The Market-Perceived Monetary Policy Rule: Evidence from Prediction Markets*

EMREHAN AKTUĞ† *Sabancı University*

ORHAN TORUL‡ *Boğaziçi University*

October 16, 2026

Abstract

How strongly do markets expect the Federal Reserve to respond to economic news? We measure this market-perceived policy rule in real time using prediction markets. Kalshi prices both the coming CPI print and the next FOMC decision, so inflation news can be measured against the market's own expectation minutes before the release, and the policy response read from prices within the same hour. Combined with fed funds futures and Kalshi contracts on later CPI prints, this simultaneous measurement links the expectation of today's print, the policy response and expectations of future prints, rather than pairing a price response with a survey consensus collected days earlier. Over 2021–2026, a 1 percentage point upside CPI surprise raises the expected rate at the next FOMC meeting by 18 basis points, and other major releases move it with the signs a policy rule implies. Measured against the survey consensus instead, the same price changes imply a response a fifth to a quarter weaker. The response builds along the policy path, from 13 basis points at one month to about 72 at eleven months, and the same news raises the expected next CPI print. A hot print thus raises expected policy far beyond the next meeting and expected near-term inflation, all within the hour.

Keywords: Monetary policy rule; perceived reaction function; prediction markets; high-frequency identification.
JEL: E43, E52, E58, G13, G14.

*All errors are our own. Replication code and data construction scripts are available from the authors.

†Corresponding author. Address: Sabancı University Department of Economics, 34956, Tuzla, Istanbul, Türkiye. E-Mail: emrehan.aktug@sabanciuniv.edu

‡Address: Boğaziçi University, Department of Economics, 34342, Bebek, Istanbul, Türkiye. E-Mail: orhan.torul@bogazici.edu.tr

1 Introduction

At 8:15 on 13 September 2022, fifteen minutes before the release, the market-implied distribution of the August CPI print had a mean of −0.14pp. The print came in at +0.1pp, an upside surprise of 0.24pp. Within forty-five minutes, the priced federal funds rate for the September FOMC meeting rose 5.5 basis points on Kalshi. Behind this repricing is the object this paper measures: the market-perceived monetary policy rule, meaning how strongly markets expect the Federal Reserve to respond to economic news. We measure that expected response in real time, trace how far it reaches along the expected policy path, and observe the revision in near-term inflation expectations that accompanies it.

That object matters whatever rule the Committee actually follows. Monetary policy works largely through expectations of future policy, so the way yields respond to news reflects how markets believe the Fed will react, which need not coincide with how it does react (Piazzesi, 2005; Bauer *et al.*, 2024). It also matters for how the profession measures monetary policy. High-frequency identification reads the price change in a narrow window around an FOMC announcement as a monetary policy surprise, which presumes that markets had already priced the Fed's systematic response to the news that came before. Bauer and Swanson (2023a,b) document that policy surprises are in fact predictable from previously public news, and attribute this to markets underestimating how strongly the Fed responds to that news, which they call the "Fed response to news" channel. How strongly markets expect the Fed to respond to news is therefore central both to how news moves yields and to how policy surprises should be read.

Yet that expectation has been hard to observe at the moment the news arrives. Asset-price responses to macroeconomic announcements have been measured within minutes of the release for decades (Fleming and Remolona, 1999; Balduzzi *et al.*, 2001; Andersen *et al.*, 2003, 2007; Faust *et al.*, 2007; Gürkaynak *et al.*, 2020), but the surprise they respond to is almost always measured against a survey consensus.[1] Estimates of the perceived rule do the same, pairing a traded response with an expectation proxy from outside the market: survey forecaster panels (Bauer *et al.*, 2024), or daily futures moves measured against a survey median (Hamilton *et al.*, 2011). The difficulty is not that surveys are inaccurate. It is that the consensus need not equal the expectation embedded in prices just before the print.

Event contracts make a real-time measurement possible. Kalshi, a CFTC-designated contract market, lists ladders of binary contracts on the federal funds target set at each FOMC meeting and on the headline values of CPI inflation, unemployment, payrolls and GDP growth at each release. A contract such as "CPI above 0.1%" pays one dollar if the print exceeds that threshold, so its price is a market-implied probability, and the full ladder traces out a distribution for the coming print whose mean is the market-implied expectation. The surprise is the print minus that expectation, read fifteen minutes before the release. The response is the revision in the policy

[1] A notable exception is Gürkaynak and Wolfers (2007), who use the short-lived Economic Derivatives auctions and find that market-based surprises predict asset-price responses somewhat better than survey-based ones.

ladder for the next FOMC meeting over the following forty-five minutes. Both sides come from the same exchange on the same clock. Identification comes from the narrowness of the window: an audit of the BLS and Census release calendars identifies the other scheduled releases inside it, and dropping the CPI windows shared with jobless claims or retail sales leaves the estimate intact. The data are the complete trade histories of the Kalshi policy and macro ladders, 883,588 trades spanning 138 scheduled data releases and 41 FOMC meetings from July 2021 to July 2026, and four million CME futures observations inside the event windows. Fed funds futures, traded on the CME and long the standard market measure of expected policy, give a second, independent read of the same expectations. Kalshi's expected rate tracks the futures-implied rate closely, in level and in its revisions within release windows (correlations of 0.99 and 0.74), and the two venues give nearly the same response to CPI news. The futures also extend the response month by month along the expected policy path.

Our main contribution is to measure the market-perceived policy rule in real time. Survey–futures estimates of the perceived rule (Hamilton *et al.*, 2011; Lapp and Pearce, 2012) relate daily futures moves to surprises measured against a survey median. Here the news and the policy response are both priced within the same hour, and that matters for the answer: measured against the market's own expectation, the policy response to inflation news is stronger than against the survey consensus. The response arrives within the hour, reaches far along the expected policy path and comes with a revision in expected inflation. Relative to daily studies of Kalshi and macroeconomic news (Diercks *et al.*, 2026), the window is one hour rather than a day, and the surprise is market-based rather than survey-based. And relative to Swanson *et al.* (2026), who use Kalshi to study how monetary policy announcements move macroeconomic expectations, **we study the other direction: how macroeconomic news moves expected policy and, in the same window, expected future inflation**. This measurement delivers two facts.[2]

Fact 1: When inflation comes in above expectations, the expected federal funds rate for the next meeting rises within the hour, and other major releases move it with policy-rule signs. The response is stronger when surprises are measured against Kalshi's real-time expectations than against Bloomberg survey forecasts. A 1 percentage point upside surprise to monthly CPI raises the expected federal funds rate after the next FOMC meeting by 18 basis points within forty-five minutes. The baseline estimate rests on 50 release windows and is statistically significant under every inference method we use, including methods designed for samples of this size. Actual surprises are much smaller (a typical one is about a tenth of a percentage point), so on an ordinary CPI day the expected rate moves by about 2 basis points. Surprises in the other statistics move the expected rate in the directions a policy rule implies: higher year-over-year inflation, stronger payrolls and stronger GDP raise it, and higher unemployment lowers it. Each is statistically significant even under our most conservative test, which also corrects for testing five statistics at once. Payrolls and unemployment arrive in the same report, however, and once both enter together the unemployment estimate

[2]Appendix F separately explores dispersion and skewness; sparse strike updating makes those estimates exploratory rather than a third fact.

becomes imprecise, so we read these signs as corroborating the inflation response rather than as separately identified rule coefficients. Nothing in the construction imposes them: the policy and macro ladders trade in separate order books, and the estimation imposes no sign restrictions.

The comparison with the survey holds the realized print and the observed price change fixed and changes only the expectation subtracted from the print. Measured against the Bloomberg consensus rather than Kalshi's real-time expectation, the same price changes imply a response 4.5 basis points weaker on both venues: 13.5 rather than 18.0 basis points on Kalshi's policy ladder, and 17.1 rather than 21.6 on fed funds futures for the same releases. Both differences are statistically significant after correcting for the two comparisons. The reason is that the part of the survey surprise that differs from the market's expectation moves expected policy significantly less than the news itself, with an effect not distinguishable from zero, so it dilutes the survey measure of news. Part of that difference is resolution: a priced distribution records expectations more finely than a consensus reported to one decimal. The expectation used to define the news therefore changes the measured strength of the perceived rule, not just the source of its regressor, extending to the perceived policy rule what Gürkaynak and Wolfers (2007) found for asset prices.

The 18 basis points are also an average over a period in which the Fed did very different things: the response is roughly three times larger in the hiking period than in the hold period. Because those periods are contiguous, we treat this as heterogeneity consistent with state dependence, not as a causal separation of the policy stance from inflation, attention or calendar time.

Fact 2: The response builds along the expected policy path, and the same news raises the expected next CPI print. Fed funds futures price the expected average federal funds rate in each calendar month out past a year, so a single CPI print shows how far into the future the market carries the news. It carries it a long way, and by more the further out one looks. On the same 54 releases at both horizons, a 1 percentage point surprise that raises the expected rate one month ahead by 13 basis points raises the rate expected eleven months ahead by about 72.[3] The increase is about five times its standard error and statistically significant under every test we run. There is no detectable further increase between eleven and twelve months, and beyond a year the releases quoting each contract thin out, so its shape there is hard to pin down. The same build-up appears on a second and far more liquid strip, three-month SOFR futures, whose one-year response on the same releases matches the fed funds estimate almost exactly. Unemployment surprises trace the same pattern with the opposite sign.

Why does the response build? One reading is gradual policy adjustment: markets expect the Fed to move in steps rather than all at once. Another is that a hot CPI print signals inflation that will itself persist, so that later policy responds to a larger expected inflation gap. Both raise expected rates at longer horizons, so the policy strip alone cannot separate them. The second reading requires that the same news raise expected *future* inflation, and Kalshi's ladders on later

[3]The one-month figure is below the 18 basis points at the next meeting because a futures contract settles on the average rate over a calendar month, which can include days before the decision. Corrected for that averaging, futures give 20 basis points at the next meeting, in line with Kalshi.

CPI releases let us observe that revision directly, in the same window, rather than infer it from the policy strip. A 1 percentage point surprise raises the expected next monthly print by about a fifth of a percentage point, a robust estimate from 38 releases. The print after that, quoted in only 14 releases, also rises, though less precisely.[4] Ladders further out are too thinly quoted to use. These revisions do not decompose the policy response into inertia and inflation persistence. They show how market-implied expectations of policy and of near-term inflation revise together when the news arrives.

Section 2 places the paper in the literature. Section 3 describes the exchanges and the ladder construction, and Section 4 lays out identification. Section 5 reports the two facts: the real-time next-meeting response, its comparison with survey-based news and its variation across policy regimes in Section 5.1, then the policy-path response and the revision in expected future CPI prints in Section 5.2. Section 6 concludes.

2 Related literature

The paper speaks to four strands. First, Gürkaynak *et al.* (2005b) show that long yields respond to macro news as if markets price a policy rule; Piazzesi (2005) models perceived policy within the term structure. Hamilton *et al.* (2011) estimate a perceived rule from daily fed funds futures moves against survey-median surprises, including responses rising with maturity, and Lapp and Pearce (2012) study how economic news changes expected policy. Those results establish precedents for policy-rule signs and a response that rises with maturity. Our increment is real-time measurement: we define news against the settling macro ladder's own pre-release mean and observe the next-meeting policy repricing in the same hour, which also shows that survey-based news weakens the measured response. In the same window we observe revisions in deferred CPI prints. Hamilton, Pruitt and Borger instead use auxiliary forecasting equations to connect news to future fundamentals. The deferred ladders measure named future prints directly, but only at the short horizons available in the data. Bauer *et al.* (2024) establish perceived policy rules as an important object using Blue Chip forecaster panels; our release-window response is a financially backed, intraday counterpart. Herbst and Page (2026) find weak evidence of economically meaningful coefficient variation once panel pooling is estimated rather than imposed. We report contiguous-regime heterogeneity as a qualification to the average, not as a causal attribution to policy stance.

Within this strand, Fact 2 speaks to the distinction between policy inertia and persistent fundamentals. Rudebusch (2002, 2006) question whether large estimated smoothing coefficients represent genuine partial adjustment, and Coibion and Gorodnichenko (2012) separate inertia from serially correlated shocks. A rising release-window policy response admits the same two interpretations. We add an observable component of the fundamentals channel: the simultaneous revision in the next named CPI print, and a less precise revision in the second. These are additional

[4] On the 14 releases quoting both prints, the effect falls from 0.28 to 0.17 percentage points, a decline that is not statistically distinguishable from zero.

empirical objects, not a decomposition of the policy response. Appendix G.2 derives a single smooth profile from a simple adjustment model. Even within that model, the strip cannot say whether policy smoothing or fading inflation sets the pace of the response, and because meeting timing and monthly averaging also shape the curve, its fitted parameters describe the profile rather than measure structural adjustment speeds.

Second, a long literature measures intraday asset-price responses to macro news, in Treasury, foreign-exchange and equity markets (Fleming and Remolona, 1999; Balduzzi *et al.*, 2001; Andersen *et al.*, 2003, 2007; Faust *et al.*, 2007), and Gürkaynak *et al.* (2020) show how to recover the effects of releases whose surprises are only partly measured. In nearly all of it the price response is intraday but the expectation it is measured against is a survey consensus. Boehm and Kroner (2026) document the global effect of U.S. releases. Closest to our sample, Kroner (2025) measures 60-minute yield and inflation-swap responses and attributes stronger CPI reactions during the inflation surge to investor attention. Our distinction is not intraday timing alone: we observe a meeting-specific policy distribution against the market's own pre-release macro expectation, alongside deferred-CPI revisions. The hiking-period slope exceeds the hold-period slope, but the hiking period is also the inflation surge. The regime result is therefore compatible with an attention channel as well as state-dependent policy perceptions.

Third, the paper relates to the high-frequency identification literature descending from Kuttner (2001) and Gürkaynak *et al.* (2005a), to the information-effect debate it opened (Nakamura and Steinsson, 2018; Bauer and Swanson, 2023a), to the measurement conventions consolidated by Altavilla *et al.* (2019), and to tests of whether policy surprises are predictable (Romer and Romer, 2000; Cieslak, 2018; Schmeling *et al.*, 2022; Bauer and Swanson, 2023b). Altavilla *et al.* (2019) read a maturity structure of responses to policy *announcements* and decompose it into target and forward-guidance components; the profile in Section 5.2 is the same kind of object measured on the other side of the rule, tracing how a *data* surprise propagates across the expected policy path. Our CPI-release regressor combines news about today's inflation with any persistence markets infer from it. Observing deferred CPI revisions in the same window supplies direct short-horizon evidence about that inference, rather than extracting it from the policy strip. Appendix H separately reports a negative association between prior CPI news and FOMC-window repricing, the opposite sign to the pre-2020 predictability that Bauer and Swanson (2023a,b) attribute to the "Fed response to news" channel. Its timing and inference vary across windows and instruments, so it is supplementary evidence rather than a main fact or a validated policy-shock correction.

Fourth, a growing literature uses event markets as macro data. Gürkaynak and Wolfers (2007) compare market- and survey-based surprises in intraday asset-price responses using Economic Derivatives auctions; our application adds meeting-specific policy repricing and revisions to expectations of named future CPI prints. Snowberg *et al.* (2013) survey earlier prediction markets. On Kalshi, Diercks *et al.* (2026) validate distributions and regress one-day changes in next-FOMC moments on realized-minus-Bloomberg surprises. We instead define the surprise from the settling ladder itself and measure the policy response in a common one-hour window. Swanson *et al.*

(2026) use front and deferred macro ladders to study how policy news moves macro expectations at daily frequency; our release design studies how current macro news moves expected policy and expected future prints. Angelini (2026) measures high-frequency CPI expectations before releases and their response to earlier survey surprises; we couple the current print's market-based surprise to simultaneous policy and deferred-print revisions. Goel (2026) studies the information in pre-release dispersion, whereas our regressor uses the mean. Bürgi *et al.* (2026) document contract-level pricing and favorite-longshot bias. Our increment is thus the real-time measurement of the perceived rule from this source, not priority in Kalshi distributions, perceived rules or deferred inflation expectations. Appendix Table A.1 sets out the comparison with each of the closest designs. Sections 3 and 4 explain the construction and inference needed to make that comparison credible.

3 Data

3.1 Event-contract ladders

Kalshi is a CFTC-designated contract market that lists binary event contracts. Each CPI release and each FOMC meeting has its own event, which lists a ladder of contracts at different thresholds, a median of eight for CPI and eleven for the policy rate. A contract such as "CPI above 0.1%" pays one dollar if the official figure exceeds its threshold and nothing otherwise. CPI contracts settle on the single-decimal monthly change published by the BLS, and policy contracts on the upper bound of the federal funds target range that the Federal Reserve publishes after the meeting. Prices are quoted in cents, so a price reads directly as a probability. A CPI event opens about two months before the release and a policy event about ten months before the meeting, and trading stops minutes before the print or the statement. Table 2 below shows one CPI ladder and one policy ladder around a release.

