Identifying Monetary Shocks: It's All in the (Orthogonalization) Timing

Francesco Amodeo*

Johns Hopkins University

November 2026

Abstract

High-frequency surprises around policy announcements are the workhorse instrument for monetary shocks, and yet they are predictable from information already public when the announcement lands. The standard fix is to purge that predictable component. Instruments for monetary policy shocks are then built in two steps: high-frequency surprises are purged of what public information already predicts, and aggregated to the frequency of analysis. A simple model shows that these steps do not commute: with several announcements in a period, aggregating first leaves predictable variation in the series, which also inflates instrumental relevance. In the orthogonalized series of Bauer and Swanson (2023a), purging at each announcement before aggregation collapses the first stage, leaving no detectable macroeconomic response. I trace their instrument's power to the leftover predictability in the 8% of months containing more than one announcement.

Keywords: Monetary Policy, Shocks, Surprises, News
JEL Codes: E31, E32, E43, E44, E52, G01

*famodeo2@jh.edu.

A previous version of this paper circulated with the name *A Corrected Measure of Orthogonalized Monetary Policy Surprises.* I am grateful to Edoardo Briganti, Wei Cui, Paula Donaldson, Domenico Giannone, James Hamilton, Nir Jaimovich, Daniel Lewis, Xinwei Ma, Valerie Ramey, Giacomo Rondina and Johannes Wieland for useful comments.

I. Introduction

High-frequency movements in interest rates around policy announcements are widely used as instruments for monetary policy shocks. Their appeal is simple: within a narrow window, asset price movements should reflect the policy announcement rather than other macroeconomic developments. Yet narrow timing does not by itself ensure exogeneity: these surprises are often predictable from information available before the announcement and may therefore combine unexpected policy variation with endogenous responses to economic conditions. Predictability poses a threat to instrumental validity, and it becomes an identification problem when the retained component is related to non-policy innovations that also move macroeconomic outcomes.

A seminal study by Ramey (2016) finds that popular instrument series for monetary shocks are serially correlated and predictable from Greenbook projections. Relatedly, Cieslak (2018) shows that the federal funds rate and recent employment growth predict target surprises ex post, with much of this predictability concentrated in unscheduled easings. Miranda-Agrippino and Ricco (2021) show that standard high-frequency instruments are predictable and autocorrelated and construct an informationally robust proxy by purging each FOMC surprise before aggregation. Bauer and Swanson (2023a) instead document that public macrofinancial variables predict FOMC surprises and interpret this relation primarily as systematic forecast errors about the Federal Reserve's response to public news. The mechanisms differ, but the common implication is that the construction of a monetary policy proxy must state which predictable components are removed, and *when* in the aggregation.

This paper studies that timing dimension. It shows that when high-frequency surprises are combined into a lower-frequency external instrument, orthogonalization and temporal aggregation do not commute. One can first aggregate the surprises and then purge the total using a single predictor vector (*sum-then-purge*), or purge each surprise using the included predictors measured immediately before it, and then aggregate the residuals (*purge-then-sum*). The two procedures agree only when purging the monthly total removes exactly what purging each announcement separately would have removed. When the predictors change between announcements, there is no reason for this condition to hold. Hence, I take the announcement as the unit at which the fitted predictable component is removed, and only then aggregate the resulting residuals at the desired frequency.

The distinction is most visible when the unit interval contains more than one announcement. A purge based only on predictors measured before the first (or any) single announcement does not remove the component of a later surprise that

is predictable. This direct channel is absent in single-announcement months.[1] Therefore, multi-announcement periods provide both the sharpest test of the timing mechanism, and the economically relevant cases in which monetary policy actions cluster.

For illustration, consider the case of October 2008. The Federal Open Market Committee (FOMC) cut rates on October 8, after an unscheduled conference call the day before, and announced another policy decision after its regularly scheduled October 28–29 meeting. Between those announcements, amidst the unfolding financial crisis, federal authorities announced a major financial-stabilization package; the Bureau of Labor Statistics (BLS) released September CPI, and financial indicators changed sharply; in the application, these events reach the purge through the financial predictors. The predictors observed before the second decision therefore differed materially from those observed before the first. Hence, a monthly purge based only on October 8 predictors removes one fitted value for the month and cannot use the predictors observed before the October 29 decision, leaving most of that decision's predictable part in the instrument. In my empirical application, I show that the sum-then-purge and purge-then-sum measures for October 2008 are 10.1 and 18.2 basis points, respectively — an absolute gap of 8.1 basis points, compared with a median absolute monthly surprise of 2.5 basis points, suggesting that the two procedures may produce sharply different results in multi-announcement months.

