# A Simple Weekly Signal for the U.S. 10-Year Treasury Yield

*A checklist for timing duration adjustments after forming a macro view*

Chirag Mirani

*Independent Researcher*

November 7, 2026 | Working paper, version 5

**Abstract**

A macro view tells a bond investor why to add or reduce duration. It still leaves a timing decision. I present a simple weekly signal for the U.S. 10-year Treasury yield to help review that decision. Fourteen market, economic and calendar readings form an equal-vote checklist. A logistic regression, refitted every 13 weeks, provides a second view; the combined signal takes a direction when the two do not conflict. Across November 2016-October 2026, the selected model gets 52.9% of 342 active calls right and scores +32.6 yield basis points per year, or +27.2 after an illustrative charge. The refit frequency and 30-year TIPS auction addition were chosen after reviewing the historical results; the paper reports the searches and benchmarks alongside the selected model. A practical example shows how an investor with a bullish macro view can use the readings to add duration, wait or build gradually. The public tool supplies the dated checklist and an accumulating record of weekly outcomes. Scores measure yield direction, not portfolio returns; allocation size remains a portfolio decision.

**AI assistance:** OpenAI Codex assisted with code development, validation, editing and presentation. The author is responsible for the analysis and final text.

**One-week-ahead outlook for the 10-year Treasury yield: Neutral.**

**Signal generated as of:** 07 November 2026, 03:48:18 PM EST (New York).
**Outlook for the week ending:** 13 November 2026 close, based on the 06 November 2026 close.

**Model 1 - checklist:** Bearish Duration, score -2 (5 add, 7 reduce, 2 neutral). **Model 2 - regression:** Bullish Duration, 56.6% for lower yields (uncalibrated).

Issued before the following trading week. See the tool for subsequent readings and outcomes.

The models disagree. Neutral means no additional weekly duration tilt.

**Signal tool and weekly record:** chiragmirani.github.io/treasury-signal/

**Preprint, code and results:** chiragmirani.github.io/treasury-signal/paper/

**Keywords:** U.S. Treasury yield forecasting; duration management; bond market timing; checklist; logistic regression; Treasury auctions; real-time data.

**Questions and comments:** chiragmirani@gmail.com

## 1. A macro view still leaves a timing decision

I may expect slower growth, lower inflation and an easier Fed to bring yields down over the next few months. That gives me a reason to own more duration. It does not tell me whether to add all of it today, wait for a better entry or build the position in stages.

This checklist is an additional tool for that decision. It brings market prices, economic readings and a known auction date into one weekly view of the U.S. 10-year Treasury yield. I want to see the balance of evidence, which readings disagree with my macro view and whether the statistical model sees the same thing.

The output is a duration view **one week ahead: Bullish Duration, Bearish Duration or Neutral**. Bullish favors lower yields; bearish favors higher yields. The investor starts with a macro position and a risk budget. Neutral adds no weekly tilt to that plan; it does not mean sell the portfolio or return to benchmark.

**Why a simple model?** In practice I also use more complex models, including nonlinear methods and ensembles. Here I want the reader to see every input, understand each vote and reproduce the calculation. A checklist and logistic regression make that possible.

## 2. Fourteen readings, two views, one weekly signal

Model 1 is the checklist. It covers equities, oil, the dollar, volatility, unemployment, inflation, the yield curve, foreign yields, December seasonality and an upcoming 30-year TIPS auction. Each reading contributes +1 for lower yields, -1 for higher yields or zero. The total is the checklist score. Appendix A gives every definition.

Model 2 is a logistic regression. It uses the same information, allowing the size of a move and the weight on each input to differ. It refits every 13 weeks using all available past history. This is a quarterly refit, not a 13-week training window. A model reading above 50% favors lower yields; below 50% favors higher yields.

When the directions agree, I keep the shared view. When they disagree, the combined signal is zero. If the checklist is exactly zero, the regression decides. The calculation is position = sign(sign(checklist score) + sign(probability - 0.5)). Positions are always +1, -1 or zero. The score and probability describe the reading; neither automatically scales exposure.

For the auction input, a 30-year TIPS auction must already have been announced before the reference Friday and occur in the following seven days. It contributes one add-duration vote and a binary regression input. Auction yields, bid-to-cover ratios and other results are never used. This is a calendar relationship selected from the data, rather than an assumption that more supply must lower yields.

## 3. How I would use it with a bullish macro view

Suppose my portfolio and benchmark both have six years of duration. My bullish macro view supports a target of seven years, within the risk budget. I set that target before looking at the checklist. The weekly reading then helps me review the entry.

***Practical example.*** *Hypothetical readings and portfolio decisions. The one-year target and 0.25-year steps are the investor's choices.*

| Decision | Weekly evidence | Implementation |
|---|---|---|
| **Defer addition** | 5 add, 7 reduce, 2 neutral; score -2. Regression 46% for lower yields. Combined: Bearish Duration. | Both views oppose my planned addition. I hold at 6.00 years and review the entry against my macro view. |
| **Add** | 7 add, 5 reduce, 2 neutral; score +2. Regression 56% for lower yields. | The reading supports my macro view. If the planned adjustment still fits, I move toward 7.00 years. |
| **Build gradually** | The same supportive reading, with a preference for staged entry. | I add 0.25 years to reach 6.25, then reassess both views before the next step. |

The last two choices deliberately use the same signal. Position size comes from the portfolio plan, conviction in the macro thesis and tolerance for an adverse move. Staging reduces initial exposure but may miss part of a rally. If the two model views disagree, I have no additional tactical support for the change.

For the week from 06 November 2026, the combined model is Neutral. Model 1 is Bearish Duration (score -2); Model 2 is Bullish Duration (56.6% for lower yields, uncalibrated).