The design uses six series. The policy outcome is the upper bound of the federal funds target range set at each FOMC meeting. The news comes from five headline statistics: monthly CPI inflation, year-over-year CPI inflation, the U-3 unemployment rate, the monthly change in nonfarm payrolls and advance real GDP growth.[5] These series account for 883,588 of the 1,270,649 trades, all with microsecond timestamps, across the 39 nonempty Kalshi series we collect. Table 1 reports coverage for each.

[5]The Kalshi tickers are KXFED for the policy rate, KXCPI for monthly CPI, KXCPIYOY for year-over-year CPI, KXU3 for unemployment, KXPAYROLLS for payrolls and KXGDP for GDP. Earlier payroll contracts traded under other tickers, which Appendix B combines into one series.

Table 1: Data coverage

Series	Ticker	First traded	Markets	Trades	Volume
Funds target upper bound	KXFED	Jul 2021	489	209,296	56.6m
Monthly CPI (% MoM)	KXCPI	Jun 2021	533	175,105	22.6m
CPI (% YoY)	KXCPIYOY	Dec 2022	692	204,585	34.2m
U-3 unemployment (%)	KXU3	Jul 2021	627	71,492	10.0m
Payrolls change (jobs)	KXPAYROLLS	Mar 2022	565	110,077	18.7m
Advance GDP (% SAAR)	KXGDP	Jun 2021	225	113,033	14.6m

Notes: "Markets" counts contracts and "Volume" contracts traded. The payroll row combines KXPAYROLLS with its predecessor KXPROLLS and the briefly traded KXUSNFP (Appendix B). Releases in the sample (July 2021 to July 2026): 59 CPI, 58 Employment Situation, 41 FOMC, 21 GDP. BLS/Census and Federal Reserve dates are validated; BEA co-releases remain the residual calendar exposure, and five windows whose dates were moved or canceled during the 2025 shutdown are excluded. CME futures, from Databento: 4,035,919 observations of fed funds (ZQ) and three-month SOFR (SR3) futures.

For the survey benchmark, we use Bloomberg consensus expectations and first-release actuals for headline, seasonally adjusted monthly CPI. The series are aligned by publication date, not the month whose prices are measured. Overlapping actuals agree with the Kalshi settlement values for 54 ordinary monthly prints. The comparison retains the same policy outcome, release dates and price window for both expectation measures. Bloomberg supplies a median consensus recorded to one decimal place; Kalshi supplies a probability-weighted distribution mean at the pre-release cutoff. The Bloomberg data do not record when respondents submitted their forecasts or when the consensus was taken.

3.2 From ladders to densities

At any instant, the price of a contract such as "CPI above 0.1%" is the market-implied probability that the print exceeds 0.1%, so the last-trade prices across a ladder trace out a survival function. Differencing adjacent prices gives the probability that the print lands between two strikes: if "above 0.0%" trades at twelve cents and "above 0.1%" at five, the implied probability of a print in $(0.0, 0.1]$ is seven percent. Because last trades are not synchronous, adjacent prices can occasionally violate the ordering a survival function requires, so we enforce monotonicity before differencing. On the one-decimal CPI grid the interval $(0.0, 0.1]$ contains only one attainable print, 0.1%. Wider observed intervals can contain several attainable outcomes; exterior tails are not located by the ladder. These distinctions matter when converting bin probabilities into moments.

Formally, let a ladder on X list strikes $s_1 < \cdots < s_K$, and let $p_i(t)$ be the price of the contract paying \$1 if $X > s_i$. Write $\widehat{p}_i(t)$ for the volume-weighted monotone projection of those prices. The baseline probability-scale convention is

$$\widetilde{P}_i(t) \equiv \widehat{p}_i(t) \approx \mathbb{P}_t^{\mathbb{Q}}(X > s_i). \tag{1}$$

Differencing across adjacent strikes, with two exterior masses closing the distribution, gives

$$q_i = \widetilde{P}_i - \widetilde{P}_{i+1} \quad (i = 1, \ldots, K-1), \qquad q_0 = 1 - \widetilde{P}_1, \qquad q_K = \widetilde{P}_K. \tag{2}$$

For CPI, year-over-year CPI, U-3 and the funds target, the baseline represents an interior bin by its upper quoted boundary. This is exact if adjacent observed strikes are one publication-grid step apart, but an approximation when missing or untraded strikes leave a wider bin. For example, $(1,2]$ on the 25bp target grid contains 1.25, 1.50, 1.75 and 2.00. A support-aware robustness variant represents the bin mean by the average of its first and last attainable values, $(s_i + g + s_{i+1})/2$ for grid step g; this corresponds to equal allocation over its attainable interior outcomes, without claiming to identify that allocation. GDP and payroll ranges use interval or bucket midpoints. The resulting discrete approximation $\{(x_i, q_i)\}$ has mean

$$\mathbb{E}_t^{\mathbb{Q}}[X] = \sum_{i=0}^{K} q_i \, x_i, \tag{3}$$

together with variance and skewness under the same support conventions. Throughout, $\mathbb{E}^{\mathbb{Q}}$ denotes this reconstructed price-implied moment, not an exactly recovered physical expectation or a uniquely located full distribution. The risk-neutral interpretation is maintained rather than guaranteed by collateralization. We use Kalshi prices as they trade, without discounting them at the policy rate; Appendix B.1 explains why, and discounting every price at the target rate leaves the CPI estimate at 0.181, against 0.180 in the baseline. Section 4 explains why moving valuation premia can still enter the release-window response.

The two objects the paper regresses follow directly. For a release at t of the statistic X_t, and with F_m the funds-rate target upper bound the FOMC will set at the next scheduled meeting m,

$$\underbrace{s_t = X_t - \mathbb{E}_{t-15}^{\mathbb{Q}}[X_t]}_{\text{the surprise}}, \qquad \underbrace{\Delta f_t = \mathbb{E}_{t+45}^{\mathbb{Q}}[F_m] - \mathbb{E}_{t-15}^{\mathbb{Q}}[F_m]}_{\text{the response}}. \tag{4}$$

The two are read from *different* ladders. The surprise comes from the ladder that settles on today's print, evaluated fifteen minutes before it; that contract stops trading minutes before the release, which is why the surprise must be measured against the print and cannot be read as a change in the ladder itself.[6] The response comes from the policy ladder for meeting m, which no settlement inside the window disturbs.

As an example, Table 2 works through a release step by step. On 13 September 2022 the August CPI ladder listed fifteen contracts. Representing each bin by its upper boundary gives a pre-release mean of −0.14pp, so the published +0.1pp print is a +0.24pp surprise. The policy ladder for the 21

[6]The obvious alternative is to difference the macro ladder across its own release. That construction compares this month's settling contract before the print with next month's live contract afterward, so it changes the event as well as the expectation. The resulting slope is −0.020, against +0.180 for the realized print minus the fixed settling ladder's pre-release mean.

September meeting moves from 3.185 to 3.240, or +5.5bp.

Table 2: One release, strike by strike: the August 2022 CPI print

Panel A: the settling CPI ladder (KXCPI, event `CPI-22AUG`) at 08:15 ET

Contract	Price	$\widehat{\mathbb{P}}^{\mathbb{Q}}(X > s)$	Implied mass
(residual below the lowest strike)	—	—	0.16
Above −0.4%	0.84	0.84	0.14
Above −0.3%	0.70	0.70	0.20
Above −0.2%	0.50	0.50	0.21
Above −0.1%	0.29	0.29	0.17
Above 0.0%	0.12	0.12	0.07
Above 0.1%	0.05	0.05	0.03
Above 0.2%	0.02	0.02	0.01
Above 0.3%	0.01	0.01	0.00
(seven further strikes to 1.0%*)*	0.01	0.01	0.01

Panel B: the policy ladder (KXFED, event `FED-22SEP`), before and after

Contract	Price at 08:15 ET	Price at 09:15 ET	Change
Above 0.75%	0.98	0.99	+0.01
Above 1.00%	0.99	0.99	0.00
Above 2.00%	0.99	0.99	0.00
Above 2.50%	0.99	0.99	0.00
Above 2.75%	0.99	0.99	0.00
Above 3.00%	0.81	0.96	+**0.15**
Above 3.25%	0.01	0.08	+**0.07**
Above 3.50%	0.01	0.01	0.00

Panel C: the resulting observation

$\mathbb{E}^{\mathbb{Q}}_{t-15}$[CPI], from Panel A	−0.14pp
realized print	+0.10pp
surprise s_t	+**0.24pp**
$\mathbb{E}^{\mathbb{Q}}_{t-15}[F_m]$, from Panel B	3.185%
$\mathbb{E}^{\mathbb{Q}}_{t+45}[F_m]$	3.240%
response Δf_t	+**5.5bp**

Notes: One observation rebuilt from traded prices. The CPI interior strikes shown are adjacent on the one-decimal grid, so their bin masses have exact interior support; exterior masses use the baseline boundary closure. The policy ladder contains wider gaps, so its mean, which represents each bin by its upper boundary, is an approximation. Support-aware interior allocation and Gaussian exterior tails are separately reported robustness variants.

Four construction choices are explicit: monotonicity, within-bin support, exterior-tail placement and a cap on the age of a strike's last trade. Appendix B records them, and Appendix D re-estimates the response under alternatives. Differencing does not automatically eliminate their effects: bin masses and tail probabilities can change inside a window, and the macro regressor uses a pre-release level. Monotonicity materially changes the headline estimate; tighter freshness caps

change both its measurement and its usable sample.

Of the 59 CPI releases, 50 have both a reconstructible pre-release CPI density and a usable next-meeting policy outcome, and these form the CPI estimation sample. The surprise ladder is fresher than the policy ladder. In that sample, half of the settling CPI ladder's probability mass sits, in the median window, on strikes that last traded within about seven minutes, and 99% on strikes that traded within the previous day. Across these same windows, the median post-release interval has 11 policy trades updating two contracts; 6 windows have no post-release policy trade and only 16 update at least three contracts. Cumulative historical volume therefore overstates event-window participation. The response remains positive in windows with observable activity: the CPI coefficient is 0.180 in all windows, 0.187 when at least one post-release trade occurs, 0.193 when at least three contracts update, and 0.201 when post-release volume exceeds 1,000 contracts. Because those active-window selections are endogenous, they diagnose sensitivity to thin trading rather than establish that every snapshot incorporates current information. Appendix Tables B.3, B.4 and B.5 report the sample construction, activity and freshness.

3.3 Fed funds futures and cross-venue validation

We use CME 30-day fed funds futures (ZQ) for external validation and the term structure, and three-month SOFR futures (SR3) to re-estimate the term structure on a second, more liquid strip; both are recorded at one-second resolution (4.0m observations inside the relevant windows). A ZQ contract settles on the *monthly average* effective rate, not the rate immediately after one meeting. For macro-release validation we therefore solve the calendar identity $DA = d\, r_{\text{pre}} + (D-d)\, r_{\text{post}}$, where A is the contract's implied monthly-average rate, D the number of days in the month, d the number of days before the decision takes effect, and r_{pre} and r_{post} the effective rates before and after it. Using the prevailing rate for r_{pre} and the exact meeting timestamp for d, this de-averages the meeting-month contract into an expected post-meeting rate. This is the standard correction for monthly averaging (Kuttner, 2001; Ottonello and Winberry, 2020): with the prevailing rate fixed inside the window, it scales the change in the contract's rate by $D/(D-d)$. When fewer than six days of the month remain after the decision, solving for r_{post} magnifies small quoting errors in A, so we read the following month's contract instead (Appendix B). If the current-month contract is unavailable, a following-month contract is used only when no later meeting changes the effective rate before that month ends. Releases are mapped to meetings by the statement's timestamp rather than its calendar day, which matters when a CPI print and a decision fall on the same day (Appendix B).

The corrected futures level and the Kalshi target midpoint correlate at 0.99 over the 49 common CPI windows, and their window changes correlate at 0.74 over the 48 with a change on both venues. The remaining level gap averages 5.9bp, with a standard deviation of 5.3bp and no window above 25bp. A gap of that size is expected: Kalshi settles on the target range, which we convert to its midpoint, while fed funds futures settle on the effective rate, which sits inside the range rather

than at its midpoint. Re-estimating the CPI response with the de-averaged futures outcome gives 0.199 on 51 releases, statistically significant and within two basis points of the estimate from Kalshi. Figure 1 shows both comparisons. Two instruments with no shared exchange, collateral rule or strike grid agree on the level of expected policy and on how it is revised inside a release window.

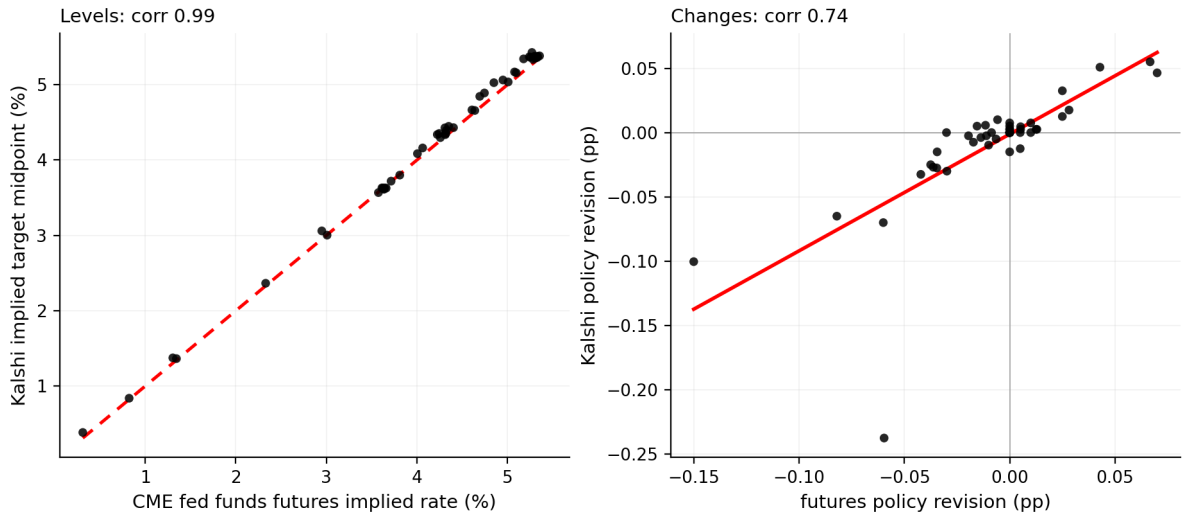


Figure 1: Cross-venue validation, measured in CPI release windows

Notes: CPI release windows on the same clock. Kalshi is expressed as the target midpoint. The CME rate is de-averaged for the exact upcoming meeting using the prevailing effective-rate proxy; a following-month contract is used only when it is not affected by another decision inside that month. The left panel has 49 common levels and the right panel 48 common changes.

4 The release-window design and identification

Let f_t denote the market-implied expected federal funds target (the upper bound of the range) set at the next FOMC meeting, that is, the mean of the KXFED ladder for that meeting, and let s_t be the release surprise: the printed statistic minus the mean of the settling ladder read fifteen minutes before the release. The estimating equation, one release family at a time, is

$$\Delta f_t = \alpha + \sum_{j \in J} \beta_j \, s_{j,t} + \varepsilon_t, \tag{5}$$

where Δf_t is the change in the implied next-meeting rate from -15 to $+45$ minutes around the release and J is the set of headline statistics represented by traded ladders for that release family. We estimate monthly CPI and year-over-year CPI separately as alternative summaries of the Consumer Price Index report, advance GDP as the headline summary of the BEA report, and U-3 and payrolls both one at a time and jointly, because both print in the Employment Situation. The coefficient β_j is the repricing associated with one unit of headline news.

Identification comes from the release calendar, and the calendar is checked rather than assumed. Appendix C reads every scheduled BLS and Census release over the sample from the agencies'

own schedules; BEA releases are not covered and remain the residual exposure. Real Earnings accompanies CPI but is an arithmetic transformation of the CPI innovation and previously released earnings. Twelve CPI windows, all Thursdays, also contain weekly jobless claims, and one further window shares its slot with advance retail sales; excluding all thirteen leaves the CPI estimate intact. Because the window closes 45 minutes after the print, later same-day news, such as 10:00 releases and afternoon Fed speeches, stays out of the outcome. The Employment Situation is the case in which two headline ladders are both available: payrolls and U-3 print together, so their joint regression reports conditional associations. Their surprises correlate at -0.28, and both marginal and joint estimates are reported.

Two features of the design deserve to be stated before any estimate, because they bound what the coefficient can mean rather than how precisely it is measured. The first is that both sides of (5) are estimated from finite trade tapes. Classical error in the surprise attenuates the slope; correlated or news-dependent construction errors can do more. We use repeated split-tape IV as a sensitivity check: rebuild each panel from a random half of the trades and instrument one half's surprise with the other's. Under independent measurement errors and instrument exclusion, this removes the component of regressor error that differs across halves.