A simple model formalizes this logic. It decomposes the predictors observed before later announcements into a component spanned by first-announcement predictors and a new innovation. Sum-then-purge retains the predictability coefficient times that innovation, whereas purge-then-sum removes the fitted predictable component event by event. Under the maintained assumption that the event-level residuals are valid proxies for the latent monetary shock, the model derives the resulting contamination term in the IV estimand. The argument applies to proxy-SVARs, LP-IV, and other popular estimators. Sum-then-purge also removes one fitted value per month, so it keeps predictable parts in months with multiple announcements even when no news arrives. Separately, the model decomposes the observed proxy gap into the "carried-forward" prediction (the predicted value of the month's first announcement, carried forward to each later announcement), news arriving between announcements, and a remainder. Whether the variation removed by announcement-level orthogonalization also supplies first-stage strength (instrumental relevance) is an empirical question, which I address empirically next.

I apply these results to a leading empirical series, the orthogonalized monetary policy instrument of Bauer and Swanson (2023a). The authors show that FOMC-level surprises are predictable from pre-announcement macrofinancial data and

[1] Although different projection coefficients can still generate smaller gaps between the two procedures.

propose residualizing the proxy against those predictors to make it more likely to satisfy the IV exclusion condition. Their orthogonalized series is obtained by first summing all surprises within a month and then purging the sum using a set of predictors observed before the *first* announcement of the month ("sum-then-purge"). Holding the raw surprises, predictor set, monthly target frequency, sample, and empirical specifications fixed, I construct the alternative purge-then-sum measure implied by the model. Using the available replication data, I reproduce their event-level coefficient estimates and asset-price responses at the reported precision, independently confirming the predictability of raw high-frequency surprises. The two monthly proxies diverge only when event-level surprises are aggregated and residualized in a different order.[2]

As anticipated by the theory, the announcement-level construction makes the residual series orthogonal to the included predictors measured before each announcement. However, changing the order has large consequences for instrument relevance. In the baseline proxy-SVAR first stage targeting the two-year Treasury yield, the F-statistic is 9.65 for the raw surprise, 2.47 for the sum-then-purge series, and 0.29 for purge-then-sum. Qualitatively similar reductions obtain for the federal funds rate and the one-year Treasury yield. With first-stage statistics below 0.3, conventional proxy-SVAR estimates and confidence bands are not reliable for causal inference; expectedly, in the reported implementation, they yield no statistically detectable macroeconomic responses. Local-projections and internal-instrument exercises show the same (null) qualitative results, although they are close analogues rather than exact replications, because the public files exclude chair speeches.

I investigate the stark differences and show that the gap between the two series is indeed concentrated in multi-announcement months. In the baseline sample, only 31 of 383 months contain more than one monetary announcement (consistently with the schedule of eight meetings per year established since 1981). In fact, I document that these episodes are a non-representative sample of monetary policy conduct, as they are concentrated in two scenarios: (*i*) they are common before the communication regime shift in 1994, when policy changes were not conveyed in a single post-meeting statement and occurred in multiple announcement windows; (*ii*) they cluster in crises, such as the Global Financial Crisis and oil-shocks/credit-crunch episodes. I show that these observations are responsible for a disproportionate share of the strength of the sum-then-purge instrument, providing an intuitive rationale for the sharp contrast between my results and Bauer and Swanson (2023a).

Three results corroborate this finding. First, the two series are close in single-announcement months but diverge systematically in months with multiple announcements, with correlations dropping from 0.99 to 0.84, and the mean abso-

[2]Data available at the FRB of San Francisco's monetary policy surprises webpage.

lute gap rising from 0.60 to 3.20 basis points. Second, the model's decomposition shows that most of the gap is the predictable part of later announcements: news arriving between announcements and the carried-forward prediction. Third, later announcements carry a predictable part of their own, which purge-then-sum removes and sum-then-purge largely keeps, so the gap is predictable variation a purge is meant to remove.