## 4. What the historical results show

The selected model combines fourteen rules with quarterly refitting. Across 521 completed weeks from November 2016 to October 2026, it is active in 342 weeks and gets 52.9% of those calls right. Its average signed yield-change score is +32.6 bp per year. The last 52 completed weeks contribute +52 bp.

***Table 1. Historical comparison.*** *November 2016-October 2026. Hit rates use active weeks. Scores and Newey-West t-statistics use all 521 weeks, including flat weeks as zero. Net subtracts an illustrative 0.25 yield bp per position unit changed.*

| Model | Active | Hit | Gross bp/yr | Net bp/yr | t |
|---|---|---|---|---|---|
| 14 rules + quarterly refit (selected) | 342 | 52.9% | +32.6 | +27.2 | 1.67 |
| 13 rules + quarterly refit | 342 | 52.6% | +30.6 | +25.4 | 1.57 |
| 13 rules + annual refit | 314 | 51.0% | +22.6 | +17.6 | 1.20 |
| Always expect higher yields | 521 | 48.4% | +31.4 | +31.4 | 1.17 |

I score a lower-yield call positively when yields fall and a higher-yield call positively when they rise. A flat week scores zero; an unchanged yield is not a winning call. These are yield-direction scores, not returns from a bond or ETF. The score excludes carry, financing, duration changes and execution prices.

The auction addition improves the quarterly thirteen-rule model by 2.0 bp/year, with a paired t-statistic of 1.27. Always expecting higher yields scores 31.4 bp/year over this period. The selected model's paired difference against that benchmark has t = 0.04; after the assumed charge, the benchmark scores more. The benefit to examine is therefore the pattern of weekly decisions as well as the average.

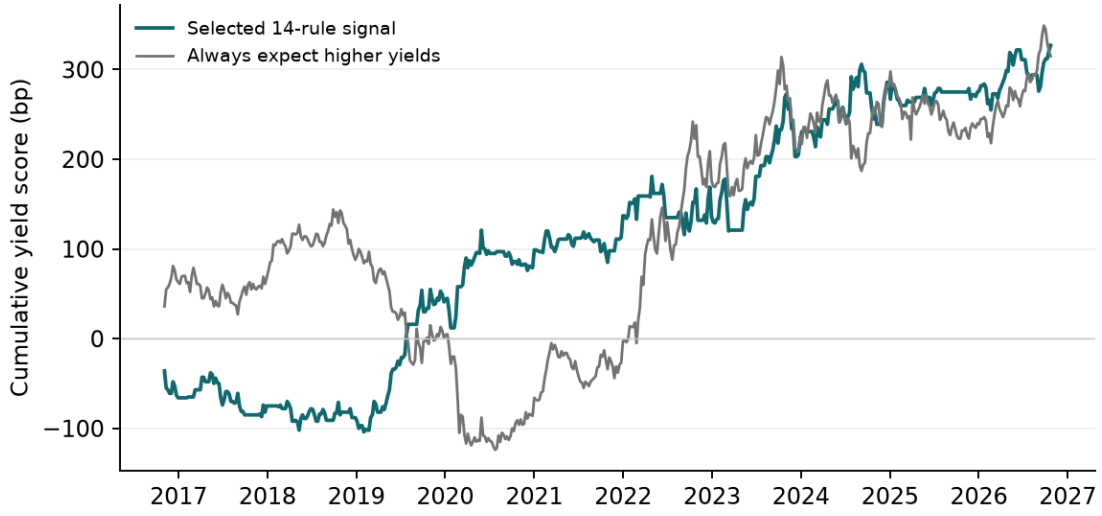



***Figure 1.*** *Selected model and constant higher-yield benchmark. Cumulative signed yield changes, November 2016-October 2026; before the illustrative charge.*

***Table 2. Selected model by year.*** *2016 begins in November; 2026 ends with the October 23 signal and October 30 outcome.*

| Year | Weeks | Active | Total score, bp |
|---|---|---|---|
| 2016 | 9 | 8 | -66 |
| 2017 | 52 | 30 | -9 |
| 2018 | 52 | 31 | -13 |
| 2019 | 52 | 41 | +129 |
| 2020 | 52 | 37 | +38 |
| 2021 | 53 | 36 | +58 |
| 2022 | 52 | 29 | -1 |
| 2023 | 52 | 38 | +86 |
| 2024 | 52 | 37 | +62 |
| 2025 | 52 | 23 | -9 |
| 2026 | 43 | 32 | +52 |

The most recent 52-week window runs from the October 31, 2025 signal through October 23, 2026, with outcomes through October 30. It has 37 active weeks, a 62.2% hit rate and a +52 bp total score. This is a fixed window in the paper, not a rolling claim about the date on which someone reads it.

***Table 3. Last 52 weeks by quarter of signal date.*** *First and last quarters are partial.*

| Quarter | Weeks | Active | Yield change, bp | Signal score, bp |
|---|---|---|---|---|
| 2025Q4 | 9 | 5 | +8 | +0 |
| 2026Q1 | 13 | 11 | +16 | +2 |
| 2026Q2 | 13 | 11 | +14 | +34 |
| 2026Q3 | 13 | 7 | +79 | -3 |
| 2026Q4 | 4 | 3 | -34 | +19 |

## 5. Finding and testing the inputs

I start with an economic reason to examine a variable, specify its reading and test it on its own. Rising oil, for example, can point to inflation pressure; a stronger dollar can tighten financial conditions. These hypotheses concern the next week, not just the contemporaneous market reaction. Appendix E explains what each input measures and where its interpretation can break down.

I then ask whether a candidate helps the combined model. The initial forward search picks two additions using the early period: Next Friday falls in the first seven days of a month (jobs calendar proxy) (yes = sell); 10-year yield below its 300-day average (yes = sell). It produces a 15-rule model with a later score of 31.9 bp/year. Its paired gain over the annual reference has t = 0.78. The full candidate list and both selection steps remain in the results package.