The second is valuation. Read as probabilities, binary prices are risk-neutral, not physical. Differencing removes an additive, locally constant wedge from the policy response but not a news-dependent one, and the surprise is measured against a pre-release level, so any wedge in the settling ladder enters it directly. The coefficients are therefore reduced-form responses of market-implied expectations.

5 Results

The results are organized around two facts. Fact 1 is the real-time next-meeting response: to inflation news, to the other major releases, and to inflation news measured against the survey consensus instead of the market's own expectation. Fact 2 traces the response along the policy path and directly observes how the same news revises the expectation of the next named CPI print.

5.1 Fact 1: When inflation comes in above expectations, the expected federal funds rate for the next meeting rises within the hour, and other major releases move it with policy-rule signs. The response is stronger when surprises are measured against Kalshi's real-time expectations than against Bloomberg survey forecasts

5.1.1 Baseline estimates

Figure 2 shows the variation the design rests on, and it is visible before any regression is run: on CPI days the priced next-meeting rate moves up with the surprise, and on Employment Situation

days it moves down with the unemployment surprise. Table 3 puts numbers on both.

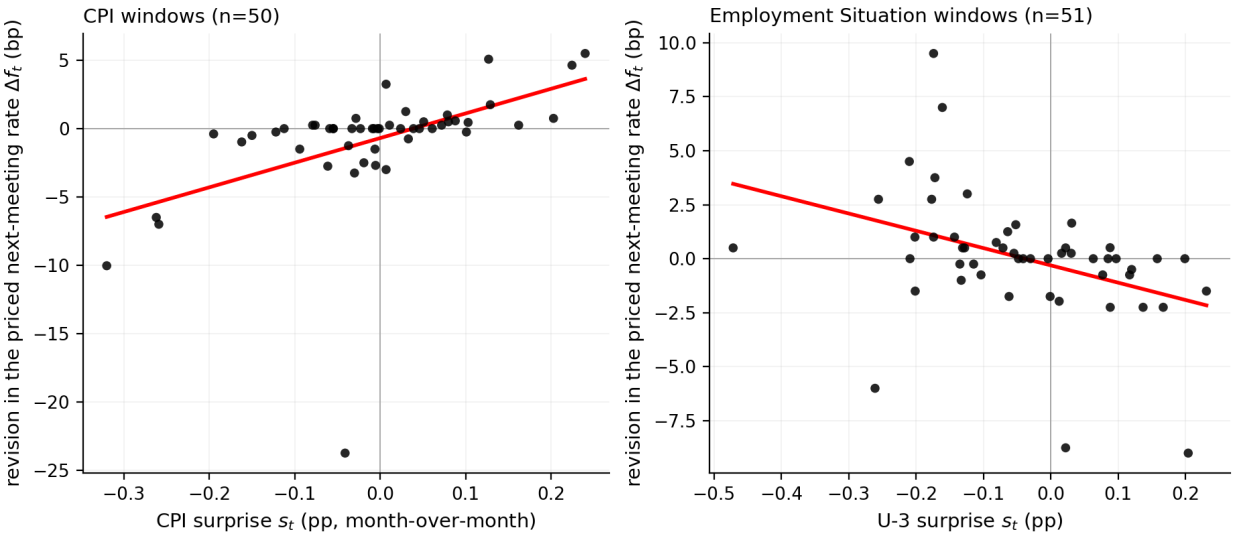


Figure 2: The identifying variation in the release windows

Notes: Each point is one release window. The vertical axis is the revision in the market-implied next-meeting funds rate, Δf_t, and the horizontal axis the release surprise, s_t, both as defined in equation (4): the surprise is the printed statistic minus the mean of the ladder settling on it fifteen minutes earlier, and the revision is the change in the mean of the policy ladder from −15 to +45 minutes around the release. Left: the 50 CPI windows. Right: the 51 Employment Situation windows, with the U-3 surprise on the horizontal axis. Red lines are ordinary least squares fits; their slopes are the β_π and β_u of Table 3. Surprises are in percentage points and revisions in basis points, so a point at $(0.1, 2)$ is a release that printed 0.1pp above the market's density and moved the priced next-meeting rate by two basis points. The single CPI observation near −24bp is 13 November 2024, when one large sell order swept a thinly traded strike of the December policy ladder inside the window (Appendix D.2). We keep the window, since no release is excluded on its outcome; dropping it lowers the CPI slope from 0.180 to 0.169, and Appendix Figure D.1 shows that no individual release drives either slope.

A 1pp upside surprise to monthly CPI raises the priced next-meeting funds rate by 18.0bp. The standard deviation of the surprise in the estimation sample is 0.116pp, so the fitted revision for a one-standard-deviation print is about 2.1bp. This is economically visible in a one-hour window but should not be compared mechanically with the inflation coefficient in a static Taylor rule (Taylor, 1993): a monthly print directly enters measured twelve-month inflation, while the coefficient here concerns one expected meeting and may also incorporate news about later inflation. The next-meeting estimate is the front of the response. Section 5.2 asks how much larger the same release-window response becomes farther along the path without assuming why it becomes larger.

Table 3: Release-window policy repricing: baseline coefficients

	Units	n	β	se	R^2	joint β
β_π	pp mo. CPI	50	$+0.180^{***}$	(0.035)	0.25	—
β_π^{yoy}	pp YoY CPI	41	$+0.080^{**}$	(0.036)	0.04	—
β_u	pp U-3	51	-0.080^{**}	(0.038)	0.15	−0.056
β_{nfp}	100k jobs	50	$+0.016^{**}$	(0.006)	0.24	+0.013
β_y	pp GDP	18	$+0.008^{**}$	(0.003)	0.46	—

Notes: Marginal rows regress the $[-15, +45]$-minute next-meeting revision on the settling-ladder surprise. Coefficients are percentage points of the expected rate per unit of surprise, so 0.180 is 18.0 basis points per percentage point of monthly CPI. HAC standard errors are in parentheses. Stars denote 1%, 5% and 10% significance using the largest HAC, restricted-wild, studentized-permutation and Romano–Wolf p-value. All five marginal rows pass 5%; monthly CPI passes 1%. Joint Employment estimates have $n = 50$, $R^2 = 0.30$ and surprise correlation −0.28, and are unstarred because the familywise exercise covers marginal specifications.

The evidence has a clear hierarchy. Monthly CPI is the headline margin: it is significant at 1% even under the conservative rule, its leave-one-release-out sign never changes, and the de-averaged CME outcome gives a statistically compatible estimate. All five marginal coefficients pass the conservative 5% rule, which takes the largest of HAC, restricted-wild, studentized-permutation and Romano–Wolf p-values, each reported in Appendix Table D.1. Employment Situation news supplies a joint sign check rather than two separate experiments. Marginally, a 1pp unemployment surprise lowers the priced rate by 8.0bp and 100,000 additional payroll jobs raise it by 1.6bp. When both simultaneous surprises enter together, the payroll coefficient is +1.3bp, while U-3 is −5.6bp and imprecise. The payroll signal is the sharper conditional margin, and the joint regression has $R^2 = 0.30$. Advance-GDP news raises the priced rate by 0.8bp per percentage point of annualized growth, on eighteen windows. Year-over-year CPI gives +8.0bp per pp and remains significant at 5% after correcting for testing five statistics at once.

These baseline estimates establish the first part of Fact 1: a positive, statistically significant response to monthly-CPI news, corroborated by the other marginal policy-rule signs. The second part holds the observed repricing fixed and asks how much the estimated response depends on the expectation used to define the surprise.

5.1.2 Kalshi news versus Bloomberg news

We compare market-based and survey-based news on identical releases, keeping the realized CPI print and the observed policy-window change fixed. Only the expectation subtracted from the print changes. Across 49 regular monthly-CPI releases, the Kalshi next-meeting response is 18.0 basis points per percentage point of Kalshi news and 13.5 per percentage point of Bloomberg news (Table 4). Fed funds futures give the same contrast on the same releases: for the 47 of them with a

futures outcome, the response is 21.6 basis points to Kalshi news and 17.1 to Bloomberg news.[7] The difference is 4.5 basis points on both outcomes, about a quarter of the Kalshi-news response on Kalshi's policy ladder and a fifth on futures. It is statistically significant on both after correcting for the two comparisons, and paired block-bootstrap intervals exclude zero. Both measures imply a positive response; the substantive difference is how strongly the market appears to respond.

The futures row matters because its outcome comes from a different venue, with its own traders, collateral and contract design, so its difference cannot reflect a pricing error shared by Kalshi's CPI and policy ladders. Its releases are those of the Kalshi row with a futures outcome, and on those 47 releases the Kalshi-row difference is nearly unchanged (notes to Table 4).

The intuition is that a print matching the consensus need not be no news to the market. Suppose, for illustration, that the consensus is 0.3%, the priced distribution has a mean of 0.27%, and the print is 0.3%. The survey records no surprise, while the traded distribution implies 0.03pp of upside news. The same observed policy revision is thus assigned different news content. A narrow response window does not remove this difference: the expectation benchmark is part of measuring the innovation to which prices respond.

Table 4: The next-meeting response to Kalshi news and to Bloomberg news

Policy outcome	n	Kalshi news	Bloomberg news	difference	95% block interval	adjusted p
Kalshi next meeting	49	18.0	13.5	$+4.5^{**}$	[1.1, 10.0]	0.037
CME next meeting	47	21.6	17.1	$+4.5^{***}$	[1.1, 7.0]	< 0.001

Notes: All coefficients and differences are basis points per 1pp monthly-CPI surprise. Within each row, the releases, realized prints, price outcome and $[-15, +45]$-minute window are identical across benchmarks. The difference is Kalshi minus Bloomberg. Its joint estimating-equation HAC test uses four lags and retains cross-regression covariance; the reported p-value and stars adjust for the two contrasts using Bonferroni, separately from the five-statistic exercise in Table 3. Intervals are unadjusted 95% paired circular-block bootstrap intervals, with blocks of four releases and 49,999 draws. The Kalshi row uses the 49 regular monthly-CPI releases with a Kalshi policy outcome, which leave out the 18 December 2025 release of a two-month change (Appendix C); the CME row uses the 47 of them for which the de-averaged futures outcome is available. On those 47 releases the Kalshi-row difference is +4.4.

The intuition can be tested directly. The Bloomberg surprise is the Kalshi surprise plus the gap between the two expectations, Kalshi's mean minus the Bloomberg consensus. Entering the two components separately asks whether the gap moves expected policy as much as news does, as it would if the survey surprise were the right measure of news. It does not. On Kalshi's policy ladder the gap's coefficient is −15 basis points per percentage point, against 17 for Kalshi news; on fed funds futures it is −14, against 21. Neither gap coefficient is statistically distinguishable from zero, while equal weights on the gap and on news are rejected at 1%. Because the Bloomberg slope is exactly a weighted average of the two loadings, with a weight of about 0.11 on the gap, the gap's

[7] After correcting for the two comparisons, the futures difference remains positive in all 47 leave-one-release-out regressions and significant at 5% in 46 of them; the remaining deletion is significant at 10%.

much lower loading accounts for the whole difference: it dilutes the survey measure of news, which is why the response to Bloomberg news is weaker.

The priced expectation also preserves finer information than a point forecast reported to one decimal. Rounding the Kalshi mean to that precision lowers the next-meeting response from 18.0 to 15.0 basis points, closer to the survey estimate. For futures it falls from 21.6 to 19.1. Timing, precision and mean-versus-median aggregation can all contribute to the gap between the two expectations; the Bloomberg data cannot separate them. Monthly forecast accuracy is similar, so this is not a claim that Kalshi invariably predicts the CPI print better. The result concerns the expectation relevant to observed policy repricing: a directly priced pre-release distribution changes the perceived response recovered from the same announcements.

Timing can be examined within the same market. We rebuild the Kalshi expectation one, three and seven days before each release with the identical construction, and split the surprise against that earlier expectation into the baseline surprise and the intervening drift. The drift has no statistically detectable loading in the release-window response on either venue, consistent with the policy price having absorbed it before the window opens. Measured against the week-old expectation, the estimated response falls to 86 percent of the baseline on Kalshi and 82 percent on CME (Appendix Table E.1). An expectation a day old changes the estimate little. This is separate evidence that aging the priced expectation changes the measured response; it complements the actual Bloomberg comparison without assigning its entire difference to timing. Appendix E.2 compares Kalshi and survey forecasts of the FOMC decision itself.

5.1.3 Robustness and threats to identification

The 30 alternative constructions in Appendix D vary the window, support and tail assumptions, freshness and sample composition. Every reported CPI estimate is positive. The CME outcome, measured on a different venue and contract, gives a response close to the Kalshi baseline. Support-aware allocation within sparse bins and a hypothetical discounting sensitivity barely move the headline CPI slope; Gaussian tails test the unobserved exterior values separately. Turning off monotonicity, which a survival function requires but asynchronous last trades can violate, lowers the slope, so that restriction is consequential rather than cosmetic.[8]

Re-measuring the surprise on core CPI gives +0.169 against the headline +0.180. Core is less liquid, so this is supportive rather than a test of which index defines the perceived response.

A first threat is simultaneous news. Excluding the twelve Thursday windows that also contain weekly claims moves the coefficient from 0.180 to 0.186, and excluding as well the one window shared with advance retail sales gives 0.190.

A second is that the surprise leaks into prices before the release. Rebuilding the panel with

[8]We restore monotonicity by volume-weighted isotonic regression. Isotonicity improves fit and changes the slope, but neither fact establishes that it removes only noise: transforming the regressor can change its covariance with the outcome without using that outcome, and the adjustment diagnostics show some association with surprise magnitude. We therefore do not interpret it as a proven measurement-error correction.

a 75-minute pre-window and regressing the pre-release drift (from −75 to −15 minutes) on the eventual surprise gives an estimated slope that is nearly zero and insignificant on every test we run. Figure 3 below repeats the check over the 30 and 15 minutes before the snapshot, on both venues.

A third is the choice of endpoint. Figure 3 holds the releases and the pre-release snapshot fixed and moves the end of the window from 15 minutes to five hours after the print. The two points before the release are a placebo: over the 30 and 15 minutes ending at the pre-release snapshot, expected policy barely moves with the coming surprise. The slopes are within 0.1 basis point of zero on Kalshi and at most 1.7 on futures, against about 20 after the print, and none is statistically significant at 5%. Fed funds futures price the news almost at once: their response is 20.4 basis points per percentage point at 15 minutes and stays between 20.4 and 21.9 through two hours, where the futures data end. The thinner Kalshi ladder catches up more slowly, from 14.6 basis points at 15 minutes to 18.0 at 45, and then holds between 17.3 and 20.3 out to five hours. After 45 minutes the response changes by less than 2.5 basis points at any endpoint on either venue, so the baseline window captures the repricing and the repricing does not reverse. Longer windows also admit later news, which is why the baseline stops at 45 minutes.

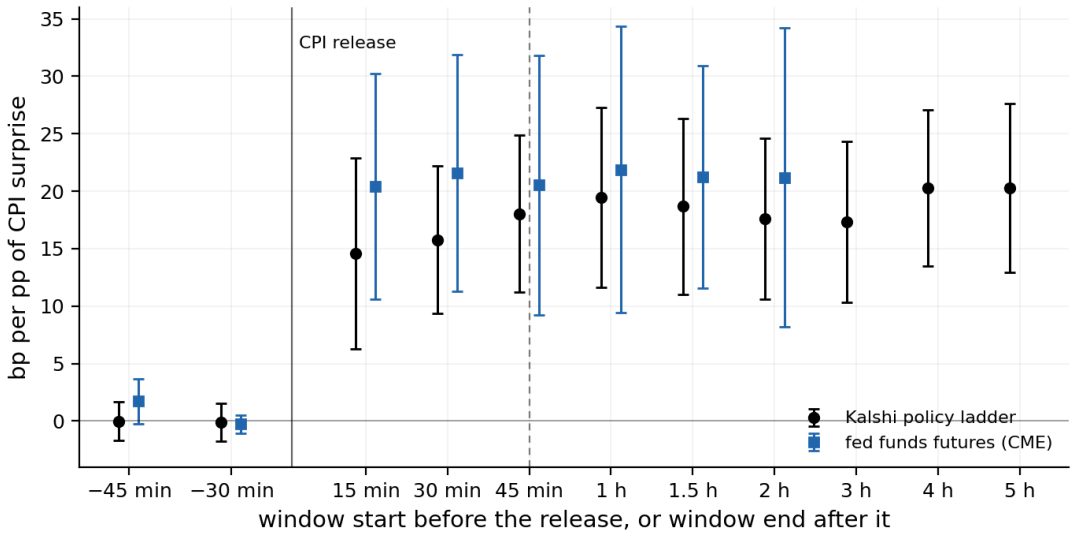


Figure 3: The next-meeting response to CPI news as the window is extended

Notes: Each point regresses the change in the expected next-meeting rate between 15 minutes before the release and the stated endpoint, 15 minutes to five hours after it, on the CPI surprise, with 95% HAC bands. The pre-release snapshot and the surprise are those of the baseline; only the endpoint changes. The points marked −45 and −30 min regress the change from that time to the pre-release snapshot on the same surprise, a placebo that is zero if the surprise was not anticipated; on futures it uses the 46 and 49 releases with a quote 45 and 30 minutes before the release. Kalshi uses the 50 baseline releases at every endpoint. Fed funds futures use the 49 releases with a de-averaged futures outcome at every endpoint from 15 to 120 minutes, where the CME data end. Windows are evenly spaced rather than to scale; the solid line marks the release and the dashed line the 45-minute baseline. Estimated release by release, the change from the 45-minute response is below 2.5 basis points at every later endpoint and insignificant under the restricted wild bootstrap. Longer windows also include later news, such as 10:00 releases and Fed speeches.