These findings indicate that multi-announcement months are the origin both of the divergence between the two orthogonalized series and of the sum-then-purge instrument power. In fact, those same months also supply the interest rate response associated with most of the proxy's relevance. On impact, the response to an orthogonalized one basis point surprise is -0.21 in single-announcement months and 2.02 in multi-announcement months, compared with a pooled response of 0.40. Multi-announcement months receive a pooled-regression weight of 27% despite representing only 8.1% of the sample; because their estimated slope is much larger and the single-announcement estimate is negative, their weighted contribution exceeds the whole pooled impact response. The same observations therefore expose the timing problem and concentrate the interest rate response underlying the monthly proxy's strength.

Finally, the same margin operates in other surprise series. I impose the BS predictor set and the baseline proxy-SVAR first stage on five leading public surprise measures, and build both orderings for each: the sum-then-purge F statistic is 2.0 to 4.7 times its purge-then-sum counterpart in every one.

Taken together, the results suggest that when high-frequency monetary surprises are converted into monthly external instruments, orthogonalization timing is part of the identifying design. The distinction is most consequential in the very episodes where event clustering is economically salient, as events of macroeconomic and financial stress where the Federal Reserve intervenes more prominently and, often, multiple times in a month. In this application to the Bauer and Swanson (2023a) series, aligning the projection with each announcement removes any residue of predictability, but it also weakens the instrument far below any threshold for statistically informative macroeconomic inference.

Related Literature This paper contributes to the broad literature, recently surveyed by Hürtgen and Uhlig (2026), on the use of interest rate surprises as instruments for monetary policy shocks, and on their limitations. Seminal in this literature is Nakamura and Steinsson (2018), who argue that information effects contaminate high-frequency surprises in the identification of monetary disturbances. Jarociński and Karadi (2020) separate monetary from information shocks using the joint response of interest rates and equity prices, while Miranda-Agrippino and Ricco (2021) show that standard high-frequency instruments combine policy and signaling components and propose a cleaner proxy. Relatedly,

Bauer and Swanson (2023b) argue that much of the documented predictability in these surprises reflects the Fed's response to public news, and attribute little of it to information effects, and Cieslak and Schrimpf (2019) show more broadly that central-bank communication often conveys non-monetary news about growth and risk premia. Karnaukh and Vokata (2022) purge each meeting's surprise of prior revisions to professional growth forecasts. This paper complements this line of work by shifting attention from the decomposition of a single announcement to the time aggregation of multiple announcements at the frequency of analysis. A complementary view on aggregation comes from Lee and Sekhposyan (2024), which study how high-frequency surprises should be mapped into low-frequency VAR instruments, and from Jacobson et al. (2026), who show that the order of estimation and aggregation can produce adverse inflation responses. Here the aggregation rule is fixed; the margin is whether residualization occurs before or after aggregation.

The paper argues that orthogonalization and time aggregation are not commutative steps. In particular, when a month contains multiple announcements, aggregating surprises before orthogonalizing them leaves later-announcement predictability embedded in the instrument. In this sense, the paper is also related to Chen (2026), who offer an alternative explanation for the predictability of monetary surprises based on the Fed's response to financial conditions and a wait-and-see attitude toward recent data, and to Aruoba and Drechsel (2024), who identify monetary shocks by conditioning policy decisions on information extracted from Fed staff documents.

More broadly, the paper speaks to the literature using high-frequency identified instruments as proxies for shocks in VARs and local projections (Gertler and Karadi (2015), Stock and Watson (2018)). Lewis and Mertens (2026) study weak-instrument bias in the corresponding just-identified impulse-response estimators; close in spirit is also Ramey (2011), where incorrect timing causes standard VAR shocks to miss the relevant fiscal innovation. Here too, timing is not an implementation detail, but part of what determines whether the resulting series can be interpreted as a valid instrument for a structural shock. My results suggest that proxy construction cannot be separated from identification: changing the timing of orthogonalization can materially affect instrument validity and the resulting macroeconomic inference.