A broader search tests 2,430 classifier settings and then 1,440 ridge-regression settings. These cover training windows, refit frequency, input sets, thresholds and scaled positions. The in-sample winners score 7.0 and 2.6 bp/year in the later period. They do not improve the reference average. Appendix D reports the grids and winners, including the unsuccessful approaches.

I also test upcoming nominal 20-year, nominal 30-year and 30-year TIPS auctions, plus an any-long-auction flag, in both directions. With annual refits, the early winner is an any-auction reduce-duration vote. With quarterly refits, the early winner is a nominal 30-year add-duration

vote, whose later score is 26.3 bp/year. I choose the quarterly 30-year TIPS add-vote version after comparing its later results. Appendices F and G retain every auction variant, so the reader can see the choice in context.

---

**What was decided when.** I developed this implementation in November 2026 after inspecting history through October. On November 7 I corrected timing, archive and curve-tie issues and reran the calculations. The first feature search and the two broader searches rank candidates on the early period; the broader searches require every ranking outcome to end before November 1, 2016. I then compare refit frequencies, choose quarterly refitting using the later results, and repeat the auction search on that baseline. Version 4 adopts the 30-year TIPS add-vote variant after reviewing those later results. The 2027 watchlist separately selects directions using 2026 outcomes. These are disclosed research choices: the later period is retrospective evaluation, not an untouched test. Version 5 leaves those rules unchanged and adds checked original-publisher market-data fallbacks. The first prospective call is generated on November 7 from November 6 reference data, for the week ending November 13, 2026.

---

## 6. Data timing and strength of evidence

I rebuild unemployment, core PCE inflation and foreign yields from dated ALFRED snapshots. The unemployment comparison uses distinct earlier months; inflation is calculated within one vintage. Foreign-yield rules are neutral before their archives begin. The dollar basket uses six raw exchange rates. Currency and oil observations are delayed seven days. For live updates, missing recent Treasury yields and VIX closes can come directly from Treasury and Cboe after twenty overlapping observations match FRED. Existing history is preserved and actual retrieval times are recorded. Auction flags use announcement dates, with a strict before-Friday cutoff.

These steps address publication timing, but current market histories can still contain corrections or backfills, and not every historical intraday publication time has been recovered. The auction audit verifies a sample of original notices, not the entire calendar. Appendix C gives the calculations and data checks; Appendices F-G describe the event tests.

For the selected model, Newey-West t-statistics are 1.67 with four lags and 1.74 with eight. The simple 95% hit-rate interval is 47.6%-58.2%. These statistics do not adjust for the searches and later-informed choices. The regression's Brier loss is 0.2505, compared with 0.2500 for a constant 50% forecast. I use the probability as a model reading, not a calibrated confidence level or a sizing formula.

The annual-reference placebo gives 43 exceedances in 500 circular shifts, or 0.0878 using (k+1)/(N+1). Repeating the first two-addition search gives 11 in 200, or 0.0597. Those diagnostics do not repeat the later refit and auction choices and are not significance tests for version 4. The input-substitution study also changes economic definitions, so it cannot isolate the effect of revised data alone.

The illustrative cost charge reduces the selected model's score to 27.2 bp/year, counting opening, closing and intermediate position changes. It is a sensitivity assumption, not a measured trading cost. A portfolio evaluation would record actual implementation prices and compare the macro strategy with and without the tactical adjustments. The public record will distinguish those decisions from the model's yield score.

## 7. What I will watch from 2026 into 2027

A separate descriptive screen ranks both directions of 27 existing rules on 43 completed 2026 weeks, January 2-October 23, with outcomes through October 30. I subtract the score of each rule's average constant tilt to isolate its timing score. The shortlist requires at least 20 active weeks, eight observations in each state and four state changes. All 54 direction trials are saved.

***Table 4. Relationships to follow in 2027.*** *Selected on completed 2026 outcomes. Next-week mean yield changes show the two states, with counts in parentheses. Timing scores are period totals before costs. Earlier columns use the 2026-selected sign.*

| Reading | True: next bp | False: next bp | 2026 timing | 2023 | 2024 | 2025 |
|---|---|---|---|---|---|---|
| 1. Four-week natural-gas change > 0 | -1.37 (19) | +4.57 (23) | +124.0 | -85.7 | -148.9 | +37.4 |
| 2. Four-week change in the 2s10s spread > 0 | -2.15 (13) | +3.59 (29) | +104.1 | -145.8 | +21.8 | -1.7 |
| 3. Four-week WTI change > 0 | +0.33 (27) | +4.13 (16) | +76.2 | -31.8 | -61.3 | +15.2 |
| 4. Four-week dollar change > 0 | +0.14 (22) | +3.43 (21) | +70.7 | +34.2 | -104.7 | +0.0 |
| 5. Latest published monthly German yield change > 0 | +0.42 (26) | +3.76 (17) | +68.7 | +112.9 | +60.4 | +2.7 |

Natural gas is a useful example of why the sign needs explaining. Higher energy prices can suggest higher inflation, yet in this selected 2026 sample rising gas is followed by an average -1.37 bp yield move, versus +4.57 bp in the other state. A reversal after information has been priced in is one possible explanation. The data do not observe the original news surprise or expectations, so they do not establish that mechanism. The sign also varies across earlier years.

I would use the list to ask what the market has already priced, whether the same relationship is persisting and what changed when it fails. It describes selected historical associations, not proven causes of market moves or a ranking of the best future predictors. Natural gas remains outside the fourteen-rule model. Keeping the watchlist separate lets the research evolve without changing the published signal's rules.