Tightening the staleness cap raises the headline coefficient, from 0.180 at the thirty-day baseline to 0.209 at seven days and 0.226 at one day, so fresher quotes, if anything, deliver a larger response. Part of the rise is sample selection: re-estimating the uncapped construction on the releases that survive each cap gives 0.190 at seven days and 0.220 at one day (Appendix Table D.3). We keep 0.180 as the baseline, since the capped estimates describe fresher but selected subsamples.

Two diagnostics for measurement error in the surprise, each valid under its own assumptions, leave the estimate essentially unchanged. Repeated split-tape IV estimates, which remove the error that differs across random halves of the trade tape, give a median of 0.182 across 50 partitions and both orientations, with 5th and 95th percentiles of 0.167 and 0.219 (Table D.4). Instrumenting the ladder surprise instead with a price-independent proxy, the realized print minus the previous available published print, gives 0.190, also statistically significant.

Appendix F reports asymmetry and stability diagnostics: upside and downside surprises give slopes of 0.195 and 0.169, and the qLL test does not reject constant coefficients, although with fifty releases neither test is powerful. The next subsection asks a sharper question about time variation.

5.1.4 Qualification to Fact 1: regime heterogeneity

The stability test reported above can miss even large smooth drift, so it neither confirms nor rules out a time-varying response. Survey panels provide a complementary benchmark: Herbst and Page (2026) find little support for economically meaningful time variation in perceived response coefficients once the pooling structure is estimated rather than imposed. A more interpretable qualification is whether the average differs across policy regimes: does the market map a hot print into a larger policy move in the hiking period than in the hold period?

The post-liftoff sample splits into three regimes at dates fixed by the Committee's own actions (the first hike, the last hike and the first cut), so the split uses no information from the left-hand side. The response is about three times larger while the Fed is hiking than while it is on hold, with the easing period in between (Table 5). Equality of the three slopes is rejected by a restricted wild bootstrap, and the ranking survives moving every boundary a month in either direction.

The split establishes that the headline coefficient is an average over a threefold range rather than a single parameter, and that the variation lines up with the policy stance, consistent with state dependence. Because the regimes are contiguous calendar periods, the design cannot attribute it to the stance alone: the hiking regime is also the high-inflation regime and the period of most intense attention to CPI releases, the explanation Kroner (2025) favors.[9]

[9] Nothing here separates a rule that depends on the policy state from a perceived rule that shifted as inflation rose. The hiking window also supplies about twice the surprise dispersion of the other two, which sharpens that slope without mechanically enlarging it.

Table 5: The release-window response by policy stance

Stance	n	β_π	se	R^2	sd of surprise
hiking (Mar 2022–Jul 2023)	16	$+0.208^{***}$	(0.033)	0.71	0.163
on hold (Aug 2023–Sep 2024)	14	$+0.069^{***}$	(0.013)	0.23	0.086
cutting (Oct 2024–Jul 2026)	19	$+0.169^{***}$	(0.059)	0.08	0.092

Notes: Each row re-estimates equation (5) on the CPI windows falling in that policy stance. Boundaries are the first hike of the cycle, the last hike and the first cut, so the classification uses no information from the dependent variable; the single pre-liftoff window is dropped. Newey–West standard errors in parentheses; *** denotes significance at 1% on the HAC test. Equality of the three slopes is rejected: the restricted wild bootstrap rejects equality at 5% over 49 releases, using 4,999 draws and imposing a common slope and regime intercepts under the null. Moving all three boundaries one month earlier or later preserves the ranking and the joint rejection at 5%, although coefficient magnitudes change. The final column reports the standard deviation of the CPI surprise within each stance, because the stances differ in how much identifying variation they supply.

5.2 Fact 2: The response builds along the expected policy path, and the same news raises the expected next CPI print

We first trace the response along the expected policy path and then show that the same news raises the expected next CPI print. The release window reprices the entire futures strip, not just the next meeting. The CME lists a separate fed funds contract for every calendar month, so for each CPI release and each contract month $h = 1, \ldots, 14$ months ahead we regress the window change in that contract's implied monthly-average rate on the CPI surprise. As in Table 3, a slope of 0.13 means 13 basis points per percentage point of surprise. Figure 4 plots the resulting curve with 95% bands; every CPI horizon is individually significant under the wild bootstrap (Appendix Table G.1). Each point is a distinct contract and a distinct regression: nothing is interpolated across horizons and nothing is fitted, so the plotted profile is the reduced-form object the release windows identify.

Individual significance does not establish that the curve rises. We therefore compare horizons on identical releases, difference the two contract changes release by release, and regress that paired difference on the same surprise. Between months one and eleven, on the 54 releases quoting both contracts, the response rises from 0.133 to 0.717: the difference of +0.584 is significant at 1% under both the restricted wild bootstrap and the studentized-permutation test. Month one is a calendar-month average that can include days before the next decision, when the rate is still the current one and cannot respond to the news. Every later month lies wholly after the next decision, so only month one is diluted in this way. We therefore also measure the build-up from the de-averaged next-meeting rate of Section 3.3: on the 51 releases with both, the eleven-month response exceeds it by 0.517, significant at 1% under HAC and the restricted wild bootstrap. The build-up continues from month six to eleven by +0.092, significant at 5% on both tests. It does not rise detectably from month eleven to month twelve: the estimated increment is nearly zero and insignificant on every test. Thus the paired data establish a build-up through month eleven and no

detectable eleven-to-twelve-month increment.

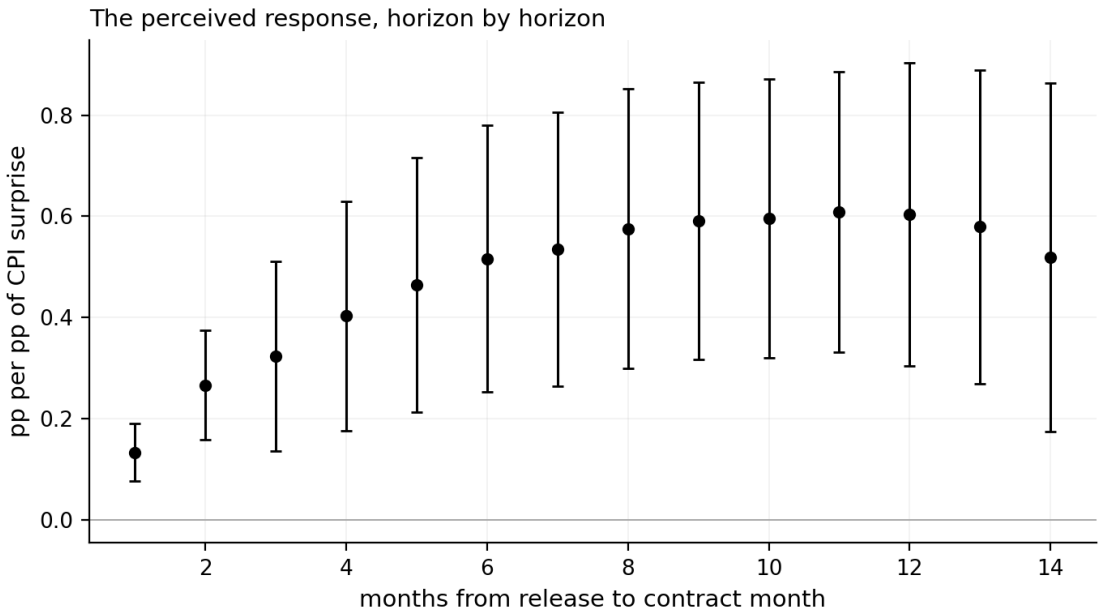


Figure 4: The term structure of the perceived response, measured in CPI release windows

Notes: Each point is a separate regression on a separate contract: the slope of the $[-15, +45]$-minute change in the implied rate of the CME fed funds contract h calendar months after the release, on the CPI surprise of that release, with pointwise 95% bands from HC3 standard errors. Nothing is interpolated across horizons and nothing is fitted. The current-month contract is excluded because it averages over days already realized. Each horizon uses every release quoting that contract, 48 to 55 through thirteen months and 25 at fourteen. From three months onward this includes 10 November 2021, which lowers the plotted level (see the text); Appendix Table G.2 re-estimates every horizon on the 21 releases quoting all fourteen.

This profile is the core of Fact 2. On the 54 releases that quote both the one- and eleven-month contracts, the response builds from 0.13 per pp of surprise at one month to 0.27 at two, 0.63 at six and 0.72 at eleven, then changes little between eleven and twelve. Figure 4 uses every release that quotes each contract, and from three months onward that adds one: 10 November 2021, when neither the one- nor the two-month contract has a usable price at both ends of the window. That release carries the largest CPI surprise in the sample, +0.39pp or more than three standard deviations, and arrived while the policy rate was at its lower bound; the eleven-month contract rose only 5 basis points. Including it lowers the eleven-month slope to 0.61, the level the figure shows. No other release comes close: dropping any other single release leaves that slope between 0.57 and 0.64. U-3 news produces the opposite-signed profile, becoming more negative with horizon, but it comes from the multidimensional Employment Situation and is corroborating rather than a second headline curve. The CPI curve is consistent with gradual policy adjustment, persistence in the inflation state that markets infer from the print, or both; monthly-average contracts also smooth neighboring horizons mechanically. We therefore identify a rising reduced-form path response, not a separate policy-inertia parameter. The curve itself has a precedent in survey–futures work (Hamilton *et al.*, 2011), and survey panels also show perceived-rule coefficients that vary with the

forecast horizon (Herbst and Page, 2026).

Three features of the curve deserve checking explicitly. The first is liquidity. Deferred fed funds contracts trade far less than the front of the strip: median window volume falls from about 29,000 contracts at two months to 200 at twelve and under 100 beyond that. The question is whether the deferred window changes are price changes or repeated stale prints. At every horizon through fourteen months, the contract trades inside the $[-15, +45]$-minute window in 100% of usable release windows, and the quote read at -15 minutes is less than two minutes old at the median through twelve months. Only the last two horizons show meaningful staleness, at five and ten minutes, and those are also the horizons at which the sample thins and whose shape we do not interpret. A second strip shows that the build-up does not depend on thin trading in deferred fed funds contracts. Three-month SOFR futures are the most liquid U.S. short-term interest-rate futures and are quoted well beyond a year. On 55 CPI releases, their response rises from 0.36 for the contract whose reference quarter begins one to three months after the release to 0.62 for the contract beginning ten to twelve months out, an increase that is significant under both small-sample tests, and then levels off (Appendix Table G.3). On the same 55 releases the fed funds strip gives 0.61 at eleven months, and on the 54 without November 2021 both strips give 0.72, so the one-year level matches across the two strips almost exactly.

The second feature is mechanical smoothness. The Committee meets eight times a year, so fourteen consecutive contract months do not price fourteen separate decisions, and neighboring window changes correlate at 0.996 at the median. This does not bias β_h, the response of the h-month-ahead expected monthly-average rate. It does mean that the fourteen points contain fewer than fourteen independent pieces of information, so the bands in Figure 4 are pointwise rather than joint, and the parameters of the smooth curve fitted in Appendix G.2 describe the profile rather than identify a policy rule. Appendix G.1 documents the meeting calendar behind this.

The third is sample composition. The build-up is a fact about the data; the shape past twelve months is partly a fact about the sample. Deferred contracts are quoted for fewer releases, with n falling from 55 at eleven months to 52, 48 and 25 at twelve, thirteen and fourteen, so a decline there is not separable from a change in sample composition. Appendix Table G.2 re-estimates the whole curve on the 21 releases that quote every horizon. The build-up survives and is if anything steeper; a clear fall-back does not. On the balanced sample the point estimate is 0.858 at thirteen months against 0.843 at twelve, then 0.783 at fourteen.

The incremental evidence on the persistence reading comes from deferred CPI ladders: inside the same release window, they reveal how today's surprise revises the expected value of a named future print. A 1 percentage point upside surprise raises the expected next monthly print by 0.224pp in the baseline window, significant under the HAC, restricted-wild and studentized-permutation tests (Table 6). The print two months out also rises, by 0.171pp, although with fourteen releases the inference methods disagree on its significance.[10] These estimates measure short-horizon revisions

[10]The permutation test rejects at 5%, whereas the restricted wild bootstrap does not. Because the later ladder is available less often, comparing the two displayed coefficients is not a horizon test. Holding the release sample fixed at

in inflation expectations, not their contribution to the policy-path coefficient.

The deferred-ladder response keeps growing after the policy response has leveled off. Across the nine alternative windows of Appendix Table D.2, the next-meeting response stays between 0.141 and 0.194, whereas the next-print response is about a tenth over $[-5,+15]$, a fifth over $[-15,+45]$ and a third over $[-15,+90]$ (Table 6). We keep the one-hour window so that the inflation and policy revisions are read on the same clock. The larger response over longer windows is consistent with slower updating in these thinner ladders, but a longer window also admits later news, so we do not treat it as the estimate. The response remains positive and significant under a seven-day staleness cap, and a placebo over $[-75,-15]$ minutes, entirely before the print, detects no relation to the surprise.

Table 6: Short-horizon revisions in inflation expectations, from deferred CPI ladders

	next print		two prints out	
Specification	n	coef. (se)	n	coef. (se)
baseline, $[-15,+45]$ min	38	$+0.224^{***}$ (0.038)	14	+0.171 (0.076)
window $[-15,+30]$	38	$+0.205^{***}$ (0.042)	14	+0.158 (0.079)
window $[-15,+90]$	38	$+0.358^{***}$ (0.048)	14	$+0.214^{*}$ (0.074)
window $[-5,+15]$	38	$+0.089^{**}$ (0.026)	14	+0.052 (0.027)
seven-day staleness cap	35	$+0.236^{***}$ (0.036)	13	+0.121 (0.049)
placebo: pre-release drift	38	−0.003 (0.012)	14	−0.010 (0.010)

Notes: Each cell regresses the mean revision in the fixed Kalshi ladder settling on a named future CPI print on the current print's surprise. The strike set is held fixed across the window. Prices and the upper-boundary support convention follow Section 3.2. HAC standard errors are in parentheses; stars use the maximum HAC, restricted-wild and studentized-permutation p-value. The placebo is entirely pre-release. On the fourteen releases quoting both ladders, the two-prints-out slope minus the next-print slope is −0.108 and insignificant under all three tests. Only two releases quote ladders further out at both ends of the window, so those horizons are not estimated.

6 Conclusion

Event-contract ladders make it possible to measure the market-perceived policy rule in real time, with both the inflation news and the policy response read from traded expectations inside the same hour. Over 2021–2026, a percentage point of upside monthly-CPI news raises the priced next-meeting rate by 18.0 basis points, and all five marginal headline coefficients pass the conservative 5% inference rule and carry policy-rule signs; monthly CPI is the most precisely estimated margin, while the conditional unemployment response is imprecise. Measured against the Bloomberg survey consensus instead, the same price changes imply a response 4.5 basis points weaker on both Kalshi's policy ladder and fed funds futures, a statistically significant difference on both. The

the fourteen dates that quote both, the coefficients are 0.279 and 0.171; their difference of −0.108 is insignificant on every test, so decay between the two horizons is not established.

expectation used to define the news therefore changes the measured strength of the perceived rule. The CPI average also hides economically important heterogeneity: the hiking-period slope is about three times the hold-period slope. Because those stances are contiguous calendar periods, the split is a qualification consistent with state dependence rather than a causal attribution to stance.

The second fact is that the response builds along the expected policy path, and the same news raises the expected next CPI print. On the 54 releases quoting both contracts, the fed funds futures response rises from 13 basis points at one month to 72 at eleven months, a precisely estimated increase, and three-month SOFR futures give the same one-year level. In the same window, Kalshi's deferred CPI ladders measure revisions in named future prints. The next-print response is positive and robust; inference for the second-print estimate is mixed, their difference is imprecise, and coverage is too sparse to trace expectations farther. These revisions directly observe a short-horizon manifestation of expected inflation persistence alongside policy repricing, without imposing a structural decomposition between persistent inflation and gradual policy adjustment.

