-breakout-audit-certificate.pdf`) reproduce el certificado de auditoría completo.

Interpretación — por qué el caso EUR/USD ancló la auditoría original

El caso EUR/USD fue el caso ancla de la auditoría original por tres razones convergentes; bajo la re-auditoría con el engine de producción (nota de re-auditoría, §6) el gate de régimen ya no vincula y el rol vigente del caso es la demostración del score público end-to-end, mientras el rol de demostración de gates pasa al caso NQ ORB del repositorio.

Primero, es el caso donde la agregación continua y la agregación con gates duros producen verdicts cualitativamente distintos. En los casos CL y BTC, el perfil continuo está estructuralmente bajo el umbral incluso antes de que los gates disparen, y el verdict No Robusta es robusto frente a la elección del esquema de agregación. En EUR/USD, el perfil continuo es marginalmente favorable y la decisión depende exactamente de si el sistema incorpora gates duros o no. Es el caso que separa la metodología TRAVIDENCE de la metodología consumer aggregator.

Segundo, el caso ilustra que un fallo *marginal* de un criterio en un *único* régimen es suficiente para disparar el gate más severo. El régimen vol-off pasa el criterio Sharpe con holgura y no satisface el criterio de drawdown del régimen. La regla AND across regimes no es contemplativa de la magnitud del fallo; el gate dispara y el cap aplica. Esta inflexibilidad es funcional al diseño audit-grade: una estrategia con asimetría detectable es una estrategia que el mercado eventualmente expone, y la metodología debe rechazarla antes del despliegue, no flexibilizar el criterio en función de la pequeñez del margen.

Tercero, el caso es operacionalmente representativo del perfil de estrategia que llega a las firmas de prop trading con la convicción de su viabilidad: Sharpe 2.67, profit factor 1.76, drawdown agregado controlado, mecanismo plausible. Sin la dimensión de régimen evaluada explícitamente, la estrategia pasa todos los filtros internos que el challenger habría aplicado por su cuenta. La detección viene de la metodología externa, audit-grade, con tercera parte independiente. Es precisamente este perfil el que motiva la propuesta de valor TRAVIDENCE al cliente retail.

§7. Comparación frente a herramientas existentes

§7.1 Tabla feature-by-feature

La Tabla 1 contrasta TRAVIDENCE con la categoría de agregadores consumer y las librerías de implementación de referencia sobre dimensiones críticas para el caso de uso de pre-validación prop firm. Las marcas ✓ y ✗ reflejan la disponibilidad de la capacidad en la oferta típica de cada columna a la fecha de publicación (v1.0).

Tabla 1. Capacidades comparadas: TRAVIDENCE frente a la categoría de agregadores consumer y las librerías de implementación (mlfinlab, RiskLabAI).

Capacidad	TRAVIDENCE	Agregador consumer (categoría)	mlfinlab	RiskLabAI
Deflated Sharpe Ratio implementado	✓	✓	✓	✓
Reporte dual n_trials strict / session	✓	✗	✗	✗
Calmar normalizado por capital (dual frame)	✓	✗	✗	✗
Sub-period monotonic decay con AND semantics	✓	✗	✗	✗
Component isolation con floor $ 1 - \sqrt{K} $	✓	✗	✗	✗
MUST-PASS regimes vol-on / vol-off	✓	✗	✗	✗
CPCV (Combinatorial Purged Cross-Validation)	parcial (Fase 5 upstream)	✗	✓	✓
PBO (Probability of Backtest Overfitting)	parcial (Fase 5 upstream)	✗	✓	✓
Sistema de gates duros (cliff-edge audit)	✓	✗	n/a	n/a

Capacidad	TRAVIDENCE	Agregador consumer (categoría)	mlfinlab	RiskLabAI
Familia pública de bandas con Abstención de Opinión para muestra insuficiente	✓	✗	n/a	n/a
Score [0,100] entregable como certificado	✓	✓	✗	✗
Calibración trade-secret por clase de activo	✓	parcial	✗	✗
Certificación firmada por tercera parte	✓	✗	✗	✗
Bilingüe ES + EN como entregable	✓	✗	✗	✗
Pricing tier comercial transparente	✓	✓	n/a	n/a
Acceso al código fuente del validador	Especificación pública completa (§4 + Apéndice A)	✗	visible (licencia comercial)	✓

La columna del agregador consumer describe la oferta típica de la categoría a la fecha de publicación, no un producto específico.

Las celdas marcadas n/a corresponden a productos cuya naturaleza (librería técnica) no contempla el feature evaluado (entrega comercial al cliente, certificación firmada). El check parcial en CPCV y PBO para TRAVIDENCE indica que las dos técnicas están contempladas como pasos upstream del orquestador de Fase 5 que alimenta el Score de Robustez; la integración tight con el cuerpo de validadores es la diferencia respecto a las librerías que lo implementan como funcionalidad standalone.

§7.2 Discusión de los diferenciadores

La lectura honesta de la Tabla 1 es que TRAVIDENCE no inventa la metodología de validación cuantitativa. El stack López de Prado (DSR, CPCV, PBO) lo inventó López de Prado y colaboradores; mlfinlab y RiskLabAI lo implementaron como librerías de referencia (RiskLabAI open-source; mlfinlab bajo licencia comercial). TRAVIDENCE hace tres cosas distintas: enriquecer el stack con cinco contribuciones novel (§4), operacionalizarlo como servicio audit-grade con certificación de tercera parte (capa de servicio ausente en las librerías), y calibrar el sistema de gates duros que distingue funcionalmente la entrega TRAVIDENCE de la entrega consumer aggregator.

