FACULTAD DE ESTUDIOS ESTADÍSTICOS

GRADO EN ESTADISTICA APLICADA

Curso 2025/2026

Trabajo de Fin de Grado

TÍTULO: Mercados de predicción y decisiones del Comité de Mercados Abiertos de la Reserva Federal de los Estados Unidos. Un estudio estadístico.

***Alumno*: Mario Gertrudix Moreno**

***Tutor*: Enrique González Arangüena**

Junio de 2026

UNIVERSIDAD COMPLUTENSE
MADRID

Índice

Resumen **3**
Abstract **3**
1. Introducción **4**
1.1. Los mercados de predicción como herramienta estadística 4
1.2. El caso de aplicación: decisiones de política monetaria del FOMC 5
1.3. Riesgos y mecanismos de control en los mercados de predicción 7
1.4. Objetivos 8
2. Estado del arte **10**
2.1. Fundamentos clásicos y primera generación 10
2.2. La nueva generación de plataformas: *Polymarket* y *Kalshi* 11
2.3. Mercados de predicción y expectativas macroeconómicas 12
3. Datos **14**
3.1. Descripción general de las fuentes 14
3.2. Período muestral: las trece reuniones del FOMC 15
3.3. Datos de *Polymarket* 16
3.4. Datos de *Kalshi* 17
3.5. Datos de *CME FedWatch* 17
3.6. Datos de *Fed Funds Futures* (ZQ) 18
3.7. Análisis descriptivo preliminar 18
4. Modelos y metodología **21**
4.1. Estructura de los contratos de evento 21
4.2. Metodología de *CME FedWatch* 23
4.3. Tratamiento de los datos y alineamiento temporal 24
4.4. Probabilidades *risk-neutral* frente a probabilidades físicas 25
4.5. Los dos análisis complementarios 26
4.6. Inferencia estadística 27
5. Resultados **29**
5.1. Precisión predictiva (H1) 29
5.2. Plataforma regulada vs descentralizada (H2) 31
5.3. Valor añadido del árbol binomial (H3) 33
5.4. Velocidad de respuesta y descubrimiento de precios (H4) 36
6. Conclusiones y futuras líneas de investigación **39**
Bibliografía **43**

Resumen

Este trabajo estudia el funcionamiento de los mercados de predicción como mecanismo de agregación de expectativas, utilizando las decisiones del *Federal Open Market Committee* (FOMC) como caso de aplicación. Se comparan cuatro fuentes de expectativas sobre las trece reuniones celebradas entre junio de 2024 y diciembre de 2025: *Polymarket*, mercado descentralizado con liquidación en criptoactivos; *Kalshi*, mercado de predicción regulado por la *Commodity Futures Trading Commission*; *CME FedWatch*, herramienta institucional que deriva probabilidades de los *Fed Funds Futures* mediante un árbol binomial; y los precios implícitos directos de dichos futuros. La comparación se realiza con el objetivo de aportar evidencia sobre la precisión, la calibración y la velocidad de respuesta de los mercados de predicción frente a los instrumentos financieros tradicionales. Los resultados muestran que los mercados de predicción ofrecen una precisión predictiva comparable a la de las fuentes institucionales, con Polymarket equiparándose al nivel de estas.

Palabras clave: mercados de predicción, FOMC, expectativas de política monetaria, *Polymarket*, *Kalshi*, *CME FedWatch*, Fed Funds Futures.

Abstract

This study examines the functioning of prediction markets as a mechanism for aggregating expectations, using the decisions of the *Federal Open Market Committee* (FOMC) as a case of application. Four sources of expectations are compared across the thirteen meetings held between June 2024 and December 2025: *Polymarket*, a decentralised market with cryptoasset settlement; *Kalshi*, a prediction market regulated by the Commodity Futures Trading Commission; *CME FedWatch*, an institutional tool that derives probabilities from *Fed Funds Futures* through a binomial tree; and the direct implied prices of these futures. The comparison aims to provide evidence on the accuracy, calibration and responsiveness of prediction markets relative to traditional financial instruments. The results show that prediction markets achieve predictive accuracy comparable to institutional sources, with Polymarket performing on par with them.

Keywords: prediction markets, FOMC, monetary policy expectations, *Polymarket*, *Kalshi*, *CME FedWatch*, Fed Funds Futures.

1. Introducción

Los mercados de predicción operan como un sistema de agregación de expectativas, lo que les hace especialmente interesantes como objeto de estudio. Su funcionamiento puede revelar información útil para estudiar y comparar su utilidad, en términos de precisión, calibración y velocidad de respuesta, con los instrumentos financieros convencionales. Para abordar su análisis, este trabajo realiza un estudio de caso a partir de las decisiones del *Federal Open Market Committee* (FOMC) durante el periodo de política monetaria comprendido entre junio de 2024 y diciembre de 2025.

En esta sección se sientan las bases conceptuales del estudio. La Subsección 1.1 presenta los mercados de predicción y su interés desde el punto de vista de la Estadística aplicada. La Subsección 1.2 describe el caso del FOMC y su idoneidad como dominio de prueba. La Subsección 1.3 examina la robustez de los mercados de predicción y los debates recientes sobre su integridad, identificando los factores que reducen significativamente estos riesgos en el caso particular de los mercados sobre decisiones del FOMC. Finalmente, la Subsección 1.4 enuncia los objetivos del trabajo.

1.1. Los mercados de predicción como herramienta estadística

Un mercado de predicción es un mecanismo de negociación en el que los participantes compran y venden contratos cuyo pago final depende de la ocurrencia de un evento futuro. El precio al que se cruzan las operaciones se interpreta como la probabilidad que el conjunto de participantes asigna de forma agregada a que dicho evento ocurra. La lógica que sustenta esta interpretación se encuadra en la tradición de la agregación de información dispersa, en la que cada participante trae consigo una porción de conocimiento privado, y el precio resultante del proceso de intercambio comprime esa información en una única señal pública observable en tiempo real.

Los fundamentos de este planteamiento están en trabajos clásicos como el de Hayek (1945) sobre el sistema de precios como mecanismo de coordinación, y tiene antecedentes más allá de las ciencias económicas, ya que Lindauer (1955) documentó cómo los enjambres de abejas melíferas eligen colectivamente la ubicación de un nuevo nido mediante un proceso descentralizado de exploración y agregación cuantitativa, sin liderazgo individual ni jerarquía. Esta idea la concretó Surowiecki (2004) bajo el concepto de *wisdom of crowds*, que la formaliza como principio aplicable a una variedad de sistemas de decisión colectiva, desde los mencionados enjambres hasta los mercados financieros. Desde un punto de vista estadístico, los mercados de predicción constituyen un estimador particular, puesto que se trata de un estimador basado en los incentivos monetarios de agentes heterogéneos que apuestan dinero real sobre los resultados. A diferencia de las encuestas, donde las respuestas no tienen consecuencias económicas directas, los precios de los mercados de predicción incorporan el esfuerzo de los participantes que tienen incentivos para procesar información y actuar sobre ella. Esta propiedad los convierte en un objeto de estudio especialmente interesante para la estadística aplicada al análisis de expectativas, dado que

combina el principio de agregación propio de los métodos de panel con la dinámica de actualización continua que es propia de los precios financieros.

El interés académico por los mercados de predicción aparece ya en el trabajo de Wolfers y Zitzewitz (2004) quienes sistematizaron las primeras décadas de experiencia con plataformas como el *Iowa Electronic Markets*, o Arrow et al. (2008) que defendieron su potencial como herramienta de política pública. Sin embargo, durante mucho tiempo el alcance de los mercados de predicción ha estado limitado por restricciones regulatorias, bajos volúmenes de negociación y escasa variedad de contratos. La situación ha cambiado notablemente en los últimos años con la aparición de nuevas plataformas que operan con volúmenes de dinero sin precedentes y con una cobertura temática claramente más amplia.

Para el análisis planteado en este trabajo, se estudian fundamentalmente dos de estas plataformas, *Polymarket* y *Kalshi*. Su elección se debe a que representan dos modelos institucionales contrapuestos que resultan especialmente interesantes desde el punto de vista comparativo. Por una parte, *Polymarket* es un ejemplo de mercado descentralizado, sin regulación específica en Estados Unidos y que opera sobre la cadena de bloques *Polygon* y liquida sus contratos con la moneda estable USDC. Por otro, *Kalshi* es un *Designated Contract Market* supervisado por la *Commodity Futures Trading Commission* (CFTC), y tiene, por tanto, un modelo regulatorio similar al que poseen instituciones tradicionales como la *Chicago Mercantile Exchange*. El hecho de que pueda aplicarse este análisis a los mismos eventos macroeconómicos permite estudiar empíricamente cómo la estructura regulatoria, la composición de los participantes y la tecnología subyacente afectan a la calidad de las probabilidades agregadas.

1.2. El caso de aplicación: decisiones de política monetaria del FOMC

Las decisiones de la Reserva Federal sobre política monetaria poseen una gran influencia en los mercados financieros globales. Las modificaciones en el rango objetivo del *Federal Funds Rate* afectan directamente a los costes de financiación de hogares y empresas, a las valoraciones de activos financieros y a los flujos internacionales de capital. Por ello, la capacidad de anticipar con precisión las decisiones del *Federal Open Market Committee* (FOMC) resulta de interés tanto para inversores institucionales y gestores de riesgos como para responsables de política económica y académicos. En este trabajo, las decisiones del FOMC no son el objeto último de estudio, sino que sirven como marco empírico en el que se observa el funcionamiento de los mercados de predicción; un dominio en el que estos mercados conviven con fuentes institucionales ampliamente establecidas, lo que facilita poder realizar una comparación sistemática.

Tradicionalmente, las expectativas del mercado sobre las decisiones del FOMC se han inferido mediante instrumentos financieros derivados, siendo los *Fed Funds Futures* el instrumento de referencia. Los contratos de *Fed Funds Futures* son derivados financieros negociados en la Bolsa de Chicago (CME) cuyo valor de liquidación se determina a partir del promedio mensual del *Effective Federal Funds Rate* (EFFR). Cada contrato refleja la expectativa del mercado sobre el nivel medio del tipo de interés de los fondos federales

durante un mes determinado. Al comparar los precios implícitos de estos contratos con el tipo vigente, es posible extraer las expectativas del mercado sobre los cambios futuros en la política monetaria. Sobre esta base, CME Group desarrolló la herramienta *FedWatch*, que transforma los precios de los futuros en probabilidades de cada resultado posible del FOMC mediante un árbol binomial.

No obstante, estas herramientas institucionales tienen limitaciones que han sido señaladas en la literatura científica reciente (Diercks, Katz y Wright, 2026). Los *Fed Funds Futures* proporcionan estimaciones del tipo medio esperado durante un mes completo, lo que, combinado con el supuesto de árbol binomial de *FedWatch*, reduce la distribución de la reunión inmediata a dos rangos de tipos contiguos. Es decir, a medida que se acerca una reunión se produce un proceso de simplificación a dos rangos. Y esto puede subestimar la incertidumbre real, especialmente en períodos de alta volatilidad o cuando el mercado contempla múltiples trayectorias de tipos simultáneamente. Además, los precios de los futuros pueden incorporar primas de riesgo que distorsionan su lectura como predicciones puras de tipos futuros, y su negociación está condicionada a horarios bursátiles determinados, lo que limita su capacidad de respuesta ante eventos con amplia repercusión fuera de sesión.

En este contexto, la popularización y adopción de los mercados de predicción proponen nuevas herramientas para inferir expectativas sobre indicadores macroeconómicos. Estos mercados, basados en el principio de la *wisdom of crowds*[1] (Surowiecki, 2004), aglutinan información dispersa entre participantes heterogéneos mediante incentivos financieros. A diferencia de los derivados tradicionales, los mercados de predicción permiten a los participantes expresar directamente sus expectativas sobre resultados discretos, generando distribuciones de probabilidad explícitas sin necesidad de supuestos intermedios. Que operen de manera continua les dota, además, de una capacidad potencial de reacción ante nueva información que los mercados de futuros con horarios restringidos y estrictas regulaciones no pueden igualar.

Otro aspecto relevante, es que el ecosistema actual de los mercados de predicción a nivel global muestra una dualidad estructural relevante. Por un lado, encontramos plataformas reguladas como *Kalshi*, la primera bolsa financiera regulada de Estados Unidos dedicada al intercambio de contratos de eventos (*event contracts*), reconocida como Mercado de Contratos Designado (*Designated Contract Market*) por la Comisión de Negociación de Futuros de Productos Básicos (*Commodity Futures Trading Commission*). Esta validación regulatoria sitúa a *Kalshi* al mismo nivel de instituciones financieras consolidadas como la Bolsa de Chicago (CME), pero con una base de participantes compuesta predominantemente por actores individuales, a diferencia del ecosistema estrictamente institucional de los mercados de futuros tradicionales (Diercks, Katz y Wright, 2026). Por otro lado, plataformas como *Polymarket* representan el modelo de finanzas descentralizadas (*Decentralized Finance*). A diferencia de *Kalshi*, *Polymarket* opera sobre *blockchain* y emplea criptoactivos (como la moneda estable USDC) para la liquidación de sus mercados. Aunque no cuenta con el mismo marco regulatorio en territorio estadounidense, *Polymarket* se ha consolidado como

[1] Traducción: Sabiduría de los grupos

uno de los mercados de predicción con mayor volumen de negociación a nivel global, con más de 124 millones de operaciones procesadas entre 2022 y 2025 (Reichenbach y Walther, 2025).

Estas plataformas han obtenido un notable crecimiento en los últimos años. Impulsado en gran medida por las elecciones presidenciales de Estados Unidos, *Polymarket* experimentó durante 2024 un crecimiento exponencial, alcanzando volúmenes de negociación mensuales de varios miles de millones de dólares y atrayendo a cientos de miles de *traders* activos. *Kalshi*, por su parte, ha registrado un crecimiento sostenido desde su lanzamiento en 2021, ampliando progresivamente su oferta de contratos y su base de usuarios, especialmente tras la integración con plataformas de inversión minorista como *Robinhood*.

Este crecimiento no ha sido homogéneo entre categorías de mercados. En *Polymarket*, las categorías de política y deportes han concentrado históricamente la mayor parte del volumen, seguidas por criptomonedas, economía y cultura. En *Kalshi*, las categorías de mayor actividad incluyen política, mercados financieros, indicadores macroeconómicos y meteorología. Los contratos relativos a la Reserva Federal han cobrado un impulso creciente dentro de la oferta macroeconómica de los mercados de predicción. *Kalshi* ha visto cómo sus mercados sobre el *federal funds rate* escalaban hasta situarse entre los más activos de su historia, superando con frecuencia el millón de contratos por evento (Diercks, Katz y Wright, 2026). Al mismo tiempo, *Polymarket* ha logrado capturar una cuota de interés creciente, consolidando sus mercados de la Fed a través de un aumento constante en la participación y el volumen de transacciones.

El creciente interés institucional en estos mercados pone de manifiesto su relevancia. *Intercontinental Exchange* (ICE) anunció en octubre de 2025 una alianza estratégica con *Polymarket* que incluye un compromiso de inversión de hasta 2.000 millones de dólares y planes para distribuir los datos de *Polymarket* entre sus clientes institucionales. *Robinhood,* por su parte, permitió a sus usuarios operar contratos de eventos a partir de octubre de 2024, y, a través de su alianza con *Kalshi*, los usuarios negociaron más de 2.000 millones de contratos hasta agosto de 2025. Estos desarrollos revelan hasta qué punto se está produciendo una convergencia progresiva y real entre los mercados de predicción y la infraestructura financiera tradicional.

1.3. Riesgos y mecanismos de control en los mercados de predicción

Los mercados de predicción, pese a sus ventajas como mecanismo de agregación de expectativas, presentan también vulnerabilidades que se han discutido en los ámbitos académico y mediático. Su creciente adopción se ha visto acompañada de constantes cuestionamientos sobre su robustez frente a posibles intentos de manipulación, derivados en buena medida de la propia estructura de las plataformas. Estos debates pueden agruparse en distintas tipologías.

