

Kalshi and the Rise of Macro Markets

Anthony M. Diercks* Jared Dean Katz †
Federal Reserve Board Northwestern University

Jonathan H. Wright ‡
Johns Hopkins University and National Bureau of Economic Research

First version: July 22, 2025
This version: September 20, 2026

Abstract

Prediction markets offer a new market-based approach to measuring macroeconomic expectations in real-time. We evaluate the accuracy of prediction market-implied forecasts from Kalshi, the largest federally regulated prediction market overseen by the CFTC. We compare Kalshi with more traditional survey and market-implied forecasts, examine how expectations respond to macroeconomic and financial news, show how policy signals are interpreted by market participants, and find that inefficiencies in some markets may be smaller than in comparable surveys. We show in a theoretical model that the strong forecast performance can be rationalized even with less informed and relatively few traders, assuming they are free of reputational concerns. Our results suggest that Kalshi markets provide a high-frequency, continuously updated, distributionally rich benchmark that is valuable to both researchers and policymakers.

JEL Codes: E3, G1, C5.

Keywords: Kalshi, Federal Reserve, prediction markets, expectations, forecasting

*The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve System. We thank Eric Engstrom, Michael Palumbo and members of the Monetary and Financial Market Analysis section in the Division of Monetary Affairs for helpful comments and suggestions. We also thank seminar participants at the CFTC and IIF MacroFor and conference participants at the Paris 10th International Workshop on Financial Markets and Nonlinear Dynamics (keynote), Midwest Macro 2026, and Theories and Methods in Macro (T2M). All errors and omissions are the sole responsibility of the authors. Diercks is the corresponding author. Website: www.anthonydiercks.com. Email: anthony.m.diercks@frb.gov.

†Email: jared.katz@kellogg.northwestern.edu.

‡Email: wrightj@jhu.edu.

1 Introduction

Managing expectations is central to modern macroeconomic policy. Yet the tools that are often relied upon—surveys and financial derivatives—have many drawbacks. Surveys can quickly become stale and ask specific questions that often provide point forecasts without any measure of uncertainty. Financial derivatives are also limited to certain contracts and trading may be thin. This paper introduces a new source of real-time, financially-backed expectations data: Kalshi macroeconomic prediction markets. Kalshi is the first federally regulated platform where traders bet directly on economic outcomes such as inflation, payrolls, and Federal Reserve decisions. Using high-frequency trading data, we show that these markets yield well-calibrated, rapidly updating density forecasts on important economic variables, including several for which alternatives are not available. Our study highlights the promise of prediction markets as a new benchmark for measuring expectations and informing monetary policy decisions.

Despite a large literature using surveys, options, and asset prices to infer macroeconomic expectations (e.g., [Gürkaynak, Sack and Swanson \(2012\)](#), [Nakamura and Steinsson \(2018\)](#), [Swanson \(2021\)](#)), real-money prediction markets remain underexplored within macro-finance. This reflects the shortcomings of earlier platforms, which were often thin, unregulated, or focused on non-economic outcomes. The availability of credible, liquid markets for macro variables now enables a direct, incentive-compatible lens on expectations—one that avoids many of the limitations of existing proxies. Our analysis is one of the first to systematically examine this new data source.

Kalshi, the largest CFTC-approved prediction market in the United States, began operating in 2021. Classified as a “Designated Contract Market”—the same category as the Chicago Mercantile Exchange—it is supported by market makers such as Susquehanna. Kalshi contracts are also accessible to retail traders through brokerage platforms like Robinhood and Webull. Each contract is a simple Arrow-Debreu security that pays one dollar if the specified outcome occurs; therefore, the full risk-neutral probability density function (pdf) of an event can be constructed by the set of contracts in the given market. Kalshi provides intradaily trade data for different measures of CPI (MoM, YoY, or calendar year), Core CPI, PCE inflation, unemployment rate, payroll releases, GDP growth, probability of recession, and federal funds rate target rates meeting-by-meeting. Several of these variables had no previous financial derivatives trading.

We begin by showing that the probability distributions implied by Kalshi markets are well-behaved and broadly consistent with those from more established financial instruments, where available. We find that in several episodes, they allocate probability mass in ways that

may reflect the range of plausible macroeconomic outcomes better than traditional financial derivative or survey-based forecasts.

Second, we document rich intraday dynamics in Kalshi markets. These probabilities respond sharply and sensibly to major developments. For example, the implied probability of a rate cut at the July 2025 FOMC rose to 25% following remarks from Governors Waller and Bowman, before falling after a stronger-than-expected June employment report. This pattern is representative of the broader high-frequency responsiveness we observe and is missed with daily data.

Third, we evaluate the forecasting performance of Kalshi markets relative to the Survey of Market Expectations (SME) run by the Federal Reserve Bank of New York (FRBNY). For the federal funds rate forecasts 150 days (3 FOMC meetings) ahead, Kalshi’s mean absolute error is very similar to that of professional forecasters. But unlike the survey—which provides a snapshot every six weeks of a modal path—Kalshi offers a continuously updating full distribution. We provide a preview of these results in Figure 1. This figure shows the mean absolute prediction error, averaging across all FOMC meetings, plotted against the number of days before the FOMC meeting. We find the Kalshi median and mode have a perfect forecast record on the day before the FOMC meeting. We also assess CPI, core CPI, and unemployment forecasts relative to the Bloomberg consensus. We find that Kalshi expectations are statistically similar for core CPI and unemployment, with forecast errors that are almost the same as the Bloomberg consensus. For headline CPI, we find that Kalshi provides a statistically significant improvement over the Bloomberg consensus forecast.

Fourth, we demonstrate that Kalshi provides real-time, distributional forecasts for macroeconomic variables such as GDP, core CPI, and unemployment—markets for which options data have historically been unavailable. While option-implied distributions exist for equities, interest rates, and CPI headline inflation swaps, Kalshi uniquely extends this to other headline macro indicators at high frequency while providing a more retail-investor perspective. Moreover, there are no currently trading options that give distributions of the federal funds rate after specific FOMC meetings, another hole that Kalshi fills.

To better understand why Kalshi-implied forecasts may outperform surveys, we extend the model of [Morris and Shin \(2002\)](#) to one where market participants and forecasters have different qualities of information and behavioral biases, and where market participants trade an asset whose price provides a forecast. The model yields testable predictions: that (1) market participants have less predictable errors (less bias) and (2) that liquidity is relatively unimportant for driving prediction accuracy, hypotheses we test and validate in the data. We show evidence that suggests that prediction markets exhibit less behavioral biases than traditional surveys and that forecast performance does not improve much with added market

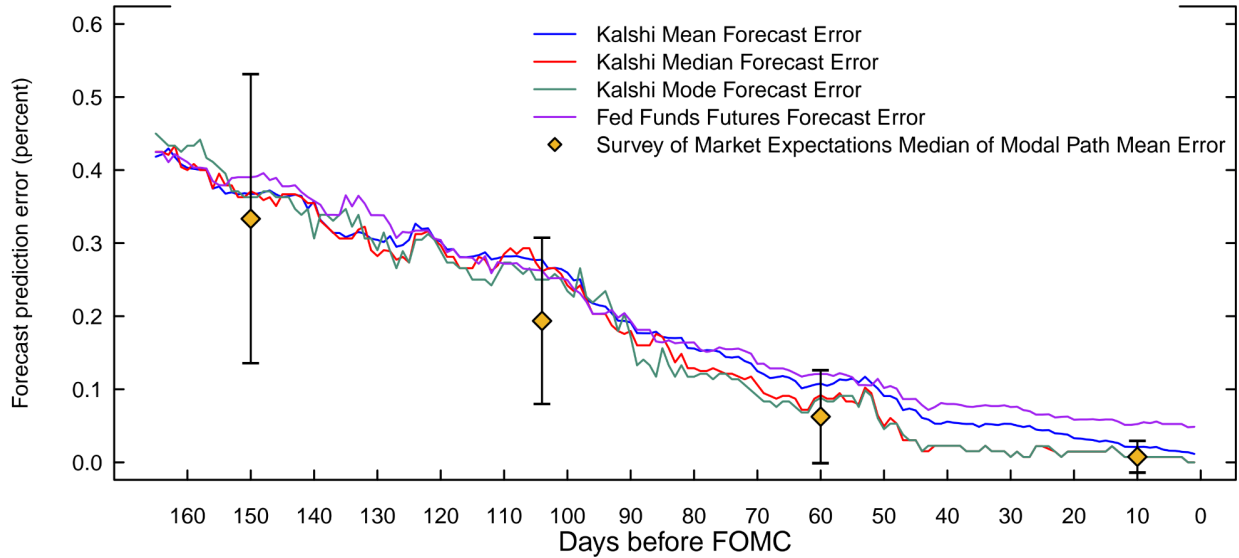


Figure 1: **FOMC Federal Funds Rate Forecast Errors Comparison**

This figure compares the mean absolute errors from 160 days out to each FOMC for the effective federal funds rate since 2022. The vertical bars reflect one-half standard deviation for the forecast errors for the SME and correspond to when the surveys were completed.

volume.

Fifth, with access to full distributions, we uncover new patterns in how macroeconomic news affects the moments of the Federal Funds rate distribution. We study how different kinds of news announcements affect the mean, variance and skewness of interest rate beliefs. We find that, during the 2022-2025 period, the variance of interest rate distributions declines substantially following data announcements, especially inflation releases. These tests can be used to learn which events markets consider the most important to policymakers at a given time, even without these events changing first moments. We split out the announcements of CPI inflation into ones which were positive surprises, negative surprises and zero surprises. We find that the variance of interest rates declines the most following a zero surprise inflation release, but also falls—though to a lesser extent—after both positive and negative surprises, consistent with a general resolution of uncertainty. Furthermore, we uncover an asymmetry with respect to the first moment, where positive shocks to CPI lead to much larger responses in the mean of the Fed Funds rate when compared to negative CPI shocks. We also examine the effects of macro surprises on the moments of the Federal Funds rate in a standard event-study regression and find many of these moves are statistically significant.

Lastly, we conduct event study regressions for the effects of monetary policy shocks on the distribution of the Federal Funds rate. We find that contractionary shocks associated

with the release of the FOMC statement have statistically significant impacts on not just the first moment, but also increase the second moment. In contrast, shocks associated with the press conference have limited effect on the first moment but a significant and negative effect on skewness, implying some potential resolution of upside uncertainty to interest rates during this window.

We intend (subject to approval) to make the distributional daily-level data publicly available via an interactive website at EconFutures.com, offering downloadable time series and visualizations for each contract, as well as trade-level data and code to further update prediction market data available on [this paper’s GitHub](#). Our goal is to facilitate further research and enable policymakers to easily monitor shifts in investors’ beliefs in real-time.

Our paper touches upon several strands of the macro-finance literature. A large body of work extracts interest rate beliefs from options prices. Seminal contributions include [Aït-Sahalia \(1996\)](#), who develops nonparametric pricing methods, and [Ball and Torous \(1999\)](#), who highlight the role of stochastic volatility. Subsequent research emphasizes unspanned volatility ([Fan, Gupta and Ritchken, 2003](#); [Li and Zhao, 2006](#)) and the pricing implications of volatility smiles ([Jarrow, Li and Zhao, 2007](#); [Amin and Ng, 1997](#)). Other studies examine the structure of state price densities ([Bakshi and Madan, 2000](#); [Li and Zhao, 2009](#)) and the pricing of caps and swaptions ([Longstaff, Santa-Clara and Schwartz, 2001](#); [Miltersen, Sandmann and Sondermann, 1997](#)), while more recent work by [Mertens and Williams \(2021\)](#) uses option prices to derive expectations related to the zero lower bound and [Wright \(2018\)](#) evaluates options for longer-maturity real interest rates. Although informative, many of these approaches rely on complex structural models that infer expectations from instruments indirectly related to the policy rate. In contrast, we use prediction market data explicitly tied to the federal funds rate and other macroeconomic variables, enabling the recovery of risk-neutral probabilities with minimal assumptions at high frequency.

Our work is also closely related to studies of monetary policy expectations, uncertainty, and skewness. [Gürkaynak, Sack and Swanson \(2012\)](#) construct measures of market-based policy expectations, while [Beber and Brandt \(2006\)](#) show that macroeconomic news affects both beliefs and preferences as reflected in options markets. More recent work explores the role of monetary policy uncertainty ([Bauer, Lakdawala and Mueller, 2022](#); [Wright, 2025](#)) and the pricing of asymmetric policy responses—often labeled the “Fed put” ([Cieslak and Vissing-Jorgensen, 2021](#)). Our analysis contributes to this literature by introducing a high-frequency, event-level measure of policy expectations distributions from prediction markets. We show that macroeconomic news significantly shifts not only the mean but also higher moments of the federal funds rate distribution in asymmetric ways. Asymmetries in interest rate distributions have also been explored for longer-dated yields as in [Bauer and Chernov](#)

(2024), while [Diercks, Tanaka and Cordova \(2024\)](#) focus on skewness in expectations of interest rates coming from the FRBNY’s SME.

A related literature examines beliefs about inflation extracted from options markets. [Kit-sul and Wright \(2013\)](#) provide an early analysis of risk-neutral inflation density functions, measuring risk of high inflation and deflation, and documenting meaningful variation in higher moments around key macroeconomic events. [Hilscher, Raviv and Reis \(2022\)](#) and [Hilscher, Raviv and Reis \(2024\)](#) explore how the pricing of extreme inflation outcomes informs debates on debt sustainability and macroeconomic fragility. [Rich and Tracy \(2010\)](#) compare market-based and survey-based measures, highlighting differences in expectations, disagreement, and uncertainty. Our analysis contributes to this literature by introducing a new source of market-based inflation expectations—derived from prediction markets—which provides direct, high-frequency measures beliefs about upcoming releases of both headline and core inflation. Whereas the existing inflation swaps and options reference headline inflation over periods of a year or more, Kalshi is different in giving information about upcoming data announcements, and also including information about core inflation which is otherwise unavailable.

Also related to our work is a growing literature that highlights the informational value of retail investor behavior. While early research often emphasized behavioral biases, recent studies show that retail trading can enhance market efficiency. [Farrell, Green, Jame and Markov \(2022\)](#) find that the rise of retail trading has increased the informativeness of prices, particularly around earnings announcements. [Boehmer, Jones, Zhang and Zhang \(2020\)](#) show that retail order flow predicts returns, and [Kelley and Tetlock \(2013\)](#) find that aggregated retail trades reflect valuable firm-level information. [Kelley and Tetlock \(2017\)](#) further show that retail short-selling contains predictive content for future stock price movements. [Chen, De, Hu and Hwang \(2014\)](#) also demonstrates that retail opinions shared via social media platforms contribute meaningfully to price discovery. Our use of prediction market data offers a new lens on retail investor expectations, capturing their beliefs directly and in real time, independent of the portfolio constraints and hedging motives that can often shape the positions of institutional investors, potentially generating sizeable risk premiums.

More broadly, we contribute to the forecasting literature that assesses the informativeness of financial market instruments and surveys, along with their distributions ([Ang, Bekaert and Wei, 2007](#); [Bakshi, Panayotov and Skoulakis, 2011](#); [Christoffersen, Jacobs and Chang, 2013](#); [Clements, 2018](#); [Goff, Kostka and Masera, 2014](#); [Duffee, 2012](#); [Faust and Wright, 2012](#)).

Lastly, and most importantly, our work builds on a substantial literature examining prediction markets and economic forecasting. [Snowberg, Wolfers and Zitzewitz \(2013\)](#) and [Wolfers and Zitzewitz \(2004\)](#) provide comprehensive reviews of prediction markets in this

context. The study closest to ours is [Gürkaynak and Wolfers \(2005\)](#), which analyzed the now-defunct Economics Derivatives market developed by Goldman Sachs and Deutsche Bank. [Hanson and Oprea \(2009\)](#) shows that even in the presence of manipulation, prediction markets can yield reliable forecasts, and [Snowberg and Wolfers \(2010\)](#) examine behavioral biases that may affect pricing. Our contribution builds on this literature by applying prediction markets to the study of monetary policy, inflation, and other macro expectations using newly available contracts with high-frequency, event-level granularity.

We also acknowledge several concurrent studies that complement our findings. [Swanson, Wang and Wu \(2025\)](#) use Kalshi data to evaluate the Fed information effect, confirming that monetary policy shocks reduce expectations for both growth and inflation. [Burgi, Deng and Whelan \(2025\)](#) examine the full range of Kalshi markets and find them to be valuable forecasters for nearly all events, many of which are unrelated to economic variables. [Eichengreen, Viswanath-Natraj, Wang and Wang \(2025\)](#) use Polymarket data to explore questions surrounding Fed independence. [Gomez Cram, Guo, Jensen and Kung \(2026\)](#) also considers the universe of Polymarket data and shows the value of informed traders, while [Gomez Cram, Guo, Jensen and Kung \(2025\)](#) validates Polymarket markets as a measure of corporate earnings expectations. We complement these contributions by focusing on extensively validating Kalshi as a forecasting tool and as a source of market-based expectations for key macroeconomic variables that are at the core of the Federal Reserve’s mission, while also documenting real-time stagflationary risks and asymmetric responses of Fed Funds rate distributions to news surprises. In addition, we provide codes and a transparent process for constructing probability distributions from underlying prediction market trades, facilitating future use by both researchers and policymakers.

Our paper is structured as follows. Section 2 reviews institutional details of prediction markets. Section 3 documents how we convert Kalshi options prices to probability distributions. Sections 4 and 5 compare the Kalshi point and density forecasts with alternatives around specific episodes. Section 6 gives a comparison of forecasts by average absolute error. Section 7 examines how Kalshi predictions respond to news announcements. Section 8 presents a model which rationalizes our findings and tests mechanisms for why Kalshi-implied forecasts may outperform alternatives, and section 9 concludes.

2 Institutional Features and Design of Kalshi Markets

2.1 Overview

Prediction markets are financial platforms where users can trade contracts that pay out based on the outcome of real-world events, such as elections, sporting events, or macroeconomic releases. These contracts function similarly to options. A trader can buy or sell an option for a price x that pays off \$1 if a given outcome occurs and nothing otherwise. The exchange naturally takes no position, and so every option has to have a buyer and a seller.

Kalshi, the largest CFTC-approved prediction market, defines a market as a single binary contract (e.g., “CPI exceeds 2.5% YoY in March 2025”), and a series as a group of related markets (e.g., CPI exceeding various thresholds across different months). Kalshi has gained significant traction, with market making provided by firms such as Susquehanna. Retail investors can access Kalshi contracts via platforms like Robinhood and Webull. This growing infrastructure makes prediction markets a viable and valuable subject for economic study.

Competing Platforms. Other platforms offering prediction markets include Polymarket, PredictIt, and Interactive Brokers. Of these, only Kalshi and Interactive Brokers operates completely under regulatory approval. Polymarket operates in a legal gray area and, along with PredictIt and Interactive Brokers, supports fewer contracts, lower liquidity, and smaller individual position limits. Kalshi’s maximum exposure per market currently reaches \$7 million.

Given these distinctions, we argue that Kalshi represents the most mature and comprehensive prediction market for economic forecasting. Accordingly, we rely on Kalshi data for the remainder of this paper. We discuss further in Appendix A.1 further technical advantages of using Kalshi data compared to Polymarket in the context of our specific analysis.

Forecast Markets on Macroeconomic and Monetary Policy Events. We identify a promising new class of prediction markets focused specifically on macroeconomic indicators and monetary policy outcomes. Table 1 lists several key Kalshi series in this space, including CPI, payroll releases, and FOMC decisions—events that typically attract substantial attention from financial markets.

Prior literature has shown that surprises in these events can significantly move asset prices, elevate market volatility, and contribute to macroeconomic uncertainty. Since central bank decisions depend heavily on incoming data and market expectations, accurate, real-time forecasting of these variables is vital for effective policy implementation.

Series	First Contract	Frequency	Theme
CPI MoM	June 2021	Monthly	Inflation
CPI YoY	November 2022	Monthly	Inflation
CPI for Year	2022	Annual	Inflation
Core CPI MoM	June 2022	Monthly	Inflation
Core CPI YoY	December 2022	Monthly	Inflation
Core CPI for Year	2025	Annual	Inflation
Unemployment Rate	July 2021	Monthly	Labor Market
Payroll Release	March 2023	Monthly	Labor Market
GDP Growth	Q2 2021	Quarterly	Growth
GDP Growth	2025	Annually	Growth
Probability of US Recession	2022	Annually	Growth
Federal Funds Rate Decision	May 2023	FOMC	Monetary Policy
Federal Funds Rate Target Rate	December 2021	FOMC	Monetary Policy

Table 1: **Kalshi Markets of Interest**

There are thousands of Kalshi markets available, but we deem these particular markets to be the most economically relevant.

Why High-Frequency Forecasts Matter. Prediction markets offer high-frequency, continuously updating forecasts that can complement central bank decision-making. High-frequency data let us apply an event study methodology to see how news shapes beliefs about macroeconomic indicators. For macroeconomic indicators like CPI and unemployment—two pillars of the Federal Reserve’s dual mandate—market-based forecasts improve the Fed’s ability to communicate policy direction and assess the market’s perceived reaction function under various scenarios.

For monetary policy itself, such as upcoming FOMC decisions, prediction markets offer a valuable check on the effectiveness of Fed communication. Importantly, these markets can also reflect the entire distribution of expectations, helping policymakers understand tail risks, asymmetries, and market uncertainty—information that is lost when only modal or median expectations are provided.

2.2 Traditional Forecasting Approaches

Table 2 summarizes commonly used forecasting tools for the economic indicators listed in Table 1. These tools fall into two categories: surveys and market-based measures.

2.2.1 Survey-Based Forecasts

Surveys offer forecasts without requiring complex models or assumptions and are not subject to risk premia since respondents do not face financial consequences for their predictions. Notable surveys include:

Source	Variables	Horizon	Frequency	Type	Format
Bloomberg Consensus Survey	CPI/Payrolls/GDP releases & Fed decisions	Next release	Once-per-event	Survey	Point-estimates
Survey of Market Expectations PDFs	Fed decisions & EoY CPI, GDP	2 FOMC Meetings	Once-per FOMC cycle	Survey	Distribution
Survey of Market Expectations Paths	Fed decisions	6 FOMC Meetings	Once-per FOMC cycle	Survey	Point-estimates
Blue Chip Economic Indicators	GDP/PCE/Unemployment	Quarterly 2-Years ahead	Once-per-month	Survey	Point-estimates
Survey of Professional Forecasters	GDP/CPI/Unemployment	Annual 2-Years ahead	Once-per-quarter	Survey	Distribution
Federal Funds Rate Futures	Fed decisions	Far	Market-based	Market-based	Point-estimates
SOFR options	Fed decisions	Far	Market-based	Market-based	Distribution
Overnight Interest Rate Swaps	Fed decisions	Far	Market-based	Market-based	Point-estimates
CPI Fixings	CPI releases	One year before	Market-based	Market-based	Point-estimates
Kalshi	CPI/Payrolls/GDP releases & Fed decisions	Next release & Next 4 FOMC meetings	Market-based	Market-based	Distribution

Table 2: **Forecast Tools**

This table shows the available forecast tools along with the relevant variables, the horizons, how frequently they are available, and whether or not they provide only point estimates or full distributions.