Las cinco contribuciones novel son ortogonales al stack base. El dual reporting de `n_trials` (§4.1) extiende el DSR estándar sin invalidarlo; el Calmar capital-frame (§4.2) extiende el Calmar tradicional sin reemplazarlo; el esquema sub-período K-chunk (§4.3) generaliza el schema 3-period que era reporting convention sin reemplazar el análisis agregado; el aislamiento de componentes con $\text{floor} \lfloor 1 - \sqrt{K} \rfloor$ (§4.4) es una primitiva estructural complementaria a los tests univariados aplicados al agregado; la taxonomía MUST-PASS vol-on / vol-off (§4.5) instancia el principio general de validación multi-régimen de López de Prado con un protocolo binario operacionalizable. Cada una de las cinco está especificada por completo en este documento (§4 y Apéndice A) y es replicable por cualquier tercero a partir de las ecuaciones publicadas.

El moat de TRAVIDENCE no es técnico-secreto sobre la metodología (que es publicable, como el presente documento demuestra). El moat es institucional: (i) la calibración trade-secret del Score de Robustez por clase de activo (knots, gate caps, band thresholds), validada empíricamente sobre población sintética inicial y con recalibración trimestral planificada sobre data anonimizada acumulada en producción; (ii) la firma de tercera parte sobre el dictamen, con disciplina procesal heredada del cuerpo auditor profesional; (iii) la marca institucional construida sobre el track record de dictámenes que se sostienen en el tiempo. El equivalente conceptual en otro mercado es la diferencia entre publicar el GAAP/IFRS (el aparato técnico, público) y operar una firma de auditoría big-four (la capa de servicio, institucional). TRAVIDENCE se posiciona en el segundo lugar.

§8. Discusión

§8.1 Implicación para pre-validación retail prop firm

La aplicación inmediata de la metodología TRAVIDENCE es la pre-validación de estrategias antes de que el challenger pague la fee de evaluación. Bajo el perfil documentado de 86% de fracaso de challenges, el cálculo esperado para el challenger es: una fee del orden de USD 300–600 por intento para una cuenta estándar de 50,000–100,000 USD, contra una probabilidad del 14% de pasar el challenge y una probabilidad del 7% (relativa al total) de recibir payout efectivo. El expected value antes de pre-validación es negativo en la inmensa mayoría de los perfiles de estrategia. Una pre-validación honesta, a un precio sustancialmente menor que el fee y con dictamen audit-grade, recorta el universo de estrategias a aquellas con probabilidad razonable de superar el challenge, transformando el cálculo esperado del challenger.

La economía del producto retail debe ser compatible con el orden de magnitud del fee del challenge. El pricing de TRAVIDENCE está calibrado para situarse por debajo del fee típico de evaluación pero por encima del nivel micro-transaccional de la oferta consumer aggregator. La razón es estructural: el grado de auditoría requiere un proceso de revisión que no es escalable a precios micro-transaccionales. La franja entre los dos extremos — escalable pero institucional — es el espacio comercial natural del producto. Precios vigentes en travidence.com.

§8.2 Implicación para pre-screening B2B white-label

La segunda aplicación es B2B: las firmas de prop trading tier-2 (firmas más pequeñas que las top-tier) que enfrentan un costo sustancial en cuentas financiadas cuando los challengers pasan la evaluación pero nunca alcanzan un payout. Para esas firmas, integrar TRAVIDENCE como filter pre-evaluation reduce su exposición a cuentas que pagarán el fee de challenge pero que estructuralmente no completarán el ciclo. La firma B2B obtiene tres beneficios concurrentes: (i) reducción del costo de cuentas financiadas que no alcanzan payout, (ii) diferenciación competitiva frente a firmas tier-1 mediante el sello de pre-validation independiente, (iii) externalización de la decisión técnica de validation a una tercera parte cuyo conflicto de interés está estructuralmente acotado.

La estructura comercial es Institutional Engagement, con scope custom y onboarding técnico para integración API. La calibración por bucket de asset class (futuros vs forex/CFD vs crypto vs equities) preserva la trade-secret de los knots mientras provee al cliente B2B un servicio configurado a su universo de challengers.

§8.3 La analogía Big Four

El paralelo conceptual entre TRAVIDENCE y la oferta Big Four de auditoría financiera no es meramente retórico; es estructural. Las firmas Big Four (Deloitte, EY, PwC, KPMG) operan en el mercado de aseguramiento financiero como tercera parte independiente que firma opiniones (Unqualified / Qualified / Adverse / Disclaimer) sobre estados financieros bajo estándares procesales (GAAP/IFRS) y de auditoría (SAS/ISA). Su valor para el mercado deriva de tres atributos: independencia (reglas que restringen los servicios susceptibles de crear conflicto con el dictamen), estándares (la disciplina procesal está codificada y es auditable retrospectivamente), y reputación (el track record del dictamen sostiene la prima del servicio).

TRAVIDENCE se propone como el equivalente conceptual en el mercado de aseguramiento de estrategias cuantitativas, con tres ajustes contextuales: (i) el objeto auditado es la metodología de validación de la estrategia, no estados financieros (la analogía es con el aseguramiento del proceso, no del resultado financiero); (ii) los estándares procesales son la metodología TRAVIDENCE documentada en el presente paper y en docs/methodology/methodology-generalized.md, inspirada en la disciplina procesal AICPA (2011) SAS-122 / IAASB (2015) ISA 705 del cuerpo auditor profesional; (iii) el mercado objetivo retail-plus-B2B-tier-2 es estructuralmente subatendido por los incumbentes Big Four, que no han considerado relevante el segmento de pre-validation de estrategias retail. La franja institucional vacía es precisamente la oportunidad que el modelo TRAVIDENCE articula.