La primera, y más sensible, es el uso de información privilegiada. En enero de 2026, una cuenta de reciente creación en *Polymarket* apostó alrededor de 33.000 dólares a que Nicolás Maduro saldría del poder antes del 31 de enero, obteniendo más de 400.000 dólares de

beneficio horas antes del anuncio público de la captura del mandatario venezolano (NPR, 2026). El *trader* resultó ser un sargento de las Fuerzas Especiales del Ejército de Estados Unidos involucrado en la planificación de la operación que, posteriormente, fue acusado por el Departamento de Justicia y la *Commodity Futures Trading Commission* por uso indebido de información clasificada (Department of Justice, 2026). En la última hora del mandato del presidente Biden, otra cuenta anónima ganó cerca de 300.000 dólares acertando los cuatro indultos presidenciales específicos que se anunciarían antes de la transición de poder (Fortune, 2026). Casos análogos se documentaron en los mercados sobre el alto el fuego entre Estados Unidos e Irán durante abril de 2026, en los que cuentas nuevas concentraron apuestas con una sincronización perfecta sobre fechas concretas (PBS News, 2026).

Un segundo tipo afecta a la fiabilidad de las fuentes de resolución de los contratos. Casos como la presunta manipulación de los sensores de temperatura del aeropuerto de París para resolver favorablemente apuestas climáticas en *Polymarket* muestran que incluso fuentes de datos consideradas objetivas pueden ser objeto de interferencia (CBC News, 2026).

Estos episodios no invalidan el valor predictivo de los mercados de predicción como categoría general, pero sí introducen matices sobre su uso como referencia de expectativas. Esto ha llevado al Senado de los Estados Unidos, en abril de 2026, a prohibir la operativa en estos mercados a sus miembros y personal (CNBC, 2026). En el caso específico de los mercados sobre decisiones del FOMC, sin embargo, varios factores reducen significativamente este conjunto de riesgos, como son la liquidez elevada de los contratos; la naturaleza pública, verificable y ampliamente anticipada del resultado, que dificulta cualquier manipulación sostenida; la inexistencia de información privilegiada en sentido estricto, dado que las decisiones del FOMC se adoptan en deliberaciones colegiadas; y la existencia, en ambas plataformas, de reglas explícitas de resolución que delimitan inequívocamente el resultado del contrato.

1.4. Objetivos

Este trabajo tiene como objetivo principal analizar empíricamente el funcionamiento de los mercados de predicción como mecanismo de agregación de expectativas, utilizando, para ello, las decisiones del FOMC como caso de estudio.

Para ello se contrastan cuatro aproximaciones diferenciadas:

1. ***CME FedWatch***. Probabilidades derivadas del árbol binomial construido a partir de Fed Funds Futures.
2. ***Fed Funds Futures*** (precios implícitos). Expectativas de tipos extraídas directamente de los precios de los contratos de futuros, sin aplicar la transformación binomial de *FedWatch*.
3. ***Polymarket*.** Mercado de predicción descentralizado operando sobre infraestructura blockchain con liquidación en USDC.
4. ***Kalshi*.** Mercado de predicción regulado por la CFTC operando en dólares estadounidenses.

Esta comparación permite evaluar la precisión predictiva de los mercados de predicción frente a los derivados institucionales, las diferencias entre plataformas reguladas y descentralizadas, y el valor añadido de la transformación binomial de *FedWatch* frente a los precios directos de los futuros.

El análisis se centra en identificar si la naturaleza de los mercados de predicción, con participación global, operativa continua y distribuciones de probabilidad explícitas, permite capturar cambios en las expectativas macroeconómicas con mayor agilidad y precisión que los mercados de futuros institucionales. Con este fin, se contrastan las expectativas de cambio en los tipos de interés reflejadas en las cuatro fuentes a lo largo de las trece reuniones del FOMC celebradas entre junio de 2024 y diciembre de 2025.

Para articular el análisis empírico, este trabajo plantea cuatro hipótesis específicas, que se contrastarán en la Sección 5 mediante las técnicas detalladas en la Sección 4.

H1 (Precisión predictiva). Los mercados de predicción (Polymarket y Kalshi) presentan una precisión predictiva al menos comparable a la de los derivados institucionales (CME FedWatch y los precios implícitos de los Fed Funds Futures) en métricas de Brier Score, RPS, MAE y RMSE.

H2 (Estructura institucional). Las dos plataformas de mercados de predicción principales producen predicciones de precisión y calibración estadísticamente equivalentes: Polymarket, descentralizada y operada sobre blockchain, y Kalshi, regulada por la CFTC.

H3 (Valor del árbol binomial). La transformación binomial aplicada por CME FedWatch sobre los precios de los Fed Funds Futures aporta valor predictivo sustancial respecto a la extracción directa de expectativas a partir del precio de un único contrato.

H4 (Descubrimiento de precios). Los mercados de predicción, gracias a su operativa continua y a su base global de participantes, lideran el descubrimiento de precios sobre los derivados institucionales en la dinámica diaria de expectativas sobre decisiones del FOMC.

El resto del trabajo se estructura de la siguiente manera. La Sección 2 presenta el estado del arte, revisando la literatura relevante sobre mercados de predicción, su precisión y sus sesgos. La Sección 3 describe los datos utilizados y presenta un primer análisis descriptivo. La Sección 4 detalla los modelos y la metodología, incluyendo la estructura de los contratos de evento, la metodología de *CME FedWatch* y las decisiones metodológicas específicas de cada fuente. La Sección 5 presenta los resultados y la Sección 6 recoge las conclusiones.

2. Estado del arte

En esta sección se revisa la literatura académica sobre mercados de predicción que sirve de marco al presente trabajo. La Subsección 2.1 presenta los fundamentos clásicos y los primeros estudios sobre estas plataformas, desarrollados principalmente entre los años 1990 y 2010. La Subsección 2.2 describe la nueva generación de plataformas a gran escala, con especial atención a *Polymarket* y *Kalshi*, que ocupan el centro de este estudio. Finalmente, la Subsección 2.3 revisa la literatura más reciente sobre el uso de mercados de predicción en el análisis de expectativas macroeconómicas, donde se inscribe la presente investigación.

2.1. Fundamentos clásicos y primera generación

Los mercados de predicción han suscitado un interés creciente en la literatura económica debido a su capacidad para agregar información dispersa entre participantes individuales mediante incentivos financieros, generando estimaciones que, en condiciones adecuadas, pueden aproximarse a las mejores predicciones disponibles sobre eventos futuros (Arrow et al., 2008). La síntesis más influyente de la primera ola de estudios fue documentada por Wolfers y Zitzewitz (2004), construida en torno al Iowa Electronic Markets y a plataformas posteriores como Intrade y TradeSports. Sus conclusiones principales fueron que los precios de los mercados de predicción agregaban información con precisión razonable y que, para eventos de suficiente interés público, sus predicciones competían favorablemente con encuestas y modelos estadísticos.

Arrow et al. (2008), en una carta colectiva publicada en *Science* y firmada por un amplio grupo de economistas y politólogos, defendieron que los mercados de predicción podían aportar valor como herramienta de política pública si se superaban las barreras regulatorias que entonces limitaban su desarrollo. Los autores insistieron en dos aspectos fundamentales de estos mercados de predicción: su capacidad para incorporar rápidamente información nueva, y su funcionamiento como fuente de incentivos para la búsqueda activa de dicha información. Estas dos propiedades justifican su estudio como alternativa a los métodos tradicionales de agregación de expectativas.

Atanasov et al. (2017) desarrollaron el primer test experimental a gran escala comparando mercados de predicción con otros métodos de *crowdsourcing*. Sus conclusiones demostraron que los precios de los mercados de predicción constituyen estimaciones generalmente precisas de las probabilidades de eventos futuros, confirmando su eficacia como mecanismo de agregación de información. Su estudio comparó mercados de predicción con encuestas estructuradas a expertos, encontrando que el desempeño predictivo es similar en muchos casos y superior en eventos con información dispersa entre los participantes. En esta línea de trabajo Tetlock y Gardner (2015) realizaron un programa más amplio sobre habilidad predictiva individual que documentó la existencia de *superforecasters*, entendidos como individuos cuyas predicciones probabilísticas sobre eventos geopolíticos y económicos superan sistemáticamente a las estimaciones agregadas por simple promedio. Esto sugiere que la heterogeneidad en la capacidad de procesar información es una propiedad relevante para entender el comportamiento de estos mecanismos de agregación. No obstante, hasta

hace pocos años, estas plataformas permanecieron como un fenómeno de nicho, siendo incapaces de materializar su potencial teórico debido a limitaciones técnicas, baja liquidez y restricciones regulatorias.

2.2. La nueva generación de plataformas: *Polymarket* y *Kalshi*

Esta situación ha cambiado sustancialmente con la aparición de una nueva generación de mercados de predicción que operan a escala global y con volúmenes de negociación sin precedentes como *Kalshi* y *Polymarket* (Reichenbach y Walther, 2025). Ambas plataformas han multiplicado en varios órdenes de magnitud los volúmenes tradicionales del sector, lo que abre la posibilidad a realizar estudios empíricos con una potencia estadística suficiente y con una cobertura temática muy amplia, desde resultados electorales hasta indicadores macroeconómicos y deportivos.

Reichenbach y Walther (2025) han analizado la precisión de los mercados descentralizados como fuente de expectativas a través de 124 millones de operaciones realizadas por cerca de un millón de traders en *Polymarket*. Los autores concluyen que los precios de la plataforma siguen razonablemente las probabilidades realizadas de los eventos, lo que respalda su utilidad como herramienta de inferencia de expectativas. Además, no encuentran evidencia de un *longshot bias* general, un sesgo frecuentemente documentado en mercados de predicción anteriores que consiste en la sobrevaloración de resultados poco probables. El estudio también pone de relieve la existencia de *traders* con habilidad predictiva persistente en el tiempo, cuyas ganancias por operación superan las que cabría esperar bajo un comportamiento puramente aleatorio, lo que sugiere que la plataforma atrae a participantes capaces de procesar información de forma eficiente.

Más allá de la precisión de una plataforma individual, la comparación entre distintos mercados de predicción resulta esencial para comprender cómo la estructura, la liquidez y la composición de participantes influyen en la calidad de las señales de precio. En esta línea, Ng, Peng, Tao y Zhou. (2026) aportan la primera evidencia empírica sobre dinámicas de descubrimiento de precios entre plataformas de predicción modernas, analizando contratos comunes negociados simultáneamente en *Polymarket*, *Kalshi*, PredictIt y Robinhood durante las elecciones presidenciales de 2024. Sus resultados revelan disparidades significativas de precios entre plataformas para contratos idénticos, y establecen que *Polymarket* lidera el proceso de descubrimiento de precios frente a *Kalshi*, particularmente cuando la liquidez y la actividad de negociación son elevadas. Asimismo, documentan que el flujo de órdenes neto procedente de operaciones de gran tamaño predice significativamente los rendimientos subsiguientes, y que la plataforma con mayor flujo direccional tiende a liderar la formación de precios. Estos hallazgos son directamente relevantes para el presente estudio, ya que sugieren que las diferencias en liquidez y estructura de mercado entre *Polymarket* y *Kalshi* podrían traducirse en diferencias observables en la precisión y la velocidad de adopción de información y la consecuente actualización de sus probabilidades implícitas sobre las decisiones del FOMC.

2.3. Mercados de predicción y expectativas macroeconómicas

En el ámbito específico de las expectativas de política monetaria, Diercks, Katz y Wright (2026) han realizado la primera evaluación sistemática de los mercados de predicción de *Kalshi* como fuente de expectativas macroeconómicas. Utilizando datos intradiarios de contratos sobre el *Federal Funds Rate*, CPI, desempleo y PIB, los autores demuestran que las distribuciones de probabilidad implícitas en *Kalshi* están bien calibradas y se actualizan de forma continua, ofreciendo predicciones de precisión comparable a la de los *forecasters* profesionales de la *Survey of Market Expectations* del Federal Reserve Bank de Nueva York. De hecho, la moda y la mediana de la distribución de *Kalshi* alcanzaron un error absoluto medio de cero el día previo a cada reunión del FOMC desde 2022, lo que supone una mejora estadísticamente significativa respecto a las predicciones derivadas de los *Fed Funds Futures*. Concluyen, así, que *Kalshi* proporciona un *benchmark* de alta frecuencia, continuamente actualizado y distribucionalmente rico, de valor tanto para investigadores como para responsables de política monetaria.

Los autores atribuyen la superioridad de *Kalshi* a dos factores. Uno, la capacidad de reflejar distribuciones multimodales cuando el mercado percibe múltiples escenarios simultáneamente, algo que el árbol binomial de *CME FedWatch* no puede capturar por construcción. Y dos, la actualización continua de los precios en respuesta a nueva información, producida incluso fuera del horario de negociación de los futuros. Ambas propiedades resultan especialmente útiles en períodos de incertidumbre elevada sobre la orientación de la política monetaria, como el que se dio en varias reuniones del período analizado en este trabajo.

Bonini, Huang y Simaan (2025) realizan la primera evaluación sistemática del *CME FedWatch* como herramienta predictiva sobre las decisiones del FOMC. Los autores detallan cómo la metodología ensemble de *FedWatch*, basada en la combinación de múltiples contratos de *Fed Funds Futures* y un árbol binomial sobre la *Effective Federal Funds Rate* implícita, alcanza un 88 % de *accuracy* en la predicción del *outcome* a 30 días vista de la reunión, frente a un 75 % del método de extracción directa a partir de un único contrato. La diferencia se atribuye a dos limitaciones del método *baseline*. Por una parte, su dependencia de un único contrato sometido al ruido idiosincrático del precio individual, y, por otra, la necesidad de una elección ad hoc del valor de la tasa de partida utilizada en la inferencia inversa, que introduce un error de estimación sistemático. Aunque ambas fuentes comparten activo subyacente, la literatura reciente sugiere que sus propiedades predictivas son sustancialmente distintas, lo que motiva el tratamiento diferenciado de *CME FedWatch* y ZQ como objetos de comparación independientes en el presente trabajo.

El presente trabajo se ubica en la intersección de estas líneas de investigación, contribuyendo a la literatura al comparar directamente cuatro fuentes de expectativas sobre decisiones del FOMC (*CME FedWatch*, precios implícitos de Fed Funds Futures, *Polymarket* y *Kalshi*) durante un período de trece reuniones entre junio de 2024 y diciembre de 2025. Este período resulta de especial relevancia al coincidir con un punto de inflexión en la política monetaria de la Reserva Federal, marcado por el inicio de los recortes de tipos tras un ciclo prolongado

de endurecimiento, un contexto en el que la capacidad predictiva de las distintas fuentes se somete a una prueba particularmente exigente. El trabajo incorpora además a *Polymarket*, un mercado de predicción que Diercks, Katz y Wright (2026) excluyen explícitamente de su análisis por razones regulatorias, pero cuya relevancia empírica es innegable dado su volumen y su relevancia institucional creciente.

3. Datos

En esta Sección se describen los datos utilizados en el estudio, sus fuentes de obtención y un primer análisis descriptivo de su comportamiento. La Subsección 3.1 presenta una visión general de las cuatro fuentes empleadas y sus características principales. La Subsección 3.2 detalla el período muestral analizado, correspondiente a las trece reuniones del FOMC celebradas entre junio de 2024 y diciembre de 2025. Las Subsecciones 3.3 a 3.6 describen, de manera detallada, los datos procedentes de cada fuente. Finalmente, la Subsección 3.7 ofrece un análisis descriptivo preliminar que ilustra las propiedades fundamentales de los datos antes de la comparación sistemática realizada en las Secciones siguientes.

3.1. Descripción general de las fuentes

Este trabajo utiliza datos procedentes de cuatro fuentes independientes, cada una con su propio formato, granularidad y régimen de actualización. La Tabla 1 resume sus características principales.