- **Survey of Market Expectations** (SME) by the Federal Reserve Bank of New York (formerly SPD/SMP), which elicits expectations for upcoming FOMC decisions and some year-end macroeconomic variables.
- **Bloomberg Consensus**, which collects modal expectations for macroeconomic announcements shortly before they are released.
- **Blue Chip Economic Indicators**, which gather monthly modal expectations for year-end GDP and CPI.

While surveys have been considered rather accurate on average ([Ang, Bekaert and Wei, 2007](#)), they are generally infrequently available and may not reflect updates in real-time. Also, numerous studies have rejected the full information rational expectations hypothesis in survey expectations ([Coibion and Gorodnichenko, 2015](#); [Farmer et al., 2024](#); [Bordalo et al., 2020](#)). Anchoring, inertia, and reputational concerns can distort survey results (see [Diercks and Jendoubi \(2023\)](#) for further discussion).

2.2.2 Market-Based Forecasts

Market-based measures offer the advantage of near-instantaneous updating in response to news. For monetary policy, key instruments include:

- **SOFR Options**, which provide liquid estimates of the distribution of SOFR over quarterly horizons. However, while SOFR futures are one of the most heavily traded interest rate derivatives, their options are much less traded. Moreover, these are options on one-month or three-month average repo rates. Translating these into forecasts for specific FOMC meetings requires strong and unreasonable assumptions. Moreover, the basis between SOFR and the effective federal funds rate can be time-varying and, at times, substantial.
- **Federal Funds Futures**, which offer mean rate forecasts by calendar month. Unfortunately, they do not give the entire distribution of outcomes unless one is willing to assume that there are only two possible outcomes, which can be problematic in times of high uncertainty or when looking beyond the next meeting. While there was a time when Fed Funds futures options were traded (see [Emmons, Lakdawala and Neely \(2006\)](#)), this came to a halt following the financial crisis back in 2008. Quotes are still provided on an indicative basis, but they do not trade.

Overall, we argue that Kalshi should be used to provide risk-neutral pdfs concerning FOMC decisions at specific meetings because SOFR is too far removed from the monetary

policy interest rate decision (both because of the basis and the different time horizon) and these are the only other short-term interest rate options that actually trade.

For inflation, **CPI Fixings** provide mean expectations for specific releases but are primarily institutional products with limited retail accessibility. More broadly, most of the markets described above are dominated by large institutions that use them primarily for hedging purposes. Outside of prediction markets, we are unaware of any market-based forecasts for other important upcoming releases such as payrolls, unemployment, GDP, or core CPI. There are headline CPI swaps and options, but these have much longer maturities (see [Diercks, Campbell, Sharpe and Soques \(2023\)](#) for analysis of the relative forecast performance of inflation swaps).

2.3 Volume

Given that Kalshi is relatively new when compared to more established markets such as Fed Funds futures, it is natural to wonder about its liquidity. Liquidity is important as it helps to ensure prices reflect real-time information from incoming news. As noted above, in recent years, quotes on options for Fed Funds futures have been provided on an indicative basis, but are not actually traded.

In contrast, Kalshi contracts for the federal funds rate have relatively large liquidity. Figure 2 shows a volume heatmap for every contract dating back to 2021 for the Federal Funds rate. In this figure, the darker the shading, the greater the volume. One can see that recent periods have had volumes greater than a million for several strikes, and more likely outcomes tend to have greater volumes.

Figure 3 shows that event volume has been growing over recent years, and peaks in times of high uncertainty. For example, the April 2025 CPI release achieved over \$11M in volume, reflecting uncertainty on trade-war related inflation. Appendix Figure A.2 documents volume across trading price. Trading occurs most at extremes—volume is exchanged on very unlikely or very likely events—mostly reflecting FOMC decisions being largely certain at the time of trading activity.

3 Converting Kalshi into Probability Distributions

Kalshi structures financial and macroeconomic contracts as a series of binary options that pay \$1 if the realized outcome exceeds a specified strike. Figure 4 illustrates this structure for the December 2025 FOMC meeting. Each contract allows users to trade “Yes” or “No” positions, depending on whether they believe the event will occur. Our objective is to

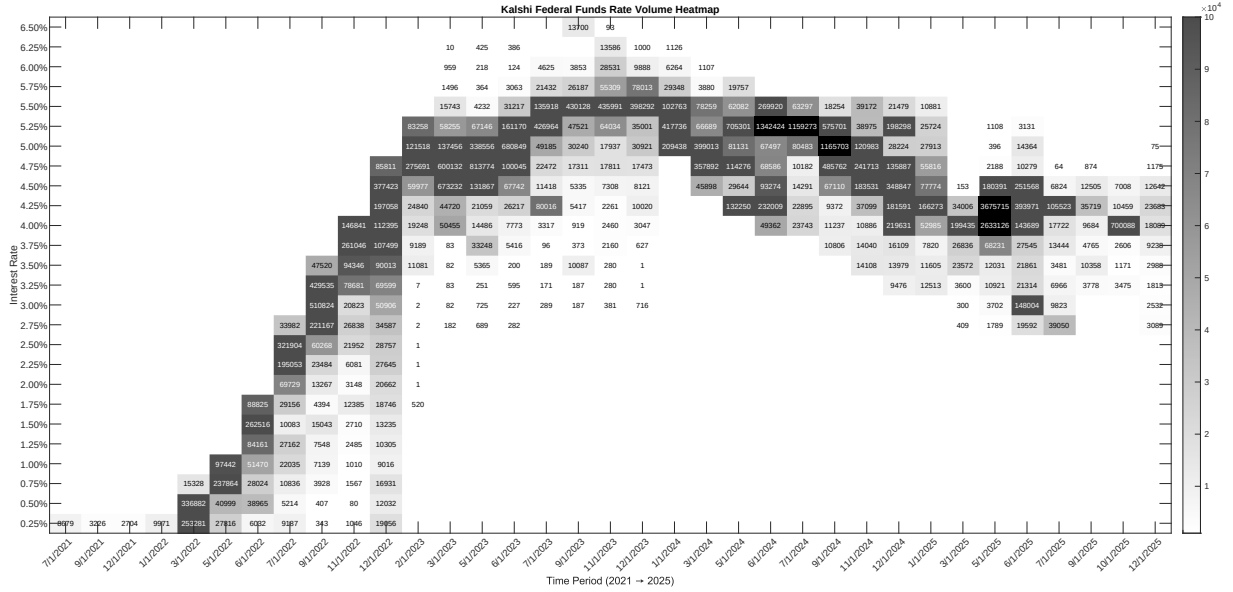


Figure 2: Volume Heatmap for Kalshi Federal Funds Rate Forecasts

This chart shows the volume for each strike price for each meeting as of June 2025. Each box is shaded based on the level of volume, with darker boxes corresponding to greater volume.

convert the market prices of these binary contracts into an implied probability distribution, updated daily, to produce high-frequency forecasts for key macroeconomic variables. We briefly outline methodology here, with a full discussion and tests of robustness to our minimal assumptions in Appendix A.3.

Methodology. We begin by scraping trade-level data from Kalshi. We then apply each day a straightforward mapping from market prices to probability mass, treating the price of a “Yes” contract as the market-implied risk-neutral probability that the event occurs. In contracts that cash as long as the outcome exceeds the strike (rather than the outcome falling in a specific bin), we calculate the probability of the outcome by subtracting from the “Yes” price of exceeding the previous strike. For instance, in Figure 4, the 4.0 strike has a last-traded price of \$0.40, while the 4.25 strike has a last-traded price of \$0.22, so we calculate an 18% probability the Fed Funds rate falls in the interval $(4.0, 4.25]$ ¹.

In the event a strike has no traded contracts on a day, we carry over the last-traded strike from a previous day. Sometimes, two strikes have the same price², and in that case we assign the probability to the outcome closer to the mode. This simple, model-free approach allows us to construct a daily-updating risk-neutral pdf over the relevant outcome space. Once we

¹Kalshi’s contract is denoted for the upper bound of the Federal Funds rate, so in practice the probability allocated between 4.0 and 4.25 being 18% implies a 18% chance on the target range being 4.0-4.25.

²This happens mostly near the tails, since prices are bounded between \$0.01 and \$0.99.

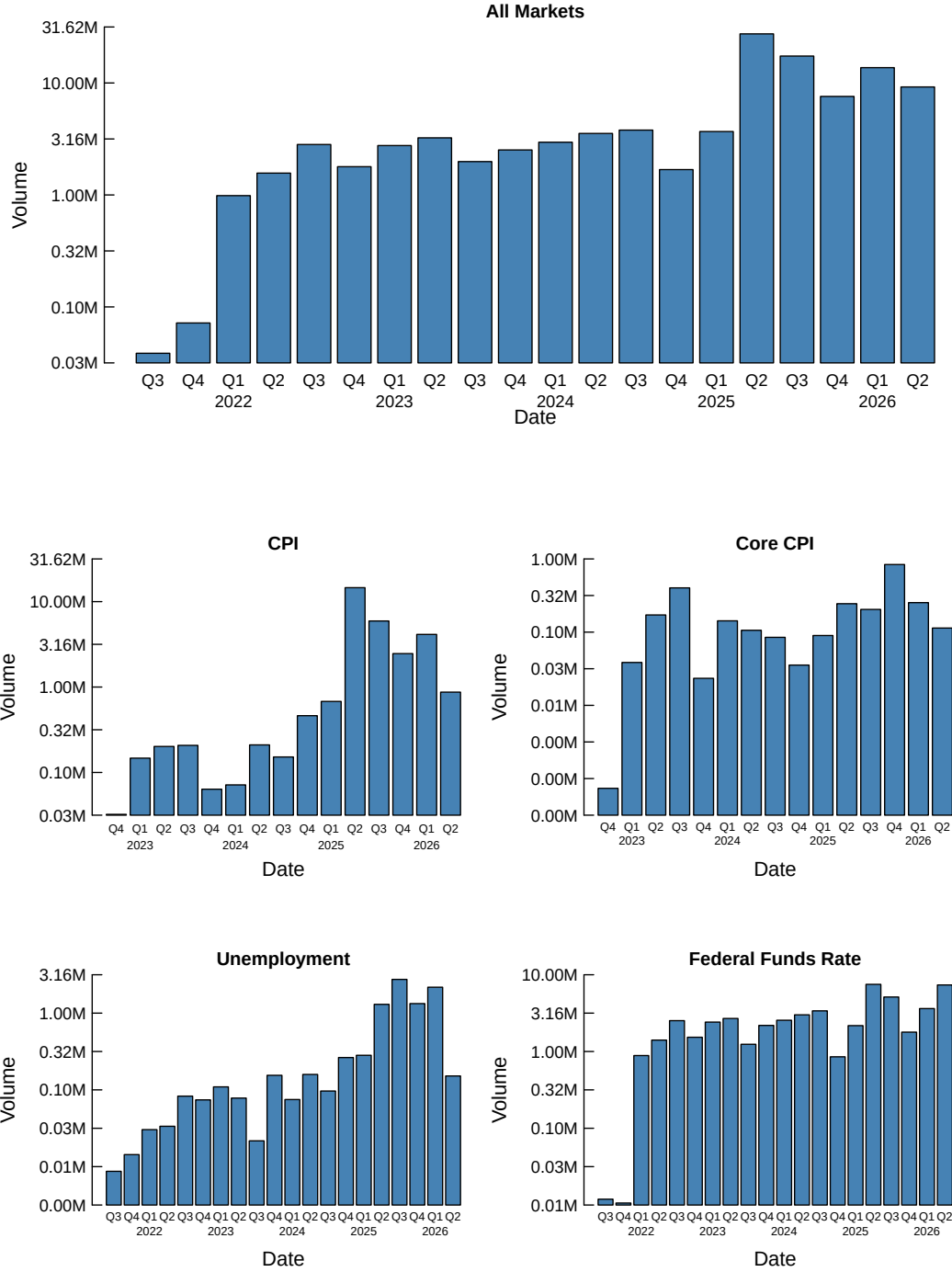


Figure 3: **Volume Time Series for Kalshi Macroeconomic Markets**

Each panel shows the total volume at a quarterly aggregate for each primary macroeconomic market we analyze from Kalshi on a log scale, dating back to the quarters each market starts. The top panel reflects the sum of CPI, Core CPI, unemployment, and the Federal Funds Rate. For the Federal Funds Rate, note that we only include volume for the Federal Funds Target Rate market, which has less volume than the Federal Funds Rate Decision market, but has a trading structure that allows us to back-out market-implied distributions more than one meeting in advance.

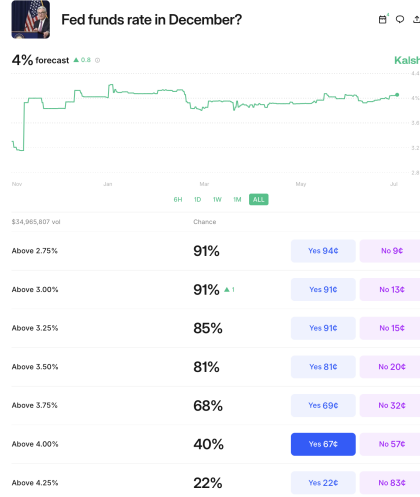


Figure 4: Fed Funds Rate for December 2025 contract

The prices shown in this figure were captured on July 22, 2025. The underlying contract is for determining the federal funds rate in December. In this market, the upper bound of the range is considered when evaluating the outcome. The “Above 4.25” outcome corresponds to the target range or no further rate cuts by December 2025, as of the time the image was captured.

have the risk-neutral pdf, we can immediately work out the implied moments, notably the mean. In subsequent sections, we demonstrate that this parsimonious approach improves forecast performance and enhances analytical flexibility over existing forecast tools. We relax the assumptions in Appendix A and get similar forecast performance.

Caveats and Discussion. While our approach is straightforward, there are some caveats worth noting. First, as with all options, it is giving risk-neutral probabilities under the $\mathbf{Q}$ measure, not actual physical probabilities under the $\mathbf{P}$ measure. In other words, the probabilities may be distorted by risk premia. The retail investor base of Kalshi might alter the risk premia properties. We will return to assessing the risk premia later. Separately, the outermost (tail) contracts often suffer from low trading volume, which can lead to stale prices and noisy estimates in the tails of the distribution—especially in illiquid markets or at dates far from the event realization. However, these issues are not unique to Kalshi. SOFR options also suffer from sparse trading in extreme outcomes and Fed Funds Futures options over the past several years have not traded at all. To mitigate tail-related distortions, we enforce monotonicity in the CDF by constructing the distribution outward from the mode toward the tails. This approach prevents stale or illiquid pricing in the tails from contaminating the rest of the pdf. We find that enforcing monotonicity of our cdf starting from the mode improves the reliability when compared to starting on one side or the other when constructing the distributions. We highlight that constructing complete distributions from Kalshi’s economic

contracts can have challenges, and we intend to make public complete code packages which do so from trade-level and bid-ask data.

Even with these limitations, the probability distributions implied by Kalshi prices compare favorably to these alternative market-based forecasts when evaluated on predictive accuracy. While refinements—such as inferring stale probabilities from the midpoint of bid-ask spreads—may further improve precision, we find such enhancements are not necessary to deliver forecasting gains.³ More importantly, Kalshi provides the fastest-updating distributions currently available for many key macroeconomic indicators, and the only reasonable distributions available for macroeconomic indicators like GDP, Unemployment, or Core CPI. It is also the only financial option with active trading which directly gives beliefs about the Federal Funds rate.

4 Time-Series Comparisons to Existing Benchmarks

A natural question is how the time-series evolution of Kalshi-implied forecasts compares to established benchmarks. In this section, we document both the similarities and differences between Kalshi and more traditional market-based measures, such as Federal Funds futures and overnight index swap (OIS) rates.

Federal Funds Rate Projections. Figure 5, left panel, displays the evolution of Federal Funds rate expectations for the December 2024 FOMC meeting. The Kalshi-implied mean forecast (red line) closely tracks both market-based expectations and the consensus from the SME. Notably, the sharp decline in forecasts at the end of July—following the FOMC’s signal of a likely rate cut in September—is more pronounced in the OIS and Fed Funds futures data. In contrast, the Kalshi mean and the SME consensus exhibit a more gradual adjustment, with broadly similar timing and magnitude.

In the right panel, we see that the projections for the June 2025 FOMC meeting were fairly similar across the different approaches. One slight difference is that the Kalshi mean did not decline as much in early April, which ended up being prescient given the developments that unfolded during that period.

Federal Funds Rate Intermeeting Period Dynamics. We can also more closely track the time series of probabilities for particular rate outcomes for a given meeting. For instance, in Figure 6, we plot the time series of the probability of the most likely rate outcome for the

³We find that midpoints of bid-ask spreads seem to introduce additional issues due to occasionally large spreads on tail outcomes. See Appendix A.4 for more discussion.

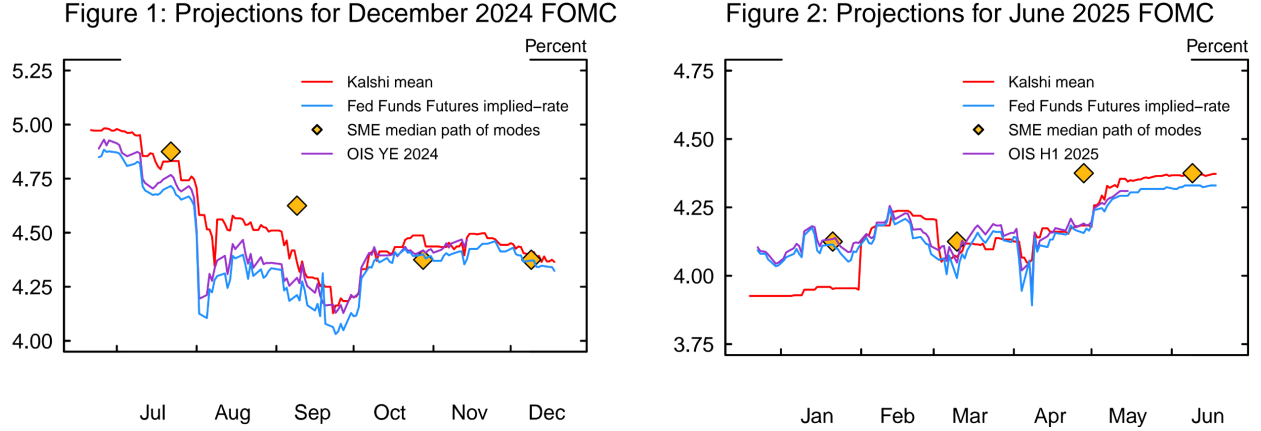


Figure 5: **FOMC Federal Funds Rate Comparisons**

The chart on the left shows the federal funds rate projections for the period for the December 2024 FOMC meeting. OIS YE 2024 corresponds to the overnight index swap rate for the end of 2024. The PDS Median path of modes corresponds to the SME median of modal path projection for the federal funds rate. The chart on the right shows similar projections for the June 2025 FOMC meeting.

July 2025 FOMC meeting (blue line)—this corresponds to the probability of no rate change. One can see that the probability of no rate change declined following several communications that were interpreted by investors as more accommodative than expected. This led the probability to fall below 80%. It wasn't until the June Nonfarm Payrolls report that the probability of no rate change jumped above 90%.

Additional examples of comparisons between Kalshi forecasts and survey counterparts are provided in the Appendix: Figures A.9, A.10 and A.11. These give forecasts for monthly CPI readings, Q4/Q4 projections for CPI and real GDP and unemployment.

5 Distributional Comparisons to Existing Benchmarks

Kalshi vs Fed Funds Futures. A key strength of the Kalshi market is the probability distribution provided by its market structure. In contrast, Fed Funds futures require binomial tree assumptions which enforce just two possible outcomes.

For instance, Figure 7 provides a comparison for the September 2025 FOMC meeting as of dates in June and July of that year. The panel on the left shows the implied probabilities based on two outcomes: no rate change and one 25 basis point cut. In contrast, the Kalshi distribution provides a richer set of outcomes as it does not face the same set of restrictions. While the Fed Funds futures suggest a weight of 0.75 on a 25 basis point cut, this is likely lumping in weight on lower rate outcomes, as shown by the Kalshi chart on the right. Instead, according to the Kalshi chart, there is much greater uncertainty as investors do not put such a large weight on the modal outcome of a quarter-point rate cut.

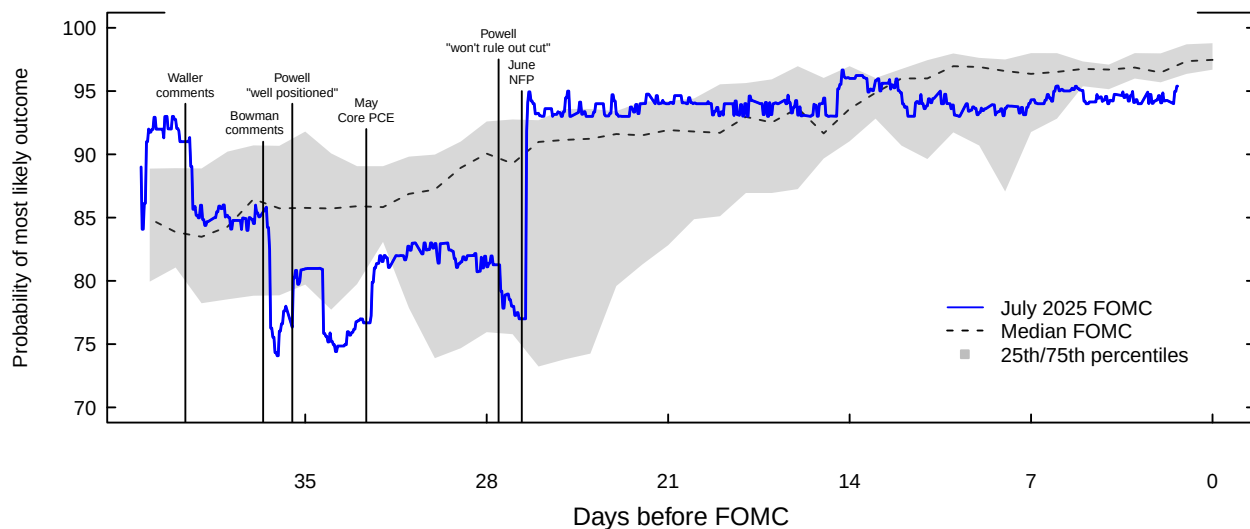


Figure 6: **Probability of Most Likely Rate Outcome Before July 2025 FOMC**

The chart shows the probability of the most likely outcome for the July 2025 FOMC meeting (blue line), in this case for no rate change. The rest of the historical distribution in terms of the 25th to 75th percentile is shown in gray, with the dashed black line reflecting the median.