§9. Limitaciones

§9.1 Estrategias sintéticas únicamente

Por la política de separación firma/fondo (§5.2), los tres casos del presente documento son estrategias textbook sobre data pública. La metodología TRAVIDENCE no se ha aplicado en este docu-

mento a estrategias propietarias de ningún cliente real ni a estrategias del fondo propio de ningún empleado o asesor de TRAVIDENCE. La validación de la metodología sobre real-world strategies de clientes ocurre en producción comercial y se acumula como track record post-launch; el presente paper es la base metodológica, no la evidencia de track record. La separación es deliberada y se mantiene como invariante institucional.

§9.2 Idiomas ES + EN únicamente

La versión 1.0 del documento existe en español y en inglés, con traducción profesional entre ambos. Otros idiomas relevantes para el mercado objetivo retail (portugués brasileño, francés, mandarín) están diferidos a futuras versiones. La doble entrega ES + EN es el mínimo viable para el mercado TRAVIDENCE inicial (LATAM hispanohablante + audiencia institucional global anglófona).

§9.3 Calibración v1.0 sobre población sintética

La calibración inicial del Score de Robustez v1.0 — knots de las mapping functions, gate cap levels, band thresholds — se realiza sobre una población sintética construida a partir de los tres case studies del presente documento más una población simulada de variantes de las mismas estrategias bajo perturbaciones controladas. Esa calibración constituyó el `calibration_vintage` inicial v1.0.0 / 2026-Q2; el vintage vigente documentado en los metadata de los certificados de auditoría es v1.1.0 / 2026-Q3 (recalibración de los criterios must-pass por régimen, 2026-07-05). La recalibración trimestral con data real anonimizada acumulada en producción (la “Capa 3” del defense stack institucional, referenciada en el spec interno) es protocolo a partir del trimestre posterior al launch. Hasta entonces, los dictámenes emitidos llevan la versión de calibración explícita y son auditables retrospectivamente bajo esa versión.

§10. Conclusión

La pre-validación de estrategias cuantitativas con grado de auditoría es una franja institucional vacante entre las librerías de implementación del stack López de Prado y la oferta retail consumer aggregator. TRAVIDENCE articula esa franja a través de tres componentes integrados: cinco contribuciones novel al stack López de Prado (reporte dual de `n_trials` en DSR, Calmar normalizado por capital, esquema sub-período con detección de decay monotónico, aislamiento de componentes con floor $|1 - \sqrt{K}|$, y taxonomía MUST-PASS de regímenes vol-on / vol-off); un score híbrido continuo-puerta — el Score de Robustez — con una banda pública de resultado inspirada en la disciplina procesal AICPA SAS-122 / ISA 705; y un perfil comercial transparente con tiers de pricing diseñados para coexistir con el fee típico de challenge prop firm.

Tres estudios de caso textbook sobre data pública (Golden Cross sobre CL, RSI(14) mean reversion sobre BTC perpetuo, Asian Range Breakout sobre EUR/USD) ilustran la aplicación del filtro; ninguno alcanza banda Robusta. Bajo la re-auditoría con el engine de producción (nota §6), CL y BTC cierran con Abstención de Opinión por muestra insuficiente, y EUR/USD emite el score

público end-to-end — 56 / 100, Robustez Limitada, sin gate vinculante. El comportamiento diferencial gate-vs-continuous que separa la metodología TRAVIDENCE de la metodología consumer aggregator — y constituye el moat metodológico defendible del producto — lo exhibe el caso NQ Opening Range Breakout del repositorio, admitido con 2,424 trades y capado por los gates de régimen y sub-período; a través de los cinco case studies del repositorio, el portafolio cubre los tres desenlaces públicos: score emitido, gated e insufficient sample.

El trabajo futuro inmediato incluye (i) la certificación SOC 2 Type I como capa de aseguramiento procesal sobre la operación misma de la firma (gate post-launch a 6–12 meses), (ii) la firma de LOIs con prop firms tier-2 para integración B2B white-label, y (iii) la calibración periódica del Score de Robustez con data real anonimizada acumulada en producción a partir del trimestre posterior al launch. La construcción del track record de dictámenes que se sostienen en el tiempo — el equivalente TRAVIDENCE del cuerpo de jurisprudencia auditora Big Four — es el eje de defensa institucional a mediano plazo.

Declaración sobre IA generativa y tecnologías asistidas por IA en el proceso de redacción

Durante la preparación de este trabajo, el autor utilizó Claude (Anthropic) para asistir la redacción y el lenguaje del manuscrito, bajo su dirección. La metodología, el análisis, las decisiones de diseño y las conclusiones son del autor. Tras el uso de esta herramienta, el autor revisó y editó el contenido y asume la plena responsabilidad por el contenido de la publicación.

Descargo de responsabilidad

TRAVIDENCE es un servicio de validación independiente. No es asesoría financiera ni una recomendación de inversión. El rendimiento pasado no garantiza resultados futuros; ninguna validación elimina el riesgo de pérdida.

Referencias

American Institute of Certified Public Accountants (AICPA). (2011). Statement on Auditing Standards No. 122 — *Statements on Auditing Standards: Clarification and Recodification*. AU-C Section 705: Modifications to the Opinion in the Independent Auditor's Report. New York: AICPA.

Arian, H., Norouzi Mobarekeh, D., y Seco, L. (2024). Backtest Overfitting in the Machine Learning Era: A Comparison of Out-of-Sample Testing Methods in a Synthetic Controlled Environment. *Knowledge-Based Systems*, 305, 112477. doi:10.1016/j.knosys.2024.112477.