Fuente	Tipo	Acceso	Granularidad	Cobertura
Polymarket	Mercado de predicción descentralizado (blockchain Polygon)	Dataset público de Jon Becker (GitHub, Parquet)	Trade a trade (continuo)	13 reuniones
Kalshi	Mercado de predicción regulado (DCM, CFTC)	Dataset público de Jon Becker (GitHub, Parquet)	Trade a trade (continuo)	13 reuniones
CME FedWatch	Probabilidades derivadas del árbol binomial	API oficial CME Group (OAuth2, endpoint /forecasts)	Snapshot diario (días hábiles)	13 reuniones
***Fed Funds Futures* (ZQ)**	Futuros de tipos a 30 días (CBOT)	API Databento (GLBX.MDP3, schema statistics, stat_type=3)	Settlement diario (días hábiles)	13 reuniones

Tabla 1. Características principales de las fuentes de datos utilizadas.

Las cuatro fuentes comparten un mismo universo observable, las decisiones del FOMC sobre el rango objetivo del *Federal Funds Rate*, pero se diferencian sustancialmente en su naturaleza. *Polymarket* y *Kalshi* son mercados de predicción propiamente dichos, con contratos binarios cuyo precio se interpreta como probabilidad. *CME FedWatch* es un producto derivado de segundo orden. No es un mercado por sí mismo, sino un modelo que transforma los precios de los *Fed Funds Futures* en probabilidades mediante un árbol binomial. Los *Fed Funds Futures*, finalmente, son un instrumento financiero clásico cuyo

precio refleja el valor esperado del EFFR durante un mes natural, no una distribución de probabilidades sobre resultados discretos.

Esta heterogeneidad de fuentes supone que ninguna métrica de comparación es automáticamente válida para las cuatro. Por ese motivo, el análisis se ha estructurado en dos aproximaciones complementarias, descritas en detalle en la Sección 4. La primera hace uso de las distribuciones de probabilidad disponibles en *Polymarket*, *Kalshi* y *CME FedWatch*. La segunda reduce cada fuente a una magnitud escalar común que permite incorporar los *Fed Funds Futures* en condiciones de igualdad, la tasa esperada implícita.

3.2. Período muestral: las trece reuniones del FOMC

El análisis se centra en las trece reuniones del FOMC celebradas entre junio de 2024 y diciembre de 2025. Este período cubre un ciclo completo de política monetaria, desde la pausa prolongada en el rango 5,25-5,50% hasta el cierre del año 2025 tras tres recortes consecutivos que dejaron el rango en 3,50-3,75%. La Tabla 2 resume las decisiones adoptadas en cada reunión y señala las disensiones documentadas en los votos del Comité.

Reunión	Tasa vigente	Decisión	Tasa resultante
12-jun-2024	5,25-5,50%	Sin cambio	5,25-5,50%
31-jul-2024	5,25-5,50%	Sin cambio	5,25-5,50%
18-sep-2024	5,25-5,50%	Recorte 50 pb	4,75-5,00%
07-nov-2024	4,75-5,00%	Recorte 25 pb	4,50-4,75%
18-dic-2024	4,50-4,75%	Recorte 25 pb	4,25-4,50%
29-ene-2025	4,25-4,50%	Sin cambio	4,25-4,50%
19-mar-2025	4,25-4,50%	Sin cambio	4,25-4,50%
07-may-2025	4,25-4,50%	Sin cambio	4,25-4,50%
18-jun-2025	4,25-4,50%	Sin cambio	4,25-4,50%
30-jul-2025	4,25-4,50%	Sin cambio	4,25-4,50%
17-sep-2025	4,25-4,50%	Recorte 25 pb	4,00-4,25%
29-oct-2025	4,00-4,25%	Recorte 25 pb	3,75-4,00%
10-dic-2025	3,75-4,00%	Recorte 25 pb	3,50-3,75%

Tabla 2. Decisiones del FOMC en el período analizado. Fuente: Elaboración propia a partir de los comunicados oficiales del Federal Reserve Board y las actas publicadas del FOMC.

Del total de trece reuniones, ocho se cerraron sin cambios en el tipo de referencia, una incluyó un recorte de 50 puntos básicos (septiembre de 2024, que abrió el ciclo de relajación) y cuatro registraron recortes de 25 puntos básicos. La concentración de decisiones de "sin cambio" entre enero y julio de 2025 refleja la pausa prolongada que caracterizó la primera mitad del segundo ciclo, asociada al contexto político y al debate sobre la independencia de la Reserva Federal tras el inicio de la segunda Administración Trump en enero de 2025.

3.3. Datos de *Polymarket*

Los datos de *Polymarket* proceden directamente de la cadena de bloques Polygon. Cada operación generada por el libro de órdenes de la plataforma (CTF Exchange y NegRisk Adapter) se registra *on-chain* y puede reconstruirse a partir de los logs de los contratos inteligentes correspondientes. Para cada reunión del FOMC, *Polymarket* ofrece un evento multi-contrato con varios sub-mercados binarios mutuamente excluyentes sobre la magnitud del cambio en el tipo: por ejemplo, recorte de 25 puntos básicos, sin cambio, subida de 25 puntos básicos o más.

Para cada sub-mercado se han obtenido todas las operaciones individuales, incluyendo precio, volumen nominal en USDC, marca temporal y dirección del *trade.* A partir de estas operaciones se ha construido un precio diario mediante el VWAP (*Volume-Weighted Average Price*) ponderado por el volumen monetario, lo que aporta robustez frente a operaciones aisladas con precios anómalos. Los días sin actividad se han completado mediante forward-fill, arrastrando el último precio observado, y las probabilidades de los sub-mercados de una misma reunión se han renormalizado por día para garantizar que sumen la unidad.

Para una de las reuniones del período analizado, la del 18 de septiembre de 2024, un error tipográfico en la regla de resolución del sub-mercado 'Cut 50+bps' generó una asimetría respecto a las demás fuentes que compromete la comparabilidad trilateral y tetralateral para ese evento (*Polymarket*, 2024). Por consistencia metodológica, esta reunión queda excluida del análisis comparativo en todas las fuentes, que se restringe por tanto a doce de las trece reuniones del período.

En las reuniones del FOMC de noviembre y diciembre de 2024, *Polymarket* ofreció, además de los sub-mercados principales sobre la magnitud del recorte o subida de tipos, un mercado residual denominado "Other" que agregaba de forma *catch-all* cualquier escenario fuera del rango principal de *outcomes* (recortes o subidas extremas no contempladas en los mercados específicos). Estos sub-mercados han sido excluidos del análisis comparativo por dos motivos. En primer lugar, su volumen es marginal respecto al total negociado en cada evento, por lo que su inclusión no altera sustancialmente las probabilidades implícitas de las categorías principales tras la normalización diaria. En segundo lugar, la comparación con *CME FedWatch* y *Kalshi* requiere un conjunto homogéneo de categorías a lo largo de las trece reuniones analizadas; mantener "Other" únicamente en dos reuniones introduciría una asimetría que dificultaría la interpretación de las series temporales y de las métricas agregadas. Además, *CME FedWatch* no contempla una categoría equivalente, lo que haría inviable el mapeo descrito en la Sección 4.

3.4. Datos de *Kalshi*

Kalshi publica datos agregados de todos los *trades* a través de su API pública. Para cada reunión del FOMC se ha utilizado la serie *Federal Funds Rate Decision*, que estructura los posibles resultados en contratos binarios sobre la magnitud del cambio (recorte o subida), en paralelo a la estructura de *Polymarket*. La plataforma ofrece además una serie complementaria denominada *Federal Funds Rate Target Rate*, que recoge contratos sobre el nivel del rango objetivo resultante en formato de función de distribución acumulada.

Este estudio utiliza exclusivamente la serie *Federal Funds Rate Decision*, por tres razones. Primero, su estructura de magnitudes discretas (*cut/no change/hike*) es directamente comparable con los sub-mercados de *Polymarket* y con el mapeo de bins de *CME FedWatch*, sin requerir transformaciones adicionales. Segundo, concentra el volumen de negociación en un número reducido de categorías mutuamente excluyentes, facilitando el cálculo del VWAP diario. Tercero, evita el solapamiento temporal entre contratos de reuniones distintas que caracteriza a la serie "Target Rate". La serie alternativa queda fuera del alcance de este trabajo, si bien se reconoce como fuente complementaria para futuras extensiones.

El procesamiento de los datos de *Kalshi* sigue un esquema paralelo al de *Polymarket*. Los precios se expresan en dólares entre 0,01 y 0,99 y se interpretan directamente como probabilidad implícita de que el evento ocurra. Para construir la serie diaria se ha calculado el VWAP ponderado por el número de contratos negociados, aplicado *forward-fill* en días sin actividad y renormalizado las probabilidades para que sumen la unidad por reunión y día.

3.5. Datos de *CME FedWatch*

Los datos de *CME FedWatch* se han obtenido a través de la API oficial de CME Group, bajo el entitlement *FedWatch REST API Production Access – Non Real Time*. Para cada día de negociación y cada reunión del FOMC, la API devuelve la distribución de probabilidades implícitas sobre los posibles rangos de tipos resultantes, en pasos de 25 puntos básicos. La descarga cubre las trece reuniones del período analizado, con series temporales que se extienden aproximadamente desde un mes antes de cada reunión hasta el día de la decisión.

CME FedWatch calcula las probabilidades a partir de los precios de liquidación (*settlement prices*) de los *Fed Funds Futures*, ticker ZQ, negociados en el *Chicago Board of Trade* (CBOT), perteneciente a CME Group. Estos contratos operan en los días hábiles de mercado estadounidense y no cotizan los fines de semana ni en los festivos del calendario oficial. Como consecuencia, las series temporales de *CME FedWatch* presentan discontinuidades sistemáticas en sábados, domingos y festivos bursátiles; en esos días, la API devuelve respuestas vacías para el snapshot correspondiente. *Polymarket* y *Kalshi*, en cambio, son plataformas de negociación continua sin pausas de fin de semana ni festivos: *Polymarket* opera íntegramente sobre infraestructura blockchain, por lo que sus mercados están disponibles de forma ininterrumpida; *Kalshi*, aunque es un *Designated Contract Market* regulado, ofrece también negociación continua en sus mercados sobre decisiones del FOMC.

3.6. Datos de *Fed Funds Futures* (ZQ)

Los precios de los *Fed Funds Futures* se han obtenido a través de Databento, proveedor autorizado de datos de CME Globex, con acceso al dataset GLBX.MDP3 y al ticker ZQ.FUT. Se utilizan los *settlement prices* oficiales publicados diariamente por CBOT para el contrato correspondiente al mes siguiente a cada reunión del FOMC, siguiendo la convención estándar en la literatura sobre expectativas de política monetaria. El *settlement price* es un precio oficial calculado por el *exchange* al cierre de cada sesión, más robusto que el último precio negociado porque integra la actividad de los últimos minutos de mercado y maneja adecuadamente los días de baja liquidez.

A diferencia de las tres fuentes anteriores, los *Fed Funds Futures* no proporcionan una distribución sobre resultados discretos sino un único precio por día y contrato. La conversión de este precio en tasa esperada es directa: la tasa media esperada para el mes del contrato es igual a 100 menos el precio del futuro. Para comparar con las distribuciones de las otras tres fuentes se utiliza en este trabajo la magnitud común de "tasa esperada implícita", calculada como promedio ponderado de los centros de cada categoría por su probabilidad en el caso de *Polymarket*, *Kalshi* y *CME FedWatch*, y directamente como 100 menos el precio del futuro en el caso de ZQ.

3.7. Análisis descriptivo preliminar

Antes de abordar la comparación sistemática, conviene ilustrar algunas propiedades generales de los datos. El volumen agregado de negociación en los mercados de predicción sobre decisiones del FOMC ha crecido de forma notable durante el período analizado. *Kalshi* ha visto cómo sus mercados sobre el *Federal Funds Rate* escalaban hasta situarse entre los más activos de su historia, superando con frecuencia el millón de contratos por evento en las reuniones más recientes (Diercks, Katz y Wright, 2026). *Polymarket*, por su parte, ha consolidado progresivamente su oferta de mercados sobre la Fed y ha registrado volúmenes diarios que se acercan a los de sus mercados tradicionalmente más líquidos (Reichenbach y Walther, 2025). Esta evolución conjunta, representada en la Figura 1, muestra cómo el período analizado corresponde con una fase de madurez creciente de ambas plataformas.

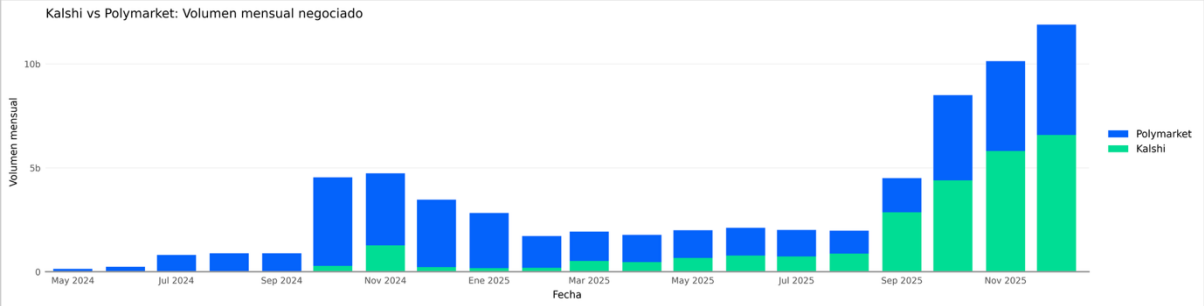


Figura 1. Evolución del volumen total de negociación mensual en Polymarket y Kalshi (junio 2024 – diciembre 2025). Fuente: Elaboración propia a partir de Dune Analytics.

En cuanto a los mercados específicos sobre la Fed, la Figura 2 y la Figura 3 representan la evolución del volumen de negociación en los contratos sobre decisiones del FOMC en *Kalshi* y *Polymarket* respectivamente, a lo largo del período analizado. Un patrón que puede reconocerse es la concentración del volumen en los días previos a cada reunión, con picos de actividad que en algunos casos superan por un orden de magnitud al volumen medio de días alejados de la fecha de decisión.

Este comportamiento confirma empíricamente una de las hipótesis centrales del estudio, y es que los mercados de predicción sobre la Fed son especialmente informativos en las ventanas temporales próximas a las reuniones, precisamente cuando la demanda de señales de expectativas es mayor.

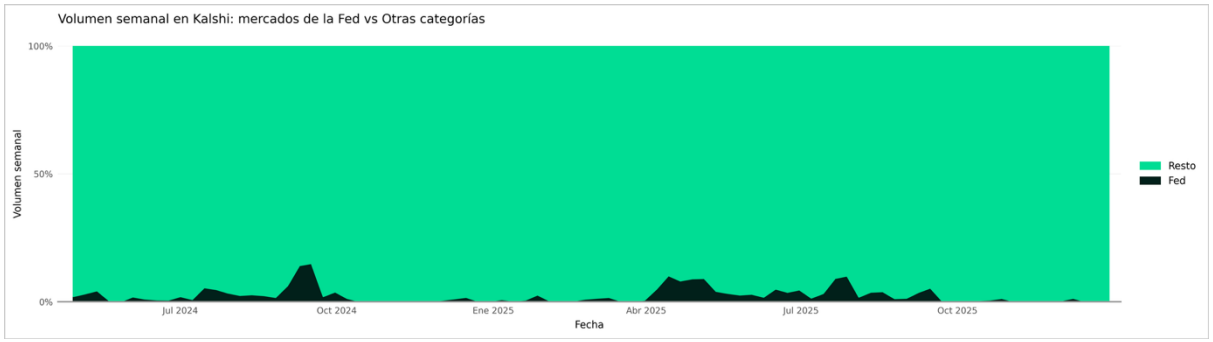


Figura 2. Evolución del volumen porcentual de negociación semanal en mercados sobre decisiones del FOMC en Kalshi (junio 2024 – diciembre 2025). Fuente: Elaboración propia a partir de Dune Analytics.