We can see a similar issue for the October 2025 FOMC meeting shown in Figure 7. The Kalshi market implies there is nonzero weight on seven different outcomes. When we instead use Fed Funds futures, we have to make an assumption of just two possible outcomes, which clearly implies too little uncertainty.

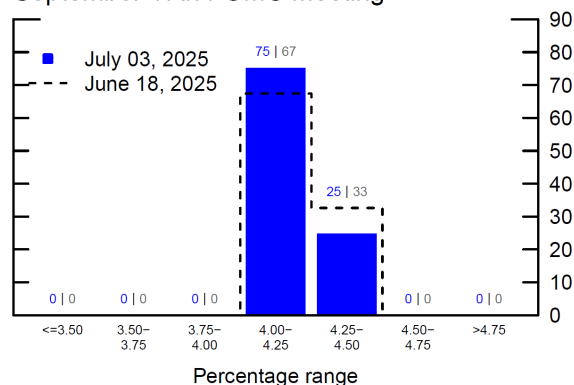
Kalshi vs SOFR Options. While the Fed Funds futures can be quite restrictive in formulating probabilities, a suitable alternative is one based on SOFR options. SOFR options have been around since January 2020. One issue with SOFR options is that the contracts that are liquid are for three month periods. This makes them difficult to interpret, as they average over multiple FOMC meetings. In contrast, the Kalshi distribution is specific to each meeting.

Another issue is that SOFR contracts settle to the repo rate which has a spread with respect to the effective federal funds rate. This spread has grown over time and is expected to continue to grow amid balance sheet runoff by the Federal Reserve. A several basis point difference may seem small, but this can dramatically affect probabilities.

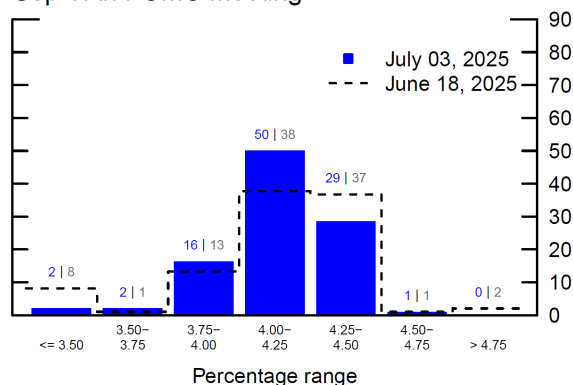
As an illustration, Figure A.12 in the Appendix compares SOFR option-implied and Kalshi density for 2025Q4 as of two dates in the summer of that year and argues that Kalshi gives the better density for the federal funds rate.

Another key distinction between the SOFR market and Kalshi is related to the investors in each market. Kalshi tends to attract the retail investor with its relatively low entry cost.

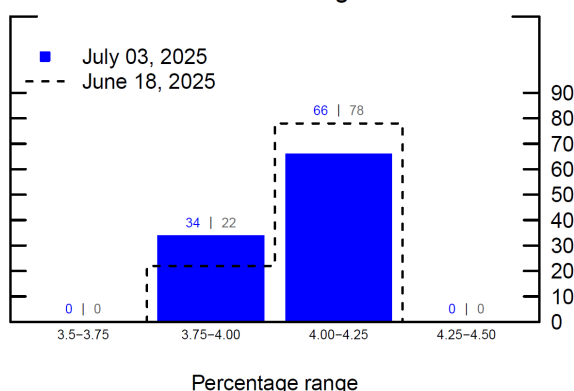
FF Futures–Implied Probability Distribution for September 17th FOMC Meeting



Kalshi–Implied Probability Distribution for Sep 17th FOMC Meeting



FF Futures–Implied Probability Distribution for October 29th FOMC Meeting



Kalshi–Implied Probability Distribution for Oct 29th FOMC Meeting

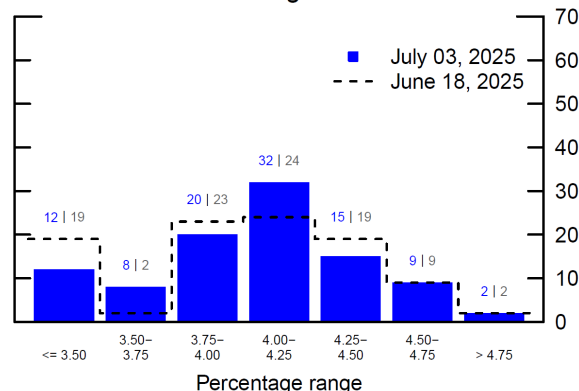


Figure 7: Fed Funds Futures vs Kalshi Sept. and Oct. 2025 FOMC Meetings

The upper left panel shows the implied probability from Fed Funds futures for the September FOMC meeting. We use the price of the Fed Funds futures contract for October to determine the probabilities which are assumed to have just two possible outcomes. The upper right panel shows the corresponding Kalshi distribution that does not enforce the assumption of just two outcomes. The lower left panel shows the implied probability from Fed Funds futures for the October FOMC meeting. The lower right panel shows the equivalent Kalshi distribution.

In contrast, the SOFR market tends to have larger institutional investors. It seems plausible that large institutional investors may have greater hedging needs and this could also play a role in the observed upward bias for this period.

The idea is that institutional firms may have large fixed income exposure and may wish to hedge against a scenario where rates remain elevated or where bond prices may fall. This would push up the demand for options for the no rate change outcome while pushing up the prices of these options, which would increase the implied probability of this outcome. But this would be more indicative of risk premiums rather than expectations. Indeed, survey-based evidence coming from the SME (not shown) also suggests greater weight on rate cuts than implied by SOFR options, which is consistent with Kalshi prices avoiding risk premia from hedging against rising rates.

2025 Inflation and GDP Growth Distributions. As previously noted, a unique aspect of Kalshi is the availability of real-time GDP growth distributions. Figure 8 shows the Kalshi probability distribution for CPI and real GDP Growth over 2025. Interestingly, it suggests that investors were placing greater weight on outcomes for growth below 1% and for CPI inflation to be above 3.5% on July 3rd, 2025 when compared to the beginning of the year. These moves seem consistent with market commentary highlighting the potential effects of trade policy developments over 2025.

The probability of growth being below 1% was close to 0.4 in July 2025. The July Blue Chip Economic Indicators consensus forecast for 2025 GDP growth was 1.4, roughly consistent with the median of the Kalshi distribution. The consensus forecast for 2025 CPI from the Blue Chip survey was 2.9, also close to the median of the Kalshi CPI distribution. However, a key difference between Kalshi and the Blue Chip surveys is the availability of the probability distribution rather than just a single point estimate. The Kalshi distribution shows how uncertain investors were about inflation and GDP growth for this year. While the Survey of Professional Forecasters also provides distributions for GDP growth and inflation, they are only available at a quarterly frequency. The advantage of high frequency data allows for analyzing shifts in the distribution of GDP outcomes after large news events.

Stagflation risks. Options-implied densities nearly always give us marginal distributions on one variable. A very unique feature of Kalshi is that some of their options are in fact joint probabilities. In particular, Kalshi trades options on unemployment being at least 5% *and* 12-month headline inflation being at least 3.5% at the end of 2026, which can be called a stagflation scenario. There are also options on “soft landing” (unemployment and inflation no higher than 5% and 3.5% respectively), “overheating” (unemployment no higher than 5%

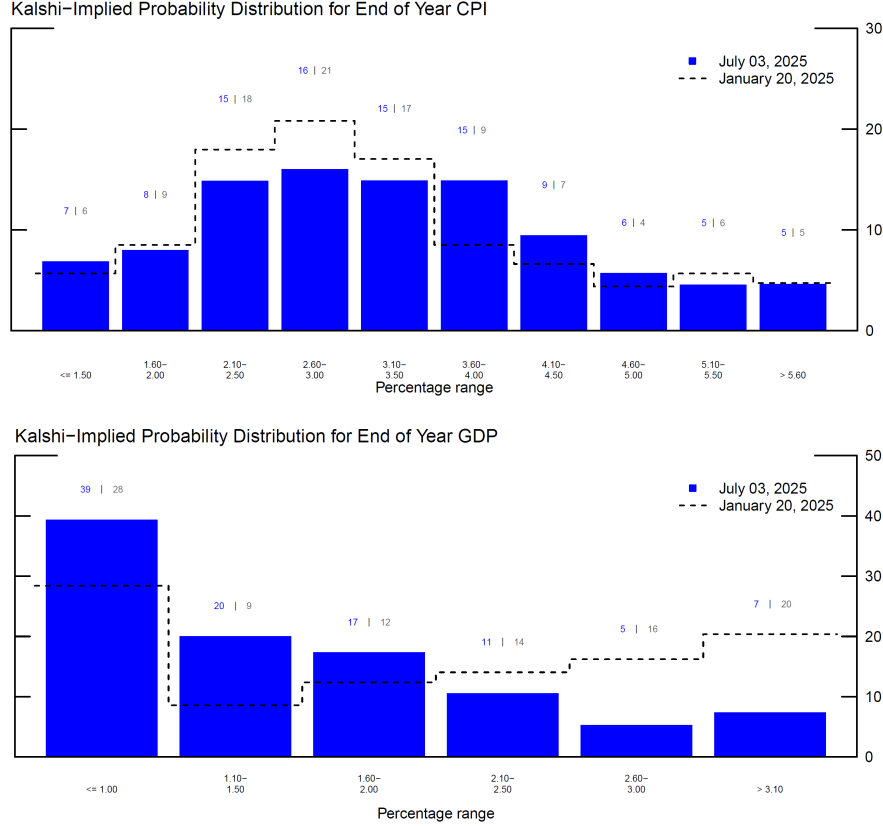


Figure 8: **End of Year 2025 Distributions for CPI and Real GDP Growth**

This chart shows the probability distributions for CPI and real GDP growth for 2025. These are calculated as December 2024 to December 2025 for CPI, and 2024Q4 to 2025Q4 for real GDP growth.

but inflation at least 3.5%) and “slack” (unemployment at least 5% but inflation no higher than 3.5%). These four options are designed so that exactly one will pay off.

Figure 9 shows the evolution of the probabilities of stagflation, soft landing and overheating scenarios in 2026 year to date (slack scenario is omitted; the probabilities sum to about 1). In April and May, amid ongoing geopolitical tensions in Iran, the odds of stagflation and overheating scenarios rose. More benign inflation data subsequently led the odds of stagflation to decline. At the end of the sample, the options-implied probability of a soft landing for 2026 is around 60 percent.

6 Forecast Performance of Kalshi

In this section, we explore the forecast performance of Kalshi forecasts of the Fed Funds rate for each FOMC meeting, year-over-year headline and core CPI inflation, and the unemployment rate for each month since 2022.

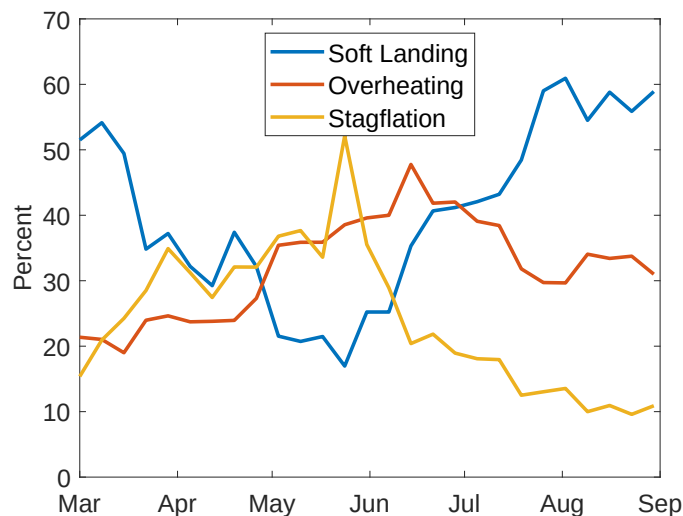


Figure 9: **End of Year 2026 Joint Risks**

This chart shows the weekly-average Kalshi-implied probabilities of the stagflation (unemployment ≥ 5 , inflation ≥ 3.5), overheating (unemployment < 5 , inflation ≥ 3.5) and soft landing (unemployment < 5 , inflation < 3.5) scenarios for 2026 over time. Unemployment is defined as the December 2026 rate; inflation is headline CPI in the 12 months to December 2026.

Federal Funds Rate. The Kalshi distribution provides a mean, median, and modal forecast based on its distribution. A natural question relates to how well these elements forecast the federal funds rate after each FOMC meeting.

Earlier, in Figure 1, we showed the mean absolute prediction error, averaging across all FOMC meetings, plotted against the number of days before the FOMC meeting. That figure also shows some comparisons. The forecast performance of Kalshi is roughly consistent with professional forecasters such as those from FRBNY’s SME. We plot one-half standard deviations of forecast errors for the survey and one can see overlap with the Kalshi-implied moments. We also plot mean absolute forecast errors coming from Fed Funds futures (purple line), and note the slight improvement with about 60 days to go for Kalshi. Also of note, the median and mode coming from Kalshi’s distribution has had a zero absolute average error on the day before the FOMC, which is in contrast to the survey and Fed Funds futures. The observation that drives this distinction was the September 2024 FOMC, in which there was substantial probability on both a 25 and 50 basis point cut, and Kalshi put greater weight on the 50 basis points that turned out to be correct.

CPI. In Figure 10, we see a similar decline in the average absolute error for CPI annualized inflation forecasts as we get closer to the release date. In contrast to Kalshi, we only observe the Bloomberg Consensus forecast just before the data release and thus cannot evaluate its

evolution over time. Thus, the Kalshi can provide a read of expectations for these monthly releases well in advance of the Bloomberg consensus.

For headline CPI, the mean absolute error from Kalshi as of the day of the release is about 7 basis points, while the Bloomberg consensus is about 8 basis points.

Unemployment. Kalshi also provides real-time distributions for expectations of the unemployment rate. As with real GDP growth, options markets do not exist for unemployment, which makes Kalshi unique in terms of providing a distribution for unemployment expectations. Figure 10 shows that the average absolute error for the mean coming from Kalshi is fairly close to the Bloomberg consensus forecast, while the mode tends to jump around as it must take values in increments of 0.1.

Significance of differences in forecast accuracy Table 3 compares the mean absolute error and the root mean square error on the day before the release for the Kalshi mean, median and mode forecasts with that from the Bloomberg consensus. Statistical significance from Diebold-Mariano statistics (Diebold and Mariano, 1995) comparing each of the Kalshi forecasts to the Bloomberg consensus is also noted. In some cases, the Kalshi forecast is a significant improvement over the Bloomberg consensus. For example, the Kalshi median and mode have significantly smaller mean absolute error than the Bloomberg consensus for headline CPI. In other cases, there is no statistically significant difference between Kalshi and Bloomberg forecast accuracy. In no case is Kalshi significantly worse than the Bloomberg consensus. The last panel of the table compares the Kalshi density forecast with the density forecast from Fed Funds futures, assuming that only two options have any mass. Here, the measure of forecast accuracy is the predictive likelihood⁴. The predictive likelihood is higher for the Kalshi forecast, indicating that it does better. However, the difference is not statistically significant using the test of Amisano and Giacomini (2007).

6.1 Probability Integral Transform

As noted earlier, the Kalshi-implied density is under the risk-neutral $\mathbf{Q}$ measure which, given risk premia, is not necessarily the same as the physical $\mathbf{P}$ measure. The extent to which Kalshi-implied pdfs are distorted by risk premia is an empirical question. If the Kalshi-implied density were indeed the physical density of the future variable, then it is well known that the cumulative distribution function (cdf), evaluated at the realized value, should be uniform on the unit interval (see, for example, Diebold et al. (1998)). We assess this in

⁴If f_t is a forecast density at time t , and y_t is the ex-post realized value then the predictive likelihood is defined as $T^{-1} \sum \ln f_t(y_t)$.

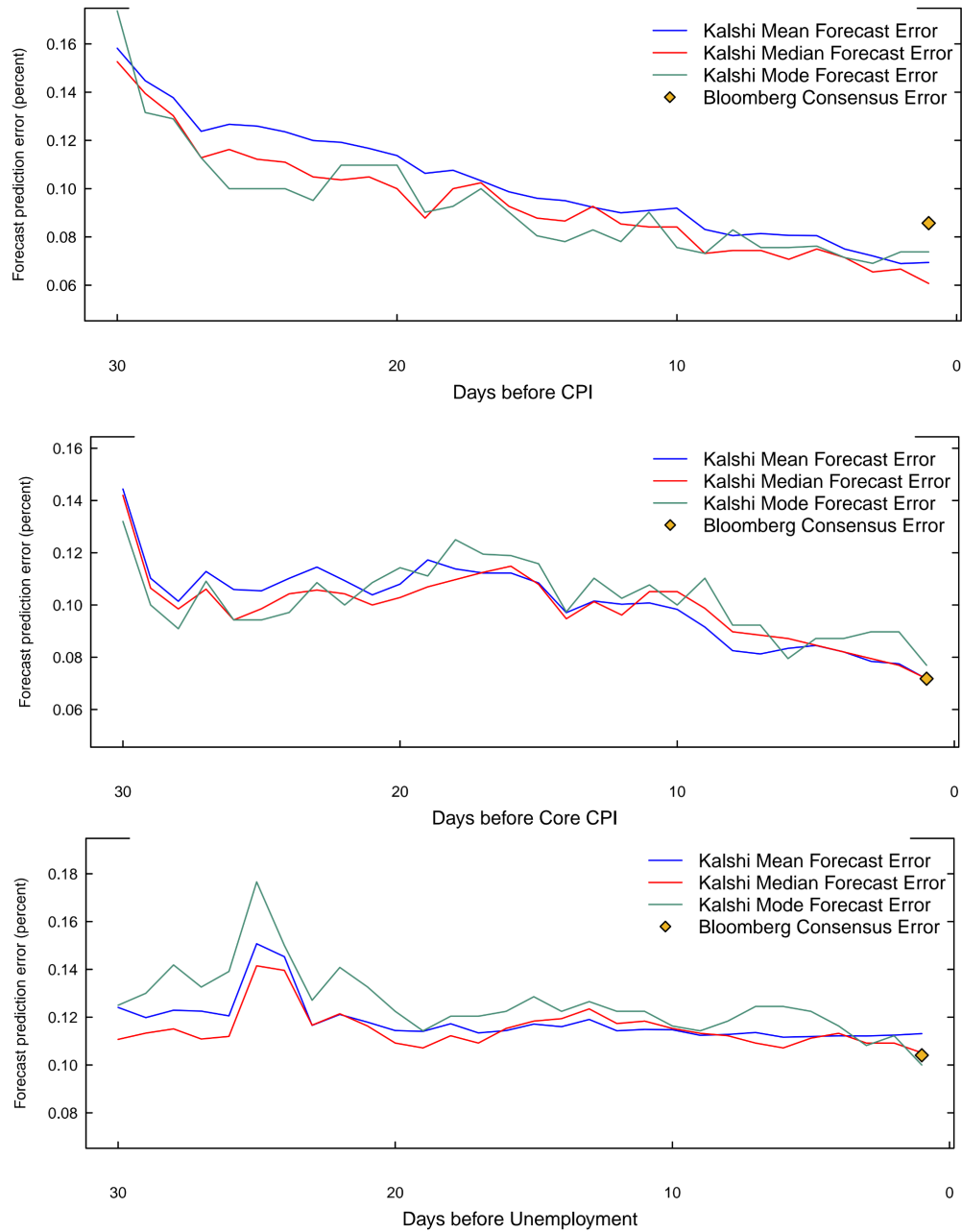


Figure 10: **Headline, Core CPI and Unemployment Forecast Errors Comparison**

This figure compares the mean absolute errors from 50 days out to each headline CPI, core CPI, and unemployment release since 2022.

Panel A: Macro Variables				
	Bloomberg	Kalshi Mean	Kalshi Median	Kalshi Mode
Headline CPI				
MAE	0.081	0.069	0.063[*]	0.063[*]
RMSE	0.100	0.080^{**}	0.083	0.087
Core CPI				
MAE	0.070	0.070	0.080	0.070
RMSE	0.100	0.090	0.110	0.100
Unemployment Rate				
MAE	0.109	0.117	0.107	0.107
RMSE	0.132	0.136	0.132	0.135

Panel B: Federal Funds Rate				
	FF Futures	Kalshi Mean	Kalshi Median	Kalshi Mode
MAE	0.010	0.010	0.000^{**}	0.000^{**}
RMSE	0.020	0.030	0.000^{**}	0.000^{**}

Panel C: Federal Funds Rate Density Forecast		
	FF Futures	Kalshi
Predictive Likelihood	-0.0982	-0.0852

Table 3: **Forecast Accuracy: Benchmark Forecasts versus Kalshi**

Panel A reports the mean absolute error (MAE) and root mean square error (RMSE) of various Kalshi point forecasts and the Bloomberg consensus forecast for CPI (headline and core) and unemployment on the day before the release. Panel B reports forecast accuracy relative to fed funds futures on the day before the FOMC. Panel C reports the predictive likelihoods for the Kalshi density forecast and that implied by the fed funds futures, assuming that there are just two possible options in the latter case. Stars indicate statistically significant differences in forecast accuracy relative to the benchmark based on [Diebold and Mariano \(1995\)](#) and [Amisano and Giacomini \(2007\)](#) tests: ^{*} $p < 0.10$, ^{**} $p < 0.05$, ^{***} $p < 0.01$.

Figure 11 for forecasts for monthly releases of unemployment, CPI (headline and core) and FOMC decisions on the target federal funds rate. These are constructed both 28 days before the release and on the day before the release.⁵ The probability integral transforms generally seem close to uniform. We can test the hypothesis that they are uniform in population using the test statistics:

$$K = \sup_{0 \leq r \leq 1} \Psi(r) \quad (1)$$

and

$$C = \int_0^1 \Psi(r)^2 dr. \quad (2)$$

where $\hat{F}(r)$ is the empirical cdf of the PIT and $\Psi(r) = T^{-1/2}(\hat{F}(r) - r)$, $0 \leq r \leq 1$ (Darling, 1957). The bootstrap p -values using the bootstrap of (Rossi and Sekhposyan, 2019) are also shown in Figure 11. For unemployment, uniformity is borderline rejected. In other cases, the null of uniformity is not rejected. Overall these probability density functions appear to be well calibrated.