The remaining sections are structured as follows. Section II introduces the two alternative ways to construct the monthly instruments, anchoring the exposition to the illustrative case of October 2008. Section III develops the identification argument in the context of a simple model explicitly considering the timing of the orthogonalization, providing a formal derivation of the bias. Section IV shows that the bulk of the power of the *sum-then-purge* instrument comes from the multi-announcement episodes and revisits the key empirical results in Bauer and Swan-

son (2023a) (*BS* henceforth), comprehensively reassessing the impulse responses and discussing the lessons that emerge. Section V concludes. Throughout, I use Arabic numerals for BS sections, and Roman numerals for this paper.

II. Orthogonalization & Timing

When it comes to residualizing raw high-frequency monetary surprises against a set of known predictors ("news variables"), and aggregate the orthogonalized surprises into a desired lower frequency of analysis (say, a month), there are two alternative approaches:

(*i*) "*sum-then-purge*": sum the within-month surprises and *then* orthogonalize the total with respect to *a* chosen measurement of the news variables;

(*ii*) "*purge-then-sum*": orthogonalize each surprise with the respective up-to-date news variables, and *then* sum the orthogonalized surprises within months.

Below, I show this distinction to be highly consequential for the identification of the effects of monetary policy.

Consider a month m with two monetary events: an early announcement, preceded by news variables X_{1m}, and a later one, preceded by news variables X_{2m}. If one first sums the two surprises and then purges the monthly total using only X_{1m} (sum-then-purge), the component of the later announcement that is predictable from the newly arrived information in X_{2m} remains in the instrument. By contrast, if each announcement-level surprise is first orthogonalized with respect to its own pre-announcement information set and the residuals are summed afterwards (purge-then-sum), that predictable component is removed by construction. This is exactly the type of mechanism we observe in the data: to illustrate, consider next the case of October 2008.

II.I Illustrative case: October 2008

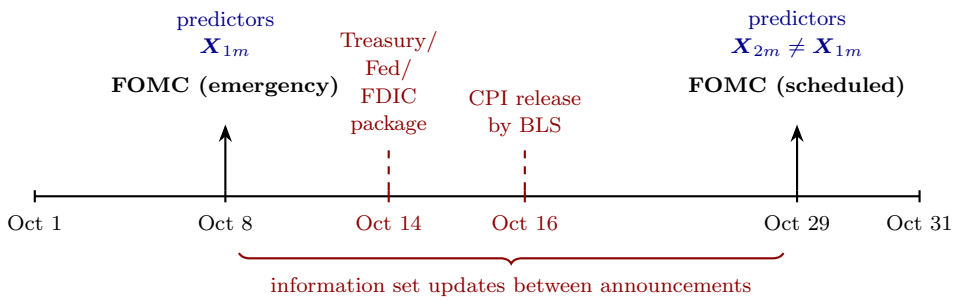


Figure I: October 2008, stylized timeline.

A useful real-world illustration comes from October 2008. The timeline is reported in Figure I below.

The Fed held an unscheduled conference call on October 7 following the developments on the unfolding financial crisis, cut rates on October 8, and reconvened at the scheduled meeting of October 28–29. On October 14, Treasury, the Federal Reserve, and the FDIC also announced a coordinated financial-stability package. It combined Treasury's TARP Capital Purchase Program (up to $250 billion of preferred-stock capital for financial institutions), the FDIC's Temporary Liquidity Guarantee Program (guarantees of newly issued senior debt and noninterest-bearing transaction accounts), and further details on the Fed's Commercial Paper Funding Facility. Between the two announcements, financial conditions also changed materially: the S&P 500 fell sharply, commodity prices collapsed, the yield curve steepened, and the Bureau of Labor Statistics released the September CPI figure on October 16. In other words, the information set observed before the second announcement differed markedly — in the application, through the financial predictors — from the information set observed before the first.

In such a month, an orthogonalization that uses only first-announcement predictors (or any single-announcement set) cannot remove the component of the later-announcement surprise that is predictable from the information released between the two meetings. The resulting monthly instrument retains that component by construction. In fact, on that occasion under study, the monthly instrument differs across sum-then-purge and purge-then-sum orthogonalization procedures (≈ 10 versus ≈ 18 basis points).[3]

Although a representative case for stress/emergency episodes, October 2008 is not a template for all multi-announcement months. For instance, July 1988 shows that the issue is not confined to crisis periods: the data contain one Fed announcement on July 1 and two later announcements during the month; between the first announcement and the later ones, the payroll surprise changed from -13.5 to $+121$ (thousand jobs); by the second announcement the option-implied Treasury skewness has risen from 0.16 to 0.24, the three-month commodity-price change has fallen from 9.6 to 7.4 log points, and the three-month S&P 500 growth from 5.0 to 4.1 log points. In that month the two measures even differ in sign (≈ 3.2 versus ≈ -3.2 basis points). Expectedly, appendix A shows that the largest gaps between the two ordering procedures all fall in months with several announcements.