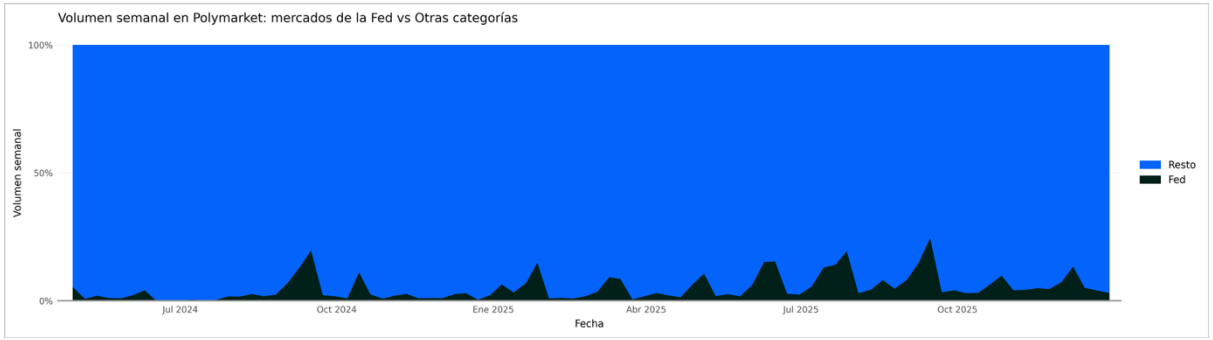


Figura 3. Evolución del volumen porcentual de negociación semanal en mercados sobre decisiones del FOMC en Polymarket (junio 2024 – diciembre 2025). Fuente: Elaboración propia a partir de Dune Analytics.

Por último, la evolución del tipo efectivo *Federal Funds Rate* durante el período analizado se representa en la Figura 4, con las fechas de las reuniones del FOMC indicadas como líneas verticales. Concretamente, se muestra el *Effective Federal Funds Rate* (EFFR), el tipo medio ponderado por volumen al que las instituciones depositarias se prestan fondos *overnight* no garantizados en el mercado interbancario estadounidense, calculado y publicado diariamente por el *Federal Reserve Bank of New York*.

Este contexto macroeconómico resulta útil para interpretar posteriormente los resultados del análisis comparativo. Las reuniones celebradas en momentos de transición de régimen monetario, como la de diciembre de 2024 (último recorte antes de la pausa prolongada de 2025), o las de octubre y diciembre de 2025 (donde se reanudó el ciclo de recortes) son especialmente útiles para analizar las divergencias entre las cuatro fuentes.

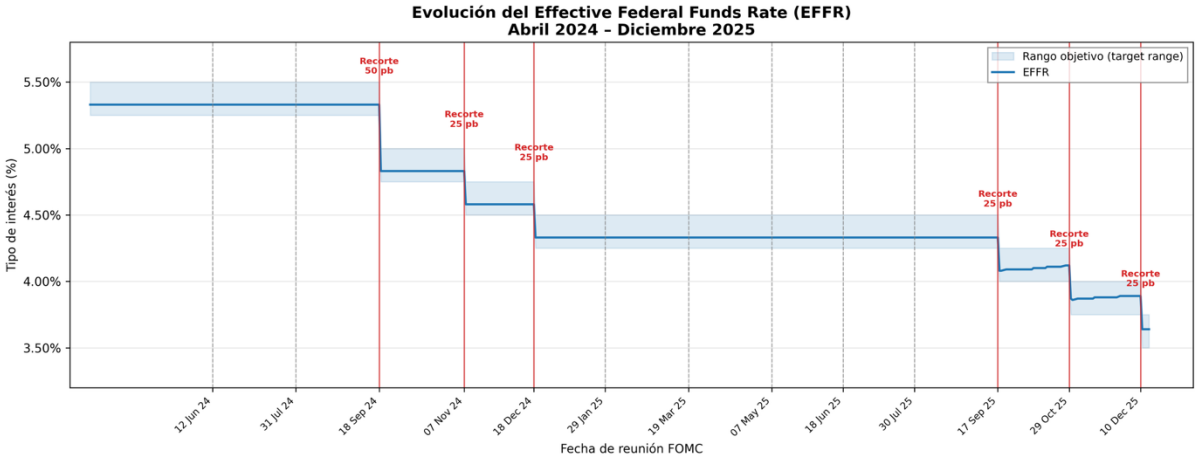


Figura 4. Evolución del Effective Federal Funds Rate (EFFR) entre junio de 2024 y diciembre de 2025. Las líneas verticales indican las fechas de las trece reuniones del FOMC analizadas. Fuente: elaboración propia a partir de datos del Federal Reserve Bank of New York

4. Modelos y metodología

Esta sección expone los fundamentos metodológicos del estudio comparativo. La Subsección 4.1 describe la estructura de los contratos de evento, la unidad básica de los mercados de predicción. La Subsección 4.2 detalla la metodología de *CME FedWatch*, herramienta institucional que actúa como referencia en el ecosistema de futuros sobre tipos. La Subsección 4.3 expone el tratamiento aplicado a los datos de *Polymarket* y *Kalshi*, incluyendo la elección del estimador diario y el procedimiento de relleno de huecos. La Subsección 4.4 discute la distinción conceptual entre probabilidades *risk-neutral* y físicas, con sus implicaciones para la interpretación de los resultados. La Subsección 4.5 presenta los dos análisis complementarios mediante los que se articula la comparación entre las cuatro fuentes. Por último, la Subsección 4.6 describe las técnicas empleadas para contrastar las diferencias.

4.1. Estructura de los contratos de evento

Los mercados de predicción se construyen sobre una unidad fundamental, el contrato de evento (*event contract*). Este constituye la unidad negociable básica de estas plataformas, y se representa a través de una pregunta binaria con resultados Sí/No. Cada contrato genera dos posiciones negociables: una posición “Yes” y una posición “No”, que funcionan como opciones binarias. Se paga 1,00 USD si el resultado correspondiente se materializa, o 0,00 USD en caso contrario. El precio de cada posición, que oscila entre 0,01 y 0,99 USD, se interpreta directamente como la probabilidad implícita de ese resultado.

Sobre esta base binaria se construyen estructuras más complejas mediante la agrupación de contratos relacionados en eventos (*events*). Un evento actúa como contenedor organizativo que agrupa uno o más contratos de eventos relacionados. Cuando un evento contiene un único contrato (*single-contract event*), el evento y el contrato son esencialmente equivalentes, creando un par simple:

```
Evento: ¿Superará la inflación de EE.UU. el 3% interanual en marzo de
2025?

└── Contrato: ¿Superará la inflación de EE.UU. el 3% interanual en
marzo de 2025? (Sí/No)
```

Para predicciones con múltiples resultados posibles, las plataformas emplean eventos multi-contrato (*multi-contract events*) que agrupan varios contratos de eventos mutuamente excluyentes.

```
Evento: ¿Cuál será la decisión de la Fed en la reunión de marzo de
2025?

├── Contrato: ¿Mantenimiento de tipos? (Sí/No)

├── Contrato: ¿Recorte de 25 pb? (Sí/No)
```

```
├── Contrato: ¿Recorte de 50 pb? (Sí/No)
├── Contrato: ¿Subida de 25 pb? (Sí/No)
└── Contrato: ¿Subida de 50 pb o más? (Sí/No)
```

Para las decisiones del FOMC, tanto *Kalshi* como *Polymarket* estructuran cada reunión como un evento multi-contrato que contiene sub-mercados independientes para cada magnitud posible de cambio en el tipo.

Cada contrato dentro del evento se negocia de forma independiente a través del libro de órdenes. En *Polymarket*, cada contrato posee identificadores únicos (*Condition ID, Question ID y Token IDs*) y genera dos tokens que se intercambian mediante el libro de órdenes centralizado (CLOB) y se liquidan en la moneda estable USDC. En *Kalshi*, los contratos de eventos operan en dólares estadounidenses bajo supervisión de la CFTC. Dado que los contratos dentro de un evento multi-contrato son mutuamente excluyentes, las probabilidades de todas las posiciones “Yes” deberían sumar aproximadamente 100%, generando una distribución de probabilidad explícita. Esta estructura proporciona mayor riqueza informativa que los instrumentos derivados tradicionales: a diferencia de los *Fed Funds Futures*, que requieren supuestos de árbol binomial para derivar probabilidades, los mercados de predicción revelan directamente la distribución de probabilidad percibida por el mercado, permitiendo asignar probabilidades no nulas a múltiples *outcomes* simultáneamente.

Más allá de su estructura económica, los contratos de evento incluyen un componente normativo igualmente relevante: cada contrato lleva asociado un conjunto explícito de reglas de resolución que especifican con precisión en qué condiciones se resolverá a favor de "Yes" o de "No" o de las opciones disponibles en dicho contrato, qué fuentes externas se considerarán válidas para verificar el resultado y bajo qué casos extremos quedarán fuera del alcance del contrato. En *Polymarket*, estas reglas son visibles públicamente en la página de cada mercado y forman parte del propio "contrato inteligente" que liquida los pagos al cierre. En *Kalshi*, las reglas las publica la plataforma como *Designated Contract Market* y están sometidas a la supervisión de la *Commodity Futures Trading Commission*. Ambas plataformas contemplan, además, mecanismos de impugnación que permiten a los participantes oponerse a una resolución cuando consideran que esta no se ajusta a los hechos verificables. En *Polymarket*, mediante un sistema de propuesta y disputa de resoluciones gestionado a través del protocolo UMA (*Universal Market Access*). En *Kalshi*, mediante un proceso de revisión interno bajo el marco regulatorio de la CFTC. Estos mecanismos no eliminan por completo los riesgos descritos al final de la Subsección 1.3, pero sí establecen una capa adicional de control sobre la integridad de la resolución, especialmente relevante en mercados sobre eventos públicos y verificables como las decisiones del FOMC, donde el resultado es inequívoco y conocido por todos los participantes en el mismo momento.

4.2. Metodología de *CME FedWatch*

CME FedWatch constituye la herramienta de referencia institucional para inferir expectativas de política monetaria a partir de los mercados de derivados. La herramienta calcula probabilidades incondicionales de los resultados de las reuniones del FOMC generando un árbol de probabilidad binario a partir de los precios de los *Fed Funds Futures* a 30 días (contratos ZQ), negociados en la Bolsa de Chicago (CME).

Los contratos de *Fed Funds Futures* incorporan las expectativas del mercado sobre el promedio diario del *Effective Federal Funds Rate* (EFFR) durante el mes del contrato. El EFFR, publicado diariamente por el Banco de la Reserva Federal de Nueva York, representa el promedio ponderado por volumen de las tasas de préstamos overnight no garantizados entre instituciones depositarias.

El cálculo de probabilidades asume que los cambios de tipos se producen siempre en múltiplos de 25 puntos básicos, como se refleja en la Figura 5.

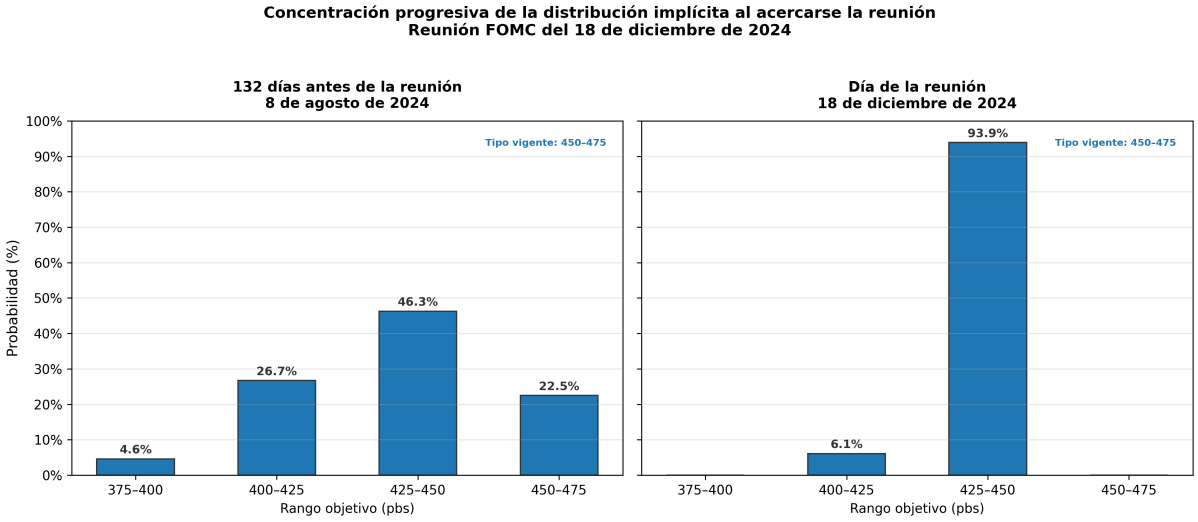


Figura 5. Representación del árbol de probabilidades de CME FedWatch para reuniones sucesivas del FOMC. Fuente: elaboración propia a partir de la metodología publicada por CME Group.

Para determinar las probabilidades de cada reunión, la metodología sigue los siguientes pasos. Primero, se identifica el mes más cercano sin reunión del FOMC como punto de anclaje, donde el EFFR promedio implícito del contrato de futuros equivale al EFFR de fin del mes anterior e inicio del mes siguiente. Segundo, para meses con reunión del FOMC, se calcula el EFFR de inicio y fin de mes utilizando el precio del contrato de futuros, el número de días antes y después de la reunión, y los valores de anclaje de los meses adyacentes. Tercero, se determina el cambio esperado en el EFFR durante el mes, definido como la diferencia entre el EFFR de fin de mes y el de inicio de mes. Cuarto, se expresa este cambio en múltiplos de 25 puntos básicos, descomponiendo el resultado en su parte entera y decimal. La parte entera determina el resultado mínimo esperado, cuya probabilidad se calcula como uno menos la parte decimal. La probabilidad del resultado inmediatamente superior

corresponde a la parte decimal. Por ejemplo, si el cambio esperado equivale a 2,9 subidas de 25 puntos básicos, *FedWatch* asigna una probabilidad del 10% a una subida de 50 puntos básicos y una probabilidad del 90% a una subida de 75 puntos básicos.

Para reuniones sucesivas, las probabilidades se combinan multiplicando las probabilidades de cada nodo del árbol binario, generando probabilidades acumuladas para los distintos escenarios de tipos. El número de resultados posibles para la reunión n es n+1, reflejando las distintas trayectorias que pueden seguir los tipos desde el punto de partida hasta dicha reunión.

Esta metodología presenta una limitación estructural que resulta relevante para el análisis comparativo. Para la reunión inmediata, el árbol binomial asigna probabilidad únicamente a dos rangos de tipos contiguos: aquel en el que se sitúa el precio implícito del futuro y el adyacente. Como consecuencia, *CME FedWatch* tiende a mostrar métricas de precisión artificialmente bajas en reuniones en las que el resultado cae en uno de esos dos rangos, pero no captura adecuadamente escenarios en los que el mercado percibe tres o más outcomes con probabilidad relevante, como ocurre típicamente en reuniones con disensiones significativas dentro del FOMC. Esta limitación constituye una de las motivaciones centrales del estudio comparativo con mercados de predicción, que sí pueden reflejar distribuciones multimodales de expectativas.

4.3. Tratamiento de los datos y alineamiento temporal

Para construir las series temporales de probabilidades implícitas de *Polymarket* y *Kalshi* se ha utilizado el VWAP (*Volume-Weighted Average Price*) diario de cada sub-mercado, calculado como la media ponderada por volumen de todos los *trades* registrados en el día. Esta elección difiere marginalmente de Diercks, Katz y Wright (2026), quienes utilizan el *last traded price* antes del cierre de los *Fed Funds Futures*. Se ha preferido el VWAP por su mayor robustez frente a *trades* aislados con precios anómalos, consideración especialmente relevante en días de baja liquidez para mercados FOMC alejados de la fecha de la reunión.

Por la distinta naturaleza de cada plataforma, el peso del VWAP no es estrictamente homogéneo entre fuentes. En *Polymarket* se pondera por el volumen monetario del *trade* (cantidad de USDC intercambiada), mientras que en *Kalshi* se pondera por el número de contratos. En ambos casos el peso refleja el tamaño económico del *trade* dentro de su propia plataforma. Las probabilidades de los sub-mercados de una misma reunión deben sumar uno, como los precios de cada sub-mercado se forman por separado, pueden aparecer pequeñas desviaciones respecto a esa suma, que se corrigen dividiendo cada probabilidad entre el total del día. Esto genera una distribución de probabilidad y garantiza la comparabilidad entre plataformas.