A different but related way of looking at the same question is to work out the historical probability of contracts expiring in the money for a given event type and number of days to maturity, for contracts with prices in a particular bin. And then we can plot these probabilities against the midpoint of the bins. These are shown in Figure 11. If the densities are well calibrated, then this should be the 45 degree line. In prediction markets, a graph like this often exhibits favorite-longshot bias (Griffith, 1949; Snowberg and Wolfers, 2010; Snowberg, Wolfers and Zitzewitz, 2013) whereby low likelihood events are overpriced and high likelihood events are underpriced. However, there is no significant deviation from the 45 degree line here, indicating that favorite-longshot bias is not a major concern for these contracts. This was also found by Gürkaynak and Wolfers (2005) in the earlier Economic Derivatives market.

The finding that risk premia do not appear to significantly distort the densities is perhaps surprising given that other financial market implied densities are well-known to be affected by risk premia. We conjecture that, for now, Kalshi markets are not being heavily used for hedging purposes and that may make them less vulnerable to distortions from risk premia.

⁵For the target funds rate, the Kalshi density nearly always puts almost all mass on one outcome that turns out to be the realized outcome. This poses a problem for defining the cdf evaluated at the realized value. Suppose, for example, that there is 96% probability of a particular outcome and 4% probability of a rate 25 basis points higher, and the outcome with 96% odds is realized. What we do then is to define the realized value of the cdf as being a uniform random number from 0 to 0.96.

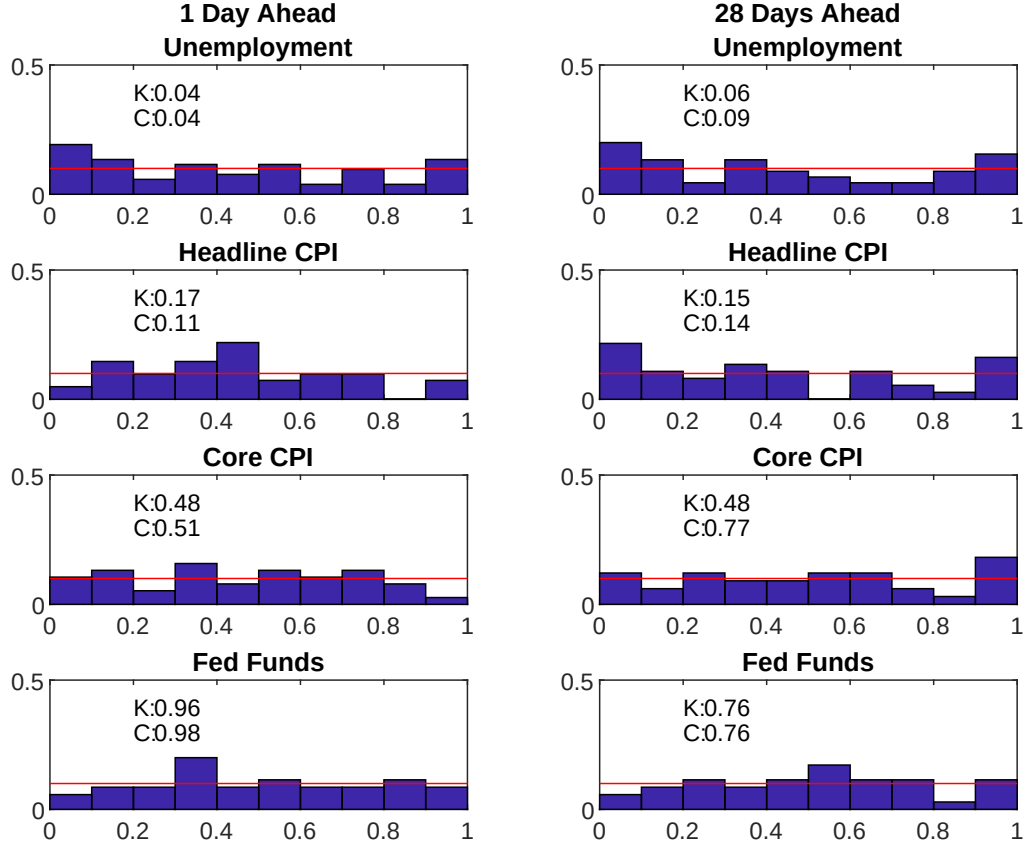


Figure 11: **Probability Integral Transform**

This figure shows the empirical cumulative distribution function as of 1 and 28 days before unemployment, CPI (headline and core) and FOMC scheduled federal funds rate announcements, all evaluated at the realized value. If the options-implied distribution were equal to the physical distribution, then this probability integral transform would be uniform in population, as shown by the lines. Each panel also reports the p -values from the test statistics k and C , defined in the text, using the bootstrap of [Rossi and Sekhposyan \(2019\)](#).

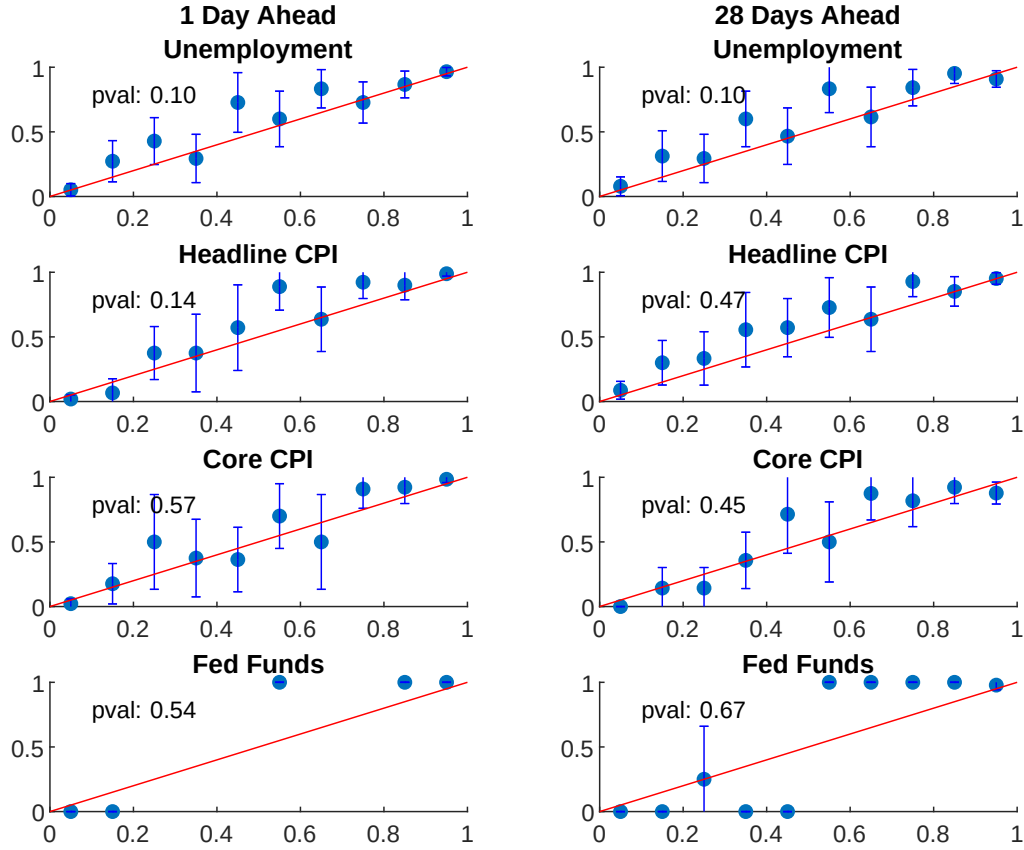


Figure 12: Plot of Probabilities against Prices

This figure shows the historical proportion of options expiring in the money in different price bins against the midpoint of the bins. Results are shown as of 1 and 28 days before unemployment, CPI (headline and core) and FOMC scheduled federal funds rate announcements. 90% confidence intervals are also included. These are based on the asymptotic variance of a binomial random variable and so collapse to a single point if the sample probability for a price bin is either exactly 0 or exactly 1. If the options prices were perfectly calibrated, they would line up on the 45 degree line. Each panel also shows the bootstrap p value testing the significance of the deviation from this line.

6.2 Why do Kalshi forecasts work?

It may seem surprising that Kalshi forecasts on economic outcomes can match or outperform traditional survey-based measures such as Bloomberg or the SME. A natural question is ‘why?’. One potential option is “wisdom of the crowd”—the idea that, since Kalshi has many more participants in a given market than respondents to the Bloomberg consensus survey, the law of large numbers is stronger, and thus forecast performance is superior. Another is “skin in the game”—that since money is at stake when forecasting in the market, forecaster incentives are better aligned towards accuracy rather than some other objective that survey respondents might have. We examine the claims of these mechanisms, and show that “skin in the game” is more important for forecast performance than “wisdom of the crowd”. That is, while increasing volume is largely unassociated with increased forecast performance, Kalshi does appear to suffer from less behavioral biases than surveys, suggesting the importance of financial stake in forecast performance.

First, we test the relationship between market volume and forecast from the model. We estimate the equation

$$|\varepsilon_{t,\text{kalshi}}| = \alpha + \beta \log(\text{volume}_t) + \gamma |\varepsilon_{t,\text{Survey}}| + u_t \quad (3)$$

where $\varepsilon_{t,\text{kalshi}}$ is the Kalshi forecast error, volume_t is the cumulative volume traded on the news event in hundred-thousands, and $\varepsilon_{t,\text{Survey}}$ is the corresponding survey’s error on the same event. The null hypothesis is that liquidity is unimportant for gains in forecast ability, or equivalently that $\beta = 0$. We control for the absolute error of the survey because events that are more unpredictable may be seen as more enticing to traders and attract more volume. The survey absolute error acts as a control for the “difficulty” of each individual macro forecast. Panel D of Table 4 shows that the addition of volume does little to improve forecast performance for our Kalshi markets of interest. While the estimated coefficient on the log of volume is generally negative, the effect is small (e.g. a 100% increase in YoY CPI volume leads to on average a 0.16 percentage point *larger* absolute error, and a 0.01 percentage point smaller absolute error for Core CPI) and statistically insignificant. Intuitively, the small effect of additional volume may be a result of the law of large numbers already biting—the baseline volume is quite high and reflects many market participants. As a result, forecast errors on macro news are not driven by a lack of forecasters, but by shared misconceptions about what event will occur, individual agents’ forecasts are not fully independent.

Now, we aim to understand whether our Kalshi forecasts for CPI and unemployment exhibit behavioral biases that often get attributed to survey responses. For example, [Coibion](#)

Table 4: Forecast Error Diagnostics

Measure	CPI	Core CPI	Unemp.	FFR (3 out)	FFR (2 out)
Panel A: Bias					
Mean	−0.02 (0.01)	−0.01 (0.01)	−0.03 (0.02)	0.12* (0.06)	0.03 (0.02)
Median	−0.01 (0.01)	−0.01 (0.01)	−0.01 (0.02)	0.10 (0.06)	−0.00 (0.03)
BB Consensus	−0.01 (0.02)	−0.01 (0.01)	0.00 (0.02)	—	—
SME Mean	—	—	—	0.09 (0.07)	0.02 (0.03)
SME Median	—	—	—	0.10* (0.06)	0.03 (0.03)
Panel B: Serial Correlation					
Mean	0.26 (0.15)	0.22 (0.16)	−0.18 (0.15)	−0.21 (0.14)	0.10 (0.16)
Median	0.19 (0.16)	0.10 (0.18)	−0.11 (0.16)	−0.27* (0.15)	0.27** (0.12)
BB Consensus	0.32** (0.15)	0.14 (0.17)	−0.02 (0.17)	—	—
SME Mean	—	—	—	−0.11 (0.15)	0.23 (0.19)
SME Median	—	—	—	−0.10 (0.12)	0.33* (0.18)
Panel C: Coibion–Gorodnichenko Predictable Errors					
Mean	0.95 (0.72)	0.64 (0.42)	−1.37 (0.95)	0.31 (0.19)	0.24** (0.10)
Median	−0.55 (0.40)	0.01 (0.26)	−0.57 (0.50)	0.15 (0.12)	0.07 (0.12)
SME Mean	—	—	—	0.52* (0.26)	0.45*** (0.08)
SME Median	—	—	—	0.24 (0.23)	0.10 (0.08)
Panel D: Cumulative Volume					
Mean	0.16 (0.27)	−0.01 (0.43)	−0.29 (0.30)	−0.95 (1.63)	−0.22 (0.66)
Median	0.30 (0.39)	−0.05 (0.73)	−0.22 (0.43)	−2.90 (1.92)	0.49 (1.42)

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns 1-3 use HC3 robust standard errors, while 4-5 use Newey-West standard errors. FFR3 and FFR2 report Federal Funds Rate tests at horizons 3 and 2 surveys out, respectively. In Panel C, FFR3 uses the horizon 4-to-3 forecast revision, and FFR2 uses the horizon 3-to-2 forecast revision. In Panel D, FFR3 uses cumulative volume until the date of the 3-meeting-out SME, and FFR2 uses cumulative volume until the date of the 2-meeting-out SME. The units in Panel D show the effects of a 1 log-unit (> 100 percentage point) increase in volume on the absolute forecast error in percentage points.

and Gorodnichenko (2015) tests whether SPF forecasts of (annual) inflation behave like full-information rational expectations (FIRE) forecasters, finds that *ex ante* forecast revisions predict *ex post* forecast errors, and argues the errors are caused by information rigidities. Bordalo et al. (2020) add to these results studying SPF and Blue Chip forecast biases for 20 macroeconomic series, finding that survey forecast revisions positively predict future forecast errors and survey averages persistently *underreact* to news. Unlike previous work, we analyze individual CPI and unemployment announcements, not quarterly or annual aggregates. We are interested in testing how Kalshi-implied forecasts compare to surveys in adhering to how one would expect a rational expectations agent to behave.

To formally test this, we aim to see if any information at time $t - 1$ can predict a time t forecast error. We estimate:

$$\varepsilon_{t,f} = \alpha + u_t \tag{4}$$

$$\varepsilon_{t,f} = \alpha + \beta \varepsilon_{t-1,f} + u_t \tag{5}$$

$$\varepsilon_{t,f} = \alpha + \gamma h_{t,t-5} + u_t \tag{6}$$

where $\varepsilon_{t,f}$ is the forecast error at time t for the forecast method $f \in \{\text{Kalshi, Survey}\}$ ⁶. Equation (4) is essentially a difference of means test for the realization and forecast. We report the coefficient and standard errors on the parameter of interest α from this regression in Panel A of Table 4. For CPI and unemployment none of the forecasts shows any significant bias. For the Federal Funds rate (FFR), both Kalshi and the SME consistently estimated the rate 3 meetings out too low by just under 12.5 bps during our sample. However, both coefficients are very similar (exactly the same, for the medians) and the result may be an artifact of the sample period rather than a difference in methodologies.

Our second test of FIRE, equation (5) asks whether we can forecast time t errors using time $t - 1$ errors, which would indicate forecasters are over- or under-correcting their error in the previous period. Here, the null hypothesis of interest is $\beta = 0$, which would indicate that previous period errors do not predict future period errors. We find that for headline CPI, we cannot reject the null for Kalshi errors, but we can reject it for the Bloomberg Consensus. A 1 percentage point underforecast of YoY CPI from the Bloomberg consensus in month $t - 1$ would predict the Bloomberg Consensus CPI estimate in month t will underforecast by 0.32 percentage points. For core CPI and unemployment, we find that both the Bloomberg consensus and Kalshi cannot reject the null of FIRE. For FFR, we find that, while means are

⁶In all specifications, we winsorize data at 5% and 95% as a result of a few outlier news releases leading to large errors that otherwise dominate results.

not statistically significant, Kalshi persistently *overreacts* to errors in its previous median forecast 3 surveys out, and *underreacts*, along with the SME, 2 surveys out. At both horizons, the coefficients are larger for the SME than Kalshi, reflecting that survey respondents are more hesitant to update their forecasts in response to news, leading to larger predictable forecast errors. It is worth noting that in *every* event besides Kalshi’s mean Core CPI forecast, the surveys are more sluggish to respond (the coefficient is larger) to past errors than the market, a key differentiator between survey- and market-based measures.

Our third test of FIRE, equation (6), is similar to the test of [Coibion and Gorodnichenko \(2015\)](#), which asks whether forecast errors are predictable by previous forecast revisions. We test this using changes in the Kalshi-implied forecast between days $t - 5$ and t , and a rejection of FIRE would manifest in $\gamma \neq 0$. We cannot run this test for the Bloomberg survey, as we do not observe any survey revisions for Bloomberg Consensus for our markets of interest. For FFR, we use movements in the Kalshi-implied and SME forecasts between SME due dates, which are spaced approximately 6 weeks apart. Panel C of Table 4 shows our results for γ using different Kalshi-implied forecasts. For CPI, core CPI and unemployment, we find no significant predictability of errors under the test. For the FFR 3 meetings out, we cannot reject FIRE using the Kalshi mean, and the coefficient estimate is smaller than the corresponding coefficient for the PDS, which we can reject FIRE at 90% confidence. For 2 meetings out, we reject FIRE for both the Kalshi and SME mean, again with Kalshi sluggishness of updating having a coefficient just over half that of the SME. Overall, these results indicate that Kalshi-implied forecasts are more resistant than surveys to sluggishness and other biases that lead to suboptimal forecasts. This is intuitive: survey respondents might be balancing incentives such as consistency or stability in their forecasts to appease their audiences, but the financial stake in a market-implied forecast aligns incentives and can lead to less predictable errors, a possibility that we expand upon in the next section.

7 A signal-extraction model of forecast performance

To better understand the mechanisms underlying why prediction markets sometimes perform better than traditional expert forecasts, we expand upon the signal extraction model presented in [Morris and Shin \(2002\)](#) (henceforth MS). MS develops a model where there is a public signal, observed by everyone, and individual private signals. Agents strategically report a prediction, given the public and their individual private signal, to maximize their conditional expected utility which includes terms for accuracy but also conformation to a consensus. We expand on this model by considering two different types of agents: (1) forecasters which face a reputational cost that enacts a herding pressure and (2) market traders

who face no such reputational cost. Then, we compare the accuracy of the aggregated forecasts among different agent types. Forecasters report individual welfare-maximizing beliefs subject to behavioral bias, while market traders buy an Arrow-Debreu security to maximize expected utility.

To preview our results, we find that in the absence of reputational concerns, the forecaster consensus performs better than the market consensus at predicting the outcome of the economic variable. However, as forecasters weight reputation more highly, the forecaster consensus is beaten by the market consensus. Increases in number of market participants—“wisdom of the crowd” is directionally helpful, but does not guarantee market consensus to beat expert forecasters. In contrast, increases in the share, not the level, of confident or accurate market participants is most helpful for improving market prediction accuracy.

7.1 Basic Model

Similar to MS, but adapted to better align with prediction markets’ microstructure, there is an underlying state of nature represented by the random variable $S \in \{0, 1\}$, either the event S happens ($S = 1$) or does not ($S = 0$). Let $y = P\{S = 1\}$, the true probability of the event.

Forecasters act by reporting their guess of y , subject to behavioral biases, while traders act by determining an equilibrium price p for the Arrow-Debreu security that pays 1 unit of consumption in the event $S = 1$. We derive a set of criteria for when surveys and markets deliver superior forecasting performance.

All agents have public and private information on the likelihood S occurs, a market price and a survey-implied probability are formed, and then S is realized.

Signal structure

Agents receive signals that are noisy, but centered around the true probability y . Agents consider the signals as information on the true probability y , and therefore we must ensure that our noise is constructed in a way such that the signal always lands in the unit interval. As a result, we construct noise by centering a symmetric Beta distribution around 0, and scale the distribution so that the largest realization of noise still produces a signal within $[0, 1]$.

To this end, for $\sigma \in (0, \frac{1}{2})$ we define:

$$\alpha(\sigma) \equiv \frac{1}{8\sigma^2} - \frac{1}{2} > 0$$

Every agent observes the same public signal:

$$s = y + m(y)u_\eta \quad u_\eta \sim \text{Beta}(\alpha(\sigma_\eta), \alpha(\sigma_\eta)) - \frac{1}{2} \quad (7)$$

where

$$m(y) \equiv 2 \min\{y, 1 - y\}.$$

The noise u_η has mean zero, support $(-\frac{1}{2}, \frac{1}{2})$ and variance σ_η^2 . As a result, after scaling by $m(y)$ we escape the problem of the signal falling outside $[0, 1]$. Scaling by $m(y)$ ensures that the noise is smallest near the tails and larger near the center.

Each agent of type τ (which indexes forecaster or trader) also observes a unique-to-them private signal:

$$x_i = y + m(y)u_{\varepsilon,i,\tau} \quad u_{\varepsilon,i,\tau} \sim \text{Beta}(\alpha(\sigma_{\varepsilon,\tau}), \alpha(\sigma_{\varepsilon,\tau})) - \frac{1}{2} \quad (8)$$

drawn independently across agents and independently of u_η , conditional on y . Again this ensures that the signal is in $[0, 1]$. This structure means that for any underlying true probability y , our signals are unbiased and have chosen variances, scaled by $m(y)$:

$$\begin{aligned} \mathbb{E}[s|y] &= \mathbb{E}[x_i|y] = y \\ \text{Var}[s|y] &= m(y)^2 \sigma_\eta^2 \\ \text{Var}[x_i|y] &= m(y)^2 \sigma_{\varepsilon,\tau}^2 \end{aligned}$$

The variance of all the signals scale together with $m(y)$, allowing us to compare forecaster and trader predictions without any dependence on the underlying y . To reiterate, this information problem has deviated from MS only in that there are two types of agents, indexed by τ , with different variances on their private signals. Specifically, MS results on optimal actions under this information problem accomodate any mean-zero noise, which our constructed Beta distributions satisfy. We further assume $\sigma_{\varepsilon,\text{forecaster}} < \sigma_{\varepsilon,\text{trader}}$, interpreted as the expert forecaster having a less noisy private signal than the trader. This can be interpreted as better research, richer information to analyze, or any other way the expert forecaster may have a less noisy private signal. Even giving the forecaster this advantage, under reasonable conditions we will show the traders' aggregate prediction can outperform the forecasters' aggregate prediction.