The binding constraints are the sample and event-window liquidity. The comparable Kalshi ladder series begins in July 2021 and covers one unusual inflation cycle. That is also too few policy decisions to compare the market-perceived response with the Fed's realized response to the same news, which is the comparison the "Fed response to news" channel of Bauer and Swanson (2023a) turns on; it awaits a longer sample. The median CPI window updates only two policy contracts, although the response remains at least as large in windows with observable post-release trading; those active-window selections are diagnostic, not exogenous. The paper therefore contributes a real-time measurement of the market-perceived policy rule and two reduced-form facts, not a time-invariant structural policy rule.

References

Altavilla, C., Brugnolini, L., Gürkaynak, R. S., Motto, R. and Ragusa, G. (2019) Measuring euro area monetary policy, *Journal of Monetary Economics*, **108**, 162–179.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Vega, C. (2003) Micro effects of macro announcements: Real-time price discovery in foreign exchange, *American Economic Review*, **93**, 38–62.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Vega, C. (2007) Real-time price discovery in global stock, bond and foreign exchange markets, *Journal of International Economics*, **73**, 251–277.

Angelini, G. (2026) The shape of macroeconomic beliefs, Preprint 2606.30040, arXiv.

Balduzzi, P., Elton, E. J. and Green, T. C. (2001) Economic news and bond prices: Evidence from the U.S. Treasury market, *Journal of Financial and Quantitative Analysis*, **36**, 523–543.

Bauer, M. D., Pflueger, C. E. and Sunderam, A. (2024) Perceptions about monetary policy, *Quarterly Journal of Economics*, **139**, 2227–2278.

Bauer, M. D. and Swanson, E. T. (2023a) An alternative explanation for the "Fed information effect", *American Economic Review*, **113**, 664–700.

Bauer, M. D. and Swanson, E. T. (2023b) A reassessment of monetary policy surprises and high-frequency identification, in *NBER Macroeconomics Annual 2022, Volume 37* (Eds.) M. Eichenbaum and E. Hurst, University of Chicago Press, pp. 87–155.

Boehm, C. E. and Kroner, T. N. (2026) The U.S., economic news, and the global financial cycle, *Review of Economic Studies*, **93**, 215–249.

Bürgi, C., Deng, W. and Whelan, K. (2026) Makers and takers: The economics of the Kalshi prediction market, CESifo Working Paper 12122, CESifo.

Cieslak, A. (2018) Short-rate expectations and unexpected returns in Treasury bonds, *Review of Financial Studies*, **31**, 3265–3306.

Coibion, O. and Gorodnichenko, Y. (2012) Why are target interest rate changes so persistent?, *American Economic Journal: Macroeconomics*, **4**, 126–162.

Crane, L. D., Haberkorn, F., Hu, J., Lee, S. J., Sokolinskiy, O., Soto, P. E. and Wolfskeil, I. (2026) Reading between the lines: LLMs and Federal Reserve communications, Working Paper 7513339, SSRN.

Diercks, A. M., Katz, J. D. and Wright, J. H. (2026) Kalshi and the rise of macro markets, Finance and Economics Discussion Series 2026-010, Board of Governors of the Federal Reserve System.

Elliott, G. and Müller, U. K. (2006) Efficient tests for general persistent time variation in regression coefficients, *Review of Economic Studies*, **73**, 907–940.

Faust, J., Rogers, J. H., Wang, S.-Y. B. and Wright, J. H. (2007) The high-frequency response of exchange rates and interest rates to macroeconomic announcements, *Journal of Monetary Economics*, **54**, 1051–1068.

Fleming, M. J. and Remolona, E. M. (1999) Price formation and liquidity in the U.S. Treasury market: The response to public information, *Journal of Finance*, **54**, 1901–1915.

Goel, A. (2026) Measuring macroeconomic surprise magnitude with prediction markets: Market-implied dispersion from Kalshi, *International Review of Economics & Finance*, **110**, 105577.

Gürkaynak, R. S., Kısacıkoğlu, B. and Wright, J. H. (2020) Missing events in event studies: Identifying the effects of partially measured news surprises, *American Economic Review*, **110**, 3871–3912.

Gürkaynak, R. S., Sack, B. and Swanson, E. T. (2005a) Do actions speak louder than words? The response of asset prices to monetary policy actions and statements, *International Journal of Central Banking*, **1**, 55–93.

Gürkaynak, R. S., Sack, B. and Swanson, E. T. (2005b) The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models, *American Economic Review*, **95**, 425–436.

Gürkaynak, R. S. and Wolfers, J. (2007) Macroeconomic derivatives: An initial analysis of market-based macro forecasts, uncertainty, and risk, in *NBER International Seminar on Macroeconomics 2005*, MIT Press, pp. 11–50.

Hamilton, J. D., Pruitt, S. and Borger, S. (2011) Estimating the market-perceived monetary policy rule, *American Economic Journal: Macroeconomics*, **3**, 1–28.

Herbst, E. and Page, K. (2026) Optimal pooling in Taylor rule estimation with multiple-horizon forecast panels, Finance and Economics Discussion Series 2026-064, Board of Governors of the Federal Reserve System.

Kroner, T. N. (2025) How markets process macro news: The importance of investor attention, Finance and Economics Discussion Series 2025-022, Board of Governors of the Federal Reserve System.

Kuttner, K. N. (2001) Monetary policy surprises and interest rates: Evidence from the Fed funds futures market, *Journal of Monetary Economics*, **47**, 523–544.

Lapp, J. S. and Pearce, D. K. (2012) The impact of economic news on expected changes in monetary policy, *Journal of Macroeconomics*, **34**, 362–379.

Nakamura, E. and Steinsson, J. (2018) High-frequency identification of monetary non-neutrality: The information effect, *Quarterly Journal of Economics*, **133**, 1283–1330.

Ottonello, P. and Winberry, T. (2020) Financial heterogeneity and the investment channel of monetary policy, *Econometrica*, **88**, 2473–2502.

Piazzesi, M. (2005) Bond yields and the Federal Reserve, *Journal of Political Economy*, **113**, 311–344.

Romer, C. D. and Romer, D. H. (2000) Federal Reserve information and the behavior of interest rates, *American Economic Review*, **90**, 429–457.

Rudebusch, G. D. (2002) Term structure evidence on interest rate smoothing and monetary policy inertia, *Journal of Monetary Economics*, **49**, 1161–1187.

Rudebusch, G. D. (2006) Monetary policy inertia: Fact or fiction?, *International Journal of Central Banking*, **2**, 85–135.

Schmeling, M., Schrimpf, A. and Steffensen, S. A. M. (2022) Monetary policy expectation errors, *Journal of Financial Economics*, **146**, 841–858.

Snowberg, E., Wolfers, J. and Zitzewitz, E. (2013) Prediction markets for economic forecasting, in *Handbook of Economic Forecasting, Volume 2*, Elsevier, pp. 657–687.

Swanson, E. T., Wang, R. and Wu, Y. (2026) The effects of monetary policy on macroeconomic expectations: High-frequency evidence from prediction markets, Working paper, University of California, Irvine.

Taylor, J. B. (1993) Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy*, **39**, 195–214.

A The empirical increment relative to the closest designs

Table A.1 compares observable objects rather than unlike structural coefficients. The increment is real-time measurement: a market-based current surprise and same-window policy repricing, which together give a stronger measured response than survey-based news, alongside named future-print revisions. No single familiar ingredient (an intraday window, a Kalshi distribution or a rising policy response) is presented as the novelty by itself.

Table A.1: What real-time measurement adds

Study	Nearby observable object	Increment in this paper
Hamilton, Pruitt and Borger (2011); Lapp and Pearce (2012)	Macro news and expected policy from survey surprises and fed funds futures	Real-time news and response, priced in the same hour; a stronger response than with survey news; directly observed deferred-CPI revision
Gürkaynak and Wolfers (2007)	Asset-price responses to market- and survey-based surprises, from Economic Derivatives auctions	The same comparison for the perceived policy rule, on a traded policy distribution, with a test of why survey-based news gives a weaker response
Diercks, Katz and Wright (2026)	Daily changes in Kalshi policy-distribution moments against Bloomberg surprises	Common one-hour window and own-ladder surprise; simultaneous policy-path and deferred-print revisions
Swanson, Wang and Wu (2026)	Policy news and front/deferred macro expectations at daily frequency	Current inflation news as the regressor; simultaneous policy and future-inflation responses
Angelini (2026)	High-frequency pre-release CPI expectations and responses to earlier survey surprises	Current released print against its own ladder, coupled to policy and future-print revisions
Kroner (2025)	Intraday yield and inflation-swap responses to CPI news	Meeting-specific policy distribution and market-based surprise alongside named future prints

Notes: The table compares designs, not coefficient magnitudes across different samples, surprise concepts or horizons.

B Construction details and data hazards

This appendix records the construction choices behind the densities, then the hazards specific to these data, and finally the sample and freshness of each density the design reads.

B.1 Construction choices

Undiscounted prices. The baseline uses Kalshi prices as they trade, without discounting them at the policy rate. Buying a Yes contract at price p is the same trade as selling a No contract at $1-p$, so the two sides always sum to one dollar, and since 10 October 2024 Kalshi has also paid interest on cash and open positions; neither supports applying a common discount to every price. Funding, fees, eligibility and position remuneration would have to enter a venue-specific pricing model, and full collateralization alone does not establish $p = e^{-rT}Q$. A hypothetical sensitivity divides prices by $e^{-r(T-t)}$, using the prevailing target midpoint, before monotonicity and differencing. This tests a common price-scale perturbation; it is not a derived Kalshi funding correction. The same convention applies to the policy, macro-surprise, deferred-CPI and FOMC measures.

Monotonicity. Trade prices printed at different instants routinely violate monotonicity of the survival function across strikes. We enforce it by volume-weighted pool-adjacent-violators, retaining the violation magnitude as a data-quality flag. It is the only construction choice that materially moves the headline slope while keeping the sample fixed (Section 5.1), and the only one with an unambiguous a priori justification: a survival function must be monotone.

Grid settlement and sparse bins. Alignment with a publication grid does not imply that observed strikes are adjacent on it. Upper-boundary support is exact for a one-step interior bin and approximate for a wider one. The support-aware variant changes only wider grid bins, representing their conditional means by the midpoint of attainable support. Neither variant identifies their within-bin shape. Gaps are measured relative to the publication-grid steps of 0.1pp for CPI and U-3 and 0.25pp for the target. GDP and payroll ranges retain midpoint approximations.

Staleness and exterior tails. Snapshots retain last trades within thirty days. For grid series, the baseline places low exterior mass at the lowest threshold and high mass one canonical step above the highest; these are boundary closures, not identified tail locations. The Gaussian variant fits a normal location and scale to the interior CDF points, uses conditional exterior means and projects them onto attainable tail grid points. It changes tail locations, not the tail probabilities implied by the end prices. Continuous-midpoint families use reflected exterior nodes in the baseline. Tail assumptions can affect means as well as higher moments. Favorite-longshot bias (Bürgi *et al.*, 2026) does not have a known sign for the regression bias, so the Gaussian and drop-extreme-strike variants are sensitivities rather than bias bounds.

Table B.1: Support and exterior-tail sensitivities

Release	n	baseline	support means	Gaussian tails
β_π	50	+0.180 (0.035)	+0.180 (0.035)	+0.188 (0.036)
β_π^{yoy}	41	+0.080 (0.036)	+0.081 (0.035)	+0.077 (0.036)
β_u	51	−0.080 (0.038)	−0.083 (0.041)	−0.086 (0.038)
β_{nfp}	50	+0.016 (0.006)	+0.016 (0.006)	+0.016 (0.006)
β_y	18	+0.008 (0.003)	+0.008 (0.003)	+0.007 (0.002)

Notes: Cells report slopes and HAC standard errors. All constructions use the same window and native prices; the samples are unchanged. Support means allocate each sparse grid bin equally across its attainable outcomes, without claiming to identify that allocation. Gaussian tails change exterior nodes, not end-price probabilities. In the regression sample, sparse settling bins occur on 5 of 50 CPI and 27 of 51 U-3 releases. Conditional on a gap, their mean affected mass is 5.4% and 18.7%, respectively. The CPI policy ladder has gaps at 20 of 50 pre- and post-window snapshots, but mean affected mass is only about 0.4% conditional on a gap. Higher moments remain conditional on unidentified within-bin shape.

B.2 Data hazards

Each hazard below produces a plausible but incorrect number rather than a visible failure, and each is covered by a regression test in the replication package. They are listed because any of them silently corrupts an analysis of event-contract ladders.

Payroll ticker and geometry changes. KXPROLLS titles and subtitles encode thresholds in jobs even when tickers abbreviate thousands, and settlements appear as both "115,000" and "178K." Some vintages are survival ladders; others are exclusive less/between/greater buckets. We convert all values to single jobs, identify the ladder type contract by contract, read exclusive-bucket prices directly as probabilities, and then scale surprises to 100,000 jobs. Because last trades are asynchronous, the bucket prices of an event need not sum to one, so we rescale them to do so. Far-tail buckets sometimes never trade at all, and the rescaling assigns them zero probability. KXPROLLS supplies 12 usable releases from April 2022 through March 2023 and KXPAYROLLS supplies 38 thereafter; a third ticker, KXUSNFP, traded briefly alongside KXPAYROLLS in 2026 and supplies no separate window. Treating only KXPAYROLLS as the series would incorrectly reduce the payroll sample from 50 to 38 releases.

Archived markets. Kalshi moves markets older than roughly three months from its live listing into a separate historical archive, and a query of the live listing for an older market returns nothing rather than an error. A download built on the live listing alone therefore looks complete but contains only the most recent quarter. Every series here is downloaded from the historical archive first.

Front-ladder settlement. The contract settling on today's release closes minutes before the print, so differencing the front macro ladder across a release measures a change of event, not of

expectation (Section 3.2). The same applies to the front FOMC ladder across a statement, so the communication-window revisions of Appendix H also fix the event before differencing.

Dating meetings by the statement time. Mapping a release to its next meeting requires the meeting's timestamp, not its calendar day. On 12 June 2024 the CPI print occurred at 08:30 Eastern and the FOMC statement at 14:00. Dating meetings at midnight would incorrectly treat the print as occurring after that meeting and map it to 31 July. We use the published 14:00 timestamp, so the release maps to the June meeting; because the June contract is missing from our futures data, July ZQ supplies the post-June rate. The 31 July decision is effective on 1 August and therefore does not contaminate July's average.

De-averaging a meeting-month contract near month-end. A ZQ contract settles on the month's average effective rate, so recovering the rate expected *after* a decision means solving $DA = d\, r_{\text{pre}} + (D-d)\, r_{\text{post}}$ for r_{post}. The solution divides by the days remaining after the decision, so a meeting late in its month multiplies every error in A by up to D. The 29 April 2026 meeting leaves one day in a thirty-day month; with A quoted on a half-basis-point grid, half a tick of quoting noise becomes 7.5bp, larger than the fitted 2.1bp revision after a one-standard-deviation CPI surprise. In this sample it produced a phantom 35bp move in the expected policy rate inside a CPI window whose surprise was nearly zero, on a morning when the Kalshi ladder, with eleven strikes and 1.7 million contracts of volume, did not move at all. Ten of the forty-one meetings amplify quoting errors more than sixfold, and without a guard the largest disagreement between the two venues reaches 137.3bp. We therefore do not de-average when fewer than six days remain after the decision and read the following month's contract instead, which gives the post-meeting rate directly whenever no later meeting falls inside it. Table B.2 reports what the threshold changes.

Table B.2: The de-averaging guard across thresholds

min. post-days	β_π	se	R^2	gap sd	max |gap|	change corr.
1	+0.245	(0.073)	0.17	23.3	137.3	0.40
3	+0.261	(0.065)	0.36	8.5	32.6	0.72
6	+0.199	(0.058)	0.40	5.3	17.7	0.74
9	+0.199	(0.058)	0.40	5.3	17.7	0.74

Notes: Each row re-attaches the futures leg under the stated guard and changes nothing else. "Min. post-days" is the number of days that must remain in the contract month after a decision before the averaging identity is solved; below it, the following month's contract is read directly, and a value of one means no guard. "Gap" is the Kalshi target midpoint minus the futures-implied level at the pre-release snapshot, in basis points; the two settle on related but distinct rates. The threshold is chosen on these conditioning diagnostics, not on the coefficient: the gap dispersion, the worst gap and the change correlation all improve as the guard tightens, and six and nine days coincide because no meeting leaves six to eight days. The coefficient stays positive and significant throughout, between 0.199 and 0.261. The paper uses six.

B.3 Samples and freshness

Table B.3 shows how the sample of every estimand is built, Table B.4 how actively the policy ladder trades in the CPI windows, and Table B.5 how recently the strikes behind each density traded when the design reads them.