Cuando un sub-mercado de *Polymarket* o *Kalshi* no registra *trades* en un día concreto, se aplica un procedimiento de *forward-fill* (arrastre del último precio observado) para mantener la continuidad de la serie temporal. Posteriormente, las probabilidades de todos los sub-mercados de una misma reunión se renormalizan por día para garantizar que sumen

exactamente la unidad, condición necesaria para la comparabilidad con la distribución de *CME FedWatch* y para el posterior cálculo de métricas como el Brier Score.

El alineamiento temporal entre las cuatro fuentes requiere un tratamiento adicional. Mientras que *Polymarket* y *Kalshi* operan de forma continua, *CME FedWatch* y los *Fed Funds Futures* no publican observaciones los fines de semana ni los festivos del calendario bursátil estadounidense. Para garantizar un calendario común que permita la comparación diaria entre las cuatro fuentes en los análisis posteriores, se ha aplicado un *forward-fill* análogo en los días sin observación de *CME FedWatch* y ZQ, arrastrando el último valor disponible hasta la siguiente publicación. Este procedimiento refleja el supuesto razonable de que, en ausencia de nueva información, las expectativas implícitas en el cierre de la última sesión bursátil siguen siendo válidas durante el fin de semana o el festivo. La alternativa sería restringir el dataset únicamente a los días con observación en CME y ZQ cuya implicación sería descartar información disponible de *Polymarket* y *Kalshi* en sábados, domingos y festivos, comprometiendo la representatividad del análisis. Esta elección no está, sin embargo, exenta de inconvenientes. El forward-fill introduce autocorrelación artificial en las series rellenadas y produce variaciones diarias nulas en CME y ZQ durante los períodos no operativos. No obstante, dado que las comunicaciones oficiales del FOMC y la publicación de los principales datos macroeconómicos se concentran en días hábiles del calendario bursátil estadounidense, el flujo de información institucional relevante para las expectativas de tipos coincide en gran medida con los días de cotización de CME y ZQ, lo que acota la magnitud práctica de estos efectos.

4.4. Probabilidades *risk-neutral* frente a probabilidades físicas

Las probabilidades implícitas extraídas de los precios de *Polymarket*, *Kalshi* y de los *Fed Funds Futures* utilizados por *CME FedWatch* son probabilidades *risk-neutral* (bajo la medida Q), no necesariamente probabilidades físicas (bajo la medida P). Esta distinción es estándar en la teoría de valoración de activos derivados: el precio de mercado de un contrato que paga un dólar condicional a un evento refleja tanto la probabilidad objetiva del evento como una prima de riesgo asociada al estado de la economía en que el pago se produce. Formalmente, el precio del contrato es igual al valor esperado del pago bajo la medida Q, que difiere del valor esperado bajo la medida P por la prima de riesgo correspondiente.

En la práctica, para contratos de corta duración y *payoffs* acotados como los analizados en este estudio - sub-mercados binarios sobre decisiones FOMC con horizonte inferior a seis meses-, la diferencia entre ambas medidas suele ser pequeña. Diercks, Katz y Wright (2026) evalúan empíricamente esta diferencia para *Kalshi* mediante el *Probability Integral Transform*, y concluyen que las distribuciones están razonablemente bien calibradas respecto a la medida física, aunque detectan evidencia modesta de primas de riesgo que sobreestiman probabilidades de inflación y desempleo altos. En consecuencia, las probabilidades reportadas a lo largo de este TFG deben interpretarse como estimaciones *risk-neutral*. Se asume, basándose en la evidencia empírica del paper citado, que la diferencia con las probabilidades físicas es suficientemente pequeña como para no alterar las conclusiones comparativas del análisis.

4.5. Los dos análisis complementarios

El estudio se articula en dos análisis complementarios. El primero compara las distribuciones de probabilidad publicadas por *Polymarket*, *Kalshi* y *CME FedWatch*, evaluando cómo cada fuente asigna masa probabilística a los distintos escenarios discretos del FOMC. Este análisis excluye los *Fed Funds Futures* porque estos publican un precio único, no una distribución, y cualquier transformación a una distribución por categorías requeriría supuestos metodológicos adicionales, como el propio árbol binomial que aplica *CME FedWatch* internamente, que introducirían artefactos en la comparación.

El segundo análisis reduce cada fuente a una única magnitud escalar, la tasa media esperada implícita, que permite incorporar los *Fed Funds Futures* como cuarta fuente sin supuestos adicionales. Para las tres fuentes con distribuciones, esta tasa se obtiene como promedio ponderado de los centros de cada categoría por su probabilidad. Para los *Fed Funds Futures*, se obtiene directamente como cien menos el precio del contrato, sin necesidad de pasar por una distribución de probabilidad intermedia.

Formalmente, si p_i denota la probabilidad asignada a la categoría i y $c_i$representa el centro en puntos básicos del rango de tipos asociado a esa categoría, la tasa esperada implícita se calcula como:

$$\hat{r} = \sum_{i=1}^{K} p_i \cdot c_i$$

donde K = 4 es el número de categorías del Mapeo distribucional (Mapeo A). Los centros $c_i$no son fijos en valor absoluto, sino que se construyen relativamente a la tasa vigente antes de cada reunión: $c_i = c_{\text{vigente}} + \Delta_i$, donde c_{vigente} es el centro del rango objetivo en vigor y Δ_i es el cambio en puntos básicos asociado a la categoría i(−50, −25, 0 o +25 pb). Para las categorías acotadas (Cut 25bps y No Change), el cambio $\Delta_i$es unívoco. Para las categorías abiertas (Cut 50+bps y Hike 25+bps), se asume el cambio mínimo del bucket (−50 y +25 pb respectivamente), supuesto razonable dado que la muestra no contiene ningún hike ni recortes superiores a 50 puntos básicos.

A modo de ilustración del cálculo, como se muestra en la Tabla 3, consideramos la víspera de la reunión del FOMC del 7 de noviembre de 2024, con la tasa vigente situada en el rango 4,75–5,00 % (centro c_{vigente}= 487,5 pb). Las probabilidades publicadas por *Polymarket* ese día se traducen en una tasa esperada implícita de 462,42 pb según el cálculo detallado en la Tabla 3.

La interpretación es clara. Con un 93,84 % de probabilidad asignada al recorte de 25 pb, la tasa esperada queda muy próxima al centro de esa categoría (462,5 pb). La decisión efectiva del FOMC fue un recorte de 25 pb que situó la tasa en el rango 4,50-4,75 % (centro 462,5 pb), por lo que el error de predicción de *Polymarket* ese día fue de apenas −0,03 pb. Este ejemplo ilustra el funcionamiento de la métrica escalar tanto en su componente probabilística como en su capacidad para aproximarse al resultado realizado.

Categoría	Probabilidad p_i	Δ_i(pb)	Centro c_i(pb)	$p_i \cdot c_i$(pb)
Cut 50+bps	3,40 %	−50	437,5	14,88
Cut 25bps	93,84 %	−25	462,5	434,01
No Change	2,47 %	0	487,5	12,04
Hike 25+bps	0,30 %	+25	512,5	1,54
Total	**100 %**			**462,47**

Tabla 3. Ejemplo de cálculo de la tasa esperada implícita en el Mapeo Escalar (Polymarket, 6 de noviembre de 2024).

La reducción de dimensionalidad del segundo análisis sacrifica información sobre la dispersión de expectativas, pero a cambio permite comparar las cuatro fuentes en igualdad de condiciones y evaluar la convergencia al valor realizado de la tasa post-reunión. Ambos análisis son, por tanto, complementarios: el primero evalúa la riqueza distribucional y la calibración; el segundo, la precisión puntual de la estimación central.

Una vez definidas las métricas y los análisis complementarios, la subsección siguiente describe las técnicas inferenciales empleadas para contrastar formalmente las diferencias observadas entre fuentes, así como las hipótesis estadísticas que sustentan su aplicación.

4.6. Inferencia estadística

La inferencia sobre las diferencias de precisión predictiva entre fuentes se realiza mediante una combinación de métodos diseñada para respetar la estructura de los datos. La fuente principal de inferencia es el test pooled de Diebold-Mariano, estimado mediante OLS sobre las diferencias diarias de losses con corrección de Harvey-Leybourne-Newbold y errores estándar cluster-robustos por reunión. La elección del clustering por reunión responde a la naturaleza del proceso subyacente: cada ventana pre-FOMC constituye un proceso de difusión de información autocontenido, en el que los precios del día t y t+1 no son independientes. Esta dependencia intra-reunión justifica el uso de *bootstrap clusterizado* (10.000 réplicas, percentil al 95 %) y de errores cluster-robust en el contraste pooled. Coherentemente con este diseño, las métricas agregadas reportadas en la Sección 5 (MAE, Brier, RPS) se calculan como media de medias por reunión, asignando peso uniforme a cada una de las doce reuniones con independencia del número de observaciones diarias disponibles en su ventana pre-FOMC.

Para verificar empíricamente la presencia de autocorrelación intra-reunión y validar el enfoque clusterizado, se aplica el test de Ljung-Box sobre los diferenciales diarios de losses $d_t = L_t(A) - L_t(B)$ dentro de cada reunión, para cada par de fuentes y cada métrica (Brier y RPS en el Mapeo distribucional, MAE en el Mapeo escalar), considerando lags de 3, 5 y 10. Los resultados agregados sobre las 12 reuniones muestran que la hipótesis nula de no autocorrelación se rechaza a lag = 5 en al menos el 30 % de las reuniones para 10 de las 12 combinaciones par-métrica analizadas (83 %), lo que justifica el enfoque clusterizado en la

mayor parte del análisis. Las dos combinaciones que muestran autocorrelación más débil (*Polymarket-Kalshi* en Brier y RPS, con rechazos en el 25 % de las reuniones) coinciden precisamente con los pares cuya inferencia conjunta no resulta significativa, por lo que la robustez del clustering en esos casos no afecta a las conclusiones.

Es importante reconocer una limitación inherente al diseño muestral. El número de clústeres disponibles ($G = 12$ reuniones del FOMC) se sitúa por debajo del umbral asintótico habitualmente recomendado para garantizar la validez plena del estimador cluster-robusto ($G \geq 30$), restricción derivada de la disponibilidad de eventos discretos del FOMC en el período cubierto por los datos. Esta limitación se mitiga mediante el uso de bootstrap clusterizado, cuya validez no depende del tamaño muestral asintótico, y se reconoce explícitamente la implementación de wild cluster bootstrap (Cameron, Gelbach y Miller, 2008) como línea futura. La hipótesis de independencia entre reuniones distintas se asume por diseño, dado que cada reunión es un evento discreto separado por aproximadamente seis semanas durante las cuales el mercado redefine sus expectativas en torno a una nueva tasa objetivo. Esta asunción es la convención estándar en la literatura aplicada que utiliza errores cluster-robustos.

El análisis de causalidad de Granger sobre la dinámica diaria de las series sigue una metodología paralela. Los tests de estacionariedad ADF y KPSS se aplican a cada combinación (reunión, fuente) sobre las series originales, y en los casos en que la abrumadora mayoría de series resultan no estacionarias se aplica la transformación de primeras diferencias a todas las series para garantizar la homogeneidad del análisis. La selección del lag óptimo se realiza mediante AIC y BIC sobre el rango $K \in \{1, \ldots, 5\}$. La inferencia se basa en el test F con errores HAC en el análisis por reunión y con errores cluster-robustos por reunión en el panel con efectos fijos. Los residuos del modelo no-restringido se verifican mediante Ljung-Box y Breusch-Pagan para confirmar la validez del F-test.

Adicionalmente, se exploró la aplicación del marco de clasificación temprana de series temporales propuesto por Mori et al. (2019), basado en el clasificador 1-NN con Dynamic Time Warping y la construcción de frentes de Pareto entre *accuracy* y earliness. La aplicación a las series del Mapeo distribucional (Mapeo A) reveló que el problema de clasificación es trivial en el régimen muestral disponible: el clasificador alcanza *accuracy* próxima al 100 % desde los primeros días con valores idénticos de hipervolumen entre fuentes, lo que impide establecer un ranking discriminativo. Este resultado es consistente con la advertencia de la propia literatura, que requiere datasets sustancialmente mayores (del orden de centenares de series por clase) para producir frentes de Pareto informativos. En consecuencia, este análisis se descartó como herramienta de comparación principal y se documenta como línea futura sobre datos intradía en la Sección 6.

5. Resultados

Esta sección presentados los resultados del trabajo y realiza el contraste de las hipótesis. La Subsección 5.1 analiza la precisión predictiva, la Subsección 5.2 compara las plataformas en función de su naturaleza regulatoria, la Subsección 5.3 evalúa los futuros frente a *CME FedWatch*, y, por último, la Subsección 5.4 analiza las dinámicas de descubrimiento de precios.

5.1. Precisión predictiva (H1)

La primera hipótesis del trabajo establece que los mercados de predicción (*Polymarket* y *Kalshi*) presentan una precisión predictiva al menos comparable a la de los derivados institucionales (*CME FedWatch* y Fed Funds Futures) en términos de Brier, RPS, MAE y RMSE. Este contraste se aborda mediante una combinación de métricas descriptivas y de inferencia formal, evaluando las predicciones tanto en su forma distribucional (Mapeo A, K = 4 categorías) como en su forma escalar (Mapeo B, tasa esperada en puntos básicos).

En el dominio distribucional, las tres fuentes evaluadas convergen hacia errores reducidos a medida que se aproxima la fecha de la reunión del FOMC, como se ilustra en la Figura 6. En el día previo a la reunión (H = 1), el Brier multiclase normalizado se sitúa en 0,007 para *Polymarket*, 0,008 para *Kalshi* y 0,005 para *CME FedWatch*, mientras que el RPS muestra valores análogos (0,004, 0,005 y 0,003 respectivamente). El intervalo de confianza bootstrap del Brier medio integral sobre las 12 reuniones, calculado mediante remuestreo clusterizado con 5.000 réplicas, sitúa a las tres fuentes en rangos solapados: *Polymarket* [0,011, 0,030], *Kalshi* [0,012, 0,036] y CME [0,009, 0,028]. Las diferencias pareadas entre fuentes confirman este patrón. El test pooled de Diebold-Mariano, estimado mediante OLS con errores cluster-robustos por reunión y corrección Harvey-Leybourne-Newbold, no rechaza la igualdad de losses entre *Polymarket* y CME (p = 0,299 en Brier; p = 0,500 en RPS), ni entre *Polymarket* y *Kalshi* (p = 0,130 en Brier; p = 0,121 en RPS). Como se ve en la Tabla 4, la única diferencia pareada estadísticamente significativa en el Mapeo distribucional (Mapeo A) se observa entre *Kalshi* y CME en Brier (p = 0,017), con CME presentando un error inferior.

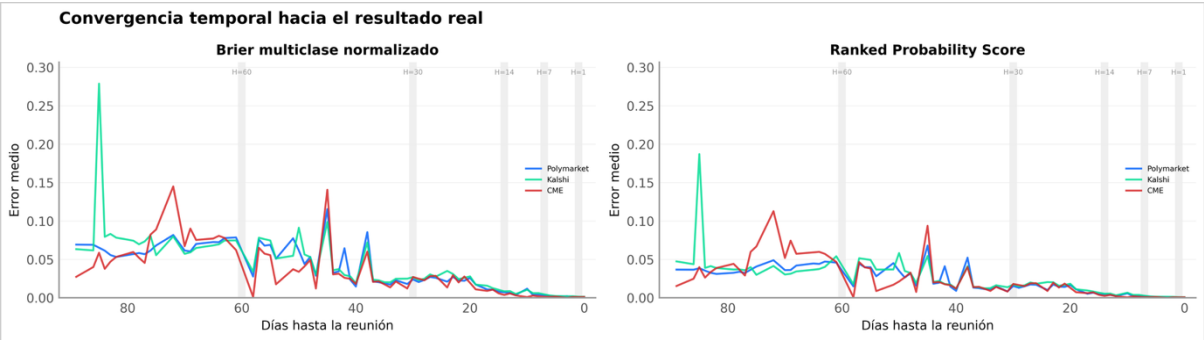


Figura 6. Convergencia temporal del Brier multiclase normalizado (izquierda) y del Ranked Probability Score (derecha) en función de los días previos a la reunión del FOMC, para las tres fuentes del Mapeo distribucional.