Now, given this information, forecasters and market participants form their beliefs about the likelihood of $S = 1$ occuring. Then, we elicit the beliefs of forecasters, who report

a likelihood of $S = 1$ to the analyst, and the market participants, who trade an Arrow-Debreu security which pays out if $S = 1$. We compare the accuracy of our forecaster and market-based estimates.

7.2 Forecaster problem:

Forecasters maximize the utility function:

$$u_i(\mathbf{a}, S) \equiv -(1-r)(a_i - S)^2 - r(L_i - \bar{L})$$

where

$$L_i \equiv \int_0^1 (a_j - a_i)^2 dj$$

$$\bar{L} \equiv \int_0^1 L_j dj$$

and $0 \leq r_\tau < 1$. This utility function is used in MS. The first term in the utility function is a standard prediction loss function, the more accurate the forecast, the higher the agent's utility. The second term is the “beauty contest” term, where agents want to report predictions close to the predictions of others. This can be internalized as anchoring, herding, or some other reputational incentive that incentivizes forecasters to report guesses close to other forecasters. The parameter r governs the relative weights on these two motives: if forecasters have $r = 0$, they do not consider the actions of others when making a prediction, and only maximize their expected prediction accuracy. If forecasters have $r \rightarrow 1$, they care only about matching the consensus.

Forecasters take their signals and utility function, and report their prediction a_i in order to maximize expected utility conditional on their information set:

$$a_i = \operatorname{argmax}_{a_i} \mathbb{E}[u_i | s, x_i]$$

MS shows that the unique Nash equilibrium to this game is to forecast:

$$a_{i,\text{forecaster}}(s, x_i) = \frac{\frac{s}{\sigma_\eta^2} + (1-r)\frac{x_i}{\sigma_{\varepsilon,i,\text{forecaster}}^2}}{\frac{1}{\sigma_\eta^2} + (1-r)\frac{1}{\sigma_{\varepsilon,i,\text{forecaster}}^2}} := \frac{sA + (1-r)x_i B_{\text{forecaster}}}{A + (1-r)B_{\text{forecaster}}} \quad (9)$$

where $A = \frac{1}{\sigma_\eta^2}$; $B_{\text{forecaster}} = \frac{1}{\sigma_{\varepsilon,i,\text{forecaster}}^2}$. The optimal action is determined by the respective noisiness of the public and private signals. When σ_η^2 ($\sigma_{\varepsilon,i,\text{forecaster}}^2$) is large, A ($B_{\text{forecaster}}$) is small, and we weight the public (private) signal less. Here the role of the parameter r is clear— it acts to *underweight* the private signal, because the agent knows that the consensus

won't have this information to act on, and doesn't want to deviate far away.

We define an estimator for the event $S = 1$ based on the simple average of the forecaster predictions:

$$\hat{y}_{\text{forecaster}} = \frac{1}{N_{\text{forecaster}}} \sum_{I(\text{forecaster})} \frac{sA + (1-r)x_i B_{\text{forecaster}}}{A + (1-r)B_{\text{forecaster}}}$$

where $N_{\text{forecaster}}$ is the number of forecasters. Finally, we can calculate analytically the finite-sample MSE for the forecaster-based estimator:

$$\begin{aligned} \mathbb{E}[\hat{y}_{\text{forecaster}}] &= \mathbb{E} \left[\frac{1}{N_{\text{forecaster}}} \sum_{I(\text{forecaster})} \frac{sA + (1-r)x_i B_{\text{forecaster}}}{A + (1-r)B_{\text{forecaster}}} \right] \\ &= \frac{1}{A + (1-r)B_{\text{forecaster}}} \mathbb{E}[sA + (1-r)B_{\text{forecaster}}\bar{x}] \\ &= P\{S = 1\} \quad (\text{by } x_i, s \text{ unbiased for } P\{S=1\}) \\ \text{Var}[\hat{y}_{\text{forecaster}}] &= m(y)^2 \left[\frac{1}{A + (1-r)B_{\text{forecaster}}} \right]^2 \left[A + \frac{(1-r)^2 B_{\text{forecaster}}}{N_{\text{forecaster}}} \right] \\ \therefore \text{MSE}[\hat{y}_{\text{forecaster}}] &= \text{Var}[\hat{y}_{\text{forecaster}}] = m(y)^2 \left[\frac{1}{A + (1-r)B_{\text{forecaster}}} \right]^2 \left[A + \frac{(1-r)^2 B_{\text{forecaster}}}{N_{\text{forecaster}}} \right] \end{aligned} \tag{10}$$

The forecast-based aggregate prediction has intuitive MSE form in equation (10). Forecasters report a precision-weighted combination of their public and private signals. Individually, their predictions are unbiased, but the overall MSE does not shrink to 0, even with infinite forecasters. The reason is the shared public signal: since every forecaster reports has beliefs based on the same public information, the A term in the variance calculation does not disappear with $n_{\text{forecaster}}$. Heightened reputational concerns, i.e. a larger r , increase the emphasis on the public signal and therefore increase MSE.

7.3 Trader Problem

Prior to the realization of S , trader i is given an endowment e_i and can buy a quantity q_i of an Arrow-Debreu security, costing p , and paying out 1 unit of consumption if $S = 1$, and 0 otherwise. The trader selects their optimal q_i to maximize expected log utility⁷ according to the trader's beliefs on the likelihood $S = 1$. The Arrow-Debreu security is in zero net supply

⁷We use log utility for simpler analytical solutions, but the core results do not change under more general utility functions.

and markets clear. Formally, the trader's problem is:

$$\max_{q_i} \mathbb{E}[\log(m_i)|P_i] \quad \text{such that} \quad (11)$$

$$m_i = \mathbf{1}\{S = 1\}(e_i + (1 - p)q_i) + \mathbf{1}\{S = 0\}(e_i - pq_i) \quad (12)$$

$$P_i = \operatorname{argmax}_{a_i} \mathbb{E}[-(a_i - S)^2 | s, x_i]. \quad (13)$$

where P_i is the probability that $S = 1$ as perceived by trader i . Equation (11) says that the traders maximize their expected utility according to their beliefs, which in equation (13) form optimally using the totality of the trader's information to minimize squared prediction error. Equation (12) specifies the returns when the asset does and does not pay out.

Notice that P_i solves the special case of the forecaster's problem in Equation (9) where $r = 0$, and so we have that:

$$P_i = \frac{\frac{s}{\sigma_\eta^2} + \frac{x_i}{\sigma_{\varepsilon_i, \text{trader}}^2}}{\frac{1}{\sigma_\eta^2} + \frac{1}{\sigma_{\varepsilon_i, \text{trader}}^2}} := \frac{sA + x_i B_{\text{trader}}}{A + B_{\text{trader}}}$$

where $A = \frac{1}{\sigma_\eta^2}$; $B_{\text{trader}} = \frac{1}{\sigma_{\varepsilon_i, \text{trader}}^2}$.

Then, we can rewrite (11) by breaking up the expectation:

$$\max_{q_i} P_i \log(e_i + (1 - p)q_i) + (1 - P_i) \log(e_i - pq_i)$$

And after taking the first-order condition with respect to q_i and solve for the asset-demand:

$$q_i = \frac{e_i(P_i - p)}{p(1 - p)} \quad (14)$$

Here, we can immediately see the intuition on how traders demand the Arrow-Debreu asset: they buy the asset if and only if they believe $S = 1$ will happen with higher probability than the price p implies, and their demand scales with the size of the mispricing and the trader's endowment. Finally, to solve for the equilibrium price we impose market clearing with N_{trader} traders:

$$\sum_{i=1}^{N_{\text{trader}}} q_i = 0.$$

which, in conjunction with equation (14) implies that:

$$p = \frac{\sum_{i=1}^{N_{\text{trader}}} e_i P_i}{\sum_{i=1}^{N_{\text{trader}}} e_i} \quad (15)$$

Thus, the equilibrium price is an endowment-weighted average of trader's beliefs. Further supposing the special case where $e_i = 1 \forall i$,

$$p = \frac{\sum_{i=1}^{N_{\text{trader}}} P_i}{N_{\text{trader}}}$$

is a raw average of the traders' beliefs. Under this special case, the MSE of the market-implied probability is:

$$\begin{aligned} \mathbb{E}[p] &= \mathbb{E} \left[\frac{1}{N_{\text{trader}}} \sum_{I(\text{trader})} \frac{sA + x_i B_{\text{trader}}}{A + B_{\text{trader}}} \right] \\ &= \frac{1}{A + B_{\text{trader}}} \mathbb{E}[sA + B_{\text{trader}} \bar{x}] \\ &= P\{S = 1\} \quad (\text{by } x_i, s \text{ unbiased}) \\ \text{Var}[p] &= m(y)^2 \left[\frac{1}{A + B_{\text{trader}}} \right]^2 \left[A + \frac{B_{\text{trader}}}{N_{\text{trader}}} \right] \\ \therefore \text{MSE}[p] &= \text{Var}[p] = m(y)^2 \left[\frac{1}{A + B_{\text{trader}}} \right]^2 \left[A + \frac{B_{\text{trader}}}{N_{\text{trader}}} \right] \end{aligned} \quad (16)$$

The MSE for the symmetric endowment case is exactly the same as the forecaster MSE when $r = 0$, except that $B_{\text{forecaster}}$ is replaced by B_{trader} .

By comparing the MSEs as $N_{\text{forecaster}}, N_{\text{trader}} \rightarrow \infty$, the market consensus under equal endowment is more accurate than the forecaster consensus if and only if

$$r > r^* \equiv 1 - \frac{\sigma_{\varepsilon, \text{forecaster}}^2}{\sigma_{\varepsilon, \text{trader}}^2} \in (0, 1).$$

Here, r^* represents the point at which the additional forecaster expertise trades-off with the herding or reputational bias. Forecasters with a private signal half as noisy in variance terms can sustain reputational weight up to $r^* = 1/2$ before the market overtakes them; beyond that, the expertise advantage is fully dissipated by the strategic underweighting of that expertise. This is because, in the limit, the most important thing for accurate forecasts is to have individuals rely on their private information. Reliance on private information, rather than public, causes the law of large numbers to bite harder by having less-correlated

individual beliefs. Increasing r undoes this by pushing people towards public signals so they do not stray far from others.

7.4 Model Discussion

The goal of this model is to allow us to think about the relative forecasting performance of markets relative to surveys, structuring our thinking around how information and biases play into forecasting, and to examine counterfactuals where we subject markets and forecasters to different contexts to examine how their forecasting ability would change. One limitation of our model is that we assume the information embedded in the market price is not used by either the trader or by the forecaster. Our model of signals is not rich enough for agents to consider the price when determining their expectation for the event, and would create a serious complication to the model. If agents all considered prices as signals, we would expect (as the number of traders grows large) for traders to converge fully on believing the market-implied price. Forecasters would also understand that the price contains everyone's private information, and place their forecasts accordingly. However, empirically we see that Kalshi and traditional forecasters often disagree, and variances in Kalshi distributions and forecaster surveys remain sizeable, both pointing to the fact that the distribution of traders and forecasters do not converge on a uniform expectation. As a result, we do not think it is unreasonable to assume that agents take prices as given, and make trades based on their non-price beliefs.

We are reassured by how the predictions of our model match the empirical results in Section 6.2, as the model also concludes that behavioral biases can be more important for model performance than the addition of traders/volume. To see this in the model, note that both estimators are unconditionally unbiased in the sense that both s and x_i are centered around the true realization of $P\{S = 1\}$. Both estimators are a weighted average of the public signal s and the mean private signal x_i . However, this has an important ramification: as the cross-section of agents increases, both estimators are *conditionally* biased with respect to s . Since every agent uses s in their forecast, if $s \neq P\{S = 1\}$ then the errors in individual predictions will not disappear in the aggregate. From the analyst's perspective, the most useful prediction for the aggregate estimator is the one that only uses x_i , not s , because then the law of large numbers would apply and both estimators would be consistent in the cross-section with respect to the number of agents, even conditional on an event's realization of s .

As a result, the key relationship to understand in the model is the ratio between $\frac{1}{\sigma_{\varepsilon_{i,\text{trader}}}^2}$ and $\frac{1-r}{\sigma_{\varepsilon_{i,\text{forecaster}}}^2}$. As the number of agents of either type increase, the average of their private

signals becomes less noisy, and converges to $P\{S = 1\}$ by the law of large numbers for i.i.d data. This is the traditional “wisdom of the crowd” result. Then, the relative accuracy of the **aggregate predictions** stems from how much each agent type weights their private signal compared to the public signal. When variance of the private signal (which depends on agent type) is small, the agent uses more of the private signal, because intuitively they know it will lead to a more-accurate guess. When reputational concerns (r) is large, the forecaster will use less of the private signal, because they know reliance on the private signal could lead them to a prediction far from others.

Importantly, the model implies that wisdom of the crowd, or $N_{\text{forecaster}}$ to be large, is not enough for a prediction market to outperform expert forecasters. Instead, wisdom of the crowd causes the private-signal component of the trader predictions to be accurate, but there is still a reliance on s which results in a conditionally biased forecast on s . To alleviate the conditional bias, the traders must be more “confident” in their private signal, and place more weight on it. In this way, the addition of traders with less noisy signals as determined by σ_{ε_i} would reduce the MSE of p beyond the affect of wisdom of the crowds. In other words, the model predicts that the share of volume supplied by informed or confident traders is more important to getting an accurate aggregate prediction than the total amount of volume or traders.

On the forecaster side, the quality of the consensus prediction is buttressed by having a smaller σ_{ε_i} , but harmed by having a reputation stake r which leads to overweighting the public signal again. While the forecaster consensus has an inherent advantage compared to the traders because of lower private-signal noise, if reputation is a big enough concern, they can weight the public signal too much and perform worse than the trader consensus.

We parameterize the model to present some features graphically in Figure 13. Panel (a) compares the consensus prediction accuracy for forecasters and traders, $\hat{y}_{\text{forecaster}}$ and $\hat{y}_{\text{trader}}$ as we vary r . With low reputation incentives, the forecasters weight their private signals more than the traders, because have more expertise (and therefore more confidence in their private signal). This leads to a better forecast performance. Under high r , the forecasters maximize utility by coordinating to near the public signal, and underweighting their expertise. As a result, under high r the market has lower prediction MSE. In panel (b), we hold fixed $r = 0.275$, and vary the number of traders in the market. While having a larger crowd does help improve $\hat{y}_{\text{trader}}$ if the market is particularly liquid, eventually the marginal value of an additional trader becomes small. This is a result of the traders reporting correlated predictions, namely they all weight s with positive weight. Conditional on a particular realization of s , our model of the prediction market forecast does not necessarily have $MSE(\hat{y}_{\text{trader}}) \rightarrow_N 0$. Decreasing the weight applied to s is key for reducing our estima-

tors’ mean squared errors. The addition of a fraction of traders who have low-noise private signals would in-turn place less reliance on the public signal, and increase the forecast accuracy. As a result, our model concludes the share of high-skill traders is the primary driver of improving prediction market accuracy. On the other hand, reputational incentives are the biggest threat to traditional forecasts’ accuracy. Our model helps rationalize concurrent work on mechanisms driving prediction market accuracy, which has shown that “informed” traders are larger drivers of price discovery than uninformed (Gomez Cram et al., 2026). It also helps to rationalize our results in Section 6.2.

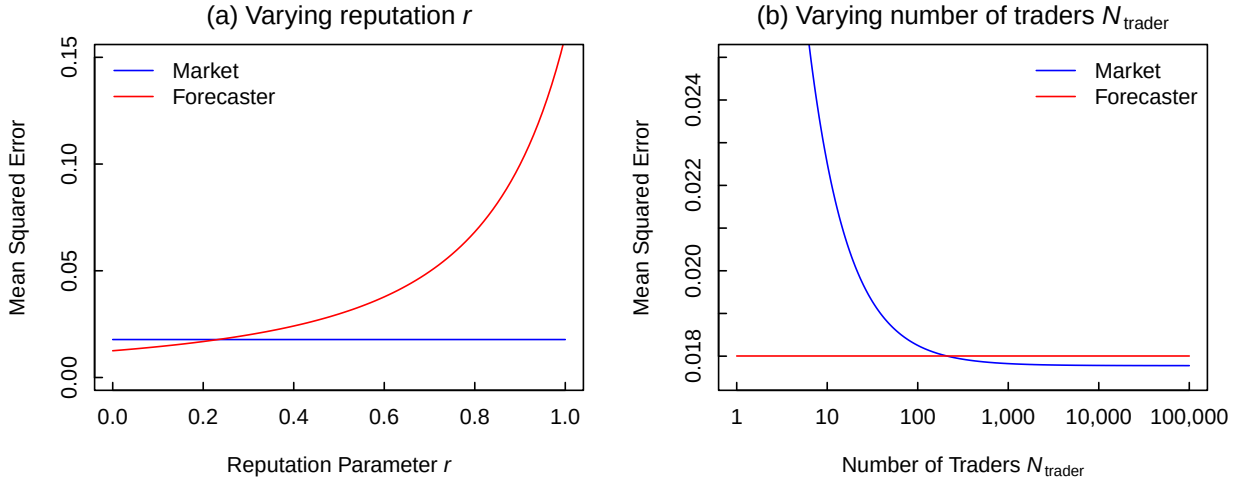


Figure 13: Model MSE Features

Note: How MSE changes as a result of changing the parameterization of our model. Panel (a) varies the reputation parameter r between 0 and 1; panel (b) varies the number of traders from 10 to 100,000. Throughout we set $\sigma_\eta^2 = 0.16$, $\sigma_{\varepsilon, \text{forecaster}}^2 = 0.06$, $\sigma_{\varepsilon, \text{trader}}^2 = 0.08$, and $N_{\text{forecaster}} = 50$; all noise parameters are reported as variances of the underlying centered Beta noise. Panel (a) sets $N_{\text{trader}} = 10,000$; panel (b) sets $r = 0.24$. Plotted values are normalized mean squared errors $\text{MSE}/m(y)^2$, which is equivalent to the $y = 0.5$ case, so that we focus on comparing forecaster and trader MSEs and our results do not depend on the underlying true probability y .

8 Distributional Responses to News

8.1 Fed Funds Rate Responses to Macro News

With access to full distributions, we can uncover new patterns in how macroeconomic news affects higher moments of expectations of the federal funds rate at the next FOMC meeting. Figure 14 shows the average change in the mean, mode, variance, and skewness of the distribution of the federal funds rate for the next FOMC on days of news releases. The news

releases we focus on are CPI, PCE inflation, non-farm payrolls, ISM manufacturing, FOMC statement, and FOMC minutes. We denote all other days as Other.

Fed funds rate mean and mode responses to news. Based on responses since 2022, we see that CPI days tend to have the largest positive effects on the first moment of the federal funds rate distribution. This could be a reflection of the sample period, as there were several announcements of higher-than-expected inflation in this particular period and is consistent with the importance of inflation for rate setting in the post-pandemic period. We also see that CPI, Non-farm Payrolls, and FOMC Minutes/Statement days on average result in the largest changes in central moments of the distribution of Federal Funds Rates, coinciding with market commentary typically placing the largest emphasis on these events. We are assured that our constructed distributions move significantly more on news release days than other days.

Fed Funds rate variance and skewness responses to news. We can also see the effects of data releases on the variance and skewness of the Fed Funds rate distributions in panel B of Figure 14. Intuitively, nearly all of the news announcements lower the variance. While variance typically decreases on any day (consistent with a general resolution of uncertainty over time), news days result in larger variance declines than other days, consistent with the idea that news days are informative for understanding future Federal Funds Rate distributions. As one might expect, the largest decline in the variance of the Fed Funds rate distributions comes on FOMC days, with CPI days generating the second largest decline.

In terms of skewness, we see that nonfarm payrolls days and FOMC days tend to have the largest positive effect. While the mode and mean tend to decline on NFP days, the rise in skewness suggests that Kalshi investors seem to also place relatively more weight on the upper tail of the Fed Funds rate distributions. In the next sections, instead of just examining the responses associated with particular news days, we will sharpen our focus on the effects of surprises for each news release.

CPI surprise effects on Fed Funds rate distributions. We also can evaluate how the distribution of the federal funds rate responds to the surprises released in each news announcement. Surprises are computed by taking the difference between the Bloomberg consensus forecast and the realized print. We first consider CPI inflation, split based on how the moments of the Federal Funds rate respond on days of positive and negative surprises, and days of zero surprise (expectations are measured to the nearest tenth of a percentage point). The results are shown in Figure 15.

As expected, the mean of the Fed Funds rate distribution (left panel) moves positively with inflation shocks. However, the magnitude is asymmetric: the response to positive CPI shocks is four times that of negative shocks. The variance of the distribution declines across all outcomes—reflecting the resolution of uncertainty—but the drop is sharpest when the release meets expectations (zero surprise). Additionally, skewness falls and kurtosis rises independent of the shock’s direction. To establish the statistical significance of these patterns, we turn to the regression analysis in the next section.

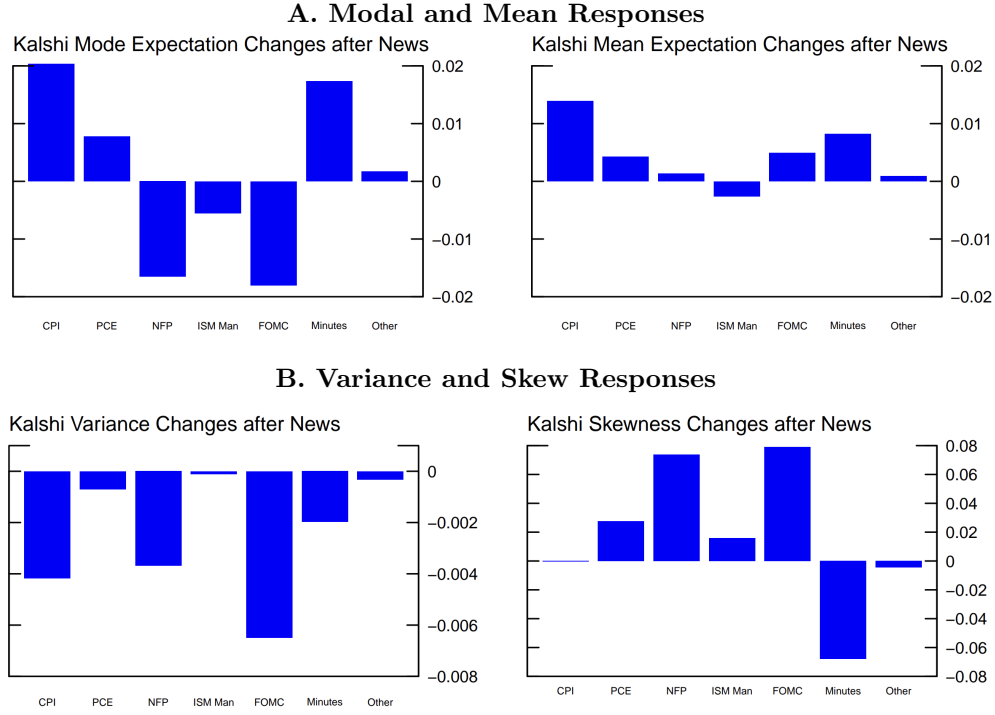


Figure 14: **Federal Funds Rate Distribution Responses to Economic News**

This figure shows the average change in moments of the federal funds rate for the next several FOMC meetings following news releases for CPI, PCE inflation, Non-farm payrolls, ISM Manufacturing, FOMC, and FOMC minutes. Other corresponds to all other days which do not fall into these categories.