- Bailey, D. H., Borwein, J. M., López de Prado, M., y Zhu, Q. J. (2017). The Probability of Backtest Overfitting. *Journal of Computational Finance*, 20(4), 39–69. doi:10.21314/jcf.2016.322.
- Bailey, D. H., y López de Prado, M. (2012). The Sharpe Ratio Efficient Frontier. *Journal of Risk*, 15(2), 3–44. doi:10.21314/JOR.2012.255.
- Bailey, D. H., y López de Prado, M. (2014). The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality. *Journal of Portfolio Management*, 40(5), 94–107.
- Brock, W., Lakonishok, J., y LeBaron, B. (1992). Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *Journal of Finance*, 47(5), 1731–1764. doi:10.1111/j.1540-6261.1992.tb04681.x.
- Chan, E. (2013). *Algorithmic Trading: Winning Strategies and Their Rationale*. Hoboken: John Wiley & Sons.
- Clenow, A. (2013). *Following the Trend: Diversified Managed Futures Trading*. Hoboken: John Wiley & Sons.
- Finance Magnates. (2024, septiembre). Exclusive: Only 7% of 300,000 Prop Trading Accounts Achieved Payouts. <https://www.financemagnates.com/forex/analysis/exclusive-only-7-of-300000-prop-trading-accounts-achieved-payouts/> (Dataset propietario de FPFx Tech; Justin Hertzberg, fundador y CEO de FPFx Tech, citado.)
- Hong, H., y Yogo, M. (2012). What Does Futures Market Interest Tell Us About the Macroeconomy and Asset Prices? *Journal of Financial Economics*, 105(3), 473–490. doi:10.1016/j.jfineco.2012.01.005.
- International Auditing and Assurance Standards Board (IAASB). (2015). ISA 700 — Forming an Opinion and Reporting on Financial Statements; ISA 705 — Modifications to the Opinion in the Independent Auditor’s Report; ISA 706 — Emphasis of Matter Paragraphs and Other Matter Paragraphs in the Independent Auditor’s Report. IFAC.
- Lo, A. W. (2002). The Statistics of Sharpe Ratios. *Financial Analysts Journal*, 58(4), 36–52. doi:10.2469/faj.v58.n4.2453.
- López de Prado, M. (2018). *Advances in Financial Machine Learning*. Hoboken: John Wiley & Sons.
- Mertens, E. (2002). Comments on the Correct Variance of Estimated Sharpe Ratios in Lo (2002, FAJ) When Returns Are IID. Working paper, University of Basel.
- Politis, D. N., y Romano, J. P. (1994). The Stationary Bootstrap. *Journal of the American Statistical Association*, 89(428), 1303–1313. doi:10.1080/01621459.1994.10476870.
- Wilder, J. W. (1978). *New Concepts in Technical Trading Systems*. Greensboro, NC: Trend Research.
- Young, T. W. (1991, octubre). Calmar Ratio: A Smoother Tool. *Futures*.

Apéndice A — Derivaciones matemáticas

A.1 Varianza del estimador del Sharpe (Mertens 2002, sobre Lo 2002)

Para una serie de retornos $\{r_1, r_2, \dots, r_T\}$ con media μ , desviación σ , skew γ_3 y kurtosis Pearson γ_4 , Mertens (2002) deriva, sobre la base de Lo (2002):

$$\hat{\sigma}_{SR}^2 = \frac{1 - \gamma_3 \cdot SR_{obs} + \frac{\gamma_4 - 1}{4} \cdot SR_{obs}^2}{T - 1}$$

donde $SR_{obs} = \mu/\sigma$ es el Sharpe muestral (no anualizado). La derivación parte del cuadrado del momento ratio bajo retornos iid no necesariamente normales y aplica el método delta sobre la transformación $f(\mu, \sigma) = \mu/\sigma$. La extensión TRAVIDENCE implementa la fórmula con la convención `scipy.stats.bias=False`, `fisher=False` para alineamiento bit-identical con la forma original de Mertens (2002).

A.2 Forma exacta de $E[\max SR \mid N]$ (Bailey & López de Prado 2014)

El paper original deriva el valor esperado del máximo Sharpe bajo el null de N trials independientes:

$$E[\max SR \mid N] = \sqrt{V[\{SR_n\}]} \cdot \left[(1 - \gamma_E) \Phi^{-1}\left(1 - \frac{1}{N}\right) + \gamma_E \Phi^{-1}\left(1 - \frac{1}{Ne}\right) \right]$$

donde $\gamma_E \approx 0.5772156649$ es la constante de Euler-Mascheroni. La derivación es la combinación convexa de los percentiles $1 - 1/N$ y $1 - 1/(Ne)$ de la distribución asintótica del máximo de N extracciones normales, con peso $(1 - \gamma_E)$ y γ_E respectivamente. La aproximación Gumbel asintótica $\sqrt{2 \log N} - \gamma_E / \sqrt{2 \log N}$ sobreestima esta forma exacta en ~9-19% ($N = 10$ -1000; la desviación decrece al crecer N); la implementación canónica TRAVIDENCE usa la forma exacta para preservar el conservadurismo declarado del test.

A.3 Identidad algebraica del inflation_ratio Calmar

Partiendo de las definiciones canónicas:

$$\text{Calmar}_{\text{margin}} = \frac{r_{ann}}{|DD_{t^*}/P_{t^*}|}, \quad \text{Calmar}_{\text{capital}} = \frac{r_{ann}}{|DD_{t^*}/\text{capital_baseline}|}$$

el ratio es:

$$\frac{\text{Calmar}_{\text{margin}}}{\text{Calmar}_{\text{capital}}} = \frac{r_{ann}/(DD_{t^*}/P_{t^*})}{r_{ann}/(DD_{t^*}/\text{capital_baseline})} = \frac{DD_{t^*}/\text{capital_baseline}}{DD_{t^*}/P_{t^*}} = \frac{P_{t^*}}{\text{capital_baseline}}$$

con cancelación de r_{ann} y DD_{t^*} . La conclusión es que `inflation_ratio` es la razón entre el rolling peak al momento del peor drawdown y el capital baseline declarado; coincide con `leverage_factor` solo en casos especiales que la práctica rara vez satisface.