Fuente	Brier IC 95%	RPS IC 95%
Polymarket	0,021 [0,011; 0,030]	0,013 [0,007; 0,018]
Kalshi	0,024 [0,012; 0,037]	0,015 [0,007; 0,023]
CME	0,019 [0,009; 0,028]	0,013 [0,006; 0,020]

Tabla 4. Brier y RPS medios por fuente (Mapeo distribucional, 12 reuniones)

En el dominio escalar, como muestra la Tabla 5, las cuatro fuentes presentan un MAE estadísticamente distinto entre sí. *Polymarket* alcanza el menor MAE de la muestra (3,18 pb**,** IC 95% [2,40; 3,96]), seguido por CME (3,38 pb) y *Kalshi* (3,47 pb**),** valores prácticamente indistinguibles en términos de bootstrap clusterizado. ZQ presenta un MAE sustancialmente mayor (8,04 pb**,** IC 95% [5,75; 10,72]), patrón que se analiza en detalle en la subsección 5.3. El test pooled de Diebold-Mariano confirma estos resultados, *Polymarket* supera a *Kalshi* con significación estadística (p = 0,038) y supera a las tres fuentes institucionales en el contraste contra ZQ (p < 0,005 en todos los casos). La evolución del error medio en función del horizonte temporal se puede ver representado en la Figura 7.

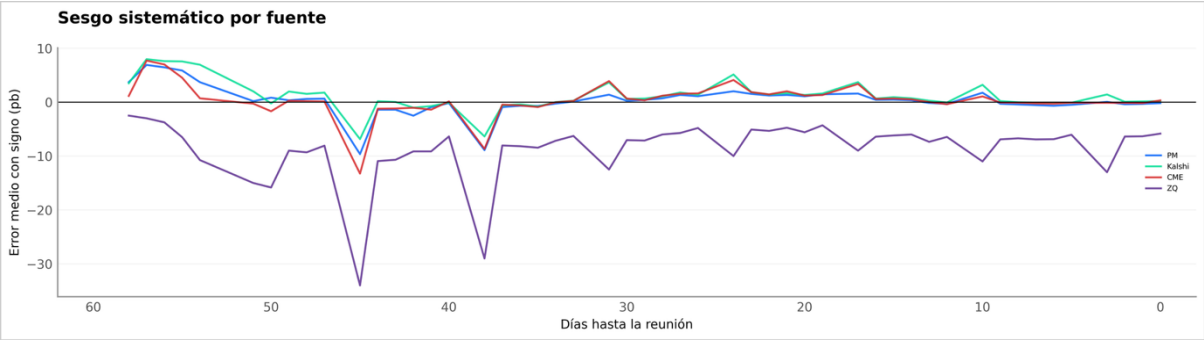


Figura 7. Error medio con signo (pb) frente a días hasta la reunión del FOMC para las cuatro fuentes del Mapeo Escalar.

Fuente	MAE (pb)	IC 95%	SS vs Persistencia
Polymarket	3,18	[2,40; 3,96]	0,69
Kalshi	3,47	[2,43; 4,55]	0,67
CME	3,38	[2,31; 4,52]	0,68
ZQ	8,04	[5,75; 10,72]	0,23

Tabla 5. MAE medio y Skill Score vs Persistencia por fuente (Mapeo Escalar, 12 reuniones)

La comparación con baselines no informativos refuerza la robustez de las conclusiones. En el Mapeo distribucional (Mapeo A), las tres fuentes baten ampliamente a los tres baselines evaluados (uniforme, no-change y empírico no condicional con leave-one-out), con BSS

medios entre 0,87 y 0,90 contra el baseline uniforme. En el Mapeo Escalar (Mapeo B), el Skill Score basado en MAE contra el baseline más exigente (Persistencia) confirma el orden: Polymarket (SS = 0,69), CME (SS = 0,68) y Kalshi (SS = 0,67) baten al baseline con significación estadística (IC 95% bootstrap no contiene cero), mientras que ZQ (SS = 0,23) no es estadísticamente distinto de Persistencia.

La evidencia sostiene la hipótesis H1, en su versión amplia: los mercados de predicción no solo son comparables a los derivados institucionales. Muestra de ello en la Figura 8, se observa que *Polymarket* presenta un MAE inferior con significación estadística respecto a *Kalshi* y a ZQ, e indistinguible de *CME FedWatch*. El ranking de precisión predictiva en el dominio escalar sitúa a *Polymarket* en primer lugar, seguido por CME y *Kalshi* con desempeño equivalente, y finalmente ZQ. En el dominio distribucional, el ranking favorece marginalmente a CME, aunque las diferencias respecto a *Polymarket* no son estadísticamente significativas.

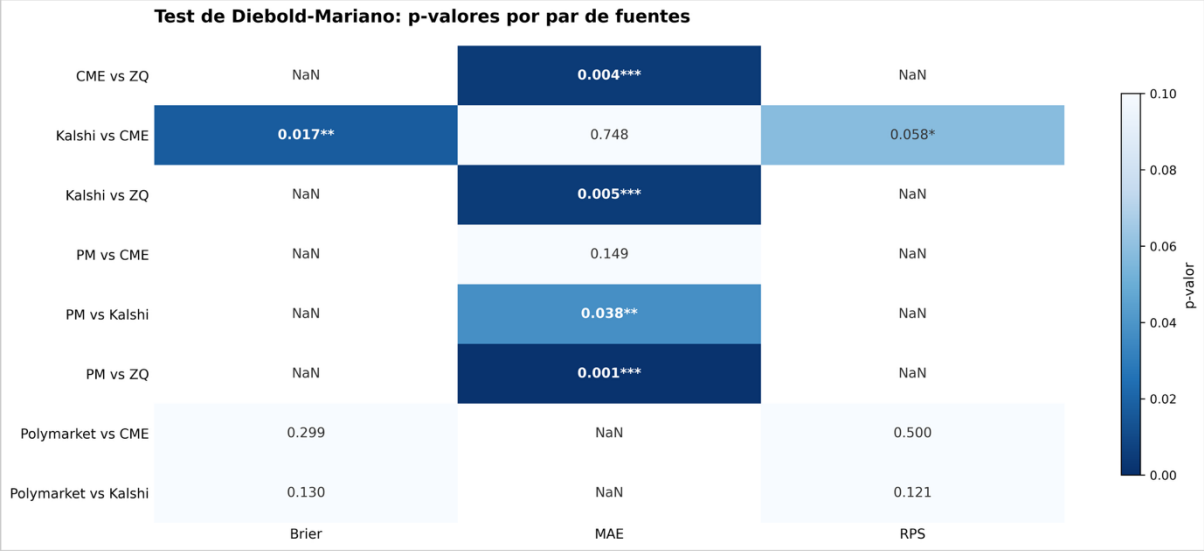


Figura 8. Mapa de p-valores del test pooled de Diebold-Mariano para todas las comparaciones pareadas entre fuentes.

5.2. Plataforma regulada vs descentralizada (H2)

La segunda hipótesis del trabajo explora si la estructura institucional de la plataforma afecta a la calidad de las probabilidades agregadas. Concretamente, se contrasta si *Kalshi* (plataforma regulada por la CFTC, con participantes residentes en Estados Unidos) y *Polymarket* (plataforma descentralizada operada sobre la red Polygon, con participantes globales en su mayoría no estadounidenses) producen predicciones de precisión y calibración equivalentes. La pregunta es relevante porque ambas plataformas operan sobre los mismos eventos del FOMC, pero difieren sustancialmente en su composición de participantes, su volumen de liquidez y los mecanismos de incentivo y arbitraje disponibles.

En términos de precisión predictiva, las dos plataformas presentan resultados próximos, pero con *Polymarket* marginalmente superior. En el Mapeo distribucional (Mapeo A), el Brier medio integral de *Polymarket* (0,021) es ligeramente inferior al de *Kalshi* (0,024), aunque la

diferencia no alcanza significación estadística según el bootstrap clusterizado (IC 95% de la diferencia: [−0,008; +0,001]) ni según el test pooled de Diebold-Mariano (p = 0,130 en Brier, p = 0,121 en RPS). En el Mapeo Escalar (Mapeo B), la diferencia se torna significativa: el MAE de *Polymarket* (3,18 pb) es inferior al de *Kalshi* (3,47 pb) con un p-valor de 0,038 en el test pooled de Diebold-Mariano sobre 5.000 réplicas bootstrap.

El análisis de calibración revela un patrón similar. El diagrama de fiabilidad de la Figura 9 ilustra esta calibración para las tres fuentes del Mapeo distribucional, mientras que las métricas detalladas se recogen en la Tabla 6. En el Mapeo distribucional (Mapeo A), el Expected Calibration Error sitúa a *Polymarket* (ECE = 0,076) ligeramente por debajo de *Kalshi* (ECE = 0,083), aunque ambas fuentes presentan calibración inferior a CME (ECE = 0,052).

La descomposición de Murphy (Brier = Reliability − Resolution + Uncertainty) muestra que en ambas plataformas el error está dominado por la componente de resolución y no por la de fiabilidad: la componente Reliability representa menos del 30% del Brier total en las dos fuentes, lo que indica que ambas plataformas producen probabilidades correctamente calibradas y que el error reside principalmente en la dificultad intrínseca del problema.

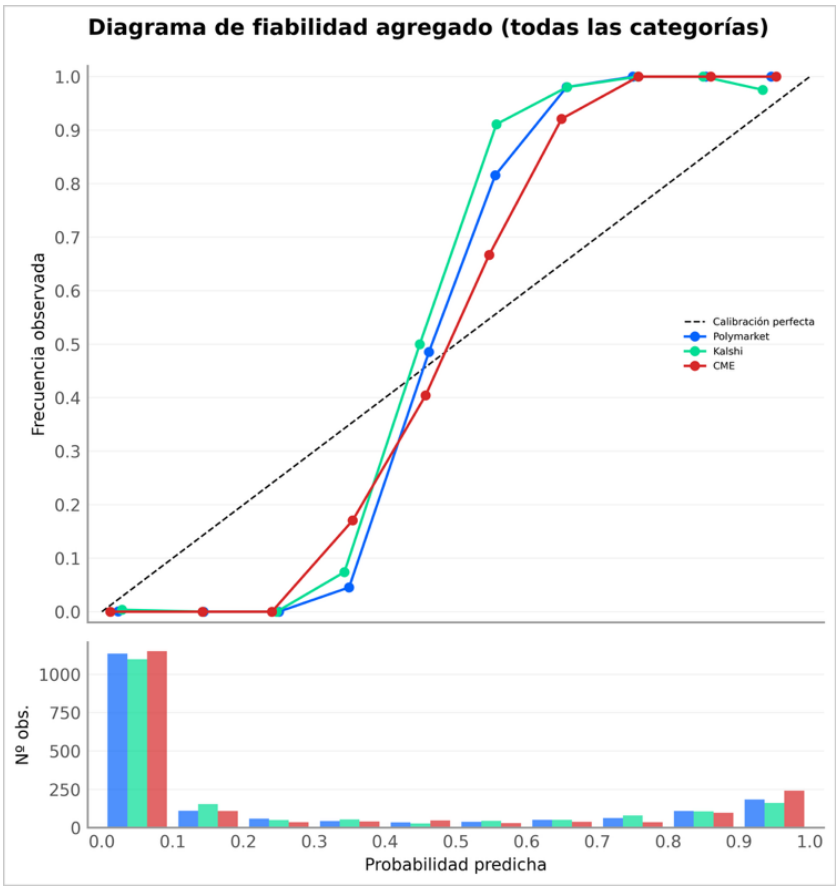


Figura 9. Diagrama de fiabilidad agregado para las tres fuentes del Mapeo distribucional, comparando probabilidad predicha y frecuencia observada. El panel inferior muestra el número de observaciones por bin.

Métrica	*Polymarket*	*Kalshi*
ECE	0,076	0,083
MCE	0,189	0,221
Reliability (Murphy)	0,006	0,007
BSS vs Uniforme	0,887	0,868
RPSS vs Uniforme	0,897	0,877

Tabla 6. Calibración comparativa Polymarket vs Kalshi.

En el dominio escalar del Mapeo B, ambas plataformas resultan estadísticamente insesgadas. El sesgo medio de *Polymarket* es prácticamente nulo (+0,03 pb, IC 95% [−0,32; +0,39], p = 0,985) y el de *Kalshi* también lo es (+0,43 pb, IC 95% [−1,74; +2,51], p = 0,694). La descomposición de Theil confirma que en ambas fuentes el error proviene predominantemente de la componente de covarianza (U_C ≈ 80% del MSE en ambas), con componentes de sesgo (U_M) marginales (2,5% en *Polymarket*, 4,9% en *Kalshi*). Este patrón es característico de predictores estadísticamente eficientes cuyo error es atribuible a ruido aleatorio y no a desplazamientos sistemáticos.

El análisis por reunión individual añade un matiz cualitativo relevante. En las tres reuniones más difíciles del período (los recortes del segundo semestre de 2025 y el inicio del ciclo en noviembre 2024), *Polymarket* mantiene un desempeño relativamente estable, mientras que *Kalshi* presenta una caída específica en la reunión de diciembre 2025. Su Brier integral es un 60% superior al de CME en esa fecha, frente a la equiparación habitual en el resto de la muestra. Este episodio sugiere una dificultad puntual de *Kalshi* para incorporar la información disponible en los días previos a esa reunión, aunque la limitación del tamaño muestral impide elevar este hallazgo a conclusión inferencial.

La evidencia sostiene una versión matizada de la hipótesis H2. Las dos plataformas son estadísticamente comparables en precisión distribucional y en calibración, pero *Polymarket* presenta una ventaja modesta y estadísticamente significativa en el dominio escalar (MAE). Esta diferencia, en un período con liquidez sustancialmente superior en *Polymarket* (volumen medio de trading de un orden de magnitud mayor) y con una base global de participantes, es coherente con la literatura sobre el papel de la liquidez en la eficiencia informacional de los mercados de predicción (Ng, Peng, Tao y Zhou, 2026). La estructura regulatoria de *Kalshi* no se traduce en una ventaja predictiva sobre la plataforma descentralizada, al menos en el período y eventos analizados.

5.3. Valor añadido del árbol binomial (H3)

La tercera hipótesis del trabajo evalúa si la transformación binomial aplicada por *CME FedWatch* sobre los *Fed Funds Futures* aporta valor predictivo respecto al uso directo de los precios implícitos de los futuros. Esta pregunta es metodológicamente relevante porque

ambas fuentes (CME y ZQ) derivan de un mismo activo subyacente, los *Fed Funds Futures* cotizados en el CME Group, pero aplican procedimientos distintos para extraer las expectativas sobre la decisión del FOMC: *CME FedWatch* utiliza un enfoque ensemble basado en múltiples contratos y un árbol binomial sobre la *Effective Federal Funds Rate* (EFFR) implícita, mientras que la extracción directa toma el precio de un único contrato y aplica la transformación `100 - precio` (Bonini, Huang y Simaan, 2025).

La diferencia metodológica se traduce en una diferencia sustancial de calidad predictiva. En el Mapeo Escalar (Mapeo B), el MAE medio de CME es de 3,38 pb (IC 95% [2,31; 4,52]) frente a 8,04 pb de ZQ (IC 95% [5,75; 10,72]), una diferencia superior al doble. El test pooled de Diebold-Mariano sobre las diferencias diarias de losses, estimado mediante OLS con errores cluster-robustos por reunión y corrección Harvey-Leybourne-Newbold, rechaza la igualdad de losses entre las dos fuentes con un p-valor de 0,0045. La diferencia es significativa además en 11 de las 12 reuniones consideradas individualmente, lo que confirma la robustez del resultado a través de toda la muestra.