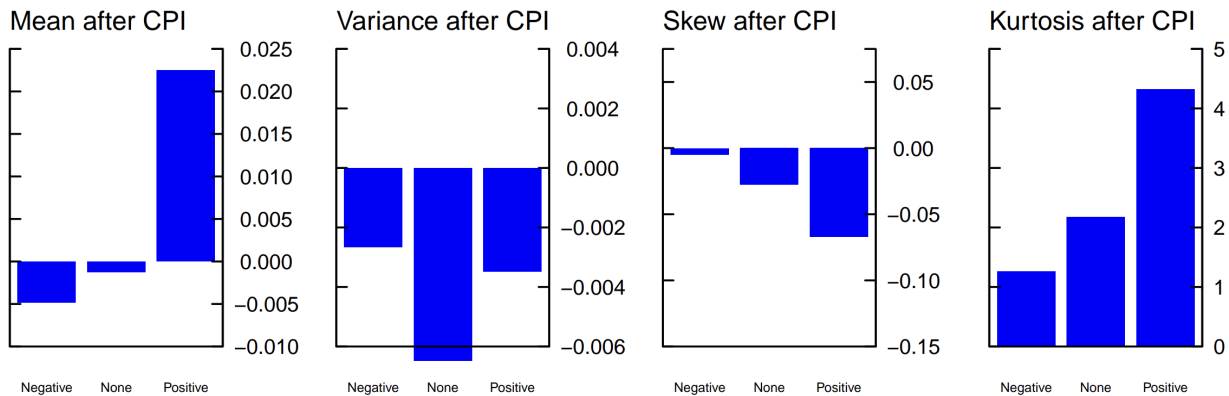


Figure 15: **Federal Funds Rate Distribution Responses to CPI Inflation Surprises**

This figure shows the average change in moments of the federal funds rate for the next several FOMC meetings following positive, negative and zero news from the CPI inflation release.

8.2 Event-study regressions for Fed Funds rate beliefs

8.2.1 Macro news surprises

We next turn to a more systematic analysis of the effects of surprises on the probability density of the federal funds rate. We consider event-study regressions of the form:

$$\Delta y_t = \beta_0 + \beta_1 S_t + \varepsilon_t \quad (17)$$

where Δy_t denotes the one day change in a moment of the federal funds distribution for the next FOMC meeting on day t and S_t denotes the surprise component of an announcement coming out on that day. The regression is run over all days of a particular announcement type. The announcement types considered are CPI, PCE and nonfarm payrolls. The constant measures the effect of the announcement in the absence of any surprise.

CPI significant effect on Fed Funds rate. The results are shown in Panel A of Table 5. The mean, median and mode of the funds rate all respond significantly to a CPI surprise with a 10 basis point surprise resulting in a 3 to 4 basis point rise in the central moments of the federal funds rate. We also see that independent of the surprise, the variance tends to fall, as the constant is negative and significant in the variance equation.

PCE less significant effect. Likewise, PCE surprises also have significant effects, but to a much lesser extent, as shown in Panel B of Table 5. This is consistent with market participants and investors paying more attention to CPI, which is more timely.

Panel A. CPI Headline YoY						
	Kalshi Fed Funds Rate Distribution					
	Mean	Median	Mode	Variance	Skewness	Kurtosis
CPI Surprise	0.320 ^{***} (0.052)	0.388 ^{***} (0.079)	0.413 ^{***} (0.079)	-0.002 (0.011)	-0.183 (0.448)	3.666 (4.676)
Constant	0.007 (0.005)	0.012 (0.008)	0.017* (0.009)	-0.004 ^{***} (0.001)	-0.028 (0.048)	1.853 ^{***} (0.540)
Observations	113	113	113	113	113	113
R^2	0.379	0.255	0.256	0.001	0.002	0.006

Panel B. PCE Headline YoY						
	Kalshi Fed Funds Rate Distribution					
	Mean	Median	Mode	Variance	Skewness	Kurtosis
PCE Surprise	0.012 ^{**} (0.004)	0.015 [*] (0.008)	0.012 [*] (0.007)	-0.001 (0.001)	0.001 (0.040)	-0.438 [*] (0.257)
Constant	-0.003 (0.006)	-0.014 [*] (0.008)	-0.017 [*] (0.009)	-0.002 [*] (0.001)	0.068 (0.043)	1.827 ^{***} (0.573)
Observations	113	113	113	113	113	113
R^2	0.060	0.040	0.023	0.007	0.000	0.007

Panel C. Nonfarm Payrolls						
	Kalshi Fed Funds Rate Distribution					
	Mean	Median	Mode	Variance	Skewness	Kurtosis
NFP Surprise	0.053 (0.032)	0.045 (0.051)	0.045 (0.051)	0.010 (0.009)	-0.061 (0.455)	-4.946 (3.440)
Constant	0.005 ^{**} (0.002)	0.004 (0.005)	0.004 (0.005)	-0.001 (0.001)	0.037 (0.030)	-0.156 (0.550)
Observations	108	108	108	108	108	108
R^2	0.056	0.008	0.008	0.011	0.000	0.007

Table 5: **Kalshi Fed Funds Rate Distribution Responses to Macro Surprises**

Notes: This table reports estimates of equation (17) using robust standard errors (HC3). Columns represent the change in moments of the Federal Funds Rate distribution on the day of the news release (end of previous day to end of day of release). Surprises are measured as the actual released value minus the Bloomberg survey expectation. Units are percentage points for CPI and PCE (year-over-year) and 100,000s of payrolls for Nonfarm Payrolls. Stars denote significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Nonfarm payrolls, no significance. In Panel C of Table 5, nonfarm payrolls show a positive relationship with the first moments of the Fed Funds rate distribution, but there is no statistical significance beyond the constant for the mean. This seems consistent with investors placing much greater emphasis on monetary policy’s potential response to inflation over the 2022-2025 period when compared to jobs, with the labor market being historically strong.

8.2.2 Monetary policy surprises

Table 6 presents the response of the federal funds rate distribution to the monetary policy shocks of Acosta et al. (2025). This separately identifies shocks in two communication windows: the statement window and the press conference window. We study the effects of both monetary policy shocks on the Kalshi-implied federal funds rate distribution at horizons out to about 6 months jointly. The results highlight a clear distinction between the two communication windows. During the statement window, the shock primarily shifts the level of rates: the mean, median, and mode all respond positively and significantly. This is not surprising as of course the statement includes the announcement about the current target for the federal funds rate. We also observe a significant rise in variance, indicating that the tighter policy generates some immediate dispersion in market views.

Press conference effect on skewness. Conversely, we find that the press conference shock has a positive, albeit statistically insignificant, effect on the central tendency (mean, median, or mode). This lack of significance may stem from the short maturity of the Kalshi Fed funds contracts, which limits the sample to a six-month horizon and the press conference obviously does not make announcements about the current funds rate target. However, the press conference significantly alters the shape of the distribution, causing a sharp decline in skewness. This negative shift could possibly reflect that more restrictive than expected communication during the press conference effectively truncates the “right tail” of the distribution. Tighter monetary policy in the press conference could be resolving upside uncertainty—capping the risk of higher rates—while leaving downside risks intact.

9 Conclusions

This paper has introduced Kalshi macroeconomic prediction markets as a novel, high-frequency, and distributionally rich source of expectations data. First we have shown that in several cases, these contracts provide unique insights—particularly for variables like GDP

	Kalshi Fed Funds Rate Distribution					
	Mean	Median	Mode	Variance	Skewness	Kurtosis
MP Statement Shock	0.872 ^{***} (0.236)	1.098 ^{**} (0.427)	1.233 [*] (0.681)	0.249 [*] (0.144)	-0.763 (2.344)	-6.609 (7.603)
MP Press Conference Shock	0.011 (0.319)	0.502 (0.318)	0.339 (0.905)	-0.032 (0.159)	-2.875 ^{**} (1.285)	-7.807 (7.299)
Constant	0.003 (0.004)	-0.004 (0.010)	-0.002 (0.013)	-0.007 ^{**} (0.003)	0.050 (0.053)	0.384 (0.320)
Observations	94	94	94	94	94	94
R^2	0.165	0.106	0.041	0.054	0.050	0.013

Table 6: **Kalshi Fed Funds Rate Distribution Responses to MP Shocks**

Notes: Robust standard errors (HC3) in parentheses. Columns report changes in moments of the federal funds rate distribution on the day of the news release for meetings more than one month ahead (end of previous day to end of release day). Monetary policy shocks are measured using the San Francisco Fed USMPD database with positive values representing tightening. ^{*} $p < 0.10$, ^{**} $p < 0.05$, ^{***} $p < 0.01$.

growth, core inflation, unemployment, and payrolls, for which no other market-based distributions currently exist. We have also argued that they provide the only credible measures of distributional beliefs about decisions at specific FOMC meetings. Second, these markets capture rich distributional dynamics—such as tail risks and asymmetries in higher moments—that are unavailable from traditional sources. For instance, we find that monetary policy press conference shocks tend to significantly reduce positive skewness in the Federal Funds rate distribution. Third, we have shown that they are comparable in forecasting accuracy to established benchmarks such as the SME and the Bloomberg consensus and indeed that Kalshi’s forecasts for the federal funds rate and CPI provide statistically significant improvements over Fed Funds futures and professional forecasters, all while providing continuously updated full distributions rather than infrequent point estimates. The mode of the Kalshi distribution, for example, has perfectly matched the realized federal funds rate by the day of each meeting since 2022, a feat not achieved by either surveys or futures. Third, the accessibility of Kalshi to retail traders introduces a perspective distinct from institutionally dominated markets, potentially offering a complementary lens on expectations formation.

Together, these findings suggest that prediction markets can serve as a valuable complement to existing forecast tools in both research and policy settings. By providing transparent, continuously updated, and economically interpretable measures of expectations with competitive forecast performance, they open new avenues for studying monetary policy transmission, market sentiment, and macroeconomic uncertainty. As these markets mature and liquidity

deepens, their potential to enhance real-time policy analysis and academic research will only grow over time.

References

- Acosta, Miguel, Andrea Ajello, Michael Bauer, Francesca Loria, and Silvia Miranda-Agrippino (2025) ‘Financial market effects of FOMC communication: Evidence from a new event-study database.’ Federal Reserve Bank of San Francisco Working Paper 2025-30
- Aït-Sahalia, Yacine (1996) ‘Nonparametric pricing of interest rate derivative securities.’ *Econometrica* 64(3), 527–560
- Amin, Kaushik, and Victor K Ng (1997) ‘Inferring future volatility from the information in implied volatility in eurodollar options: A new approach.’ *Review of Financial Studies* 10(2), 333–367
- Amisano, Gianni, and Raffaella Giacomini (2007) ‘Comparing density forecasts via weighted likelihood ratio tests.’ *Journal of Business & Economic Statistics* 25, 177–190
- Ang, Andrew, Geert Bekaert, and Min Wei (2007) ‘Do macro variables, asset markets, or surveys forecast inflation better?’ *Journal of Monetary Economics* 54(4), 1163–1212
- Bakshi, Gurdip, and Dilip B Madan (2000) ‘Spanning and derivative-security valuation.’ *Journal of Financial Economics* 55(2), 205–238
- Bakshi, Gurdip, George Panayotov, and George Skoulakis (2011) ‘Improving the predictability of real economic activity and asset returns with forward variances inferred from option portfolios.’ *Journal of Financial Economics* 100(2), 340–357
- Ball, Clifford A, and Walter N Torous (1999) ‘The stochastic volatility of short-term interest rates: Some international evidence.’ *Journal of Finance* 54(1), 233–252
- Bauer, Michael D, Aeimit Lakdawala, and Philippe Mueller (2022) ‘Market-based monetary policy uncertainty.’ *The Economic Journal* 132(647), 1291–1316
- Bauer, Michael D, and Mikhail Chernov (2024) ‘Interest rate skewness and biased beliefs.’ *Journal of Finance* 79, 173–217

- Beber, Alessandro, and Michael W Brandt (2006) ‘The effects of macroeconomic news on beliefs and preferences: Evidence from the options market.’ *Journal of Monetary Economics* 53(8), 1997–2039
- Boehmer, Ekkehart, Charles M Jones, Xiaoyan Zhang, and Xi Zhang (2020) ‘Tracking retail investor activity.’ *Journal of Finance* 75(5), 2249–2300
- Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2020) ‘Overreaction in macroeconomic expectations.’ *American Economic Review* 110(9), 2748–82
- Burgi, Constantin, Wanying Deng, and Karl Whelan (2025) ‘Makers and takers: The economics of the kalshi prediction market.’ Working Paper
- Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang (2014) ‘Wisdom of the crowds: The value of stock opinions transmitted through social media.’ *Review of Financial Studies* 27(5), 1367–1403
- Christoffersen, Peter, Kris Jacobs, and Bin Chang (2013) ‘Forecasting with option-implied information.’ In ‘Handbook of Economic Forecasting,’ vol. 2 (Elsevier) pp. 581–656
- Cieslak, Anna, and Annette Vissing-Jorgensen (2021) ‘The economics of the fed put.’ *Review of Financial Studies* 34(11), 5228–5265
- Clements, Michael P (2018) ‘Are macroeconomic density forecasts informative?’ *International Journal of Forecasting* 34(2), 197–206
- Coibion, Olivier, and Yuriy Gorodnichenko (2015) ‘Information rigidity and the expectations formation process: A simple framework and new facts.’ *American Economic Review* 105(8), 2644–78
- Darling, Donald A. (1957) ‘The Kolmogorov-Smirnov, Cramer-von Mises tests.’ *The Annals of Mathematical Statistics* 28, 823–838
- Diebold, Francis X, and Robert S Mariano (1995) ‘Comparing predictive accuracy.’ *Journal of Business & Economic Statistics* 13, 253–263
- Diebold, Francis X, Todd A Gunther, and Anthony S Tay (1998) ‘Evaluating density forecasts, with applications to financial risk management.’ *International Economic Review* 39, 863–883
- Diercks, Anthony M, and Haitham Jendoubi (2023) ‘Expectations of financial market participants.’ In ‘Handbook of Economic Expectations’ (Elsevier) pp. 385–410

- Diercks, Anthony M, Colin Campbell, Steven A Sharpe, and Daniel Soques (2023) ‘The swaps strike back: Evaluating expectations of one-year inflation’
- Diercks, Anthony M., Hiroatsu Tanaka, and Paul Cordova (2024) ‘Asymmetric monetary policy expectations.’ SSRN Working Paper
- Duffee, Gregory R (2012) ‘Forecasting interest rates.’ In ‘Handbook of Economic Forecasting,’ vol. 2 (Elsevier) pp. 105–140
- Eichengreen, Barry, Ganesh Viswanath-Natraj, Junxuan Wang, and Zijie Wang (2025) ‘Under pressure? Central bank independence meets blockchain prediction markets.’ Working Paper
- Emmons, William R, Aeimit K Lakdawala, and Christopher J Neely (2006) ‘What are the odds? Option-based forecasts of FOMC target changes.’ *Federal Reserve Bank of St. Louis Review* 88, 543–562
- Fan, Rongrong, Anurag Gupta, and Peter Ritchken (2003) ‘Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets.’ *Journal of Finance* 58(5), 2219–2248
- Farmer, Leland E., Emi Nakamura, and Jón Steinsson (2024) ‘Learning about the long run.’ *Journal of Political Economy* 132(10), 3334–3377
- Farrell, Matthew, Jeremiah Green, Russell Jame, and Stanimir Markov (2022) ‘The democratization of investment research and the informativeness of retail investor trading.’ *Journal of Financial Economics* 143(2), 697–720
- Faust, Jon, and Jonathan H Wright (2012) ‘Forecasting inflation.’ In ‘Handbook of Economic Forecasting,’ vol. 2 (Elsevier) pp. 2–56
- Goff, Stephen, Thomas Kostka, and Federico Masera (2014) ‘How informative are the subjective density forecasts of macroeconomists?’ *Journal of Forecasting* 33(8), 622–636
- Gomez Cram, Roberto, Yunhan Guo, Theis Ingerslev Jensen, and Howard Kung (2025) ‘Financial prediction markets: A new measure of earnings expectations.’ *SSRN Electronic Journal*
- (2026) ‘Prediction market accuracy: Crowd wisdom or informed minority?’ *SSRN Electronic Journal*

- Griffith, Richard M (1949) ‘Odds adjustments by american horse-race bettors.’ *The American Journal of Psychology* 62(2), 290–294
- Gürkaynak, Refet S, and Justin Wolfers (2005) ‘Macroeconomic derivatives: An initial analysis of market-based macro forecasts, uncertainty, and risk.’ In ‘NBER International Seminar on Macroeconomics’ University of Chicago Press pp. 11–50
- Gürkaynak, Refet S, Brian Sack, and Eric T Swanson (2012) ‘Market-based measures of monetary policy expectations.’ *Journal of Business & Economic Statistics* 30(2), 175–188
- Hanson, Robin, and Ryan Oprea (2009) ‘A manipulator can aid prediction market accuracy.’ *Economica* 76(302), 304–314
- Hilscher, Jens, Alon Raviv, and Ricardo Reis (2022) ‘Inflating away the public debt? An empirical assessment.’ *Review of Financial Studies* 35(2), 805–852
- (2024) ‘How likely is an inflation disaster?’ Working paper
- Jarrow, Robert A, Hao Li, and Feng Zhao (2007) ‘Interest rate caps “smile” too! but can the libor market models capture the smile?’ *Journal of Finance* 62(1), 345–382
- Kelley, Eric, and Paul C Tetlock (2013) ‘How wise are crowds? insights from retail orders and stock returns.’ *Journal of Finance* 68(3), 1229–1275
- (2017) ‘Retail short selling and stock prices.’ *Review of Financial Studies* 30(3), 801–834
- Kitsul, Yuriy, and Jonathan H Wright (2013) ‘The economics of options-implied inflation probability density functions.’ *Journal of Financial Economics* 110(3), 696–710
- Li, Hao, and Feng Zhao (2006) ‘Unspanned stochastic volatility: Evidence from hedging interest rate derivatives.’ *Journal of Finance* 61(5), 2213–2230
- (2009) ‘Nonparametric estimation of state-price densities implicit in interest rate cap prices.’ *Review of Financial Studies* 22(10), 4335–4376
- Longstaff, Francis A, Pedro Santa-Clara, and Eduardo S Schwartz (2001) ‘The relative valuation of caps and swaptions: Theory and empirical evidence.’ *Journal of Finance* 56(5), 2067–2109
- Mertens, Thomas M, and John C Williams (2021) ‘What to expect from the lower bound on interest rates: Evidence from derivatives prices.’ *American Economic Review* 111(6), 1616–1665

- Miltersen, Kristian R, Klaus Sandmann, and Dieter Sondermann (1997) ‘Closed-form solutions for term structure derivatives with lognormal interest rates.’ *Journal of Finance* 52(1), 409–430
- Morris, Stephen, and Hyun Song Shin (2002) ‘Social value of public information.’ *American Economic Review* 92(5), 1521–1534
- Nakamura, Emi, and Jón Steinsson (2018) ‘High-frequency identification of monetary non-neutrality: the information effect.’ *Quarterly Journal of Economics* 133(3), 1283–1330
- Rich, Robert, and Joseph Tracy (2010) ‘The relationships among expected inflation, disagreement, and uncertainty: Evidence from matched point and density forecasts.’ *Review of Economics and Statistics* 92(1), 200–207
- Rossi, Barbara, and Tatevik Sekhposyan (2019) ‘Alternative tests for correct specification of conditional predictive densities.’ *Journal of Econometrics* 208, 638–657
- Snowberg, Erik, and Justin Wolfers (2010) ‘Explaining the favorite–longshot bias: Is it risk-love or misperceptions?’ *Journal of Political Economy* 118(4), 723–746
- Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz (2013) ‘Prediction markets for economic forecasting.’ In ‘Handbook of Economic Forecasting’ (Elsevier)
- Swanson, Eric T (2021) ‘Measuring the effects of federal reserve forward guidance and asset purchases on financial markets.’ *Journal of Monetary Economics* 118, 32–53
- Swanson, Eric T, Renxuan Wang, and Yanbin Wu (2025) ‘The effects of monetary policy on macroeconomic expectations: High-frequency evidence from traded event contracts.’ Working Paper
- Wolfers, Justin, and Eric Zitzewitz (2004) ‘Prediction markets.’ *Journal of Economic Perspectives* 18(2), 107–126
- Wright, Jonathan H (2018) ‘Options-implied probability density functions for real interest rates.’ *International Journal of Central Banking* 12, 129–149
- (2025) ‘Event-day options.’ *Journal of Time Series Analysis* 46, 1085–1097

Appendix

A Kalshi Data

This section will track and explain the decisions we make in order to come up with a prediction-market benchline in which we can run our analysis. Since the markets are new, and research in them is relatively scarce, we make a variety of decisions on how to gather, aggregate, and clean our data. We hope this section makes our decisions more transparent and provides a template for future prediction markets datawork.

A.1 Kalshi vs. Alternatives

We must note that to analyze the accuracy of predictions markets, we use data from Kalshi. While Kalshi is one of the largest prediction market operators, we understand it is not the only one, and chose to use it for a variety of factors. First, Kalshi is CFTC-regulated, unlike (at time of writing) Polymarket, its largest competitor. There are other CFTC regulated prediction markets, for example Interactive Brokers, but they have currently have less liquidity, making Kalshi appealing.