## 8. The weekly tool and how to follow it

The public signal tool leads with the latest complete dated reading, the balance of add and reduce votes, the regression view and the resulting duration interpretation. The factor table shows the individual votes. The preprint page provides this paper, the code and downloadable historical results.

The page can be checked or refreshed on any day. Its tested horizon remains Friday to Friday. A Monday visit retrieves the existing weekly reading; it does not silently create a Monday-to-Monday forecast. New issues require fresh complete inputs inside the Friday-evening-to-Saturday issue window. Missing inputs hold back a new issue while the last complete reading stays visible with its original dates.

Each published forecast records its actual issue time and model version. Reruns preserve it. Outcomes are added separately as they become available, and the tracker distinguishes historical reconstruction from forecasts actually issued before the outcome. Weekly results are signed yield changes, with flat and unavailable weeks identified separately. The first call uses November 6 reference data and is generated on November 7, 2026, before the following trading week. Its benchmark starts at the Friday reference yield, not at an executable price at issue time.

I record the macro view, target duration, weekly evidence and action taken to see whether the checklist changed the timing. Questions are welcome at chiragmirani@gmail.com.

## Appendix A. Rule definitions

***Table A1.*** *Direction when the condition is true; otherwise the opposite direction applies unless a neutral case is listed. Unavailable research inputs are neutral; production requires the declared inputs to be fresh. Continuous regression inputs are standardized; December and the auction flag stay binary.*

| # | Condition | True direction | Neutral | Timing | Regression input |
|---|---|---|---|---|---|
| 1 | Nasdaq Composite 4-week return > 0 | Sell | Return exactly 0 | Friday close | 4-week log return |
| 2 | Nasdaq Emerging Markets 4-week return > 0 | Sell | Return exactly 0 | Friday close | 4-week log return |
| 3 | WTI oil 4-week return > 0 | Sell | Return exactly 0 | Observation + 7 days | 4-week log return |
| 4 | Six-currency dollar index 4-week change > 0 | Buy | Change exactly 0 | Observation + 7 days | 4-week change |
| 5 | VIX 4-week change > 0 | Buy | Change exactly 0 | Friday close | 4-week change |
| 6 | Unemployment earliest archived monthly value > mean of 60 earlier monthly values | Buy (else sell) | Not yet published | Release date (archive) | Gap |
| 7 | Core PCE inflation, 12-month (within vintage) < 2% | Buy (else sell) | Not yet published | Release date (archive) | Inflation minus 2 |
| 8 | 10-year yield minus core PCE inflation < 0 | Sell (else buy) | Inflation unknown | Friday close + archive | Level |
| 9 | 2s10s spread 4-week change > 0 | Buy | Change exactly 0 | Friday close | 4-week change |
| 10 | German 10-year yield, latest monthly change > 0 | Sell | Change exactly 0 | Release date (archive from 2013) | Change |
| 11 | Japanese 10-year yield, latest monthly change > 0 | Sell | Change exactly 0 | Release date (archive from 2013) | Change |
| 12 | UK 10-year yield, latest monthly change > 0 | Sell | Change exactly 0 | Release date (archive from 2013) | Change |
| 13 | Signal date is in December | Sell | Any other month | Calendar | 0/1, not standardized |
| 14 | 30-year TIPS auction in next 7 days, announced before Friday | Buy | No eligible event | Announcement strictly before Friday | 0/1, not standardized |

The inflation-adjusted measure is nominal DGS10 minus trailing core PCE inflation. It is not a TIPS real yield or an estimate of expected real returns, term premium or fair value. The unemployment average is a recent-history benchmark, not a natural-rate estimate. The dollar basket uses EUR, JPY, GBP, CAD, CHF and AUD with equal log weights and fixed quotation orientation. Oil and currency changes compare their delayed observations four weeks apart.

## Appendix B. Each rule on its own

***Table B1.*** *Hit rates use active weeks; scores and HAC t-statistics include all calendar weeks. Additional-rule directions are chosen on the early sample, so their early results are selected results. Early coverage differs across rules.*

| Group | Rule | 2001-Oct 2016 | | | Nov 2016-Oct 2026 | | |
|---|---|---|---|---|---|---|---|
| | | Hit | bp/yr | t | Hit | bp/yr | t |
| Cross-asset moves | Nasdaq Composite 4-week return > 0 | 47.9% | +1 | +0.03 | 50.1% | +31 | +1.40 |
| Cross-asset moves | Nasdaq Emerging Markets 4-week return > 0 | 51.9% | +31 | +1.33 | 47.6% | +16 | +0.67 |
| Cross-asset moves | WTI oil 4-week return > 0 (one-week-delayed observations) | 48.9% | +11 | +0.47 | 49.4% | +29 | +1.12 |
| Cross-asset moves | Six-currency dollar index 4-week change > 0 | 50.3% | +4 | +0.18 | 49.5% | +3 | +0.13 |
| Cross-asset moves | VIX 4-week change > 0 | 48.3% | -22 | -1.00 | 47.9% | -1 | -0.04 |
| Macro state | Unemployment earliest archived monthly value > mean of 60 earlier monthly values | 49.0% | +12 | +0.58 | 48.2% | -4 | -0.14 |
| Macro state | Core PCE inflation, 12-month (within vintage) < 2% | 52.0% | -4 | -0.17 | 51.2% | +36 | +1.34 |
| Rates-market state | 10-year yield minus core PCE inflation < 0 | 52.3% | +25 | +1.06 | 51.6% | +33 | +1.25 |
| Rates-market state | 2s10s spread 4-week change > 0 | 49.2% | -10 | -0.45 | 49.3% | +7 | +0.28 |
| Foreign bond markets | German 10-year yield, latest monthly change (as published) > 0 | 46.9% | -1 | -0.13 | 46.8% | -24 | -0.89 |
| Foreign bond markets | Japanese 10-year yield, latest monthly change (as published) > 0 | 48.3% | +1 | +0.16 | 47.0% | +8 | +0.29 |
| Foreign bond markets | UK 10-year yield, latest monthly change (as published) > 0 | 47.8% | +1 | +0.18 | 45.9% | -29 | -1.17 |
| Calendar | The signal date is in December | 61.5% | +4 | +0.66 | 48.9% | +6 | +0.89 |
| Energy | WTI oil 52-week return > 0 [yes = sell] | 48.5% | +10 | +0.45 | 50.5% | +27 | +1.00 |
| Energy | Retail gasoline 4-week change > 0 [yes = sell] | 49.4% | +35 | +1.53 | 46.9% | +15 | +0.55 |
| Energy | Natural gas 4-week return > 0 (one-week-delayed observations) [yes = buy] | 51.1% | +15 | +0.76 | 48.6% | -10 | -0.41 |
| Energy | Oil volatility (OVX) 4-week change > 0 [yes = buy] | 49.1% | +5 | +0.26 | 49.3% | +17 | +0.64 |
| Inflation expectations | 10-year breakeven 4-week change > 0 [yes = sell] | 50.1% | +26 | +1.23 | 51.2% | +52 | +2.11 |
| Supply | Fed balance sheet 4-week change > 0 [yes = sell] | 49.4% | +15 | +0.69 | 48.4% | +8 | +0.31 |
| Yield dynamics | 10-year yield below its 300-day average [yes = sell] | 48.1% | +17 | +0.75 | 45.5% | -30 | -1.10 |
| Yield dynamics | 10-year yield more than 2.5 average daily moves above its 10-day low [yes = buy] | 52.1% | +12 | +0.93 | 45.9% | -22 | -1.23 |