Table B.3: Sample construction by release family

Estimand	eligible dates	macro density	policy outcome	intersection
Monthly CPI	59	55	50	50
Year-over-year CPI	59	42	50	41
U-3 unemployment	58	52	51	51
Nonfarm payrolls	58	50	51	50
Advance GDP	21	19	18	18

Notes: Eligible dates pass the published-calendar check; five 2025 dates that Kalshi listed but on which no release occurred (three CPI, two Employment) are removed first. Macro density and policy outcome are independently usable counts before intersection, and there is no minimum within-window trade screen. For monthly CPI, the four July–October 2021 releases have no reconstructible pre-release CPI density. Five further releases, November 2021 through February 2022 and March 2025, have a CPI density but no usable policy outcome: fewer than three policy strikes had a usable price at the window endpoints, and a policy density needs at least three. Widening the response endpoint to +60 or +90 minutes leaves $n = 50$, because it cannot recover a missing pre-window density. FOMC samples vary with the window endpoint and are reported in Table H.1.

Table B.4: Policy-ladder activity in the 50 CPI windows

	Windows
No post-release policy trade	6
One contract updated	16
Two contracts updated	12
Three or four contracts updated	14
Five or more contracts updated	2
	Median window
Post-release policy trades	11
Contracts updated	2
Post-release volume (contracts)	3,527

Notes: Activity is measured on the next-meeting policy ladder from the CPI release to +45 minutes. An "updated contract" receives a new executed trade and hence a new last-trade price; the tape does not record every order-book quote revision.

Table B.5: How current the densities are when the design reads them

Window	Ladder	n	median age	p90 age	1 hour	1 day	7 days
CPI	settling CPI	50	0.1h	8.7h	0.84	0.99	1.00
Employment	settling U-3	51	0.2h	14.2h	0.64	0.98	1.00
Employment	settling payrolls	50	0.1h	10.3h	0.92	0.99	1.00
GDP	settling GDP	18	0.2h	20.3h	0.79	0.99	1.00
CPI	policy (KXFED)	50	11.3h	99.8h	0.41	0.94	0.99

Notes: Exact settling events are reconstructed at the −15-minute cutoff on each reported regression sample; U-3 and payrolls are separate ladders. Strike ages are weighted by adjacent probability mass. Median age is the median across releases of the age by which half the mass has traded; p90 is its 90th percentile. The final columns report medians across releases of shares of mass traded within each age cutoff. The median across releases of the alternative mass-weighted mean age is 4.1 hours for settling CPI and 17.8 hours for its policy outcome, versus median-mass ages of 0.1 and 11.3 hours. Tables D.2 and D.3 report cap sensitivities.

C The release windows, audited against the published calendar

The design of Section 4 seeks CPI windows not confounded by a separate scheduled release. The trade tape cannot check that condition, because it sees only the release families Kalshi lists: a simultaneous PPI or retail-sales print would be invisible to it. This appendix checks the condition against the agencies' own schedules instead.

We read the Bureau of Labor Statistics' *Schedule of Selected Releases* and the Census Bureau's *Economic Indicator Release Schedule* for every year of the sample. Both publish, for each release, the calendar date, the Eastern release time, and the program name. Weekly unemployment-insurance claims are added by calendar rule, since Thursday is their usual 8:30 slot; we flag CPI windows on that weekday rather than claiming a complete claims-publication calendar. Bureau of Economic Analysis releases (GDP, Personal Income and Outlays), other publishers and unscheduled news are outside this coverage. The audit therefore checks a defined set of potential co-releases rather than certifying the absence of all other news. The schedules list 1,899 releases over the sample, 924 of them at 8:30.

What shares a CPI window. Every CPI release in the sample is accompanied at the same instant by *Real Earnings*, which deflates the average hourly earnings from the latest Employment Situation by the CPI just released. We classify this arithmetic transformation as a derived companion rather than an independent scheduled innovation. This calendar audit addresses separate scheduled releases, not distinct components within the CPI report itself. Beyond that, the covered collisions are rare: twelve of the fifty estimated windows fall on a Thursday and are flagged for potential weekly claims, and advance retail sales shared an 8:30 slot with CPI exactly once, on 15 May 2024. CPI and PPI do not share a date in the sample period. Table C.1 re-estimates the headline coefficient as these

potential co-releases are progressively excluded.

Table C.1: The headline coefficient under progressive isolation of the CPI window

Sample	n	β_π	se	R^2	t (HAC)
all CPI windows	50	+0.180	(0.035)	0.25	5.2
excluding Thursdays	38	+0.186	(0.029)	0.22	6.4
no independent covered co-release	37	+0.190	(0.028)	0.22	6.7

Notes: Each row re-estimates the CPI coefficient after removing covered simultaneous releases: the second row drops the twelve Thursday windows with weekly claims, and the third also drops the window shared with advance retail sales. Real Earnings restates CPI and already-published earnings and is not treated as separate news. Differences across these selected samples do not identify whether co-releases add noise or correlated signal.

A hazard only the published calendar reveals. Kalshi's listed close and expiration times need not move when an agency reschedules a release. During the 2025 shutdown, five listed release times (three CPI and two Employment) had no release at all; checking against the agency calendars removes them before the eligible-date counts. Delayed September CPI was published on 24 October rather than 15 October, after its original contract closed, and October CPI was not published as a standalone release. Without a pre-release density for the actual publication time, those nominal windows cannot be recovered by widening the response endpoint. The exclusion concerns the availability and dating of expectations, not a minimum trading screen. It may still change sample composition; the attrition and calendar tables make that selection explicit. The shutdown also changed one release that remains in the sample. November CPI, published on 18 December, carried no one-month change: BLS reported only the two-month change from September, 0.2%, and the Kalshi contract settled on that figure. Bloomberg has no monthly consensus for it, since there was no monthly change to forecast, so the survey comparison of Section 5.1.2 uses the other 49 baseline releases. Dropping this release leaves the baseline estimate at 18.0 basis points.

D Inference, robustness and influence

This appendix collects the inference and robustness tables behind the estimates of Section 5.1, and then reports three checks of a different kind.

D.1 Supporting tables

Table D.1 reports every inference method for the five baseline coefficients and Table D.2 the 30 constructions of the headline coefficient. Table D.3 shows why the staleness cap moves it, and Table D.4 reports the measurement-error diagnostics.

Table D.1: The baseline coefficients under every inference method

	n	β	se	t (HAC)	R^2	p (wild)	p (perm)	p (RW)	p (Bonf.)
β_π	50	+0.180	0.035	+5.2	0.25	0.001	< 0.001	< 0.001	0.003
β_π^{yoy}	41	+0.080	0.036	+2.2	0.04	0.011	0.008	0.044	0.060
β_u	51	−0.080	0.038	−2.1	0.15	0.013	0.013	0.044	0.037
β_{nfp}	50	+0.016	0.006	+2.8	0.24	< 0.001	0.003	0.044	0.002
β_y	18	+0.008	0.003	+2.8	0.46	0.042	0.008	0.044	0.218

Notes: Wild denotes the restricted Rademacher bootstrap. Permutation uses the HC1-studentized slope and reassigns surprises across releases, requiring exchangeability for a randomization interpretation; a zero slope alone does not establish it. Romano–Wolf uses common release-aligned multipliers across all five equations and stepdown maxima. Bonferroni multiplies each raw marginal bootstrap p-value by five, not its already adjusted stepdown value. These bootstrap tests maintain release-level disturbance assumptions and are sensitivities alongside HAC inference, not exact finite-sample guarantees. Table 3 stars use the maximum HAC, wild, permutation and RW p-value; Bonferroni is reported separately.

Table D.2: The headline coefficient across 30 constructions

Family	Variants	β_π range	min $\lvert t \rvert$ (HAC)
Baseline	1	0.180	5.16
Window	9	0.141–0.194	3.54
Construction	7	0.138–0.188	2.60
Freshness	2	0.209–0.226	5.85
Sample	10	0.128–0.202	2.83
External	1	0.199	3.44

Notes: The baseline and 29 alternative constructions retain the positive CPI sign and have HAC $p < 0.05$. These are correlated sensitivities, not independent confirmations. The construction variants include support-aware bin means, Gaussian exterior means and hypothetical price scaling; Table B.1 reports the first two for all five releases. Freshness caps drop individual strikes and change sample composition (Table D.3).

Table D.3: Why the staleness cap moves the coefficient

Construction	n	β_π	se	What it changes
baseline, 30-day cap	50	0.180	0.035	—
Seven-day cap				
both sides capped	43	0.209	0.036	regressor, outcome and sample
surprise ladder only	50	0.203	0.039	regressor only
policy ladder only	43	0.185	0.031	outcome and sample
uncapped, on capped sample	43	0.190	0.032	sample only
One-day cap				
both sides capped	24	0.226	0.023	regressor, outcome and sample
surprise ladder only	50	0.164	0.041	regressor only
policy ladder only	24	0.227	0.023	outcome and sample
uncapped, on capped sample	24	0.220	0.031	sample only

Notes: Joint caps change the regressor, outcome and surviving sample. Re-estimating the uncapped construction on cap survivors isolates sample selection from construction changes: 0.190 versus 0.209 at seven days, and 0.220 versus 0.226 at one day. At seven days selection accounts for about one-third of the increase from 0.180, and at one day for most of it. One-sided caps diagnose additional construction effects, not an additive causal decomposition. Removing strikes can coarsen support. Under classical measurement error, regressor noise attenuates the slope while independent outcome noise only reduces precision; stale last trades need not be classical, because partial updating can attenuate the observed response and both errors can depend on the news.

Table D.4: Conditional measurement-error diagnostics

Statistic	median	5th percentile	95th percentile
Panel A: 50 partitions, both orientations (100 estimates)			
split-tape OLS slope	0.182	0.166	0.221
split-tape IV slope	0.182	0.167	0.219
IV HAC standard error	0.038	0.033	0.052
OLS/IV ratio (conditional reliability)	0.999	0.845	1.152
robust first-stage F	13555	1538	30230
Panel B: previous-print instrument (coefficient and HAC se)			
random-walk news, reduced form	+0.038	(0.012)	$n = 50$
ladder surprise instrumented by prior-print news	+0.190	(0.039)	$n = 50$

Notes: Panel A summarizes 50 seeded partitions and both IV orientations, with 48–50 observations per fit. Each rebuilds both sides and uses structural-residual HAC IV standard errors and a heteroskedasticity-robust first-stage F. Percentiles measure sensitivity to partition, not sampling confidence intervals. OLS/IV is a reliability ratio only under independent errors and exclusion; 48% of these ratios exceed one. Panel B uses the current print minus the previous available published print, not necessarily the immediately previous month during publication gaps. The reduced form has no Kalshi regressor; the IV still instruments the Kalshi surprise. Its robust first-stage F is 42.4. Exclusion is a maintained condition, not a finding: it requires that the proxy affect policy repricing only through the measured current surprise, which predictable inflation and other report components can violate. The proxy is selected before policy-window and expectation-ladder exclusions, and common errors can survive the split in Panel A.

D.2 Microstructure and influence checks

This subsection reports three checks outside the 30 constructions of Table D.2: whether the measurement is contaminated by how trades print, whether the results depend on summarizing each density by its mean, and whether any of the five baseline coefficients rests on a single release.

Order flow and the bid–ask bounce. Every ladder is assembled from last-trade prices, so a change in the last trade can reflect which side of the book a trade hit rather than a change in expectations. Bounce unrelated to the news only reduces precision; the concern is bounce correlated with the surprise. Table D.5, Panel A, shows that trade direction does respond to the news: the share of post-release policy-ladder trades initiated by yes-takers rises with the CPI surprise, from roughly one half to roughly seven tenths after a one-standard-deviation upside print. Such execution effects could inflate the slope, although order flow alone does not reveal their size.

Panel B therefore rebuilds every ladder with each strike's last-trade price replaced by the average of its last yes-initiated and last no-initiated trades, holding the strike set, the staleness filter and the sample fixed. Ninety percent of strikes trade on both sides within the staleness cap; the rest keep their last-trade price. The two sides' last prices differ by 2.4 cents on average, but because prices can move between the two executions, their average is not a quote midpoint. The coefficient

falls from 0.180 to 0.163, about a tenth, and remains significant on 50 releases. Because averaging asynchronous trades can also smooth genuine repricing, we keep the last-trade baseline and report this as a sensitivity rather than a correction.

Mean or smoothed quantile. The baseline summarizes each density by its mean. A smoothed median from probability-midpoint interpolation gives 0.186 against 0.180; the two surprises correlate at 0.995 and differ by 0.010pp on average. The similarity does not imply that the densities are symmetric, and the interpolated quantile is not the exact median of the discrete support.

Table D.5: Microstructure: order flow, the bid-ask bounce, and the summary of the density

	n	estimate	se	t	R^2
Panel A: post-release order flow					
yes-taker share on CPI surprise	38	+1.748	0.187	+9.3	0.56
volume-weighted	38	+1.641	0.209	+7.8	0.40
mean yes-taker share	38	0.492			
opposite-side execution price gap, $		0.024			
share of strikes traded on both sides		0.90			
Panel B: alternative price and summary rules					
last-trade prices, density mean	50	+0.180	0.035	+5.2	0.25
opposite-side last-trade average	50	+0.163	0.038	+4.3	0.32
smoothed 50% quantile rather than mean	50	+0.186	0.038	+4.9	0.24

Notes: Panel A uses windows with at least ten policy trades in [0, +45] minutes; HAC standard errors. The price gap is the mean difference between last opposite-side executions on settling CPI strikes at −15 minutes, not a quote spread. Panel B rebuilds both sides. Opposite-side averaging falls back to the last trade where only one side is available and does not time-match executions. The smoothed 50% quantile uses probability-midpoint interpolation, not exact inversion of the discrete CDF.

Does one release carry any of the baseline coefficients? Table D.6 drops each release in turn, and Figure D.1 plots every leave-one-out estimate, which shows whether a wide range reflects one outlying release or genuine dispersion. No coefficient changes sign under any drop, and each remains significant at 10% on the HAC test in every leave-one-out sample; monthly CPI, U-3 and payrolls remain significant at 5% in every drop. Monthly CPI is the most stable proportionally, staying within [0.160, 0.192]. Year-over-year CPI is the least stable, with a maximum shift of 40% and a range of [0.048, 0.096], so its magnitude is less firmly established than its sign.

The most visible point in Figure 2 is the CPI window of 13 November 2024, with a next-meeting revision near −24bp. At 8:59 ET a large order to sell swept the thinly traded 4.25% strike of the December ladder, and that strike's last print inside the window (9:06 ET) implied a 68% chance of a 50bp cut; it traded back at 95 cents at 9:58 ET, after the window closed, while the CME de-averaged

rate fell 6bp. We keep the window because no release is excluded on its outcome. Dropping it lowers the CPI slope from 0.180 to 0.169, inside the leave-one-out range above.

Table D.6: Leave-one-release-out for the baseline coefficients

	n	β	leave-one-out range	max shift	10% significant	min $\lvert t \rvert$
β_π	50	+0.180	[+0.160, +0.192]	11%	50 of 50	4.27
β_π^{yoy}	41	+0.080	[+0.048, +0.096]	40%	41 of 41	1.78
β_u	51	−0.080	[−0.098, −0.061]	23%	51 of 51	1.97
β_{nfp}	50	+0.016	[+0.012, +0.019]	25%	50 of 50	2.43
β_y	18	+0.008	[+0.005, +0.009]	29%	18 of 18	1.71

Notes: Each row drops one release and re-estimates equation (5). Counts use the 10% HAC threshold, not adjusted inference. At 5%, monthly CPI, U-3 and payrolls stay significant in every leave-one-out sample, year-over-year CPI in 40 of 41 and GDP in 17 of 18. No coefficient changes sign.