II.II Multi-announcement months

As argued above, multi-announcement months are especially subject to the orthogonalization timing underpinning the two possible procedures described at the

[3]For scale, the median absolute monthly surprise is 2.5 basis points.

start of this section.[4] Table I lists all 31 multi-announcement months compactly, providing some historical context.

Table I: Multi-announcement months in the baseline sample

Month	Ann.	Unsch.	Historical context
Panel A. Post-1987-crash normalization & anti-inflation tightening			
1988:02	3	2	Post-crash easing ends at 6.5%, ratified at the February meeting.
1988:04	3	3	No meeting or policy action; the funds rate drifts above 6.75%.
1988:05	7	4	Two firmings to 7.25% as unemployment falls to 5.4%.
1988:06	2	2	Drought lifts commodity prices; the funds rate firms to 7.5%.
1988:07	3	2	No vote taken; the Desk holds funds near 7.75% amid dollar strength.
1988:08	3	2	Discount rate raised to 6.5%; the funds rate target passes 8%.
1988:11	3	2	Policy held with a tightening bias; payroll report on November 4.
1989:01	3	3	Second leg of the December firming lifts the funds rate to 9%.
1989:02	5	4	Discount rate raised to 7% on inflation concerns; funds reach 9.75%.
Panel B. Pivot from inflation fight to easing as growth slowed			
1989:07	4	3	Payroll gains moderate; two July cuts return funds to about 9%.
1989:10	3	2	Reserves added after the October 13 break, then a cut to 8.75%.
1989:11	4	3	A surprise early-November easing takes the funds rate to 8.5%.
Panel C. Recession, oil shock, credit crunch, and weak recovery			
1990:07	2	1	First move in seven months: funds rate to 8% as credit tightens.
1990:10	2	1	Late-October easing to 7.75% follows a large deficit-reduction accord.
1990:11	3	0	October payrolls fall again; the Committee cuts the funds rate to 7.5%.
1990:12	3	1	Two more cuts to 7%, and the first discount rate cut since 1986.
1991:02	3	2	Both rates cut half a point, to 6% and 6.25%, as activity declines.
1991:03	2	1	Easing continues to 6% in the month NBER later dated the trough.
1991:04	2	2	Discount rate to 5.5%, funds rate to 5.75%, on industrial weakness.
1991:08	2	1	Sluggish recovery and weak money bring the funds rate to 5.5%.
1991:09	2	2	Half-point discount rate cut to 5%; the funds rate follows to 5.25%.
1991:10	2	1	A further quarter point, to 5%, as recovery and confidence flag.
1991:11	2	0	Both rates cut again: discount rate to 4.5%, funds rate to 4.75%.
1991:12	4	3	First full-point discount rate cut since 1981, to 3.5%; funds rate 4%.
1992:04	2	1	Broad money contracts, demand softens, funds rate down to 3.75%.
Panel D. Post-1994 stress episodes			
2001:01	2	1	Two half-point cuts, one between meetings, take funds to 5.5%.
2007:08	3	2	Target held at 5.25%; discount rate cut to 5.75% as funding seizes.
2008:01	2	1	Emergency 75 basis point cut, then 50 more at month end, to 3%.
2008:03	2	1	Term securities lending facility opened; 75 basis point cut to 2.25%.
2008:10	2	1	FOMC joins a coordinated global cut, then eases again, to 1%.
2019:10	2	1	Repo turmoil restarts bill purchases; funds rate cut to 1.50–1.75%.

Notes: Multi-announcement months are months with at least two entries; there are 31 such months, 25 of them before February 4, 1994, when the FOMC did not publicly announce policy decisions; markets inferred them from New York Fed Open Market Desk operations and Board discount-rate announcements. After that, the Fed began announcing policy changes and subsequently expanded this into a statement after every FOMC meeting. The unscheduled count for 1989:10 follows BS's 2023 definitions.