| Group | Rule | 2001-Oct 2016 | | | Nov 2016-Oct 2026 | | |
|---|---|---|---|---|---|---|---|
| | | Hit | bp/yr | t | Hit | bp/yr | t |
| Yield dynamics | Friday's 10-year yield in the top 30% of the week's range [yes = buy] | 52.1% | +22 | +1.59 | 47.4% | -12 | -0.74 |
| Calendar events | A scheduled Fed decision falls in the next 7 days [yes = buy] | 52.8% | +11 | +1.04 | 53.8% | +13 | +1.26 |
| Calendar events | Next Friday falls in the first seven days of a month (jobs calendar proxy) [yes = sell] | 52.1% | +17 | +1.38 | 50.4% | +0 | +0.00 |
| Calendar events | A Fed decision falls in the next 7 days and the VIX rose this week [yes = sell] | 54.0% | +3 | +0.52 | 46.2% | +4 | +0.71 |
| Calendar events | A Fed decision or jobs calendar proxy falls next week and the VIX rose this week [yes = sell] | 54.4% | +17 | +1.64 | 50.7% | +8 | +0.71 |
| Calendar events | VIX above 3-month VIX [yes = buy] | 47.6% | +7 | +0.74 | 51.6% | -5 | -0.74 |

## Appendix C. Data, calculations and reproduction

**Timing.** ALFRED dates identify source, provider or archive availability; they are not a guarantee of each original intraday release. Core PCE growth uses the latest month and the same month a year earlier in one snapshot, including revisions already present there. The unemployment average excludes the current month and uses 60 distinct earlier available months. Before there is enough history, the rule is neutral. Foreign-yield rules are neutral before June 2013. The early model therefore has fewer than thirteen informative inputs.

**Archive checks.** I reject nonfinite values, duplicate or nonconsecutive months, observations ahead of their vintage, very old latest observations and backwards jumps in the latest month. One German snapshot dated August 1, 2013 ended in December 1969 and is excluded. No revised-data fallback fills the early foreign-yield gaps. Current market histories can still contain revisions or backfills. Treasury reference-yield quotations are around 3:30 p.m., while some other inputs close later.

**Calendar rules.** The supplied Fed dates are scheduled decisions, but historical announcement-time schedule revisions have not been fully audited. The jobs rule means that next Friday falls in the first seven days of a month. It is a calendar proxy and never reads future realized jobs-release dates. Neither event rule is in the thirteen-rule specification.

**Training and scoring.** Continuous inputs use an expanding mean and standard deviation, including the current observable value, with at least 52 observations; standardized values are clipped at four and missing values become zero. December and the selected auction flag stay 0/1. Four-week curve changes are rounded to ten decimal places in percentage points before taking a sign, so exact quoted ties stay neutral. At each refit, the regression uses only training weeks whose outcomes have ended by that Friday. Unchanged targets are classified as not down. HAC statistics use an intercept-only regression with Bartlett weights and four lags; eight-lag results are also shown. Volatility and SR use all calendar weeks and annualize with 52. Wilson hit-rate intervals assume independent observations and do not adjust for selection.

**Sample boundaries.** The original tables define the early period by signal date: the last early signal is October 28, 2016 and its outcome is November 4, before that Friday's next forecast. This convention also applies to the initial two-addition search. The two broader searches impose a stricter

information cutoff: all ranking outcomes end before November 1, leaving 564 early signal weeks. The later evaluation has 521 completed weekly outcomes. Circular-shift diagnostics use offsets of at least 52 weeks and compare unrounded statistics; their reported fraction is (k+1)/(N+1).

**Reproduction.** The package contains the research model, downloader, input-sensitivity check, weekly script, frozen specification, Fed-date file, dependencies and manuscript builder. The improvement_study folder contains both protocols, both search scripts and every trial. The watchlist_2027 folder contains its dated protocol, all 54 direction trials, every rule's ranking and selected weekly scores. The economic_review folder contains interpretation notes for all 27 rules and the post-discussion reversal and curve checks. The auction_study folder contains the additional protocol, official calendar records, announcement audit, timing tests and all eight variants. The separate refit comparison is in improvement_study/refit_comparison.json. README.txt gives the commands. Downloads create a new snapshot and do not overwrite the archived local study inputs. A research run cannot overwrite the production specification. Raw market and macro files are excluded; their hashes identify the archived study inputs. A later download can change the historical results and need not reproduce the paper exactly.