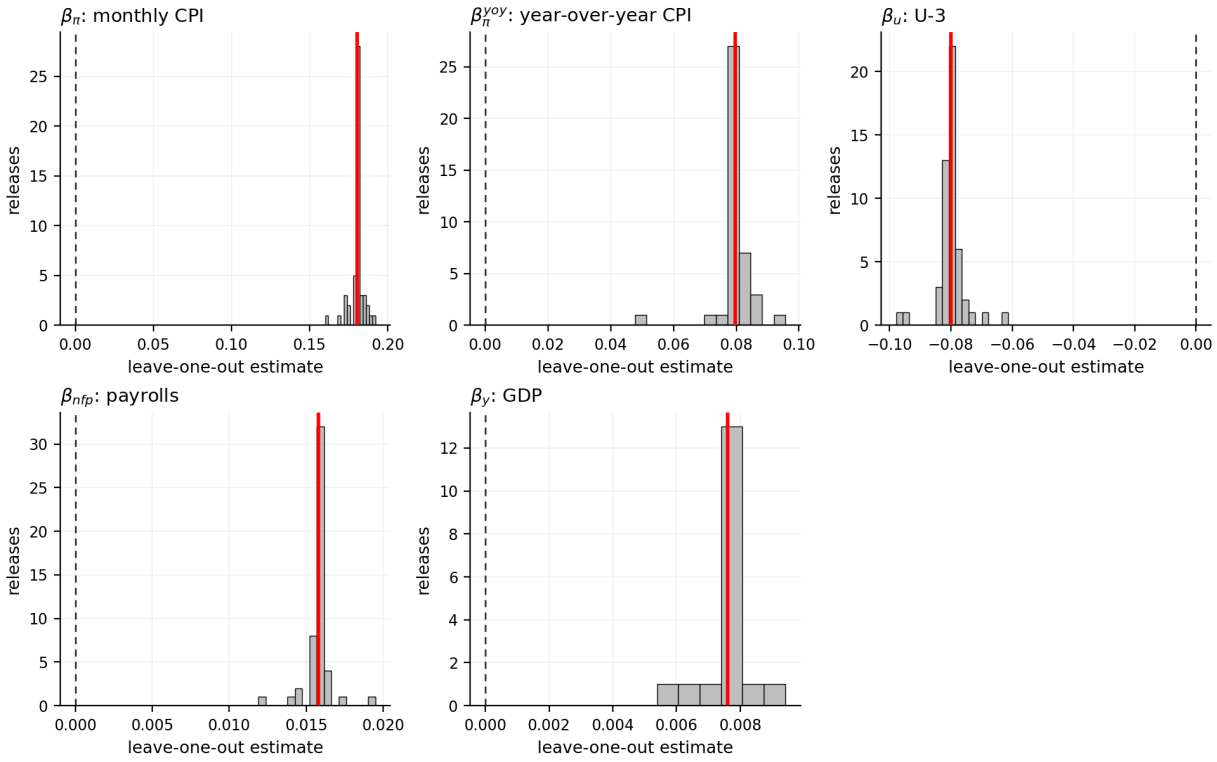


Figure D.1: Leave-one-release-out distributions for the baseline coefficients

Notes: Each panel reports all leave-one-release-out coefficients. The red line is the full-sample estimate and the black dashed line is zero. Horizontal scales differ across panels; Table D.6 reports proportional shifts.

E Kalshi and survey expectations

Section 5.1.2 compares the policy response to market-based and survey-based news. This appendix reports two related comparisons: how the age of the market's expectation affects the estimated

response, and how Kalshi and the survey forecast the FOMC decision itself.

E.1 The timing of the expectation

Table E.1 reports the timing comparison summarized in Section 5.1.2. Each row measures the surprise against an older Kalshi expectation, built with the baseline construction.

Table E.1: How the timing of the expectation affects the estimate

			Kalshi next meeting		CME next meeting	
Expectation read	sd of drift	variance ratio	slope	% of baseline	slope	% of baseline
15 minutes (baseline)	—	—	0.180	100%	0.215	100%
1 day before	0.019	0.02	0.178	99%	0.200	93%
3 days before	0.028	0.05	0.166	92%	0.191	89%
7 days before	0.046	0.12	0.154	86%	0.177	82%

Notes: Each row rebuilds the settling CPI ladder's mean at the stated time before the release, with the baseline construction, and regresses the release-window next-meeting revision on the surprise against that earlier expectation. The drift is the revision in the market's expectation between the earlier snapshot and the baseline cutoff fifteen minutes before the print; the variance ratio is the drift variance divided by the variance of the earlier surprise. Kalshi rows use the 50 baseline windows and CME rows the 48 with both outcomes, so the CME baseline differs from the 51-window estimate in the text. In a regression on the baseline surprise and the drift together, the drift coefficient is insignificant in every row under the wild bootstrap. The drift is positively correlated with the eventual surprise, so the market's expectation moves in the direction of the coming print but not all the way. The fifteen-minute snapshot reproduces the baseline surprise exactly.

E.2 Forecasts of the policy decision

Figure E.1 compares Kalshi and survey forecasts of each FOMC decision with the announced target upper bound. On the same 35 meetings, the Kalshi distribution read at 13:45 has a root-mean-squared forecast error of 3.1 basis points, compared with 4.9 for Bloomberg's survey average. The same Kalshi expectation read seven days earlier has an RMSE of 6.4 basis points. Both current forecasts closely track the announced rate, while the traded distribution lets us observe the updating before the decision. This complements the macroeconomic forecast comparisons of Swanson *et al.* (2026), using meeting-specific policy forecasts and Bloomberg survey averages.

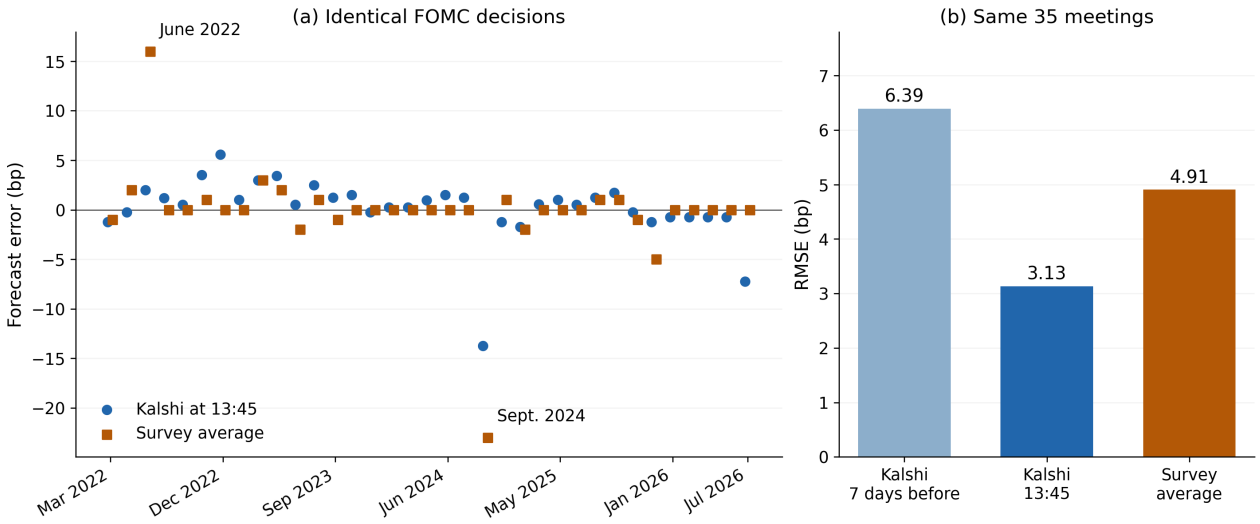


Figure E.1: Forecasts of the announced policy decision: Kalshi and survey expectations

Notes: Forecast errors are the announced current-meeting target upper bound minus the forecast, in basis points. Every comparison uses the same 35 FOMC decisions, March 2022–July 2026, with meetings ordered chronologically in the left panel. Kalshi means use the same settling event at 13:45 Eastern or 168 hours before the statement, following the baseline last-trade construction and 30-day strike-age cap; the snapshots need not consist of newly traded prices. Bloomberg supplies the survey average for the decision, with no historical survey cutoff timestamp. Of 41 meetings with survey records, six lack a usable same-decision ladder with at least three strikes. Actual rates are checked against the official FOMC statements. Mean absolute errors are 1.87bp for the 13:45 Kalshi forecast and 1.80bp for the survey. The RMSE ranking depends principally on June 2022 and September 2024; paired squared- and absolute-loss HAC tests of the 13:45 Kalshi forecast against the survey do not reject equal accuracy at 5%. These are decision-forecast errors, not a decomposition of communication surprises into target and forward-guidance shocks.

June 2022 makes the distinction tangible. Bloomberg's average forecast was 1.59%, while Kalshi priced 1.73% fifteen minutes before the statement; the announced upper bound was 1.75%. The decision was therefore 16 basis points above the survey average but only 2 above the priced expectation. Over all 35 meetings, absolute errors are similar and paired loss tests find no accuracy difference between Kalshi and Bloomberg at 5%, so the figure shows how the traded expectation updates before the decision rather than a general forecasting advantage.

F The shape of the response

Table F.1 reports how the CPI response varies with the sign of the surprise, with the policy state and the time to the meeting, and how the news moves the dispersion and skewness of the policy distribution. The last part of this appendix assesses the power of the stability tests.

No slope asymmetry is detected, but meaningful asymmetry remains possible. The upside and downside slopes are 0.195 and 0.169, and their difference is statistically insignificant. The test is not sharp: at fifty releases, the difference it detects with 80% power is about 1.4 times the baseline response, so we do not infer equal responses from this non-rejection.

Table F.1: The shape of the response (CPI windows)

	Estimate	se	Reading
Panel A: asymmetry			
slope on upside surprises	$+0.195^{***}$	(0.060)	
slope on downside surprises	$+0.169^{***}$	(0.056)	
difference	+0.025	(0.092)	no detected difference
detectable at 80% power	0.259		1.4× the response
ratio of mean |move|, up/down	0.58		[0.23, 1.67]; percentile interval
Panel B: state interactions and timing			
surprise × (YoY expectation − 2)	-0.034^{**}	(0.016)	not confirmed by bootstrap
surprise × 1(rate above median)	−0.041	(0.061)	imprecise
surprise × days-to-meeting	$+0.0029^{***}$	(0.0011)	timing association
Panel C: constructed higher moments			
Δsd on surprise (pp)	+0.023	(0.014)	imprecise
Δsd on |surprise| (pp)	+0.007	(0.024)	imprecise
Δsd, constant (pp)	-0.0053^{***}	(0.0020)	zero-surprise intercept
Δskew on surprise	−0.957	(0.804)	imprecise
Δskew on |surprise|	+0.765	(1.518)	imprecise

Notes: Newey–West standard errors in parentheses; stars use the HAC test, not the conservative multi-method rule of the headline estimates. Panel A estimates separate slopes on the positive and negative parts of the surprise; the power row is the difference detectable with 80% power in a two-sided 5% test, 2.80 times its HAC standard error. The ratio compares mean absolute policy revisions after upside and downside surprises, with a 95% percentile interval from a stratified bootstrap of the two groups. Panel B adds one interaction at a time. Panel C enters signed and absolute surprises jointly; its constant is the fitted change at a zero surprise. Higher moments depend on within-bin and tail support assumptions.

For comparison, Diercks *et al.* (2026) report much larger daily policy moves after positive CPI surprises than after negative ones. In our one-hour windows, the ratio of mean absolute revisions after upside and downside surprises is 0.58, with a 95% percentile interval of [0.23, 1.67]. This ratio also depends on surprise sizes and sample composition, and their daily design uses Bloomberg-consensus surprises, so the two results are not a controlled comparison.

Higher-moment evidence is exploratory. With signed and absolute surprises entered jointly, neither standard-deviation slope is statistically significant, and the skewness estimates are less precise still. The intercept implies a small narrowing of the distribution at a zero surprise. In the sixteen windows that update at least three policy contracts, the signed standard-deviation slope is positive and significant on the HAC test, but that subset is selected on activity. Because the median window updates only two policy contracts and within-bin support is unresolved, these moments are weakly measured, and we draw no conclusion about how news reshapes the distribution.

State and timing interactions. The interaction with the pre-release year-over-year expectation

is negative and significant on the HAC test, but the restricted wild bootstrap does not confirm it, so we do not rely on it. The days-to-meeting interaction is positive and significant on the HAC test, consistent with horizon-sensitive repricing; it does not identify a causal effect of waiting an additional day.

Stability tests have little power here. The qLL test of Elliott and Müller (2006) does not reject constant coefficients: its statistics are −12.2 for monthly CPI, −9.7 for U-3 and −8.1 for payrolls, against a 5% critical value of −17.6. To assess its power, we simulate the realized design, keeping the 50 releases and the regressor, resampling residuals and injecting a linear drift in the slope across releases. Across 400 simulations, qLL rejects 13% of the time when the slope drifts from half to one and a half times the estimate, 32% from a quarter to one and three quarters, and 54% from zero to twice the estimate, with a size of 1.5%. Even large smooth drift often goes undetected, so these non-rejections do not establish stability; Section 5.1.4 examines time variation directly.

G The response along the policy path

This appendix reports the full horizon estimates behind Section 5.2 and then a simple model for reading their shape.

G.1 Full horizon estimates

Table G.1 reports the slope at every horizon of the fed funds strip, for CPI and for unemployment surprises. Table G.2 re-estimates the CPI curve on the 21 releases that quote every horizon, and Table G.3 repeats the exercise on three-month SOFR futures.

Neighboring horizons are not independent pieces of information. Over the sample the strip spans 8.3 decisions on average, and for 40% of adjacent horizon pairs no scheduled meeting falls between the contract months, so the two points price the same set of decisions and differ only in calendar averaging. In addition, 59% of contract months contain a meeting partway through and therefore average pre- and post-decision rates. Accordingly, neighboring window changes correlate at 0.996 at the median.

Table G.1: Full horizon estimates

	CPI surprise				U-3 surprise			
h	n	β_h	se	p (wild)	n	β_h	se	p (wild)
1	54	+0.133	0.029	< 0.001	52	−0.049	0.022	0.013
2	54	+0.267	0.055	< 0.001	52	−0.101	0.036	0.003
3	55	+0.324	0.096	< 0.001	52	−0.170	0.056	0.002
4	55	+0.403	0.115	< 0.001	52	−0.209	0.069	0.003
5	55	+0.464	0.129	< 0.001	52	−0.229	0.074	0.003
6	55	+0.516	0.135	< 0.001	52	−0.252	0.080	0.004
7	55	+0.536	0.138	< 0.001	51	−0.285	0.085	0.004
8	55	+0.576	0.141	< 0.001	52	−0.274	0.091	0.006
9	55	+0.591	0.140	< 0.001	52	−0.278	0.092	0.009
10	55	+0.595	0.141	< 0.001	52	−0.278	0.093	0.011
11	55	+0.609	0.141	< 0.001	52	−0.282	0.096	0.013
12	52	+0.604	0.153	< 0.001	50	−0.296	0.098	0.012
13	48	+0.579	0.158	< 0.001	47	−0.232	0.105	0.057
14	25	+0.518	0.176	0.004	—	—	—	—

Notes: Each slope uses the window revision in a ZQ contract h calendar months after the release and the family's surprise. HC3 standard errors and restricted-wild p-values (1,999 draws). The current-month contract is excluded because it averages already realized days.

Table G.2: Horizon estimates on the 21 releases quoting every horizon

h	β_h	se
1	+0.177	0.043
2	+0.360	0.113
4	+0.594	0.162
6	+0.740	0.177
8	+0.793	0.188
10	+0.823	0.189
11	+0.828	0.191
12	+0.843	0.201
13	+0.858	0.203
14	+0.783	0.199

Notes: Each row estimates the CPI slope at horizon h on the same 21 releases, those that quote all fourteen contract months; Table G.1 reports every available release. HC3 standard errors. The point estimate rises through thirteen months and falls at fourteen, but a level standard error alone is not a paired test of that fall.

Table G.3: The CPI response along the three-month SOFR futures strip

Contract month (months after release)	n	β_q	se	p (wild)
1–3	55	+0.364	0.110	< 0.001
4–6	55	+0.514	0.136	< 0.001
7–9	55	+0.588	0.141	< 0.001
10–12	55	+0.617	0.136	< 0.001
13–15	55	+0.617	0.134	< 0.001
16–18	55	+0.605	0.126	< 0.001
19–21	55	+0.586	0.120	< 0.001
22–23	36	+0.514	0.136	< 0.001

Notes: Each row regresses the $[-15, +45]$-minute change in the implied rate (100 minus price) of a quarterly three-month SOFR futures contract on the CPI surprise of the release. Rows are quarters: the q-th row uses, for each release, the quarterly contract (March, June, September or December) dated the stated number of months after the release month. Each contract settles on compounded SOFR over the three-month reference quarter beginning in its contract month, so a row averages a quarter of expected policy rather than one calendar month; the contract dated in the release month is dropped, as on the fed funds strip. The last row is quoted for 36 releases. HC3 standard errors and restricted-wild p-values use 1,999 draws for the bootstrap. On common releases, the paired increase from row 1 to row 4 is +0.253 and significant at 1% under both small-sample tests. The increase from row 1 to row 8 is +0.226 and significant at 5% under both tests. The difference between rows 4 and 8 is insignificant.

G.2 A simple model for reading the response profile

A simple adjustment rule illustrates the build-up without identifying its cause. Suppose, for this illustration, that the market's expected point-in-time policy rate follows

$$i_t = \rho\, i_{t-1} + (1-\rho)\, i_t^*, \qquad i_t^* = c + \phi\, x_t, \tag{6}$$

at monthly frequency, where i_t is the expected policy rate, i^* the desired rate and x_t the inflation measure the rate responds to. Suppose further that a CPI surprise raises x by one unit and that its contribution then fades geometrically, by a factor λ each month. Each month policy closes a fraction $1-\rho$ of the remaining gap to the desired rate, so the response builds while the gap is large and fades as the inflation contribution fades. With no initial policy revision, the expected-rate response h months ahead per unit of surprise is

$$\beta_h = (1-\rho)\,\phi\,\lambda\,\frac{\lambda^h - \rho^h}{\lambda - \rho}, \tag{7}$$

a single smooth curve that rises at one speed and fades at the other.