El análisis del bias y la descomposición de Theil, representada gráficamente en la Figura 10 y su resumen de las métricas comparativas en la Tabla 7, aportan la información cualitativa más relevante sobre el origen de la diferencia. CME es estadísticamente insesgada (bias = +0,41 pb, IC 95% [−1,62; +2,42], p = 0,712), mientras que ZQ presenta un sesgo significativo de −6,81 pb (IC 95% [−10,10; −3,52], p = 0,003). La descomposición de Theil revela que en CME la componente de sesgo (U_M) representa apenas el 4,6% del error cuadrático medio, mientras que en ZQ alcanza el 39,2%. Este contraste es informativo: el error de CME proviene principalmente de ruido aleatorio (componente U_C ≈ 77%), mientras que el error de ZQ proviene predominantemente de un desplazamiento sistemático de la curva de pronóstico respecto a la realización.

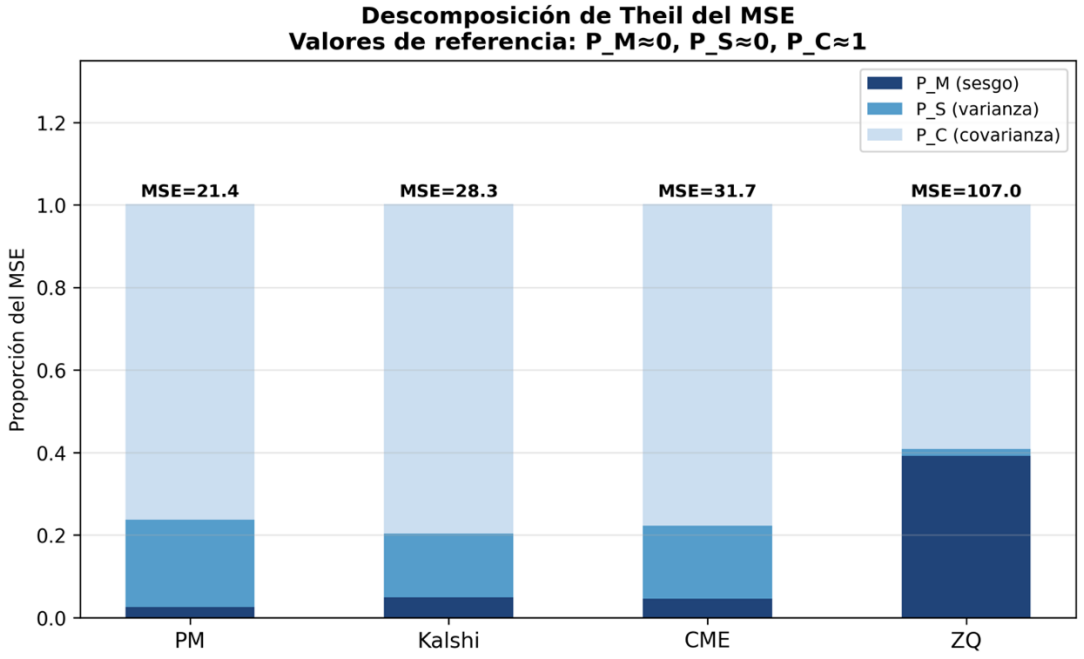


Figura 10. Descomposición de Theil del MSE en sus componentes de sesgo (U_M), varianza (U_S) y covarianza (U_C) para las cuatro fuentes del Mapeo Escalar.

Métrica	CME	ZQ
MAE (pb)	3,38	8,04
RMSE (pb)	4,32	8,71
Bias medio (pb)	+0,41 (p = 0,712)	−6,81 (p = 0,003)
Theil U_M (% del MSE)	4,6 %	39,2 %
Theil U_S (% del MSE)	18,4 %	1,7 %
Theil U_C (% del MSE)	77,0 %	59,1 %
SS_MAE vs Persistencia	0,68 (sig.)	0,23 (no sig.)

Tabla 7. Comparación CME vs ZQ en el Mapeo Escalar.

La comparación con baselines no informativos confirma la magnitud de la diferencia. El *Skill Score* basado en MAE contra el baseline de Persistencia es de 0,68 para CME, con un IC 95% bootstrap clusterizado [0,34; 0,77] que excluye el cero. La asimetría del intervalo refleja el techo natural del Skill Score: una reunión particularmente bien predicha apenas mueve el valor agregado al alza, mientras que una reunión mal predicha lo arrastra considerablemente. ZQ presenta un SS de 0,23 con IC 95% [−0,98; +0,58] que sí contiene el cero, lo que indica que la información que aporta sobre la decisión del FOMC, tras descontar lo que aporta el simple predictor de Persistencia, no es estadísticamente distinguible de cero.

El origen estructural del sesgo de ZQ admite una doble explicación, ambas documentadas en la literatura. En primer lugar, el activo subyacente liquida contra el promedio mensual de la EFFR, no contra el centro del rango objetivo. En el régimen del corredor de tasas vigente desde 2008, la EFFR se sitúa habitualmente cerca del techo del rango por la mecánica del Interest on Reserve Balances (IORB), lo que introduce un spread sistemático de aproximadamente 4-5 pb entre la tasa que el activo predice y la tasa contra la que se compara. En segundo lugar, el método de extracción directa a partir de un único contrato (utilizado en este trabajo para construir las predicciones de ZQ) presenta un error de estimación sistemático adicional respecto al método ensemble que aplica *CME FedWatch*. Bonini, Huang y Simaan (2025) documentan que *CME FedWatch* alcanza una *accuracy* del 88% en predecir las decisiones del FOMC a 30 días, frente al 75% del método de extracción directa, atribuyendo la diferencia precisamente a la dependencia del método baseline en un único contrato y en una elección ad hoc del valor de la tasa de partida r_0.

En el Mapeo distribucional (Mapeo A), CME presenta el menor Expected Calibration Error de las tres fuentes evaluadas (ECE = 0,052, frente a 0,076 de *Polymarket* y 0,083 de *Kalshi*). Sin embargo, esta ventaja debe interpretarse con cautela. El árbol binomial concentra la masa probabilística en una o dos categorías, produciendo distribuciones puntuales que son menos sensibles a errores de calibración que las distribuciones más dispersas de los mercados de predicción. En el dominio escalar del Mapeo B esta ventaja desaparece, y son *Polymarket* y CME las que comparten el primer lugar en términos de bias prácticamente nulo.

La evidencia sostiene plenamente la hipótesis H3. La transformación binomial de *CME FedWatch* aporta valor predictivo sustancial respecto al uso directo de los precios implícitos de los futuros. Los resultados de este trabajo replican y extienden la evidencia documentada por Bonini, Huang y Simaan (2025) en dos direcciones complementarias: confirman cuantitativamente la superioridad de CME sobre ZQ en una muestra más reciente (2024-2025) usando métricas escalares (MAE, RMSE) en lugar de *accuracy* binaria, e identifican mediante la descomposición de Theil que la naturaleza del error de ZQ es estructural (bias) y no atribuible a mayor varianza o ruido aleatorio.

5.4. Velocidad de respuesta y descubrimiento de precios (H4)

La cuarta hipótesis del trabajo establece que los mercados de predicción, gracias a su operativa continua y a su base global de participantes, capturan los cambios en expectativas con mayor agilidad que los mercados de futuros institucionales, y que concretamente *Polymarket* y *Kalshi* pueden liderar el descubrimiento de precios sobre *CME FedWatch*. Para contrastar esta hipótesis se aplica el test de causalidad de Granger pareado en sus dos enfoques complementarios: análisis por reunión individual y análisis en panel con efectos fijos por reunión.

La metodología se aplica en ambos mapeos. En el Mapeo distribucional (Mapeo A), se modela la dinámica diaria de la probabilidad asignada al outcome real por cada fuente (p_outcome), considerando los seis pares direccionales entre *Polymarket*, *Kalshi* y CME. En el Mapeo Escalar (Mapeo B), se modela la dinámica diaria de la tasa esperada en puntos básicos por cada fuente, considerando los doce pares direccionales entre las cuatro fuentes. En ambos casos, los tests de estacionariedad ADF y KPSS aplicados a las series originales rechazan la hipótesis de estacionariedad en la abrumadora mayoría de combinaciones (35 de 36 en Mapeo distribucional, 45 de 48 en Mapeo Escalar), por lo que el análisis se realiza sobre primeras diferencias. La selección del lag óptimo se realiza mediante AIC y BIC sobre el rango $K \in \{1,...,5\}$, y la inferencia se basa en el test F con errores HAC en el análisis por reunión y con errores cluster-robustos por reunión en el panel.

Los resultados del Mapeo distribucional (Mapeo A) revelan una relación de causalidad débil entre las tres fuentes a escala diaria. De los seis pares direccionales evaluados, únicamente CME → *Polymarket* es estadísticamente significativo en el panel ($F = 5{,}29$, $p = 0{,}023$). Las restantes cinco relaciones no rechazan la hipótesis nula de no causalidad: *Polymarket* → CME ($p = 0{,}572$), *Kalshi* → CME ($p = 0{,}637$), CME → *Kalshi* ($p = 0{,}201$), *Polymarket* → *Kalshi* ($p = 0{,}874$) y *Kalshi* → *Polymarket* ($p = 0{,}475$). Este patrón sugiere que, en el dominio distribucional, las tres fuentes son mayoritariamente independientes a escala diaria, con la única excepción de un efecto de arrastre de CME sobre *Polymarket* en la probabilidad asignada al outcome real.

El Mapeo Escalar (Mapeo B), con cuatro fuentes y mayor volumen de información diaria, produce un panorama considerablemente más rico. La Tabla 8 presenta la matriz completa de resultados del test de Granger, de los doce pares direccionales evaluados, seis resultan estadísticamente significativos en el panel. CME causa-Granger a las tres demás fuentes con

significación elevada: CME → *Polymarket* (F = 72,3, p < 0,001), CME → *Kalshi* (F = 17,8, p = 0,0001) y CME → ZQ (F = 3,5, p = 0,039). ZQ causa-Granger a los dos mercados de predicción: ZQ → *Polymarket* (F = 78,7, p < 0,001) y ZQ → *Kalshi* (F = 13,8, p = 0,0002). *Polymarket* causa-Granger a *Kalshi* (F = 24,3, p < 0,001). Las demás direcciones no son significativas, incluyendo notablemente *Polymarket* → CME (p = 0,125), *Kalshi* → CME (p = 0,468) y *Kalshi* → *Polymarket* (p = 0,087, marginal).

Fuente → Fuente	F	p-valor	Decisión
CME → *Polymarket*	72,26	<0,001	Significativa
ZQ → *Polymarket*	78,66	<0,001	Significativa
CME → *Kalshi*	17,84	0,0001	Significativa
ZQ → *Kalshi*	13,79	0,0002	Significativa
Polymarket* → *Kalshi	24,30	<0,001	Significativa
CME → ZQ	3,50	0,039	Significativa
ZQ → CME	3,12	0,054	Marginal
Kalshi* → *Polymarket	2,60	0,087	No significativa
***Polymarket* → CME**	2,23	0,125	No significativa
***Polymarket* → ZQ**	2,07	0,146	No significativa
***Kalshi* → ZQ**	1,53	0,259	No significativa
***Kalshi* → CME**	0,99	0,468	No significativa

Tabla 8. Matriz de causalidad de Granger en panel (Mapeo Escalar, 12 pares direccionales)

La jerarquía de price discovery que emerge del Mapeo Escalar (Mapeo B) es clara y consistente. CME y ZQ co-lideran el descubrimiento de precios en el dominio escalar, lo cual es coherente con el hecho de que ambas fuentes derivan del mismo activo subyacente (Fed Funds Futures): sus dinámicas diarias son casi simultáneas y juntas lideran a los dos mercados de predicción. *Polymarket* ocupa una posición intermedia, liderando a *Kalshi* pero siguiendo a las fuentes institucionales. *Kalshi* cierra la jerarquía como receptor de información sin causar a ninguna otra fuente con significación al 5%.

La comparativa entre los dos mapeos añade un matiz importante. La jerarquía direccional queda reflejada por la Figura 11. Tres de los seis pares comunes entre Mapeo distribucional (Mapeo A) y Mapeo Escalar (Mapeo B) mantienen la misma conclusión inferencial (PM → CME no significativa, *Kalshi* → PM no significativa, CME → PM significativa), mientras que dos pares (PM → *Kalshi* y CME → *Kalshi*) pasan de no significativos en Mapeo distribucional (Mapeo A) a significativos en Mapeo Escalar (Mapeo B). Esta divergencia se explica por dos factores: la mayor potencia estadística del Mapeo Escalar (Mapeo B) (que incorpora la dinámica continua de la tasa esperada en lugar de únicamente la probabilidad del outcome real) y la inclusión de ZQ, que estabiliza la inferencia en el panel.

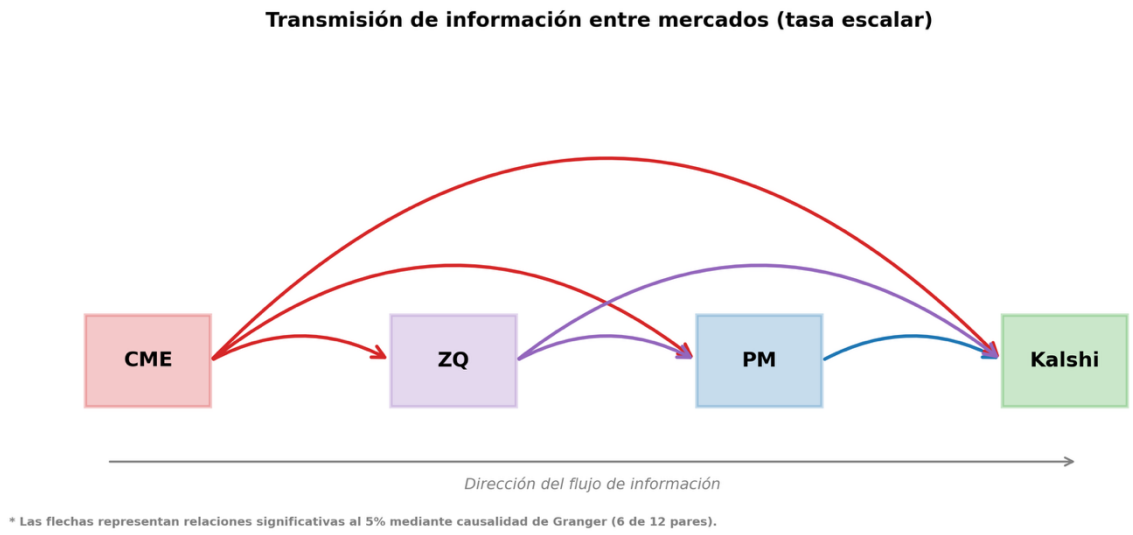


Figura 11. Diagrama de causalidad de Granger en panel para el Mapeo Escalar. Las flechas representan las relaciones direccionales estadísticamente significativas al 5%.

La evidencia refuta la versión original de la hipótesis H4. Los mercados de predicción no lideran el descubrimiento de precios sobre los derivados institucionales; al contrario, ambas fuentes basadas en *Fed Funds Futures* (CME y ZQ) lideran a *Polymarket* y a *Kalshi* con significación estadística. No obstante, una versión matizada de H4 sí encuentra apoyo empírico: dentro del ecosistema de mercados de predicción, *Polymarket* lidera a *Kalshi* en la dinámica diaria de tasas esperadas, lo cual es coherente con la mayor liquidez y la base global de participantes de la plataforma descentralizada. Este hallazgo se alinea con la evidencia de Ng, Peng, Tao y Zhou (2026), que documentan que *Polymarket* lidera el proceso de descubrimiento de precios frente a *Kalshi* en mercados de predicción sobre eventos electorales estadounidenses.

Este patrón aporta un matiz cualitativo importante a los resultados de precisión predictiva discutidos en la Sección 5.1. Aunque *Polymarket* y CME presentan MAE estadísticamente equivalentes, el análisis de Granger sobre la dinámica diaria revela una asimetría direccional clara. Los valores de CME aportan información predictiva sobre los valores futuros de *Polymarket* (CME → PM significativa en ambos mapeos), mientras que la dirección inversa no es estadísticamente significativa (PM → CME no significativa). Esta asimetría sugiere que la equivalencia en precisión final no implica equivalencia en el mecanismo informacional subyacente.