**References and source documentation**

Croushore, D. and T. Stark (2001). A real-time data set for macroeconomists. *Journal of Econometrics* 105(1).

Harvey, C. R., Y. Liu and H. Zhu (2016). ...and the cross-section of expected returns. *Review of Financial Studies* 29(1).

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3).

Federal Reserve Bank of St. Louis. ALFRED documentation.

Federal Reserve Board. H.10 publication schedule and revisions.

U.S. Treasury. Interest-rate statistics.

FRED. Nasdaq Composite and Nasdaq Emerging Markets source and redistribution notices.

Scikit-learn. Probability calibration and interpretation of Brier loss.

## Appendix D. Broader searches and the 2027 watchlist

The five input sets are: the main thirteen; those thirteen without the three foreign yields; seven inputs covering US and EM equities, oil, dollar, VIX, curve and December; thirteen plus breakeven inflation; and thirteen plus the jobs proxy and 300-day yield trend. The last two are named candidates carried into this search, not new discoveries independent of earlier results.

The classification grid uses C = 0.01, 0.1 or 1; all history, 520 weeks or 260 weeks of training; 52- or 13-week refits; minimum absolute checklist scores of 0, 1 or 3; probability distances from 0.5 of 0, 0.025 or 0.05; and three sizing rules. Probability sizing is min(abs(p-0.5)/0.10, 1); checklist sizing is abs(score)/number of rules. The existing combination supplies the direction. These choices give 2,430 settings.

The ridge grid uses the same five input sets, training windows and refit intervals; alpha = 1, 10, 100 or 1,000; with or without the checklist non-conflict condition; minimum absolute predicted moves of 0, 1 or 2 bp; and either a unit position or min(abs(predicted move)/5, 1). These choices give 1,440 settings. Positive forecasts mean lower yields.

For each grid, I rank the early calendar-week SR after a 0.25 bp charge per position unit. A candidate must be active in at least 25% of weeks, average at least 0.10 absolute exposure, and have positive cost-adjusted scores in both 2006-2010 and 2011-October 2016. The overall winner and the best setting within each input set are saved before their later results are calculated. The table reports every input-set winner, including failures. The second grid was specified after reviewing the first grid's result.

***Table D1.*** *Each input set's selected model, or failure to qualify. Early SR and later net scores include the illustrative charge. Exposure is mean absolute position. These statistics have not been adjusted for selection.*

| Search / inputs | Early net SR | Later bp/yr | Later net bp/yr | Later t | Exposure |
|---|---|---|---|---|---|
| Classification: Main 13 | 0.52 | +3.0 | -1.0 | 0.18 | 0.46 |
| Classification: No foreign yields (10) | 0.45 | -3.6 | -7.4 | -0.22 | 0.44 |
| Classification: Market inputs (7) | 0.46 | -1.4 | -2.9 | -0.16 | 0.18 |
| Classification: With breakeven (14) | 0.59 | +13.8 | +9.7 | 0.84 | 0.43 |
| Classification: With calendar and trend (15) | 0.78 | +7.0 | +2.6 | 0.46 | 0.36 |
| Ridge: Main 13 | 0.35 | -1.9 | -5.1 | -0.13 | 0.31 |
| Ridge: No foreign yields (10) | 0.37 | -4.3 | -8.3 | -0.22 | 0.57 |
| Ridge: Market inputs (7) | No setting passed the early-sample eligibility rules | | | | |
| Ridge: With breakeven (14) | 0.43 | +2.6 | -0.6 | 0.17 | 0.40 |
| Ridge: With calendar and trend (15) | 0.37 | +1.5 | -1.8 | 0.11 | 0.38 |

Detailed parameter choices, annual blocks, paired differences and full weekly positions are in the accompanying JSON and CSV files. The first grid's overall winner has a paired later-period difference of -15.6 bp/year versus the annual reference model (t = -1.22); the ridge winner's difference is -20.0 (t = -1.12). Neither this table nor a larger search can turn already-inspected history into a prospective test.

**2026 timing attribution.** With p denoting the duration view and dy the following week's yield change in bp, the total timing score is sum[-(p - mean(p)) dy]. Each displayed year or subperiod uses its own average tilt; subperiod timing scores need not add to the full-window timing score. The constant comparator is constructed after observing the window. No t-statistic is attached to this selected ranking. The 2023-2025 columns use the 2026-selected direction unchanged, with years assigned by signal date. Same-week comparisons use unlagged observation histories, log changes for equities and dollar, and level changes for WTI and VIX; they are never forecast inputs. Monthly foreign yields can repeat one reading for several Fridays, so week counts overstate independent information.

## Appendix E. The economic reasoning behind the votes

**Start with the horizon.** A release's level, its surprise and the price response already behind us are different observations. A weekly forecast asks what happens next. The same information can motivate continuation if adjustment is incomplete or reversal if prices overshot. The sign alone cannot identify which occurred. News research typically measures the surprise relative to prior expectations (New York Fed); this checklist does not observe that surprise.

**Curve compensation.** The add-duration vote after steepening is motivated by the possibility that the extra yield farther out the curve attracts demand. Carry and roll are the intended intuition. But a change in 2s10s does not calculate either return, and a constant-maturity ten-year yield does not follow a bond as it ages. Steepening can occur with rising yields, falling yields, or a curve that remains inverted. It does not by itself imply policy easing. Curve positions and outright duration are distinct choices (CFA Institute).