A.4 Floor $|1 - \sqrt{K}|$ del component isolation

Bajo aditividad ($r_{full} = \sum_i r_i$) e independencia (componentes IID con media μ y varianza σ^2), el agregado tiene media $K\mu$ y varianza $K\sigma^2$ (varianzas suman bajo independencia). Por construcción:

$$SR_{full} = \frac{K\mu}{\sqrt{K}\sigma} = \sqrt{K} \cdot \frac{\mu}{\sigma}$$

y la suma de Sharpes standalone es:

$$\sum_i SR_i^{\text{standalone}} = K \cdot \frac{\mu}{\sigma}$$

por lo tanto:

$$\frac{\rho_{\text{int}}}{SR_{full}} = \frac{SR_{full} - \sum_i SR_i}{SR_{full}} = 1 - \frac{K}{\sqrt{K}} = 1 - \sqrt{K}$$

El floor es determinístico bajo aditividad e independencia y crece monótonicamente con K . La calibración de `interaction_tolerance` debe acomodar este floor más un buffer para slack muestral finito.

A.5 Taxonomía MUST-PASS por clase de activo

El corte vol-on / vol-off es **derivado de la propia serie de la estrategia**: es el p50 (mediana) de la volatilidad rolling anualizada de esa serie (`vol_window` = 21, comparación estricta >), calculado in-engine sin ningún valor calibrado por clase de activo. Por ser un corte en la mediana es auto-referencial: es invariante a la anualización (el factor de anualización escala por igual la volatilidad rolling y su cuantil, y se cancela) y produce un split ~50/50 por construcción, de modo que ambos regímenes tienen poblaciones observacionales no degeneradas (por encima del floor `min_observations_per_regime`) sin necesidad de tuning. **No existe un `vol_threshold` per-bucket ni un campo `vol_threshold` en el output del validador** — el corte es una regla pública del framework, no un valor calibrado trade-secret. *(Nota histórica: los tres case studies del presente paper se produjeron bajo la calibración per-bucket anterior, ya retirada, cuyos cortes de volatilidad por bucket son trade-secret y no se publican; sus conteos de observación congelados reflejan ese diseño retirado, no la regla derivada vigente.)*

Apéndice B — Certificados de auditoría como figuras

Los cuatro certificados de auditoría — los tres correspondientes a los case studies §6.1–§6.3 más el caso adicional del repositorio referenciado en la nota de re-auditoría de §6 — se incluyen como figuras del documento. La inclusión binaria de los PDF en el render final del paper se realiza en tiempo de compilación por el render pipeline (pandoc/LaTeX). El markdown source-of-truth referencia los PDFs por path relativo; las versiones canónicas se mantienen en `case_studies/*/audit_certificates/` con versionado v1.1.0 (calibration vintage 2026-Q3).

- **Figura 1.** Certificado de auditoría TRAVIDENCE para Caso 01 — Golden Cross sobre CL futuros (2000–2025). Path: `../..../case_studies/01-golden-cross-cl/aud`

it_certificates/01-cl-golden-cross-audit-certificate.pdf. Banda final:
Abstención de Opinión — Evidencia Insuficiente.

ESTRATEGIA	CL Crude Oil Golden Cross (50/200 SMA)
CLASE DE ACTIVO	Commodity Futures — CL Crude Oil
PERÍODO CUBIERTO	2000-08-23 to 2025-05-16
ID DE CERTIFICADO	TRV-E8A6-E8E8
CALIFICADO BAJO	Score de Robustez v1.1.0 (2026-Q3)
EMITIDO	2026-08-17 04:42 UTC

SCORE DE ROBUSTEZ

N/A

Abstención de Opinión

VEREDICTO

Esta estrategia no puede certificarse bajo los estándares de TRAVIDENCE: $n_{trades} = 18$ por debajo del piso de admisibilidad. Genera datos de operaciones adicionales y vuelve a enviarla.

ESTADO DE PUERTAS

PUERTA	ESTADO	EFFECTO
Consistencia entre regímenes de volatilidad	APROBADA	—
Estabilidad por subperíodos	APROBADA	—
Aporte por componente	APROBADA	—
Combinado	Resultado final: sin puertas activadas	

DETALLE DE VALIDADORES

Ratio de Sharpe Deflactado

Variante DSR	DSR	n_trials	E[max SR] (anualizado)
Estricta	0.964	2	0.111
Por sesión	0.964	2	0.111

Calmar Normalizado por Capital

Marco	Calmar	Notas
Marco de margen	0.295	línea base = 5,555.56 (apalancamiento 9x)
Marco de capital	0.105	ratio de inflación = 2.808

Deterioro Monotónico por Subperíodos

Campo	Valor
Estado (sin deterioro)	REPROBADO
Métrica	sharpe
Por subperíodo (5 tramos)	0.57, 0.54, -0.72, -0.42, 0.37
Caída consecutiva máxima	234.59%
Deterioro total	34.49%

Regímenes Must-Pass

Régimen	Estado	min Sharpe	max DD	n obs
vol_off	REPROBADO	nan	0.00%	5205
vol_on	REPROBADO	1.130	17.36%	338

Aislamiento de Componentes: N/A — estrategia de un solo componente

REFERENCIAS METODOLÓGICAS

Bailey, D. H., & López de Prado, M. (2014). *The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality*. Journal of Portfolio Management 40(5).

Lo, A. W. (2002). *The Statistics of Sharpe Ratios*. Financial Analysts Journal 58(4): 36-52.

Young, T. W. (1991). *Calmar Ratio: A Smoother Tool*. Futures Magazine, October.

AICPA SAS-122 / IAASB ISA 705 & 706. *Modifications to the Opinion in the Independent Auditor's Report*.