Equation (7) carries the identification problem of Section 5.2 in miniature. It is symmetric in ρ and λ up to the scale ϕ: exchanging the two speeds and rescaling ϕ leaves every β_h unchanged. Even if the model were literally true, the strip could therefore pin down at most the pair of speeds, never whether policy smoothing or the fading of inflation sets each one. The deferred CPI ladders speak to the second: Table 6 shows that the inflation contribution does not vanish after one month. Because those revisions are measured in the same window as the policy response, the two can also be related release by release. Holding the surprise fixed, releases with a larger next-print revision show a larger revision of the mean expected rate three to twelve months ahead, 0.53 per percentage point, but not of the next meeting, −0.43. With 35 and 37 releases, respectively, neither association is statistically distinguishable from zero.

Fitted to the estimated profile, the two speed parameters coincide, and the curve reduces to its equal-speed limit

$$\beta_h = \kappa\, h\, \rho^h, \qquad \kappa = (1-\rho)\,\phi, \tag{8}$$

which rises almost linearly at short horizons, peaks at $h^* = -1/\ln\rho$ and then declines gently, with no imposed breakpoint.

There is a second distinction between the illustration and the data. Equation (6) describes a point-in-time rate adjusting every model month. ZQ contracts price calendar-month averages around eight irregularly spaced FOMC decisions a year; a CPI reference month, its release month and a contract month are not the same clock. A structural estimate would have to propagate a rule through each release's actual meeting calendar and monthly averaging identity. We do not do that here. Fitting (8) to the estimated strip therefore makes κ a scale parameter and ρ a curve-shape parameter, not identified inflation-response and policy-inertia coefficients.

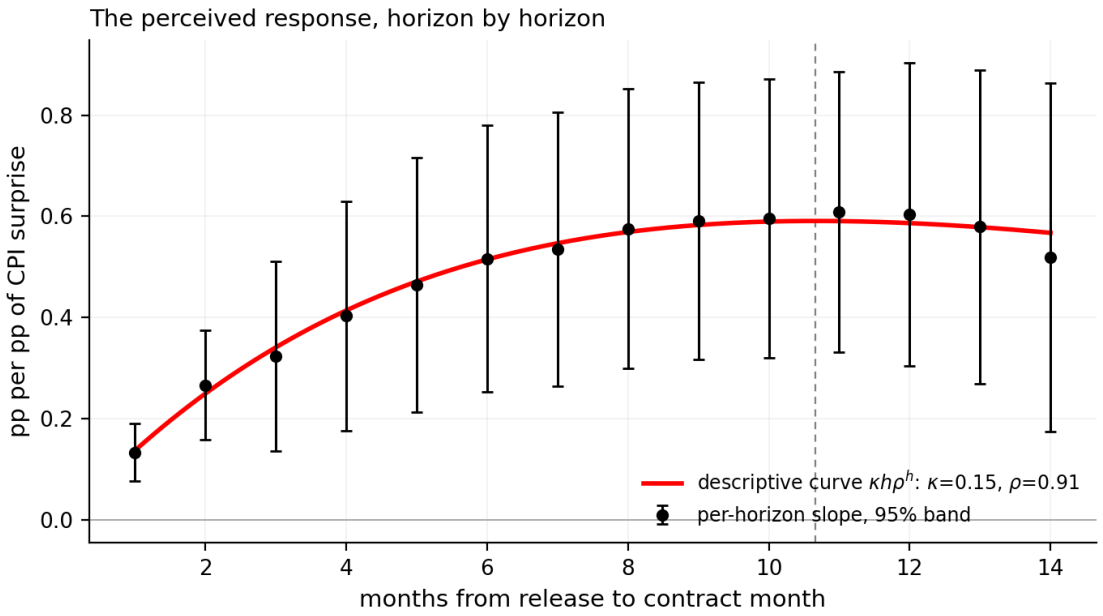


Figure G.1: A descriptive adjustment curve overlaid on the estimated profile

Notes: The black points and bands reproduce Figure 4. The red curve fits the two-parameter form in equation (8) by precision-weighted nonlinear least squares. The dashed line marks the fitted peak, $-1/\ln\hat{\rho}$, which is estimated rather than imposed. The fit summarizes scale and shape; it does not undo ZQ monthly averaging or distinguish gradual policy adjustment from expected inflation persistence.

Precision-weighted NLS on the fourteen-point curve gives

$$\hat{\kappa} = 0.151, \qquad \hat{\rho} = 0.911, \qquad R^2 = 0.99,$$

so the fitted response peaks at $h^* = 10.7$ months, where it equals 0.59. Figure G.1 draws that profile over the estimated points; it reproduces both the steep early build-up and the flattening around eleven months. The interval calculation resamples releases jointly, recomputes every horizon slope with its HC3 precision weight, and refits the curve. This preserves the dependence between horizons that a fit treating fourteen slopes as independent would miss. The 95% bootstrap interval for the peak runs from 8.0 to 17.3 months (Table G.4); it describes uncertainty in this representation, not uncertainty about structurally identified policy inertia.

The build-up is what these data establish; the long-run level is not. The balanced-panel curve gives $\hat{\kappa} = 0.207$, $\hat{\rho} = 0.913$ and a peak at 11.0 months. The Kalshi next-meeting coefficient of 0.180 lies numerically between the one- and two-month strip responses, although its meeting-specific upper-bound payoff is not the same object as either monthly-average EFFR payoff.

Across specifications the fitted peak ranges from 7.4 months on the $[-5, +15]$ window to 15.1 months when 2022 is dropped (Table G.4). This is why the paper treats the horizon profile, not the fitted pair, as the empirical object.

Table G.4: The smooth profile across specifications

Specification	$\hat{\kappa}$	$\hat{\rho}$	peak (months)	curve R^2
precision-weighted (baseline)	0.151	0.911	10.7	0.99
unweighted	0.151	0.910	10.6	0.99
$[-5,+15]$-minute window	0.165	0.874	7.4	0.47
dropping 2022	0.080	0.936	15.1	0.95
balanced release panel	0.207	0.913	11.0	0.99
range	0.080–0.207	0.874–0.936	7.4–15.1	
release-level bootstrap, 95%	[0.089, 0.213]	[0.883, 0.944]	[8.0, 17.3]	

Notes: Precision-weighted NLS of equation (8) on the horizon curve, under the stated change; the peak is $-1/\ln\hat{\rho}$. The bootstrap row resamples the 55 CPI release windows, recomputes the entire curve and refits per draw (999 draws), which is the only interval here that propagates the correlation of the horizon slopes. Allowing separate adjustment and fade speeds, equation (7), leaves the fit unchanged: the unrestricted speeds are 0.910 and 0.911. Curve R^2 is the unweighted share of cross-horizon variation in the slopes that the curve reproduces; the $[-5,+15]$-minute window gives a noisier curve that the profile fits less well.

The model thus shows why an adjustment process can produce a response that builds and then fades, and why the strip alone cannot say which speed belongs to policy. The empirical result, the paired increase in expected rates after the same releases, does not depend on it, and the fitted scale and speed should not be compared with a textbook Taylor coefficient or with structural smoothing estimates.

H Announcement repricing and prior inflation news

This appendix asks whether FOMC-window repricing is associated with CPI news released since the previous meeting. Under rational expectations, the announcement-window revision should be orthogonal to previously public information, although efficient prices still react to the announcement itself. We report the association as a supplementary diagnostic, not a third fact: it is clearest in the combined statement and opening press-conference window, but inference differs across samples and endpoints.

The sample has 41 scheduled FOMC meetings. Cumulated CPI news since the previous meeting is defined for 36 of them (two early meetings precede usable CPI ladders, and two late-2025 periods contain no CPI release because of the shutdown), and 35 of those also have a measurable futures window. To avoid de-averaging across several unknown decisions, the CME outcome is the ZQ monthly-average rate for the month after the following scheduled meeting; in 17 of the 35 observations that month contains another meeting, so we also report the 18 "clean" months separately. On Kalshi, the outcome is the ladder for the following meeting itself. The baseline window runs from 13:45 to 14:45 Eastern, covering the statement and the first fifteen minutes of

the press conference; we split it at 14:30 and extend the endpoint to 16:00 on futures, where the CME data end, and to 17:00 on Kalshi.

The pooled 13:45–14:45 CME slope is −0.074 and significant under all three tests. Its statement component is −0.064, while the pooled 14:30–14:45 increment is near zero. In the 18 clean months, the statement-only slope is −0.026 and insignificant, while the combined slope of −0.069 is significant only at 10% under the conservative rule. The exact-event Kalshi combined-window slope is −0.122 and significant at 5% under all three tests. The two cleaner constructions therefore agree in sign but not in significance, and the clean statement-only endpoint does not establish the association.

Longer windows do not strengthen inference monotonically (Table H.1 and Figure H.1). By 15:30 the pooled CME coefficient is −0.079 and remains significant, the clean-month estimate is −0.078 but is not significant at 5% under either small-sample test, and the Kalshi estimate is −0.155, significant under the permutation test but not the wild bootstrap. Kalshi reaches its largest point estimate at 15:00 and then partially fades. Longer windows add trades but little coverage: by 14:30 the median meeting has traded 13 times on the policy ladder across three strikes, and by 17:00 43 times across three or four. They also admit more communication and later news, so their response is a different object rather than a more precise estimate of the statement effect.

Table H.1: Prior CPI news and FOMC-window short-path repricing

Window or sample	n	coef.	se	p (HAC)	p (wild)	p (perm)
CME pooled, through 14:30	34	−0.064	0.029	0.024	0.032	0.028
CME pooled, through 14:45	35	−0.074	0.025	0.003	< 0.001	0.010
CME pooled, through 15:30	34	−0.079	0.033	0.015	0.001	0.045
CME clean, through 14:30	18	−0.026	0.027	0.350	0.391	0.474
CME clean, through 14:45	18	−0.069	0.030	0.022	0.005	0.099
CME clean, through 15:30	18	−0.078	0.049	0.113	0.056	0.180
Kalshi exact, through 14:30	35	−0.079	0.034	0.022	0.112	0.037
Kalshi exact, through 14:45	35	−0.122	0.046	0.007	0.039	0.016
Kalshi exact, through 15:30	35	−0.155	0.065	0.017	0.161	0.038

Notes: Each row regresses the cumulative window change from 13:45 to the stated endpoint on CPI surprises accumulated strictly since the previous scheduled meeting. CME rows use the monthly-average ZQ contract after the following meeting. "Clean" means no additional FOMC decision occurs inside that target month. Kalshi fixes the first event after the announcement on both sides of the window. No stars are used because the inference methods disagree in the small subsamples.

To gauge the economic size of the association, we take a position at 13:45 on each announcement day in a fed funds contract against the CPI news cumulated since the previous meeting, so that it gains if the expected rate falls after upside news or rises after downside news, and close it at 14:45. Table H.2 reports all eight combinations of contract horizon and sizing rule. Only the twelve-month sign-sized position clears a nominal 5% threshold after an assumed one-tick cost, with a net mean of 3.31bp; linear sizing gives 2.74bp and does not. With a median of only about 142 contracts traded in the window at that horizon, and no adjustment for the eight comparisons, this is not evidence of an implementable strategy, nor of a deliberate reversal by the Committee.

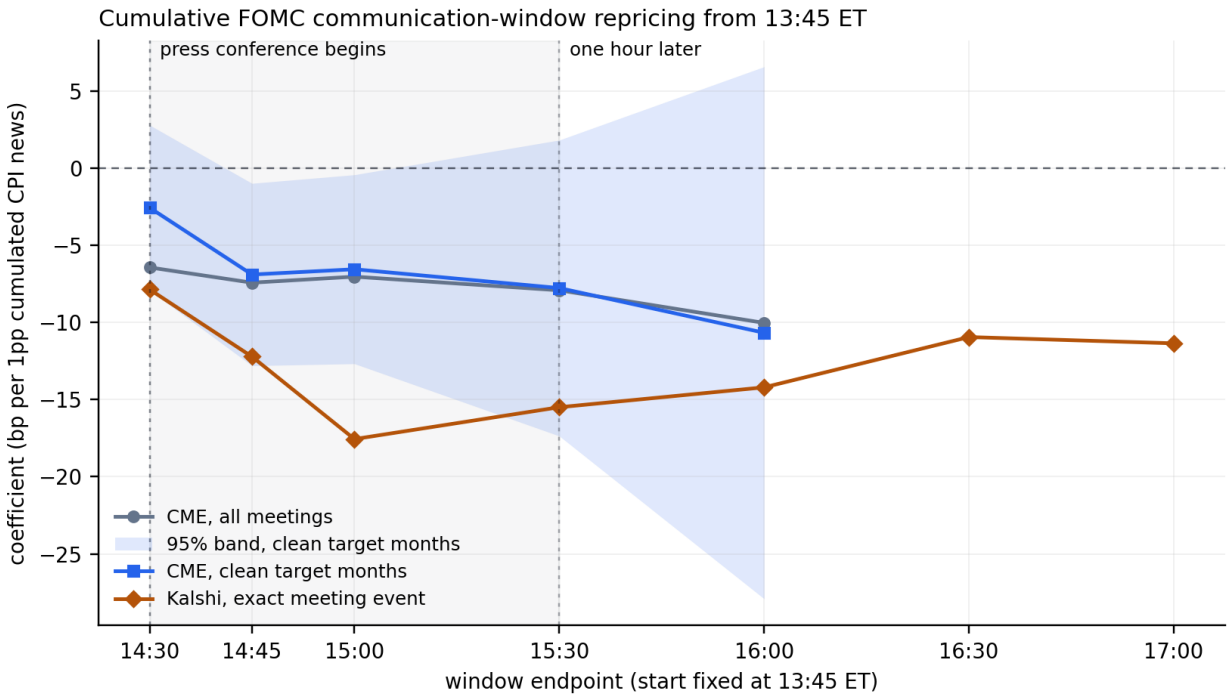


Figure H.1: Prior CPI news and cumulative FOMC communication-window repricing

Notes: Each point is the slope of the cumulative price change from 13:45 Eastern to the displayed endpoint on CPI surprises accumulated since the previous scheduled meeting; the shaded region is the pointwise 95% HAC interval for the clean-target CME series, the construction in which no second decision enters the target month; bands for the other two series are omitted for legibility. "Clean" CME target months contain no additional FOMC meeting, and Kalshi fixes the exact next meeting event. The futures data cover only the announcement windows and end at 16:00, so the two CME lines stop there; Kalshi continues to 16:30 and 17:00.

Crane *et al.* (2026) find, at a much longer horizon, that the qualitative content of FOMC statements predicts Treasury-yield and inflation-compensation changes over the following 10–15 trading days and is nearly uncorrelated with high-frequency policy surprises. Together with this diagnostic, their result suggests looking at communication beyond a single statement endpoint, although the two do not identify the same mechanism.

Relation to the "Fed response to news" channel. Bauer and Swanson (2023a,b) show that, in samples ending in 2019, news published before an FOMC announcement predicts the high-frequency policy surprise with a positive sign: stronger data-release surprises, stock prices and commodity prices are followed by hawkish surprises. They attribute this to markets underestimating how strongly the Fed would respond to that news. Hotter inflation news belongs to the same class, so under their channel the coefficients in Table H.1 would be positive. They are negative: in 2021–2026, more upside CPI news since the previous meeting is followed by a dovish revision of the expected policy path while the Committee communicates. Read at face value, markets in this cycle did not underestimate the Committee's response to inflation news ahead of its meetings; if anything, they priced more of a response than the communication then confirmed. The comparison has clear limits. The regressors differ (cumulated CPI surprises here; payroll and other release surprises and financial-market changes there), and the sample has 35 meetings. In the window closest to their

Table H.2: Fading the priced path into the meeting

Contract	Sizing	n	gross bp	net bp	net t	hit rate	median volume
2 months	sign	36	0.89	0.39	0.86	0.56	32,838
2 months	linear	36	0.63	0.20	0.36	0.56	32,838
3 months	sign	36	1.14	0.64	1.21	0.67	8,854
3 months	linear	36	0.71	0.28	0.42	0.67	8,854
6 months	sign	36	2.07	1.57	1.59	0.69	2,446
6 months	linear	36	1.53	1.10	0.99	0.69	2,446
12 months	sign	32	3.81	3.31	2.71	0.72	142
12 months	linear	32	3.18	2.74	1.77	0.72	142

Notes: Net returns subtract an assumed one-tick (0.5bp) round-trip cost per position unit, not measured execution costs or market impact. Sign sizing takes one unit against cumulated news; linear sizing scales with the news and is capped at two units. Positions open at 13:45 and close at 14:45 ET. The hit rate uses gross returns; volume is the median in the 45 minutes after the statement. The eight comparisons are in-sample and not adjusted for multiplicity.

30-minute statement window, 13:45 to 14:30, the pooled CME coefficient is negative and significant but the clean-target coefficient is not. A premium that builds with inflation news and resolves at the statement would produce the same sign. We therefore read the sign difference as a feature of this cycle that merits a longer sample, not as evidence against their channel in their sample.