**Inflation and employment.** Lower core inflation and unemployment above its recent norm are treated as conditions that may favor a lower future policy path. The fixed 2% core-PCE cutoff is a research convention. The formal Fed target concerns headline PCE; the numerical goal was announced in 2012, and maximum employment has no corresponding fixed numerical goal (Federal Reserve). The five-year unemployment average does not estimate the natural rate. Neither rule measures the next policy decision's surprise.

**Yield compared with past inflation.** I retain nominal yield minus trailing core inflation as a simple comparison with experienced inflation. The ten-year yield and past twelve-month inflation have different horizons. Its sign does not establish expected purchasing-power returns or whether a bond is cheap. The retained directional vote must stand on its empirical results.

**Market prices and risk.** Equity gains may signal easier conditions or stronger growth, motivating higher-yield votes. They do not demonstrate money leaving bonds. The dollar vote instead hypothesizes tighter conditions and less inflation pressure; a repricing toward higher US policy rates can lift both the dollar and yields. VIX is option-implied equity volatility (Cboe), not Treasury buying. Its add-duration vote assumes a demand-for-safety channel that need not dominate inflation or liquidity shocks.

**Energy and international rates.** Rising energy prices motivate a continuation hypothesis about inflation pressure, while the 2026 watchlist explores different subsequent associations. Foreign-yield votes assume common rate forces or spillovers can persist. Archived monthly foreign yields are lagged states, not contemporaneous market quotes, observed Treasury purchases, or a demonstrated lead.

**Calendar and additional readings.** December is a seasonal hypothesis; the calendar does not measure liquidity or imply a sell-off. Its later record contains 45 weeks from 10 Decembers. Fed and first-Friday flags identify events, not the signs of surprises. The selected below-300-day-average rule bets on a yield rebound, so its name should not be mistaken for trend following. Breakeven inflation includes risk and liquidity premia; total Fed assets include different balance-sheet operations. Volatility and price-pattern candidates similarly require explicit hypotheses. All 27 are discussed in the accompanying notes.

**From a view to a portfolio.** A higher probability of a yield fall need not imply a positive expected score if the less frequent rises are larger. For example, 60% odds of a 1 bp fall and 40% odds of a 3 bp rise give a long-direction expected score of -0.6 bp. The classifier's 50% cutoff is a direction rule, not an optimal duration allocation. A ten-year point forecast also does not measure a portfolio's other maturity exposures, credit spreads, carry, financing or execution costs.

## Appendix F. Does next week's long-bond auction help?

An upcoming auction is information an investor can know before acting. The hypothesis is that investors may require a yield concession to absorb supply. That motivates a reduce-duration vote, but an auction can also remove uncertainty or attract demand. This weekly test covers both the period before the auction and the adjustment afterward. It cannot identify the mechanism or an auction-day trading return.

I specified four flags before calculating their results: a 20-year nominal auction, a 30-year nominal auction, a 30-year TIPS auction, or any of the three. Reopenings count by original maturity. A flag is one only when the auction falls after the signal Friday and no later than the following Friday, and its official announcement is dated strictly before the signal Friday. Same-Friday announcements are excluded. I use no auction result, bid-to-cover ratio, tail, eventual demand or backfilled tentative schedule. Each flag is tried as one add or reduce vote alongside the thirteen rules and as a 0/1 regression input. All eight combinations keep annual refits and unit positions.

The early ranking uses the highest calendar-week gross t-statistic, with at least 20 early event weeks. Settings and results are saved before the later comparisons. This is a disclosed November 2026 extension after seeing the earlier research, not an untouched holdout. The 20-year nominal bond returned in May 2020 and has no early events, so neither of its signs qualifies for early selection. The any-auction flag changes composition when that maturity enters; its early history tests 30-year nominal bonds and TIPS only.

***Table F1.*** *Auction weeks and mean following-week changes in the 10-year yield, in bp. Positive means yields rose. Early: 564 weeks, 2006-October 2016. Later: 521 weeks, November 2016-October 2026, except the 20-year row, which uses 337 weeks from May 15, 2020. Counts are event weeks, not independent economic shocks.*

| Auction flag | Early events | Early event mean | Later events | Later event mean | Later non-event mean |
|---|---|---|---|---|---|
| 20-year nominal | 0 | n/a | 78 | +3.17 | +0.71 |
| 30-year nominal | 98 | -1.76 | 120 | +1.82 | +0.24 |
| 30-year TIPS | 20 | +1.30 | 22 | +0.50 | +0.61 |
| Any of the three | 118 | -1.24 | 207 | +2.25 | -0.48 |

The conditional sign is not stable: nominal 30-year auction weeks have falling yields on average in the early sample and rising yields later. The TIPS comparison has only 20 early and 22 later event weeks, including just two in 2026. It cannot support a strong standalone claim.

***Table F2.*** *Every combined-model variant. Gross, net and difference are annualized yield bp. Net deducts the illustrative 0.25 bp per position unit, including entry and exit. Differences and their four-lag HAC t-statistics compare each variant with the annual reference model over the same dates. The asterisk marks the early-sample winner. These t-statistics do not adjust for selection.*

| Added flag | Vote | Window | Gross | Net | Difference | Paired t |
|---|---|---|---|---|---|---|
| 20-year nominal | Reduce | 2020-26 | +23.1 | +18.0 | +8.9 | 1.71 |
| 20-year nominal | Add | 2020-26 | +19.6 | +14.2 | +5.4 | 2.13 |
| 30-year nominal | Reduce | 2016-26 | +21.8 | +16.7 | -0.8 | -0.12 |
| 30-year nominal | Add | 2016-26 | +23.2 | +17.8 | +0.6 | 0.08 |
| 30-year TIPS | Reduce | 2016-26 | +23.8 | +18.7 | +1.2 | 0.74 |
| 30-year TIPS | Add | 2016-26 | +24.3 | +19.2 | +1.7 | 1.02 |
| Any of the three* | Reduce | 2016-26 | +33.5 | +28.4 | +11.0 | 1.81 |
| Any of the three | Add | 2016-26 | +32.5 | +27.1 | +10.0 | 1.90 |
| Annual reference | - | 2016-26 | +22.6 | +17.6 | 0.0 | - |
| Annual reference | - | 2020-26 | +14.2 | +8.9 | 0.0 | - |