These episodes do not constitute a representative sample of monetary policy con-

[4] Although the two instruments need not coincide exactly even in single-announcement months, as shown in Appendix A.I.

duct. They cluster in two scenarios. First, they are common before the communication regime shift in February 1994, when policy changes were not conveyed in a single post-meeting statement and could occur in multiple announcement windows within a month. Second, they occur in exceptional episodes such as the Global Financial Crisis and the 1990–92 turmoils including the recession/oil-shocks/credit-crunch occurrences.

Next, section III describes generally the timing mechanism, encapsulating formally the multi-announcement month structures. This delivers a sharp analytical decomposition of the gap between the two monthly instruments, aiding intuition and clarifying the identification concerns inherent in the sum-then-purge instrument.

III. Identification: Timing & Monetary Shocks

This section formalizes why the sum-then-purge orthogonalization protocol is problematic from an identification perspective. The point is as simple as it is consequential: the information set that is relevant for orthogonalization can change between the first and a later announcement of the same month. Even when it does not, a single monthly orthogonalization removes one fitted value per month whatever the number of announcements, whereas each announcement carries its own predictable part. A monthly orthogonalization that uses only the first-announcement information set cannot, by construction, remove the component of later-announcement surprises that is predictable from the information that arrives afterwards. That leftover predictable component is then carried into every downstream exercise in which the orthogonalized series is used as an instrument, contaminating the instrument and biasing estimated responses to the latent monetary policy shock, as the next sections show in the empirical application.

III.I Econometric Framing

Let months be indexed by $m = 1, \ldots, M$ and within-month announcements by $t = 1, \ldots, T_m$, where T_m denotes the number of monetary policy announcements in month m.[5] Denote the raw high-frequency surprise around announcement (t, m) by $\texttt{MPS}_{tm} \in \mathbb{R}$. Let $\boldsymbol{X}_{tm} \in \mathbb{R}^k$ be the vector of news variables observed just before that announcement, BS's $\boldsymbol{X}_{t-}$ (their eq. (14)) with the month index added.

The Central Bank sets i_{tm} according to a rule of the form $i_{tm} = \alpha_{tm}\, x_{tm} + \varepsilon_{tm}$, where x_{tm} is a scalar representing the state of the economy and ε_{tm} is the *true* monetary policy shock.[6] BS present a compelling case that monetary surprises

[5] For months with no monetary announcement, $T_m = 0$, so that $z_m = \tilde{z}_m = 0$ by convention.

[6] For simplicity of exposition, I adopt BS notation, but clarify that x_{tm} and $\boldsymbol{X}_{tm}$ are distinct objects: the former is a theoretical representation of a policy-relevant variable (e.g., the output gap), the latter represents the empirical k-by-1 news variables vector ("predictors").

might capture both the intended shock and *"information that is publicly available prior to the FOMC"* (p. 100). Enriching their previous result (Bauer and Swanson (2023b)) with robust econometric evidence of predictability (which I replicate coefficient by coefficient in Section IV.II), they recommend orthogonalization of monetary surprises.

Then, I specify

$$\texttt{MPS}_{tm} = \boldsymbol{\beta}' \boldsymbol{X}_{tm} + u_{tm}, \qquad \mathbb{E}[u_{tm} \mid \boldsymbol{X}_{1m}, \ldots, \boldsymbol{X}_{tm}] = 0, \tag{1}$$

where u_{tm} denotes the residual from the event-level predictability regression.[7] Let us see what the two orthogonalization protocols imply. The purge-then-sum uses:

$$\tilde{z}_m := \sum_{t=1}^{T_m} u_{tm}, \tag{2}$$

while sum-then-purge first aggregates and *then* orthogonalizes using only the first occurrence of the news variables within each month:

$$z_m := \sum_{t=1}^{T_m} \texttt{MPS}_{tm} - \boldsymbol{\gamma}' \boldsymbol{X}_{1m}, \tag{3}$$

estimating $\boldsymbol{\gamma}$ with OLS. Defining $\boldsymbol{\delta} := \boldsymbol{\beta} - \boldsymbol{\gamma}$ and using (2) yields

$$z_m - \tilde{z}_m \;=\; \underbrace{(T_m - 1)\,\boldsymbol{\beta}' \boldsymbol{X}_{1m}}_{\substack{\textit{carried-forward} \\ \text{prediction}}} \;+\; \underbrace{\boldsymbol{\beta}' \sum_{t=2}^{T_m} (\boldsymbol{X}_{tm} - \boldsymbol{X}_{1m})}_{\text{news}} \;+\; \underbrace{\boldsymbol{\delta}' \boldsymbol{X}_{1m}}_{\text{remainder}} \tag{4}$$