Hash de Reproducibilidad (completo): sha256:e8a6e8e8fd238f82fa65da2c9eb2d20f1e8a9d8635b489e8ad9c752966d1de8c · esquema de hash 2

Nota de limitación sigma v1.1.0: Esta versión usa un manejo conservador de sigma-cero para la propagación del intervalo de confianza. Las entradas de Sharpe y Calmar de una sola realización tienen sigma=0 por convención; la propagación completa por el método delta se difiere a una versión futura. Los IC mostrados reflejan esta cota conservadora; intervalos más estrechos requieren remuestreo bootstrap o block-bootstrap.

Divulgación de calibración: Este certificado se emitió bajo v1.1.0 (2026-Q3), que representa una calibración mejor alcanzable del sistema Score de Robustez. La calibración es conservadora por diseño: las estrategias que reciben No Robusta pueden incluir algunas que podrían clasificarse como Robustez Limitada bajo una calibración refinada. La recalibración posterior con datos de producción está planificada como protocolo trimestral.

Generado por el motor de auditoría TRAVIDENCE. Calificado bajo v1.1.0 | calibración 2026-Q3.

Emitido por Travidence, LLC · Wilmington, Delaware, EE. UU.

Travidence, LLC mantiene un seguro de responsabilidad profesional tecnológica y ciberriesgo (Tech E&O/Cyber) en vigor.

- **Figura 2.** Certificado de auditoría TRAVIDENCE para Caso 02 — RSI(14) mean reversion sobre BTC perpetuo (2020–2025). Path: ../../../../case_studies/02-rsi-mean-reversion-btc/audit_certificates/02-btc-rsi-mean-reversion-audit-certificate.pdf. Banda final: Abstención de Opinión — Evidencia Insuficiente.

ESTRATEGIA	BTC Perpetual RSI 14 Mean Reversion
CLASE DE ACTIVO	Crypto Perpetual — BTC/USDT
PERÍODO CUBIERTO	2020-01-02 to 2025-05-19
ID DE CERTIFICADO	TRV-6A8E-058C
CALIFICADO BAJO	Score de Robustez v1.1.0 (2026-Q3)
EMITIDO	2026-08-17 04:42 UTC

SCORE DE ROBUSTEZ

N/A

Abstención de Opinión

VEREDICTO

Esta estrategia no puede certificarse bajo los estándares de TRAVIDENCE: $n_{trades} = 31$ por debajo del piso de admisibilidad. Genera datos de operaciones adicionales y vuelve a enviarla.

ESTADO DE PUERTAS

PUERTA	ESTADO	EFFECTO
Consistencia entre regímenes de volatilidad	APROBADA	—
Estabilidad por subperíodos	APROBADA	—
Aporte por componente	APROBADA	—
Combinado	Resultado final: sin puertas activadas	

DETALLE DE VALIDADORES

Ratio de Sharpe Deflactado

Variante DSR	DSR	n_trials	E[max SR] (anualizado)
Estricta	0.691	3	0.380
Por sesión	0.691	3	0.380

Calmar Normalizado por Capital

Marco	Calmar	Notas
Marco de margen	0.789	línea base = 1,500.00 (apalancamiento 20x)
Marco de capital	0.572	ratio de inflación = 1.379

Deterioro Monotónico por Subperíodos

Campo	Valor
Estado (sin deterioro)	APROBADO
Métrica	sharpe
Por subperíodo (5 tramos)	0.39, 0.98, -0.12, 1.01, 1.71
Caída consecutiva máxima	112.66%
Deterioro total	-332.04%

Regímenes Must-Pass

Régimen	Estado	min Sharpe	max DD	n obs
vol_off	REPROBADO	nan	0.00%	1228
vol_on	APROBADO	1.569	14.67%	596

Aislamiento de Componentes: N/A — estrategia de un solo componente

REFERENCIAS METODOLÓGICAS

Bailey, D. H., & López de Prado, M. (2014). *The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality*. Journal of Portfolio Management 40(5).

Lo, A. W. (2002). *The Statistics of Sharpe Ratios*. Financial Analysts Journal 58(4): 36-52.

Young, T. W. (1991). *Calmar Ratio: A Smoother Tool*. Futures Magazine, October.

AICPA SAS-122 / IAASB ISA 705 & 706. *Modifications to the Opinion in the Independent Auditor's Report*.

Hash de Reproducibilidad (completo): sha256:6a8e058c61179010c5be84e68f846e758acd4de1453d37b9b1539b317b5babcb6 · esquema de hash 2

Nota de limitación sigma v1.1.0: Esta versión usa un manejo conservador de sigma-cero para la propagación del intervalo de confianza. Las entradas de Sharpe y Calmar de una sola realización tienen sigma=0 por convención; la propagación completa por el método delta se difiere a una versión futura. Los IC mostrados reflejan esta cota conservadora; intervalos más estrechos requieren remuestreo bootstrap o block-bootstrap.

Divulgación de calibración: Este certificado se emitió bajo v1.1.0 (2026-Q3), que representa una calibración mejor alcanzable del sistema Score de Robustez. La calibración es conservadora por diseño: las estrategias que reciben No Robusta pueden incluir algunas que podrían clasificarse como Robustez Limitada bajo una calibración refinada. La recalibración posterior con datos de producción está planificada como protocolo trimestral.

Generado por el motor de auditoría TRAVIDENCE. Calificado bajo v1.1.0 | calibración 2026-Q3.

Emitido por Travidence, LLC · Wilmington, Delaware, EE. UU.

Travidence, LLC mantiene un seguro de responsabilidad profesional tecnológica y ciberriesgo (Tech E&O/Cyber) en vigor.

- **Figura 3.** Certificado de auditoría TRAVIDENCE para Caso 03 — Asian Range Breakout sobre EUR/USD (2015–2025). Path: ../../../../case_studies/03-asian-range-breakout-eurusd/audit_certificates/03-eurusd-asian-range-breakout-audit-certificate.pdf. Banda final: Robustez Limitada, score público 56 / 100, sin gate vinculante. Caso del score público end-to-end — la primera emisión real de score non-gated del framework.