6. Conclusiones y futuras líneas de investigación

Este trabajo ha explorado la capacidad predictiva de cuatro fuentes de expectativas sobre las decisiones del *Federal Open Market Committee* (FOMC) durante un período de doce reuniones entre junio de 2024 y diciembre de 2025, comparando dos mercados de predicción (*Polymarket* y *Kalshi*) con dos fuentes derivadas de los *Fed Funds Futures* (*CME FedWatch* y ZQ baseline). El estudio se ha articulado en torno a dos mapeos complementarios sobre los mismos datos: distribucional (Mapeo A, basado en cuatro categorías de outcome) y escalar (Mapeo B, basado en la tasa esperada en puntos básicos). Planteando cuatro hipótesis específicas sobre precisión, estructura institucional, valor del árbol binomial y *price discovery*.

A lo largo del análisis se ha verificado que los cuatro flujos de información ofrecen señales empíricamente útiles sobre las decisiones del FOMC, si bien con diferencias sistemáticas que merecen una interpretación cuidadosa. Una de las principales conclusiones del trabajo radica en que **la dualidad metodológica de mapeos permite distinguir entre aspectos de calibración que dependen de la representación elegida (distribucional o escalar) y patrones robustos que se mantienen entre ambas**. *CME FedWatch* presenta la mejor calibración distribucional (ECE = 0,052), pero esa ventaja se atribuye, en buena medida, al efecto del árbol binomial sobre los *reliability diagrams*. En el dominio escalar la ventaja se diluye y son *Polymarket* y CME quienes se alternan en el primer puesto según la métrica considerada. Este hallazgo recalca la importancia de no juzgar la calidad de un predictor probabilístico por una única métrica, sino mediante un conjunto coherente de evaluaciones complementarias que cubran tanto la dimensión nominal como la ordinal y la escalar.

Un segundo hallazgo central del trabajo es la **clara superioridad de *CME FedWatch* sobre la extracción directa de expectativas a partir de los *Fed Funds Futures* (ZQ)**. En el Mapeo Escalar (Mapeo B), CME alcanza un MAE de 3,38 pb frente a 8,04 pb de ZQ, diferencia estadísticamente significativa según el test pooled de Diebold-Mariano ($p = 0{,}0045$) y consistente en 11 de las 12 reuniones evaluadas individualmente. La descomposición de Theil identifica la naturaleza estructural del error de ZQ: el 39,2 % del error cuadrático medio proviene de un sesgo sistemático (componente U_M), mientras que en CME esta componente representa apenas el 4,6 %. Este resultado replica cualitativamente la evidencia documentada por Bonini, Huang y Simaan (2025) en una muestra histórica más amplia (1994-2024, 232 reuniones), extendiéndola al dominio escalar y a una muestra reciente caracterizada por el inicio del ciclo de recortes de tipos.

Un tercer hallazgo, que es contraintuitivo respecto a la formulación original de las hipótesis, es que **los mercados de predicción no lideran el descubrimiento de precios sobre las fuentes institucionales**. El análisis de causalidad de Granger en panel sobre el Mapeo Escalar (Mapeo B) revela que CME y ZQ co-lideran el price discovery sobre *Polymarket* y *Kalshi* con significación estadística elevada ($p < 0{,}001$ en las relaciones CME → PM y ZQ → PM). La hipótesis original H4, que postulaba el liderazgo de los mercados de predicción gracias a su operativa continua y su base global, queda refutada por los datos en su versión amplia. No obstante, se sostiene una versión matizada de la hipótesis. Dentro del ecosistema de mercados de predicción, *Polymarket* lidera a *Kalshi* en la dinámica diaria de tasas

esperadas ($F = 24{,}3$, $p < 0{,}001$), hallazgo coherente con la evidencia de Ng, Peng, Tao y Zhou (2026) sobre la mayor eficiencia informacional de *Polymarket* en mercados de predicción sobre eventos electorales estadounidenses. Esta asimetría direccional aporta una lectura matizada sobre la equivalencia en MAE entre *Polymarket* y CME documentada en la Sección 5.1, la precisión final equivalente coexiste con una asimetría en el flujo informacional diario, en el que CME aporta información predictiva sobre *Polymarket* pero no a la inversa.

Desde el punto de vista metodológico, el trabajo identifica varias limitaciones que conviene reconocer explícitamente. La muestra disponible (doce reuniones del FOMC) se sitúa por debajo del umbral asintótico habitualmente recomendado para la inferencia con errores cluster-robustos ($G \geq 30$), limitación inherente a la disponibilidad de eventos discretos del FOMC en el período de estudio. Esta restricción se mitiga mediante el uso de bootstrap clusterizado de 5.000 réplicas, cuya validez no depende del tamaño muestral asintótico, y mediante la validación empírica de la estructura de cluster con el test de Ljung-Box, que confirma la presencia de autocorrelación intra-reunión en 10 de las 12 combinaciones par-métrica analizadas. Asimismo, el periodo de estudio coincide con un único régimen monetario (transición desde el ciclo de subidas hacia el ciclo de recortes), lo que limita la generalización de las conclusiones a regímenes alternativos. Por último, las categorías agregadas del Mapeo A (Cut 50+bps y Hike 25+bps) se evalúan asumiendo el centro del primer bucket, simplificación razonable dado que la muestra no contiene ningún hike ni recortes superiores a 50 puntos básicos, pero que en muestras con mayor heterogeneidad de outcomes requeriría una formulación más sofisticada.

A pesar de estas limitaciones, el trabajo aporta tres contribuciones específicas a la literatura. En primer lugar, integra en un mismo marco analítico cuatro fuentes de expectativas que la literatura previa había estudiado por separado, con métricas no directamente comparables. En segundo lugar, propone una dualidad metodológica (Mapeo A distribucional y Mapeo B escalar) que permite identificar artefactos de la representación frente a patrones robustos. En tercer lugar, replica y extiende los hallazgos de Bonini, Huang y Simaan (2025) sobre la superioridad del método ensemble de *CME FedWatch* sobre la extracción directa de *Fed Funds Futures*, añadiendo evidencia sobre la naturaleza estructural del sesgo a través de la descomposición de Theil.

En síntesis, **las cuatro hipótesis del trabajo encuentran apoyo empírico diferenciado**. La hipótesis H1 (precisión predictiva al menos comparable entre mercados de predicción y derivados institucionales) queda validada, con *Polymarket* presentando incluso un MAE estadísticamente inferior al de *Kalshi* y al de ZQ. La hipótesis H2 (equivalencia entre plataforma regulada y descentralizada) se sostiene en términos de calibración distribucional, pero queda parcialmente matizada por la ligera ventaja de *Polymarket* en el dominio escalar. La hipótesis H3 (valor añadido del árbol binomial sobre la extracción directa) queda plenamente confirmada, con CME superando a ZQ por un margen sustancial. La hipótesis H4 (los mercados de predicción lideran el price discovery) queda refutada en su versión original, pero apoyada en su versión matizada, y es que *Polymarket* lidera a *Kalshi* dentro del ecosistema de mercados de predicción.

Para continuar profundizando en la cuestión planteada por este trabajo, se proponen las siguientes líneas de investigación o mejora del marco analítico:

- **Ampliación temporal del análisis**. Extender la muestra a un período más largo (por ejemplo, desde 2022 inclusive) permitiría incorporar reuniones del ciclo de subidas además del ciclo de recortes, mejorando la generalización de las conclusiones a regímenes monetarios distintos y aumentando el número de clusters disponibles para la inferencia.
- **Inclusión de fuentes adicionales**. Incorporar las probabilidades del *Market Probability Tracker* del *Federal Reserve Bank of Atlanta* (basado en SOFR futures) y las *World Interest Rate Probabilities* de Bloomberg permitiría una comparación más completa del ecosistema de fuentes de expectativas, contrastando además las distintas metodologías de extracción empleadas por cada una.
- **Análisis intradía y *price discovery* a alta frecuencia**. La utilización de datos intradía de *Polymarket* y *Kalshi*, junto con tick-data de los Fed Funds Futures, permitiría aplicar técnicas de microestructura de mercados financieros (información implícita, Hasbrouck information shares, Gonzalo-Granger *price discovery shares*) para cuantificar de forma precisa qué fuente incorpora primero la información en momentos críticos.
- **Implementación de wild cluster Bootstrap**. La limitación de G = 12 clusters podría mitigarse mediante el uso de wild cluster bootstrap (Cameron, Gelbach y Miller, 2008), específicamente diseñado para entornos con pocos clusters, lo que reforzaría la robustez de las conclusiones inferenciales.
- **Modelos de detección de régimen**. La incorporación de un esquema de detección automática del régimen monetario (subida, bajada, estabilidad) permitiría aplicar modelos especializados por régimen y contrastar si la jerarquía de calidad predictiva entre fuentes se mantiene en condiciones macroeconómicas diversas.

Una de las líneas de extensión más prometedoras a corto plazo consiste en la aplicación del marco de clasificación temprana de series temporales propuesto por Mori et al. (2019), basado en el clasificador *1-Nearest Neighbor* con *Dynamic Time Warping* y la construcción de frentes de Pareto entre precisión (*accuracy*) y precocidad (*earliness*). La aplicación exploratoria de este marco en el presente trabajo reveló que el problema de clasificación es trivial en el régimen muestral disponible. El clasificador alcanza accuracy próxima al 100 % desde los primeros días con valores idénticos de hipervolumen entre fuentes, lo que impide establecer un ranking discriminativo. Este resultado es consistente con la advertencia de la propia literatura, que requiere datasets sustancialmente mayores (del orden de centenares de series por clase) para producir frentes de Pareto informativos.

La extensión natural de esta línea consiste en aplicar el mismo marco sobre datos intradía de *Polymarket*, donde el volumen de observaciones (millones de transacciones a lo largo del período de estudio) permitiría explotar el método en su régimen natural de aplicación. La cuestión empírica que abriría este análisis es de gran interés práctico para los participantes

en mercados financieros: ¿en qué momento exacto, dentro de la ventana pre-FOMC, las probabilidades implícitas en los precios alcanzan un grado de estabilidad suficiente para sustentar decisiones de inversión con confianza estadística? La respuesta requeriría no solo replicar el marco de Mori et al. (2019) sobre datos de mucha mayor frecuencia, sino además contrastar la robustez del frente de Pareto a especificaciones alternativas del clasificador (1-NN frente a *k-NN* con voto ponderado, *Dynamic Time Warping* con ventana restringida, y métricas de earliness alternativas) y verificar la estabilidad de los resultados a través de submuestras de regímenes monetarios distintos. Esta línea representa una continuación natural del trabajo realizado y una contribución potencial relevante a la literatura sobre detección de información implícita en mercados de predicción de alta frecuencia.

Bibliografía

Arrow, K. J. *et al.* (2008). "The promise of prediction markets". *Science*, 320(5878), pp. 877-878. Disponible en: https://mason.gmu.edu/~rhanson/PromisePredMkt.pdf [Consultado 08-05-2026].

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., Ungar, L. y Mellers, B. (2017). "Distilling the wisdom of crowds: prediction markets vs. prediction polls". *Management Science*, 63(3), pp. 691-706.

Bonini, S., Huang, S. y Simaan, M. (2025). "Watching the *FedWatch*". *Journal of Futures Markets*. Disponible en: https://doi.org/10.1002/fut.70077 [Consultado 09-05-2026].

Cameron, A. C., Gelbach, J. B. y Miller, D. L. (2008). "Bootstrap-based improvements for inference with clustered errors". The Review of Economics and Statistics, 90(3), pp. 414-427. Disponible en: https://www.nber.org/system/files/working_papers/t0344/t0344.pdf [Consultado 04-05-2026].

CBC News (2026). "How multi-million dollar speculative betting could convert climate change skeptics". *CBC News*, abril. Disponible en: https://www.cbc.ca/news/science/betting-climate-*Polymarket*-*Kalshi*-prediction-9.7181757 [Consultado 30-04-2026].

CNBC (2026). "U.S. senators ban themselves from prediction markets trading". *CNBC*, 30 de abril. Disponible en: https://www.cnbc.com/2026/04/30/senate-prediction-markets-trading-ban-*Kalshi*-*Polymarket*.html [Consultado 30-04-2026].

Department of Justice (2026). *U.S. Soldier Charged With Using Classified Information To Profit From Prediction Market Bets*. Office of Public Affairs, 23 de abril. Disponible en: https://www.justice.gov/opa/pr/us-soldier-charged-using-classified-information-profit-prediction-market-bets [Consultado 30-04-2026].

Diercks, A. M., Katz, J. D. y Wright, J. H. (2026). "*Kalshi* and the rise of macro markets". *Finance and Economics Discussion Series*, 2026-010. Washington: Board of Governors of the Federal Reserve System. Disponible en: https://doi.org/10.17016/FEDS.2026.010 [Consultado 30-04-2026].

Fortune (2026). "*Kalshi* and *Polymarket* are racing to ban insider trading". *Fortune*, 26 de abril. Disponible en: https://fortune.com/2026/04/26/prediction-markets-insider-trading-illegal-*Kalshi*-*Polymarket*-robin-hanson-economist/ [Consultado 30-04-2026].

Hayek, F. A. (1945). "The use of knowledge in society". American Economic Review, 35(4), pp. 519-530. Disponible en: https://www.jstor.org/stable/1809376 [Consultado 29-04-2026].

Lindauer, M. (1955). "Schwarmbienen auf Wohnungssuche". *Zeitschrift für vergleichende Physiologie*, 37(4), pp. 263-324. Traducción al inglés ("House-hunting by honey bee swarms") por P. K. Visscher, K. Behrens y S. Kuenholz, disponible como material suplementario en Visscher (2007). Disponible en: https://doi.org/10.1007/BF00303153 [Consultado 30-04-2026].

Mori, U., Mendiburu, A., Miranda, I. M. y Lozano, J. A. (2019). "Early classification of time series using multi-objective optimization techniques". *Information Sciences*, 492, pp. 204-218. Disponible en: https://doi.org/10.1016/j.ins.2019.04.024 [Consultado 30-04-2026].

Ng, H., Peng, L., Tao, Y. y Zhou, D. (2026). "Price discovery and trading in modern prediction markets". *SSRN Working Paper*, No. 5331995. Disponible en: https://ssrn.com/abstract=5331995 [Consultado 30-04-2026].

NPR (2026). "A $400,000 profit on Maduro's capture raises insider trading questions on *Polymarket*". *NPR*, 5 de enero. Disponible en: https://www.npr.org/2026/01/05/nx-s1-5667232/*Polymarket*-maduro-bet-insider-trading [Consultado 30-04-2026].

PBS News (2026). "Senate bans members and staff from betting in prediction markets". *PBS News*, 30 de abril. Disponible en: https://www.pbs.org/newshour/politics/senate-bans-members-and-staff-from-betting-in-prediction-markets [Consultado 30-04-2026].

Polymarket (2024). *Fed Rate Decision in September 2024 - Resolution Notice*. Disponible en: https://*Polymarket*.com/event/fed-decision-in-september-2024 [Consultado 30-04-2026].

Reichenbach, F. y Walther, M. (2025). "Exploring decentralized prediction markets: accuracy, skill, and bias on *Polymarket*". *SSRN Working Paper*, No. 5910522. Disponible en: 10.2139/ssrn.5910522 [Consultado 30-04-2026].

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. Nueva York: Doubleday. Disponible en: https://sentry.rmu.edu/SentryHTML/pdf/lib_finn_DISC8710_wisdom_of_crowds.pdf [Consultado 26-04-2026].

Tetlock, P. E. y Gardner, D. (2015). *Superforecasting: The Art and Science of Prediction*. Nueva York: Crown Publishers.

Wolfers, J. y Zitzewitz, E. (2004). "Prediction markets". *Journal of Economic Perspectives*, 18(2), pp. 107-126. Disponible en: https://www.aeaweb.org/articles?id=10.1257/0895330041371321 [Consultado 29-04-2026].