The gap between the two orderings has three parts. The first, which I call the *carried-forward* prediction, assigns the predicted value of the month's first announcement to each later announcement, as if the news variables had kept their values from before the first announcement; news adds the change in the news variables since then; and the remainder reflects that $\boldsymbol{\gamma}$ differs from $\boldsymbol{\beta}$.

Intuitively, the sum-then-purge approach constitutes an imperfect orthogonalization, as the aggregation of surprises within a month is projected onto the information set available before the first announcement, leaving subsequent predictable information untouched and potentially also contaminating the purging of the first announcement. The two approaches coincide exactly when the monthly fitted value equals the sum of the predictable parts of the month's announcements. A single monthly regression achieves this in every month when the first announcement's news variables span the later ones and every month has the same number of announcements.

[7] Note that equation (1) is the same event-level regression estimated by BS (Table 1, p. 105).

Reading the gap as predictable variation assumes that the news variables of later announcements do not respond to the month's earlier surprises. Financial news variables can respond to an earlier announcement in the same month. Section IV.I shows that their response carries a very small part of the gap.[8]

Illustration: Two-Announcement Month Without loss of generality, I restrict the exposition to a single news variable ($k = 1$); the multivariate extension is straightforward. Consider the simple case of a month m with two announcements, matching closely to the case-study of October 2008 considered earlier:

$$\texttt{MPS}_{1m} = \beta\, X_{1m} + u_{1m}, \qquad \texttt{MPS}_{2m} = \beta\, X_{2m} + u_{2m} \tag{5}$$

Purge-then-sum delivers

$$\tilde{z}_m = (\texttt{MPS}_{1m} - \beta\, X_{1m}) + (\texttt{MPS}_{2m} - \beta\, X_{2m}) = u_{1m} + u_{2m} \tag{6}$$

Sum-then-purge delivers

$$\begin{aligned} z_m &= \texttt{MPS}_m - \gamma\, X_{1m} \\ \text{with} \quad \texttt{MPS}_m &= \beta\, X_{1m} + \beta\, X_{2m} + (u_{1m} + u_{2m}), \end{aligned} \tag{7}$$

where OLS estimates

$$\gamma = \frac{\mathrm{Cov}(\texttt{MPS}_m, X_{1m})}{\mathrm{Var}(X_{1m})} = \beta + \beta\, \frac{\mathrm{Cov}(X_{2m}, X_{1m})}{\mathrm{Var}(X_{1m})} \equiv \beta + \beta\rho.$$

Both series are orthogonal to the first announcement's news variable, z_m by construction of the monthly regression and $\tilde{z}_m$ because each u_{tm} is. If later news variables do not respond to the month's earlier surprises, only $\tilde{z}_m$ is also orthogonal to the second announcement's. Plugging γ in z_m yields

$$\begin{aligned} z_m &= [\beta\, X_{1m} + \beta\, X_{2m} + (u_{1m} + u_{2m})] - (\beta + \beta\, \rho)\, X_{1m} \\ &= \tilde{z}_m + \beta\, [X_{2m} - \rho\, X_{1m}] \end{aligned} \tag{8}$$

so z_m retains a predictable component unless $\beta = 0$ (i.e., no predictability to start with) or $X_{2m} \equiv \rho\, X_{1m}$ (i.e., the second-announcement information is fully spanned by first-announcement), both of which the data reject.[9]

In the illustration, with two announcements in every month, γ absorbs the carried-forward prediction and equation (8) keeps only the news that X_{1m} does not predict. In BS's sample the number of announcements varies, so no single γ absorbs the carried-forward prediction, which stays in the gap as predictable variation (Section IV.I). Appendix F.IV derives the gap for any number of announcements.

[8] Such a response leaves purge-then-sum a sum of purged surprises and adds a multiple of the earlier purged surprise to the gap (Appendix Table B.2, Panel A).