ESTRATEGIA	EUR/USD Asian Range Breakout
CLASE DE ACTIVO	Forex Major — EUR/USD
PERÍODO CUBIERTO	2015-01-01 to 2025-05-19
ID DE CERTIFICADO	TRV-AEC6-AD46
CALIFICADO BAJO	Score de Robustez v1.1.0 (2026-Q3)
EMITIDO	2026-08-17 04:42 UTC

SCORE DE ROBUSTEZ

56 / 100

Robustez Limitada

VEREDICTO

La estrategia muestra calidad continua por debajo del umbral de Robustez y exhibe debilidades materiales en uno o más validadores. Se requieren cambios estructurales antes del desafío.

ESTADO DE PUERTAS

PUERTA	ESTADO	EFFECTO
Consistencia entre regímenes de volatilidad	APROBADA	—
Estabilidad por subperíodos	APROBADA	—
Aporte por componente	APROBADA	—
Combinado	Resultado final: sin puertas activadas	

DETALLE DE VALIDADORES

Ratio de Sharpe Deflactado

Variante DSR	DSR	n_trials	E[max SR] (anualizado)
Estricta	1.000	3	0.260
Por sesión	1.000	3	0.260

Calmar Normalizado por Capital

Marco	Calmar	Notas
Marco de margen	2.360	línea base = 3,333.33 (apalancamiento 30x)
Marco de capital	1.728	ratio de inflación = 1.366

Deterioro Monotónico por Subperíodos

Campo	Valor
Estado (sin deterioro)	APROBADO
Métrica	sharpe
Por subperíodo (5 tramos)	3.17, 2.11, 3.54, 4.46, 2.74
Caída consecutiva máxima	38.49%
Deterioro total	13.42%

Regímenes Must-Pass

Régimen	Estado	min Sharpe	max DD	n obs
vol_off	APROBADO	3.762	1.85%	1343
vol_on	APROBADO	2.782	4.83%	1343

Aislamiento de Componentes: N/A — estrategia de un solo componente

REFERENCIAS METODOLÓGICAS

Bailey, D. H., & López de Prado, M. (2014). *The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality*. Journal of Portfolio Management 40(5).

Lo, A. W. (2002). *The Statistics of Sharpe Ratios*. Financial Analysts Journal 58(4): 36-52.

Young, T. W. (1991). *Calmar Ratio: A Smoother Tool*. Futures Magazine, October.

AICPA SAS-122 / IAASB ISA 705 & 706. *Modifications to the Opinion in the Independent Auditor's Report*.

Hash de Reproducibilidad (completo):
sha256:aec6ad462f11f53609e946864136edfb953551bcc892906aa713243e1d2db5d · esquema de hash 2

Nota de limitación sigma v1.1.0: Esta versión usa un manejo conservador de sigma-cero para la propagación del intervalo de confianza. Las entradas de Sharpe y Calmar de una sola realización tienen sigma=0 por convención; la propagación completa por el método delta se difiere a una versión futura. Los IC mostrados reflejan esta cota conservadora; intervalos más estrechos requieren remuestreo bootstrap o block-bootstrap.

Divulgación de calibración: Este certificado se emitió bajo v1.1.0 (2026-Q3), que representa una calibración mejor alcanzable del sistema Score de Robustez. La calibración es conservadora por diseño: las estrategias que reciben No Robusta pueden incluir algunas que podrían clasificarse como Robustez Limitada bajo una calibración refinada. La recalibración posterior con datos de producción está planificada como protocolo trimestral.

Generado por el motor de auditoría TRAVIDENCE. Calificado bajo v1.1.0 | calibración 2026-Q3.

Emitido por Travidence, LLC · Wilmington, Delaware, EE. UU.

Travidence, LLC mantiene un seguro de responsabilidad profesional tecnológica y ciberriesgo (Tech E&O/Cyber) en vigor.

Nota sobre el panel “Estado de Puertas” (Figuras 1–3): en un dictamen con Abstención de Opinión la evaluación de puertas nunca ocurre — el control de admisibilidad detiene el flujo antes de que exista un score compuesto que una puerta pueda capar. El panel reporta esa no-vinculación; los resultados subyacentes por validador, dos de los cuales no se satisfacen en los Casos 01 y 02, se reportan en la tabla “Detalle de Validadores” de la misma página. La plantilla del certificado carece de un tercer estado visual y rotula la no-vinculación como “APROBADA”; la misma limitación de dos estados aplica a la fila del aporte por componente cuando el test se reporta N/A por tratarse de estrategias de un solo componente (Figuras 1–3). Esta limitación de presentación queda registrada para corrección en una versión futura de la plantilla del certificado.

- **Figura 4.** Certificado de auditoría TRAVIDENCE para Caso 04 (repositorio) — NQ Opening Range Breakout (2015–2025), evaluado bajo el engine de producción vigente. Path: `../..../..../case_studies/04-opening-range-breakout-nq/audit_certificates/04-nq-opening-range-breakout-audit-certificate.pdf`. Banda final: No Robusta — gate vinculante [regime, subperiod]. Caso de demostración de gate vinculante referenciado en la nota de re-auditoría de §6.

ESTRATEGIA	NQ Opening Range Breakout (ORB)
CLASE DE ACTIVO	Equity Index Futures — E-mini Nasdaq-100 (NQ)
PERÍODO CUBIERTO	2015-01-02 to 2025-12-31
ID DE CERTIFICADO	TRV-DBCF-E8D7
CALIFICADO BAJO	Score de Robustez v1.1.0 (2026-Q3)
EMITIDO	2026-08-17 04:42 UTC

SCORE DE ROBUSTEZ

limitado por la puerta de régimen

No Robusta

VEREDICTO

La estrategia no mantiene su desempeño en ambos regímenes de volatilidad (alta y baja). Score limitado por la puerta de régimen. La estrategia NO es desplegable en su forma actual para desafíos de prop firm. Se recomiendan cambios significativos antes de la revalidación.