Another appeal of Kalshi, compared to Polymarket, was the structure of the contracts. Our primary variable of interest in the paper is a market-implied expectation for the federal funds rate, how tight or loose people think monetary policy will be. Polymarket Federal Funds Rate contracts are set up with markets for each *action* the Fed could take at a meeting, not each target rate/range for after the meeting. As a result, arriving at a Polymarket FFR forecast more than one meeting out is difficult. Take the example of Figure [A.1](#): does a “no change” in July mean that rates are where they currently are in April, or does that imply an interim cut? For our variable of interest in this example, expectations for the stance of monetary policy in July, using this contract structure would require assumptions about the relationship between consecutive FOMC actions. By contrast, the Kalshi markets we select avoid this problem by having outcomes around the rate *after* the FOMC meeting. Kalshi does also offer contracts on the Fed decision (similar to Polymarket’s), but we do not use them for our analysis for the same reason we do not use Polymarket’s markets. That said—we do think that understanding the differences between Kalshi and Polymarket’s forecasts are important, and leave it as a question for future research.

Federal Reserve rates

What will the Fed do at the upcoming meetings?

Apr 28 Jun 16 **Jul 28**

EXPECTED DECISION

No Change



Meeting in 105 DAYS 2 HRS 17 MIN 43 SEC

Fed Decision Probabilities

Tue Jul 28, 2026 FOMC Meeting



Fed funds rate after July meeting?



Fed Rate After Meeting

Search

Fed Rate After Meeting

Above 2.75%	93%	Yes 97¢	No 7¢
Above 3.00%	93% ▲ 2	Yes 94¢	No 7¢
Above 3.25%	80%	Yes 84¢	No 20¢
Above 3.50%	60% ▲ 1	Yes 60¢	No 41¢
Above 3.75%	16% ▼ 1	Yes 15¢	No 89¢
Above 4.00%	10% ▲ 4	Yes 9¢	No 97¢
Above 4.25%	9% ▲ 5	Yes 8¢	No 97¢
Above 4.50%	3%	Yes 7¢	No 97¢

Figure A.1: Polymarket (top) and Kalshi (bottom) FFR Market Structures for July 28th 2026 FOMC meeting (taken on April 14th)

A.2 Volume

In order to have faith that Kalshi markets have prices which accurately reflect investor beliefs, it is important that volume is sufficiently large, so that prices are properly formed to reflect real outcomes. Figure 3 shows aggregate volumes for each type of macro event we examine, and Figure 2 shows volume for individual markets as part of the Fed Funds Rate series. Here, we document some additional facts about volume in our Kalshi markets of interest. First, Figure A.2 subdivides the volume from Figure 3 into bins depending on the price. Volume across all markets is largely symmetric and convex on the price. The convexity is driven largely by Federal Funds Rates, where there is little uncertainty and so trading typically takes place at tails.

Another alternative to showing the activity in our markets of interest is showing how open-interest and bid-ask spreads evolve over the lifecycle of the market. We present that in Figure A.3. We find that as the market approaches resolution, open interest increases and bid-ask spreads tend to decrease. Since 2025, open-interest has increased dramatically in CPI and Unemployment markets over the entire lifecycle of the market.

A.3 Gathering Data

Kalshi subdivides their exchange into subdivisions “series”, “event”, “market” where a *series* is a recurring ‘type’ of event (ex: CPI YoY releases), *event* is a particular realization of that event (ex: the April 2026 CPI YoY release), and *market* is a particular Arrow-Debreu security and outcome you can trade on for that event (ex: whether the April 2026 CPI YoY release will be above 3.2%). We follow these definitions below, which can help clarify how and when we aggregate data.

The general overview of the construction of our Kalshi dataset is:

1. Having Kalshi series we are interested in studying
2. Gathering the ticker for each Kalshi market corresponding to our series of interest
3. Constructing trade-level price data for each market gathered in step 2
4. Aggregating trade-level price data to daily price data for each market
5. Aggregate daily-level price data for each market to the event-level. Outputs **panel data distributions** for each event of interest.
6. Construct relevant moments from panel distributions. Outputs **timeseries moments data** for each event of interest.

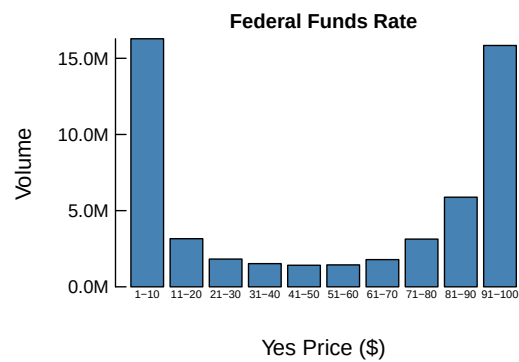
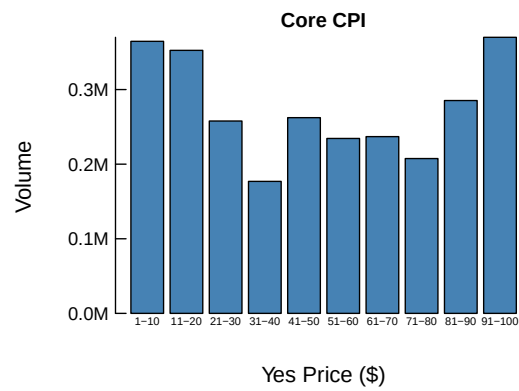
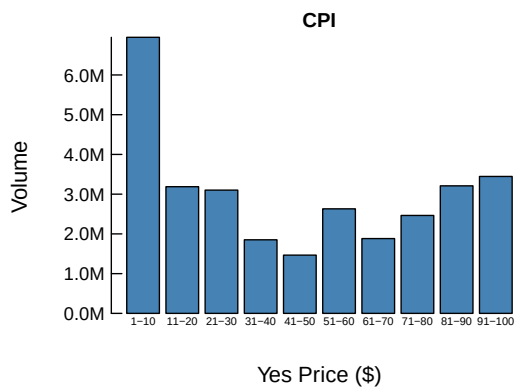
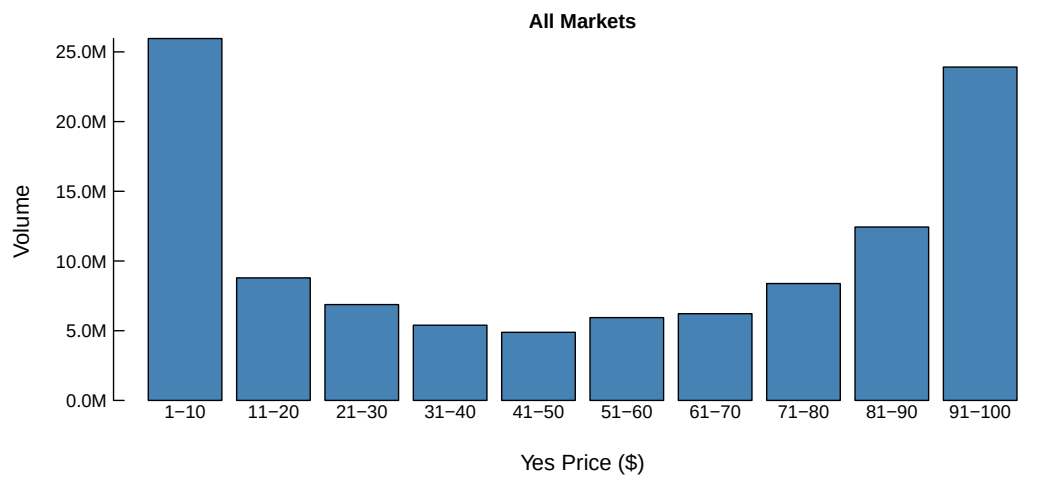


Figure A.2: Volume by Price Bucket

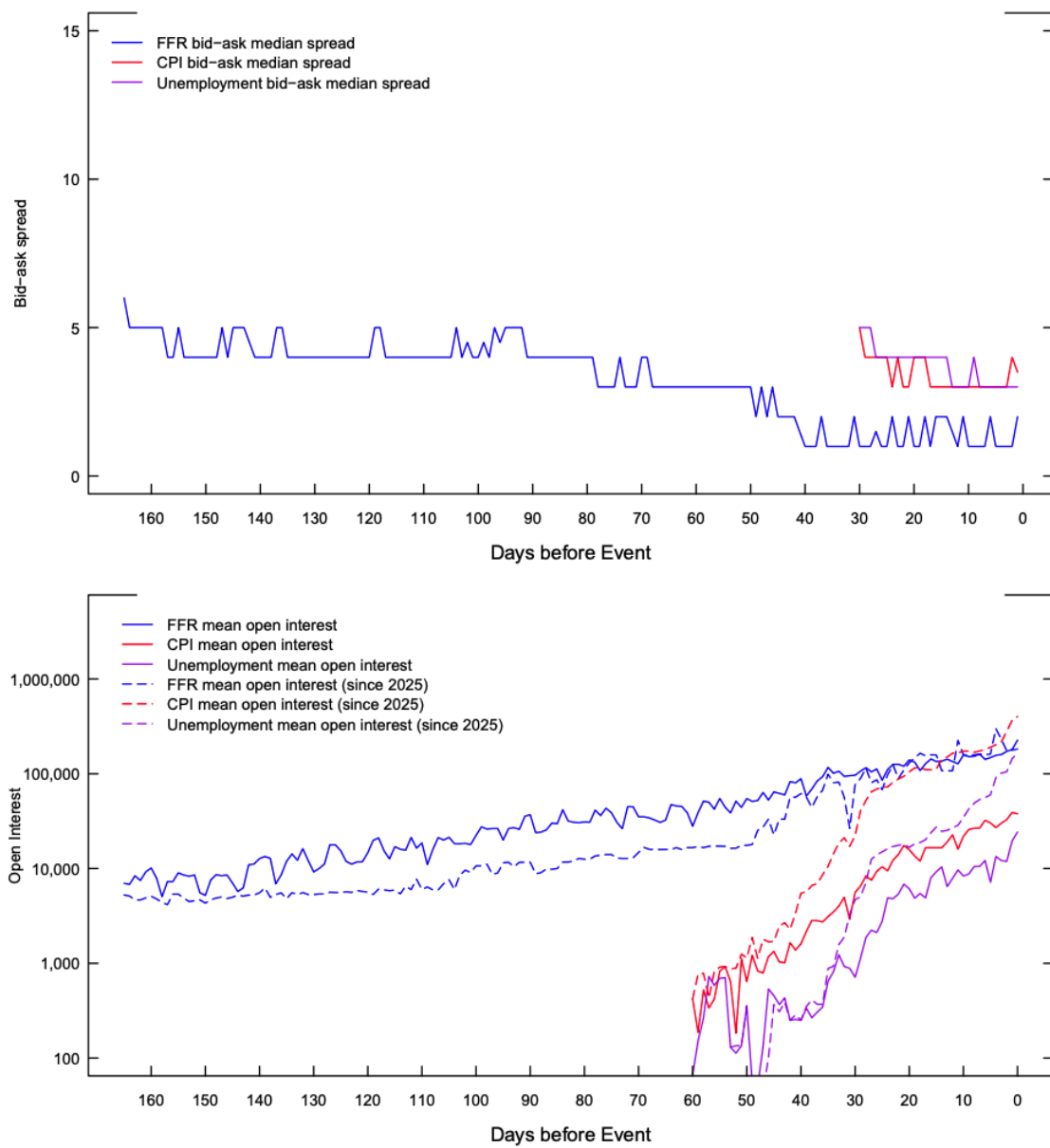


Figure A.3: Bid-ask spreads and open-interest over Market Lifecycle

Question	Ticker	First Event	Data analog
Federal Funds Rate after [month] meeting?	KXFED	July 2021 FOMC	Federal funds rate upper bound on target range
Inflation in [month] (CPI YoY)	KXCPIYOY	November 2022 CPI	BEA Headline CPI YoY
Core inflation in [month] (Core CPI YoY)	KXCPICOREYOY	December 2022 CPI	BEA Core CPI YoY
Unemployment in [month] [year]	KXU3	July 2021	BLS U3 Unemployment Rate
Inflation in [year]	KXACPI	2022	Annual Headline CPI
GDP growth in [year]?	KXACPI	2025	BEA GDP sum of Q1-Q4

Table 7: Kalshi series appearing in the paper

Since Kalshi updates their API continually, we cannot guarantee that our code/process will always work to get the most recent data⁸, we aim to be transparent and make our code to generate our Kalshi dataset public on the [GitHub](#) for the paper. We document the decisions we make in each step below. You may need your own API key from Kalshi to run the code, which you can get [here](#).

A.3.1 Kalshi Series we Study

We limit our study to looking only at a select amount of economic markets that we think may be most economically valuable to economic policymakers or market participants. We found these using the Kalshi website, and on the website looked for their corresponding series. We document these in Table 7. We are interested in these markets because they are the most monetary-policy relevant of the Kalshi markets, and their content can provide additional useful insights into variables that monetary policymakers might find interesting.

A.3.2 Gathering tickers for each Kalshi market corresponding to our series of interest

One difference between prediction markets and traditional derivative markets is that new assets are tied to specific event dates, rather than an arbitrary expiry date (ie 30 or 90 days). For example, prediction markets do not lump together a June 2026 CPI release with a July 2026 CPI release, they are entirely different. Moreover, they divide further each bin into its own market: whether June 2026 CPI exceeds 3.0% and whether it exceeds 3.1% are different markets that happen to be very similar, in contradiction to how other derivatives might work. Complicating this matter are the facts that the naming and ID conventions of markets have changed throughout the sample, and the Kalshi API separates “historical” (older than 3 month data) from “live” (more recent than 3 month data).

As a result, one immediate challenge to getting price data is figuring out what markets have historically even existed. To resolve this, we use the Kalshi API to first query for a list of all events related to each series. Then, we query further each market that corresponds to

⁸Changes to the API did break our code multiple times while writing this paper.

each event. We complete this process for our historical and our live data, and ensure there are no duplicates. Finally, we manually checked the created list of tickers with the Kalshi website to ensure we were not missing any markets. Our code to complete this process is documented in `tickers.py`, in the paper’s GitHub repository.

A.3.3 Constructing trade-level price data for each market

Given each market, we now use Kalshi’s API, specifically the `\trades` and `\historical\trades` endpoints to construct trade-level price data. We again link the trades and historical trades data ensuring no duplicates. No data cleaning or filtering has occurred at any step at this point. Our code is located in `scrape_kalshi_trades.py`.

We conduct some checks on the trade-level data to understand potential inefficiencies in Kalshi data that could lead to skepticism about the quality of the data. First, we look at the average returns, defined as the winning percentage minus the price, for taking the “Yes” side of a contract compared to the “No” side. While “Yes” does consistently lose more than the price warrants in our sample, an optimism bias, it is not significant in any market when clustering at the event level. We also look at the average returns for market makers compared to takers. In some markets makers consistently win more than their prices warrant, but in others they lose. In any case, it does not look from our trade-level data that in our markets of interest one group is consistently exploiting others.

Furthermore, we report for each Yes price the share of makers and takers that were on the Yes and No sides of the trade. We see that across all markets, makers and takers are not lopsided in their sides at any particular price.

A.3.4 Aggregating trade-level price data to daily price data

Prior to this point, our construction of our dataset has been straightforward and has not involved any major decisions. However, converting Kalshi trading data into forecasts requires some decisions that we make clear here. Figure [A.7](#) shows that forecast performance in the aggregate is little-changed, and our decisions are only made to help interpretability of individual probability distributions.

First, we convert our prices from cents (0.01-0.99) to dollars (1-99) which are associated with percentage outcomes. No trades occur at \$0 or \$1, so the largest possible trade-implied probability at this step is 99%, and the smallest possible (if the market has been traded) is 1%. Next, we use market ticker information to extract the strike price and the event that is associated with each trade. Dates are defined using UTC, so a date ends at 7:00pm EST.

To convert our trade-level data to daily-level price data, we make two main choices.

market	yes_return	no_return	yes_p_value	no_p_value	n_trades	n_clusters
CPI	-0.0808	0.0808	0.1232	0.1232	30010459	41
Core CPI	-0.0749	0.0749	0.3823	0.3823	2713217	40
Federal Funds Rate	0.0189	-0.0189	0.5760	0.5760	44459923	38
Unemployment	-0.0690	0.0690	0.2600	0.2600	9185885	57
All Markets	-0.0280	0.0280	0.3492	0.3492	86369484	176

Figure A.4: Yes vs. No Returns

market	maker_return	taker_return	maker_p_value	taker_p_value	n_trades	n_clusters
CPI	0.0242	-0.0242	0.0023	0.0023	30010459	41
Core CPI	-0.0895	0.0895	0.0634	0.0634	2713217	40
Federal Funds Rate	-0.0011	0.0011	0.7431	0.7431	44459923	38
Unemployment	-0.0227	0.0227	0.0535	0.0535	9185885	57
All Markets	0.0026	-0.0026	0.6289	0.6289	86369484	176

Figure A.5: Makers vs. Takers Returns

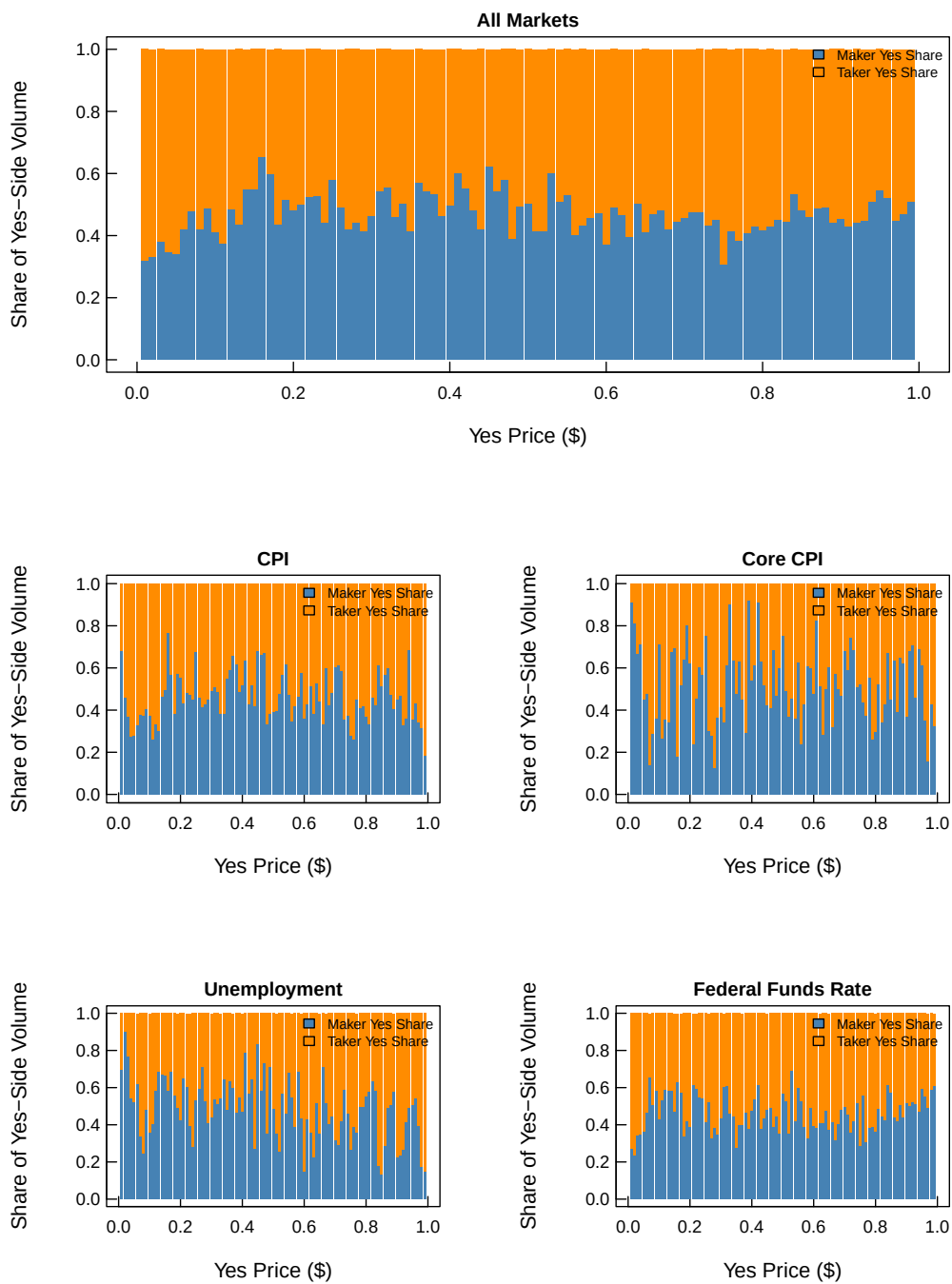


Figure A.6: Maker and Taker Shares

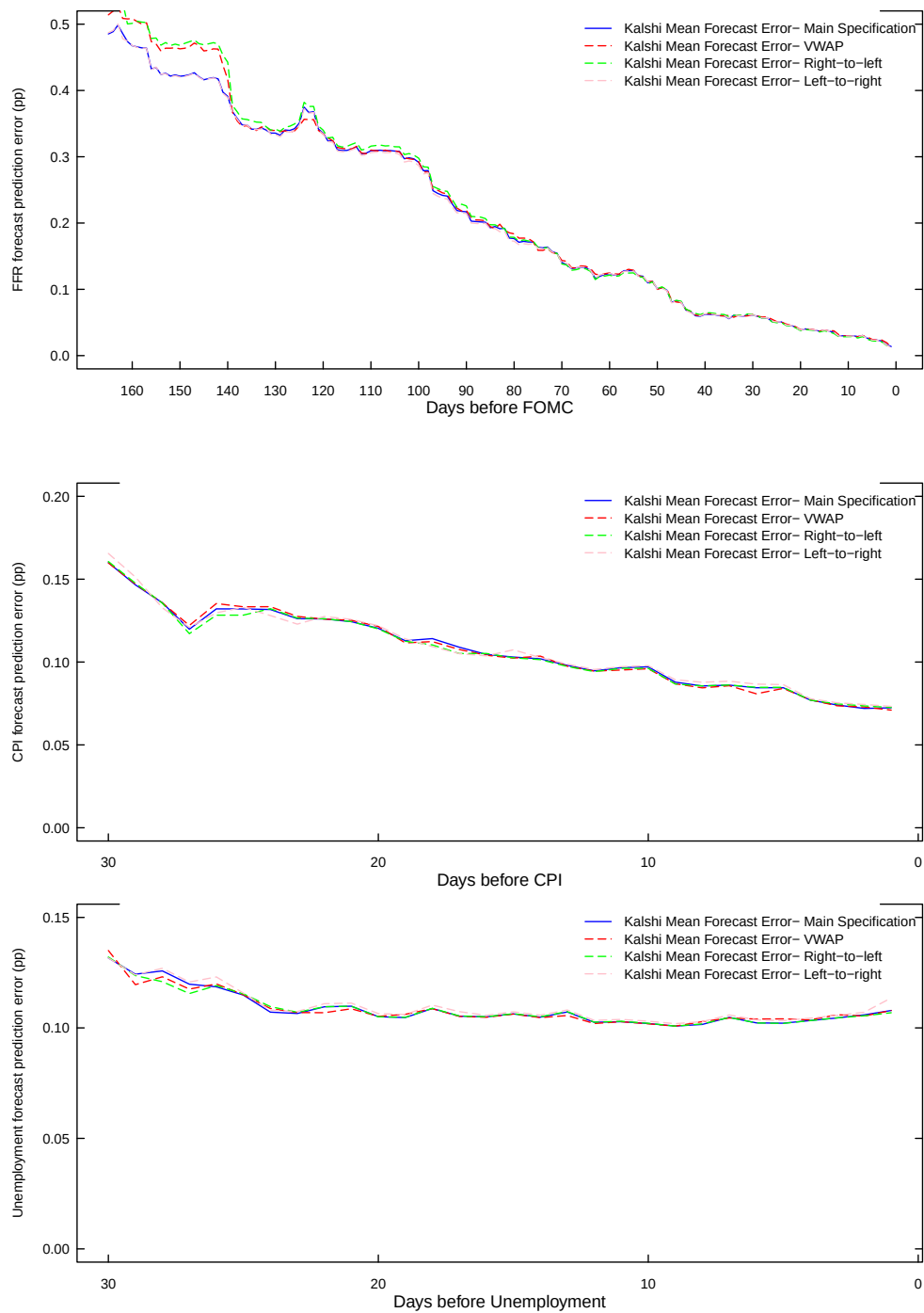


Figure A.7: Robustness of forecast performance to methodological decisions

First, we use the last trade of the day as the market’s daily price. We choose the last trade to have an analog to other financial markets, where using last trades will allow for cleaner identification of market responses to news. We also test an alternative construction which uses volume-weighted daily average prices, whose construction we leave in the code. We do not see significant changes in results.