The selected any-auction, reduce-vote variant improves the early gross score from 18.3 to 29.4 bp/year. Later it scores 33.5 gross and 28.4 after the assumed charge. Its paired later gain is 11.0 bp/year (t = 1.81). In the last 52 weeks it scores 69 bp, versus 58; over the 43 completed 2026 weeks, 72 versus 58. The constant higher-yield benchmark scores 31.4 bp/year over the full later period. Against that benchmark, the selected variant's paired gain is 2.1 bp/year (t = 0.07). The close benchmark and the additional search temper the result.

Both vote signs can improve a combined model because the new input also changes the regression's coefficients and its views on non-auction weeks. That is not evidence for two opposite economic explanations.

***Table F3.*** *Event-only positions, flat otherwise. Periods match Table F1. The reduce-duration direction follows the supply hypothesis; the last column discloses its opposite. These calendar-week statistics are descriptive and not search-adjusted. The 20-year model has no early test.*

| Flag | Early reduce bp/yr | Later reduce bp/yr | Later reduce net | Later reduce t | Later add bp/yr |
|---|---|---|---|---|---|
| 20-year nominal | n/a | +38.1 | +32.1 | 2.75 | -38.1 |
| 30-year nominal | -15.9 | +21.9 | +15.9 | 1.55 | -21.9 |
| 30-year TIPS | +2.4 | +1.1 | +0.0 | 0.34 | -1.1 |
| Any of the three | -13.5 | +46.5 | +39.3 | 2.78 | -46.5 |

Simply reducing duration in any-auction weeks has a larger later score than the combined model, but loses in the early sample. That reversal and the introduction of the 20-year bond prevent treating the later number as a stable edge. The accompanying files also report hit-rate ranges, eight-lag checks, turnover, all early trials and weekly positions.

**Source and limits.** I retrieved identifying and calendar fields from the Treasury auction database on November 7, 2026. An outcome-independent sample of 23 original Treasury announcements confirms their announcement dates, auction dates and CUSIPs. This checks a sample, not every announcement or historical calendar revision. The data retain that limitation, as well as the paper's market-history limitations. The 20-year restart is documented by the New York Fed. This annual any-auction variant is retained as a research comparison. Appendix G tests all additions against the subsequently selected quarterly baseline and identifies the adopted TIPS variant.

## Appendix G. Auction additions with quarterly refitting

After inspecting the annual auction results and choosing quarterly refitting, I repeated the same eight additions with 13-week refits for both the baseline and augmented models. The dates, flags, votes, early ranking and minimum event count stay unchanged. The second-round protocol was recorded before these calculations. This adds eight trials to the first eight; it is informed by earlier research and is not an independent confirmation. Standalone event-only scores are identical to Appendix F.

***Table G1.*** *All quarterly-refit auction variants. Differences use the quarterly baseline over matching dates. Gross, net and difference are annualized yield bp; net uses the same illustrative charge. Hit rate uses active weeks. The asterisk marks the early-sample winner. The 20-year rows have no early selection test. Paired t-statistics are descriptive and not search-adjusted.*

| Added flag | Vote | Window | Hit | Gross | Net | Difference | Paired t |
|---|---|---|---|---|---|---|---|
| 20-year nominal | Reduce | 2020-26 | 52.4% | +24.1 | +18.6 | -8.6 | -1.25 |
| 20-year nominal | Add | 2020-26 | 53.5% | +19.1 | +13.6 | -13.6 | -1.47 |
| 30-year nominal | Reduce | 2016-26 | 52.0% | +25.0 | +19.7 | -5.7 | -0.96 |
| 30-year nominal* | Add | 2016-26 | 52.8% | +26.3 | +20.9 | -4.3 | -0.72 |
| 30-year TIPS | Reduce | 2016-26 | 52.6% | +32.1 | +26.8 | +1.5 | 0.99 |
| 30-year TIPS | Add | 2016-26 | 52.9% | +32.6 | +27.2 | +2.0 | 1.27 |
| Any of the three | Reduce | 2016-26 | 51.7% | +25.5 | +20.3 | -5.2 | -0.75 |
| Any of the three | Add | 2016-26 | 53.1% | +23.6 | +18.2 | -7.1 | -0.84 |
| Quarterly baseline | - | 2016-26 | 52.6% | +30.6 | +25.4 | 0.0 | - |
| Quarterly baseline | - | 2020-26 | 53.9% | +32.7 | +27.2 | 0.0 | - |

The early winner adds a nominal 30-year auction flag with an add-duration vote. Its early score rises from 15.4 to 23.0 bp/year. Later it scores 26.3, against 30.6 for quarterly refitting alone: a paired difference of -4.3 bp/year (t = -0.72). Its hit rate is slightly higher, 52.8% versus 52.6%, illustrating why hit rate alone cannot determine usefulness.

The two TIPS variants score 32.1-32.6 bp/year, with paired t-statistics of 0.99-1.27. Neither is the early winner. After reviewing the later results, I adopt the add-vote variant in version 4: 32.6 bp/year, or 2.0 above the quarterly baseline (paired t = 1.27). There are only 22 later TIPS event weeks. The addition also changes fitted weights in non-event weeks, so its effect is not an auction-specific causal estimate.