ESTADO DE PUERTAS

PUERTA	ESTADO	EFFECTO
Consistencia entre regímenes de volatilidad	REPROBADA	límite vinculante
Estabilidad por subperíodos	REPROBADA	límite no vinculante
Aporte por componente	APROBADA	—
Combinado	Resultado final: limitado por la puerta de régimen	

DETALLE DE VALIDADORES

Ratio de Sharpe Deflactado

Variante DSR	DSR	n_trials	E[max SR] (anualizado)
Estricta	0.066	48	0.670
Por sesión	0.462	3	0.253

Calmar Normalizado por Capital

Marco	Calmar	Notas
Marco de margen	0.043	línea base = 5,555.56 (apalancamiento 9x)
Marco de capital	0.041	ratio de inflación = 1.032

Deterioro Monotónico por Subperíodos

Campo	Valor
Estado (sin deterioro)	REPROBADO
Métrica	sharpe
Por subperíodo (5 tramos)	-1.15, 0.16, -0.28, 1.25, 1.15
Caída consecutiva máxima	276.72%
Deterioro total	nan%

Regímenes Must-Pass

Régimen	Estado	min Sharpe	max DD	n obs
vol_off	REPROBADO	0.011	19.66%	1425
vol_on	REPROBADO	0.326	40.53%	1424

Aislamiento de Componentes: N/A — estrategia de un solo componente

REFERENCIAS METODOLÓGICAS

Bailey, D. H., & López de Prado, M. (2014). *The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality*. Journal of Portfolio Management 40(5).

Lo, A. W. (2002). *The Statistics of Sharpe Ratios*. Financial Analysts Journal 58(4): 36-52.

Young, T. W. (1991). *Calmar Ratio: A Smoother Tool*. Futures Magazine, October.

AICPA SAS-122 / IAASB ISA 705 & 706. *Modifications to the Opinion in the Independent Auditor's Report*.

Hash de Reproducibilidad (completo): sha256:dbcfe8d71e2b03d508a926047dbe061b6f2453f7835277d06757137060996ccb · esquema de hash 2

Nota de limitación sigma v1.1.0: Esta versión usa un manejo conservador de sigma-cero para la propagación del intervalo de confianza. Las entradas de Sharpe y Calmar de una sola realización tienen sigma=0 por convención; la propagación completa por el método delta se difiere a una versión futura. Los IC mostrados reflejan esta cota conservadora; intervalos más estrechos requieren remuestreo bootstrap o block-bootstrap.

Divulgación de calibración: Este certificado se emitió bajo v1.1.0 (2026-Q3), que representa una calibración mejor alcanzable del sistema Score de Robustez. La calibración es conservadora por diseño: las estrategias que reciben No Robusta pueden incluir algunas que podrían clasificarse como Robustez Limitada bajo una calibración refinada. La recalibración posterior con datos de producción está planificada como protocolo trimestral.

Generado por el motor de auditoría TRAVIDENCE. Calificado bajo v1.1.0 | calibración 2026-Q3.

Emitido por Travidence, LLC · Wilmington, Delaware, EE. UU.

Travidence, LLC mantiene un seguro de responsabilidad profesional tecnológica y ciberriesgo (Tech E&O/Cyber) en vigor.

Apéndice C — Hash de reproducibilidad y metadata de versión

Item	Valor
Versión del paper	1.0
Versión de la metodología canónica	v1.5 (documento interno de TRAVIDENCE, no publicado)
Versión del Score de Robustez	v1.1.0
Calibration vintage	v1.1.0 / 2026-Q3
Versión del schema de resultados de validación	v1.3
Versión del repositorio fuente	Ver CITATION.cff (campo version) en la raíz del repositorio
Reproducibility hash Caso 01 (esquema 2)	sha256:e8a6e8e8fd238f82fa65da2c9eb2d20f1e8a9d8635b489e8ad9c752966d1de8c
Reproducibility hash Caso 02 (esquema 2)	sha256:6a8e058c61179010c5be84e68f846e758acd4de1453d37b9b1539b317b5babc6
Reproducibility hash Caso 03 (esquema 2)	sha256:aec6ad462f11f53609e946864136edfb953551bcca892906aa713243e1d2db5d
Reproducibility hash Caso 04 — repositorio (esquema 2)	sha256:dbcfe8d71e2b03d508a926047dbe061b6f2453f7835277d06757137060996ccb
Versiones de dependencias clave	Python 3.11+; numpy>=1.26; scipy>=1.13; pandas>=2.1; arch>=6.3 (Stationary Bootstrap; Politis y Romano, 1994); statsmodels>=0.14. Lock detallado en uv.lock del repo. (Referencia adicional: verificación cruzada puntual contra mlfinlab>=2.0 y empyrical-reloaded==0.5.12 el 2026-05-20 — no son dependencias pineadas en uv.lock.)

Los cuatro hash de reproducibilidad provienen del campo `reproducibility_hash` del Score de Robustez output en `case_studies/*/results/metrics.json` y se verifican como invariantes del run con los mismos inputs sobre los mismos commits de validadores. Cambios en cualquiera de los validadores producen hash diferente; el campo se usa para trazabilidad procesal post-emisión del dictamen.

Nota de revisión (2026-08): la tabla anterior refleja los hashes de esquema 2 vigentes (calibración v1.1.0, esquema de hash 2, spec §9.4). Versiones anteriores de este documento citaban los hashes de esquema 1 (motor $\leq$ v0.8.x), que siguen siendo anclas de verificación válidas para los certificados emitidos bajo ese esquema.

Fin del documento.