Second, because of the multitude of Kalshi markets, not every market will have a trade each day. This arises in two situations in particular: (1) tail markets which are deemed very unlikely to happen; (2) markets which resolve far in the future. Similar cases occur in options markets, where far out-the-money options have limited trade volume compared to in-the-money options. In these less-liquid markets, we need a method to impute prices for contracts on days that do not have trades. One option would be to say any market that does not have a trade on that day has an implied-probability of \$0. However, oftentimes this situation occurs on a day after there is a non-zero implied-probability. As a result, pushing probabilities arbitrarily towards 0 because a market was not traded can bias estimates towards liquid, but not necessarily correct, benchmarks. We instead assume that if a market was not traded, it’s price did not change in the intermeeting period, and therefore carry yesterday’s price forward. Importantly, we do **not** fill prices backwards based on future prices, as that would give an unfair advantage to the prediction markets. We believe that carrying prices forwards is an agnostic approach, and that prices remaining unchanged if no trades occur is a minimal assumption to get a viable forecast for each bin. In markets that have *never* had a trade across the existence of the event, we assign probability 0. These are reserved for far-tails of distributions, or more central bins in the early period of a market, before much money has entered the market. As a result, when conducting our analysis we will trim the start of our markets to allow prices to establish prior to testing forecasting ability.

A.3.5 Cleaning daily-level price data for each market

While our daily-level price data is useful for gleaning changes in prices in specific markets, there are some problems we consider worthy of additional attention in order to arrive at a prediction market-implied density function. First, trades in different markets occur at different times, which can make it hard to square observed “last-trade-of-the-day” prices. For instance, the last trade in the June 2025 CPI YoY above 3.0% market might take place at 2:30pm, while the 3.1% market might have a trade at 4:00pm. If underlying probabilities of events changed in that time, then our distribution might reflect that, rather than the true density. This problem can be exacerbated by filling data forwards when there are no trades. A second problem is that there are occasional “anomaly” trades— where a trader typically conducts a small volume trade at a price significantly departing from where the

market’s consensus price has been. Because of the structure of most Kalshi markets we study having payouts if a news release *exceeds* some threshold, these anomaly trades can be strictly dominated by another trade that was available to the user: getting a contract that paid out in a strictly larger set of events than what was purchased, available at a cheaper price. While we cannot completely mitigate these issues, and we make a point in not throwing away any prediction-market data that we think represent “bad trades”. Instead, we seek to construct distributions that are robust to these concerns and rely on processes that can create distributions in real-time.

As a result, we impose that prices must be monotonic decreasing, i.e. the contract that pays out if CPI exceeds 3.1% must be less expensive than the contract that pays out if CPI exceeds 3.0%. We think that this is a reasonable assumption that alleviates our concerns by correcting stale prices on tails or early in samples and ensures our distributions can be constructed such that every event has nonnegative probability. One way to enforce monotonicity would be to start at the lowest bin and work upwards. In our CPI example, if exceeding 2.7% had a price of \$97, and exceeding 2.8% had a price of \$98, we would assume the 2.8% bin is mispriced since it covers a smaller set of outcomes at a higher price, and correct it’s price to \$97, the same as the 2.7% bin. However, this assumes that we know the 2.7% threshold is the market that’s correctly priced, when it’s possible the 2.8% market is right. The alternative would be to work from high-to-low bin, rather than low-to-high like in the last example. In this case, encountering the same CPI situation, we would correct the price of the 2.7% threshold to \$98, while the 2.8% market would remain unchanged. However, using a right-to-left methodology, correcting an issue on the left-tail, imposes a similar problem on the right-tail of our set of markets. We want a method that is robust to both of these issues. Once we have monotonicity in prices, we convert our “cdf”-like objects to discrete *probability density functions* by subtracting the price of one threshold from the price of the next threshold. For example, in [A.1](#), the probability assigned to the (3.25, 3.5] bin is essentially the probability the FFR is above 3.25 but not above 3.5, i.e. $80 - 60 = 20\%$. If we impose monotonicity in a reasonable way, we can apply this directly to construct our distributions.

As a result, we created an algorithm to operate these adjustments from the “middle-out”. This takes advantage of the fact that liquidity in the bins closer to the expected outcome are greatest, and therefore intuitively the prices update more regularly and appear to be more tightly calibrated. The algorithm is conceptually simple: we identify the median outcome from our prices, and then ensure that bins to its left have lower prices, and bins to its right have higher prices (thus imposing monotonicity). Then, we construct our probability distributions by subtracting the price of one threshold from the price of the previous threshold.

If two bins have the same price when doing this subtraction, we assign the probability mass to the bin that is closer to the identified median. Two markets having the same price imply that the additional outcomes that trigger a “yes” in the more-encompassing market are probability 0 outcomes, *or* they are artifacts of low liquidity. We interpret it as the latter, and assign the probability to the outcomes that are closer to the median expectation, giving us connected distributions⁹. Finally, while Kalshi only has prices between 0.01-0.99, we want probabilities between $[0, 1]$, so we normalize accordingly so that our probabilities for each bin-date-event sum to 100. We are left with a probability for each bin-date-event, which we use as our time-varying probability distributions for each event of interest.

In figure A.7 we replicate our main result from 1 and 10, showing the forecast performance for our series of interest using a variety of alternate construction methods for our trade-level data is consistent with our main results. When examining the calibration of our distributions, we cut off the beginning of the sample, since often the first one or two markets in a category on Kalshi have near-zero volume or insufficient bin granularity to infer moments. We begin our samples in March 2022 for FOMC, April 2022 for CPI and Unemployment, and March 2023 for Core CPI. We end in July 2026 for all events.

A.3.6 Constructing relevant moments from panel distributions

Finally, we construct moments on Kalshi-implied expectations from our constructed panel distributions. We calculate the mean by *center* of the bin. We calculate the median as the center of the bin which contains the 50th percentile of the distribution. We calculate the mode as the center of the bin with the highest probability mass. We calculate the variance as the probability-weighted average squared distance from the mean. We calculate the skewness as a weighted-analog of Pearson skewness, which is bounded between $[-1, 1]$ where -1 indicates left-skewness. We prefer this bounded skewness to ensure our results do not blow up when using this as an object in regressions. We calculate a weighted Kurtosis using the typical textbook definition¹⁰.

A.4 Midpoint of Bid-Ask Spreads

Another approach to constructing the PDFs for the federal funds rate expectations is to take the midpoint of the bid-ask spread, as opposed to using the prices based on the last or volume-weighted trades. Our forecast error results are shown in Figure A.8 based on this approach. One can see that the forecast errors spike quite a bit more, indicating some

⁹Almost always. Technically, a distribution can be disconnected if multiple outcomes have the median price, which does not happen in the sample of our paper, but has been observed since.

¹⁰Which happens to also be Stata’s implementation.

reliability issues, most likely due to issues surrounding the tails of the distribution. Because of these issues, we maintain focus on the last trade for our main results.

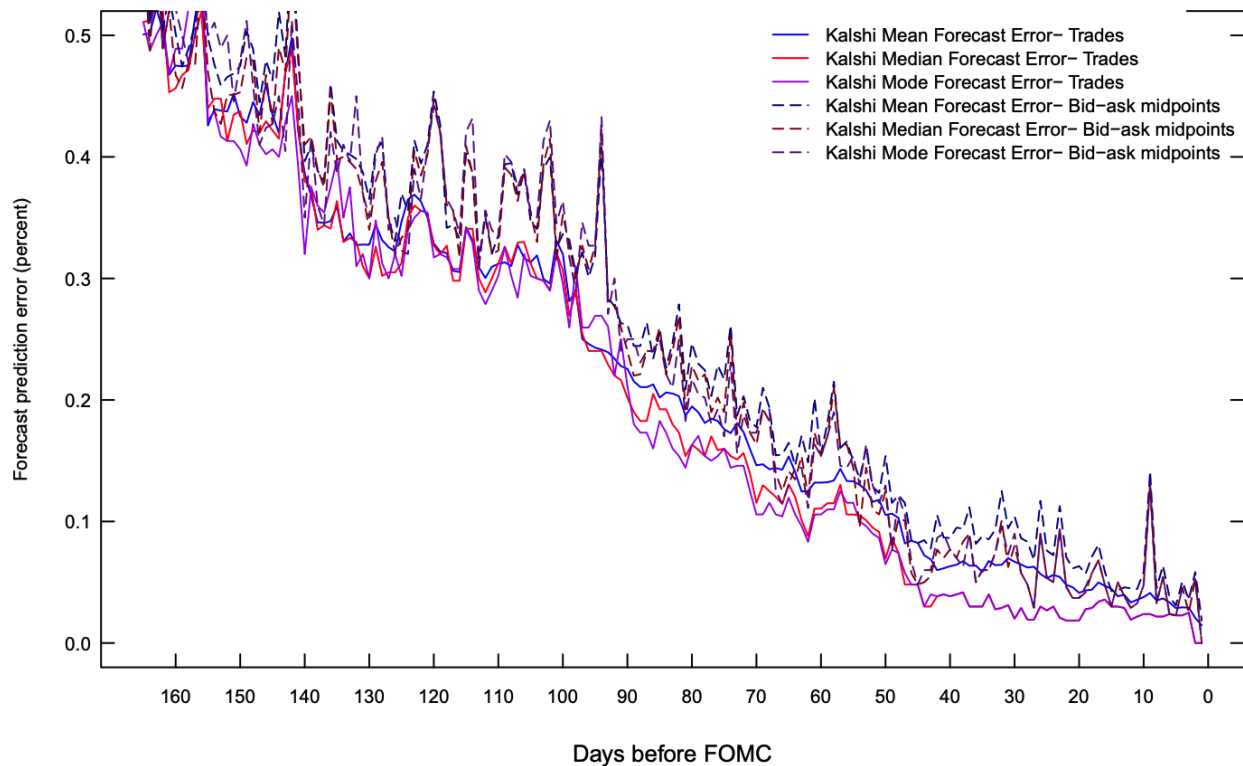


Figure A.8: **Midpoint of Bid-Ask Spread: Fed Funds Rate Forecast Errors**

This figure compares the mean absolute errors from 160 days out to each FOMC for the effective federal funds rate since 2022. Both the last trade and bid-ask midpoints are plotted.

A.5 Additional Comparisons

This part of the appendix gives additional comparisons between Kalshi and alternative forecasts (point and distributional), building on the results in Sections 4 and 5.

CPI Inflation Projections. Kalshi provides mean and median forecasts for CPI. The chart on the left of Figure A.9 shows that the March 2026 CPI expectations had drifted up following the release of the February 2026 CPI, oscillated between 3.4 and 3.3, and ultimately resolved to the right outcome while the Bloomberg consensus was slightly off. On the right panel, in the following month, both Kalshi and Bloomberg undershot expectations for the April 2026 CPI release. These figures demonstrate that both the Bloomberg consensus and the Kalshi-implied mean can at times disagree or line up closely. However, a key difference

is that Bloomberg consensus has no time series and is only available in the moments before a release. With the complete time series, we learn that developments in mid-March and again in late April caused a rise in expectations for near-term inflation. The evolution of these dynamics are not possible to observe for the Bloomberg Consensus.

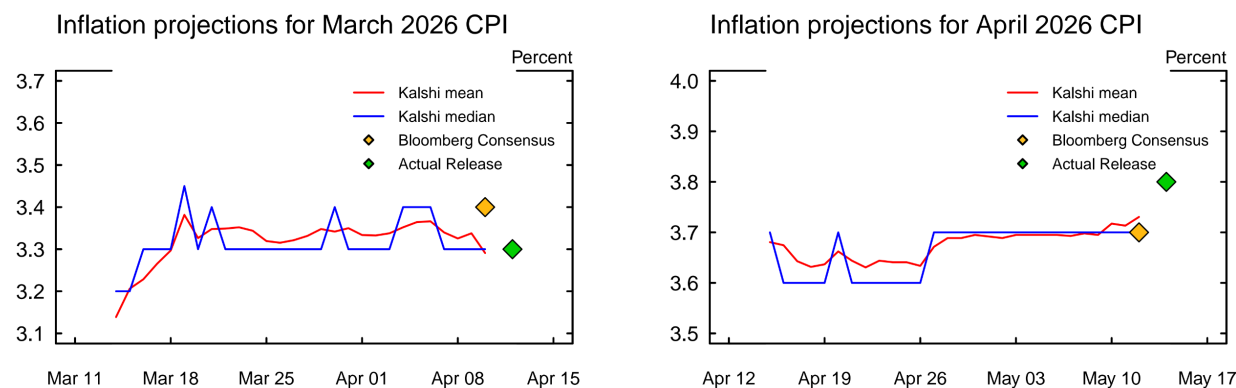


Figure A.9: **Kalshi CPI Comparisons to Bloomberg Consensus**

The chart on the left shows the year-over-year March 2026 CPI projection coming from the Kalshi market and the Bloomberg consensus. The chart on the right shows the year-over-year April 2026 CPI projection.

Year-end Projections for CPI and GDP. In addition to year-over-year monthly readings, Kalshi also tracks Q4/Q4 projections for CPI and GDP. The chart on the left of Figure A.10 shows that the one-year inflation swap (solid black line) lines up fairly closely with the Kalshi-implied mean (red line). The Blue Chip Economic Indicators 2025 inflation, which is available only at lower frequency, also lines up very closely with the other two approaches.

The chart on the right is the one that is unique to Kalshi. Previously, there did not exist a market-implied forecast of GDP growth. The right panel shows that the Kalshi-implied mean tracked very closely the SME. We also include the TIPS 2-year real yield as another proxy for growth. All of the series show a decline from the beginning of the year by close to one percentage point, with recent dynamics moving roughly sideways.

Projections for Unemployment. Kalshi offers a unique source of market-based forecasts for labor market indicators such as unemployment and payrolls—variables for which no financial markets currently exist. Aside from prediction markets, forecasts for these releases are primarily derived from surveys. However, survey-based estimates are not updated in real time and are only available at low frequency.

The chart on the left of Figure A.11 shows that both the Bloomberg Consensus and the Kalshi-implied mean accurately anticipated the May 2025 unemployment release. For June

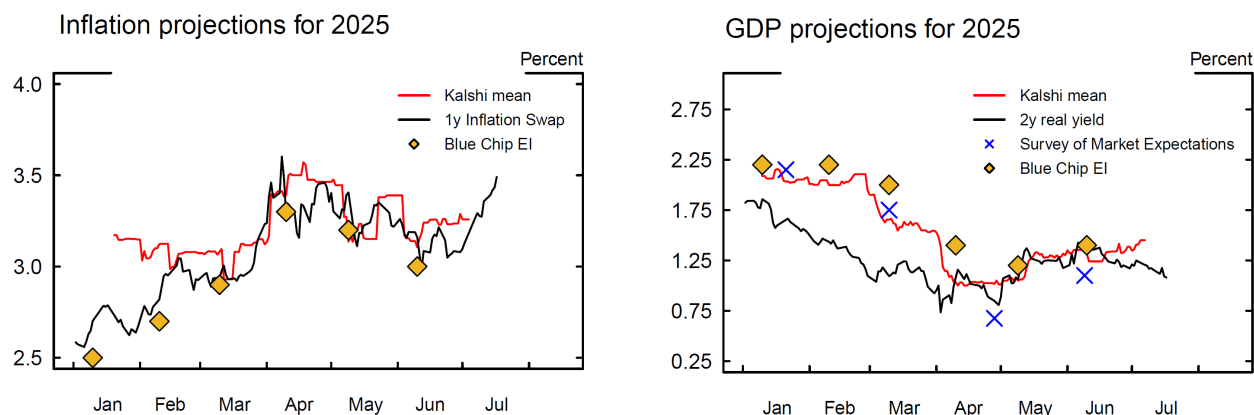


Figure A.10: Year of 2025 Comparisons

The chart on the left shows the year of 2025 CPI inflation projection for Kalshi, a one-year inflation swap, and the Blue Chip Economic Indicators consensus expectation for the year 2025. The chart on the right shows the real GDP growth projections for 2025 from Kalshi, the SME, and the Blue Chip Economic Indicators. Also included is the TIPS real 2-year yield as a proxy for real growth.

2025, both forecasts overestimated the actual unemployment rate, but the Kalshi mean was closer to the realized value than the Bloomberg Consensus.

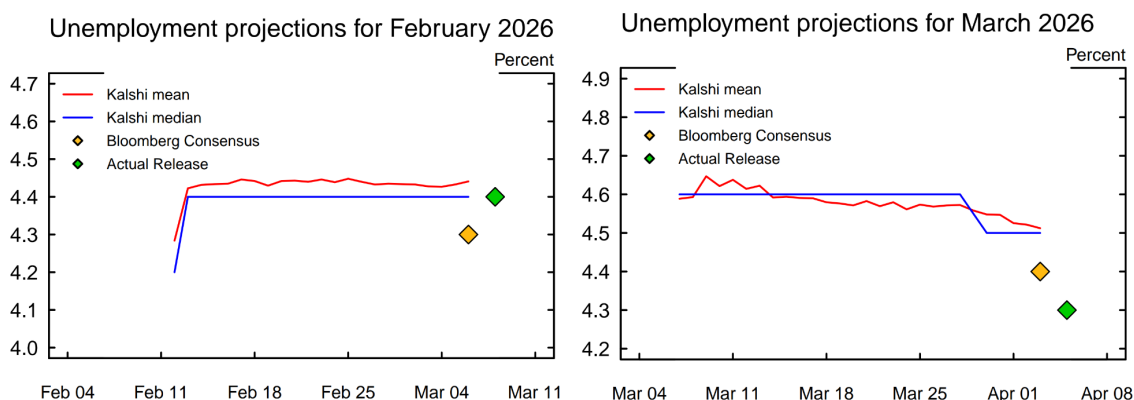


Figure A.11: Unemployment Comparisons

The chart on the left shows the unemployment projections for May 2025. The chart on the right shows the unemployment projections for June 2025.

Kalshi and SOFR comparison Figure A.12 shows that the SOFR option-implied distribution for Q4 2025 (the three months to mid-December) as of the summer of that year implied a modal outcome of no rate cuts. In contrast, the Kalshi distribution, constructed as the average of the September and October densities, along with Fed Funds futures, implied greater weight on a 25 basis point rate cut. This is likely driven, in part by the roughly

6 basis point spread between SOFR and the effective federal funds rate. While one could consider a crude lateral shift of the SOFR distribution, there's little evidence to suggest the SOFR-EFFR spread follows a normal distribution, so this adjustment could introduce further noise. The fact that the underlying SOFR rate is meaningfully higher implies an upward bias in the probabilities implied by the SOFR distribution. Since the Kalshi market is directly tied to the federal funds rate, it does not suffer from this issue making it the best available density forecast for the funds rate.

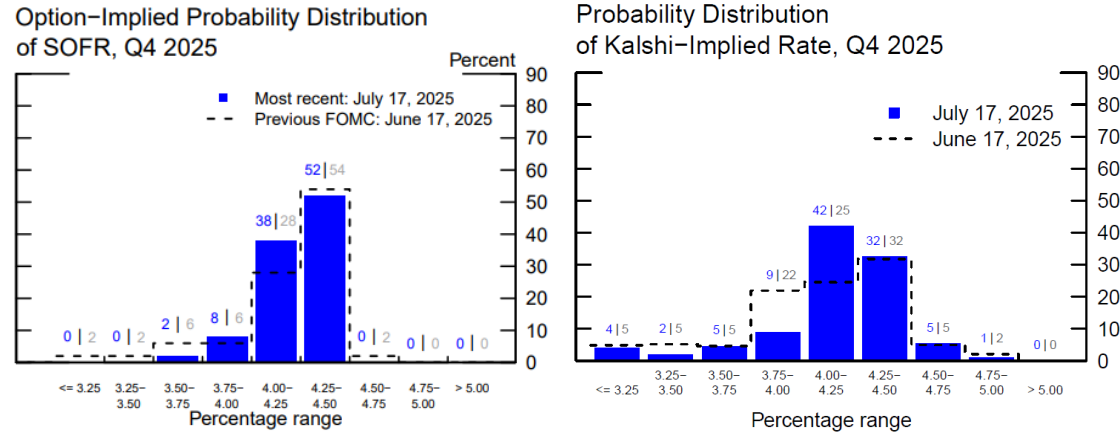


Figure A.12: **SOFR Options vs Kalshi Q4 2025 FOMC Meeting**

The left panel shows the implied probability from SOFR options for the fourth quarter of 2025. This is defined as mid-September to mid-December 2025. The right panel shows the equivalent Kalshi distribution, which is computed by averaging over the September and October 2025 FOMC meetings.