[9] See Section IV.II for the rejection of $\boldsymbol{\beta} = 0$; Section IV.I for the rejection of $X_{2m} \equiv \rho\, X_{1m}$; Appendix F gives the full derivation.

III.II Implications for Impulse Responses & Inference

To connect the instrument decomposition to the estimation of impulse responses, consider the reduced-form obtained by projecting the macro outcome Y_{m+h} onto the structural monetary shock ε_m and the pre-announcement news variables, after conditioning on VAR dynamics:[10]

$$\begin{aligned} Y_{m+h} &= \Theta_h\, \varepsilon_m + \Psi_{1h}\, X_{1m} + \Psi_{2h}\, X_{2m} + \xi_{m+h}, \\ v_m &= \varepsilon_m + \Psi_1^v\, X_{1m} + \Psi_2^v\, X_{2m} + \xi_m^v \end{aligned} \tag{9}$$

where $h \in [0, H]$ is the impulse response horizon, Θ_h is the structural response of Y to a unit monetary policy shock — the object targeted by BS' proxy-SVAR, as we will see — and $\varepsilon_m := \sum_{t=1}^{T_m} u_{tm}$ is the sum of the event-level unpredictable components; ξ_{m+h} and ξ_m^v are uncorrelated with ε_m, X_{1m} and X_{2m}. Both news variables may influence the outcome. The second equation is the same projection for v_m, the residual of the two-year yield that BS use as policy indicator, with its response to the shock normalized to one, so that Θ_h is the response to a shock that raises the yield by one unit. Because the contamination enters through the instrument, the following result applies to the proxy-SVAR, to local projections, and to any IV estimator that uses z_m or $\tilde{z}_m$ as an instrument for ε_m. The probability limits of the two IV estimators are:

$$\widehat{\Theta}_h^{(\tilde{z})} = \Theta_h, \qquad \widehat{\Theta}_h^{(z)} = \Theta_h + \beta\,(\Psi_{2h} - \Psi_2^v\,\Theta_h)\,\frac{\mathrm{Var}(X_{2m} - \rho\, X_{1m})}{\mathrm{Cov}(v_m, z_m)}. \tag{10}$$

Equation (10) says that the sum-then-purge IV estimator recovers the true impulse response Θ_h plus a contamination term.[11] The bias originates from the extra term $\beta\,(X_{2m} - \rho\, X_{1m})$ in z_m, i.e., the leftover news that is not spanned by the first announcement's variables. Because z_m retains this predictable component, it transmits the effect of later-announcement news on the outcome into the estimated impulse response, and the effect of the same news on the yield into the first stage:

$$\begin{aligned} &\mathrm{Cov}(v_m, \tilde{z}_m) = \mathrm{Var}(\varepsilon_m), \\ &\mathrm{Cov}(v_m, z_m) = \mathrm{Var}(\varepsilon_m) + \Psi_2^v\,\beta\ \mathrm{Var}(X_{2m} - \rho\, X_{1m}). \end{aligned} \tag{11}$$

When good economic news predicts tighter policy ($\beta > 0$), raises the yield ($\Psi_2^v > 0$) and is associated with stronger output ($\Psi_{2h} > 0$), the bias term is positive and attenuates the estimated contractionary response ($\Theta_h < 0$), consistent with the discussion in BS (p. 121). The magnitude depends on how much "new information" arrives between announcements relative to the covariance between z_m and

[10] In the proxy-SVAR of Bauer and Swanson (2023a), Y_{m+h} is governed by the structural moving-average representation $Y_{m+h} = \sum_{j=0}^{h} C_j\, \varepsilon_{m+h-j} + f(Y_{m-1}, \dots)$, where $C_j = \Phi^j S$. The structural impulse response to a monetary shock at horizon h is the first column of C_h, which is Θ_h below. Because the news variables X may forecast Y_{m+h} — whether through non-monetary shocks or reduced-form predictability — they enter the projection with coefficients Ψ_{1h}, Ψ_{2h}.

[11] Full derivation in Appendix F.III.

v_m: when later-announcement news is independent of first-announcement news ($\mathrm{Var}(X_{2m} - \rho\, X_{1m})$ is sizable relative to $\mathrm{Cov}(v_m, z_m)$), the contamination is large; in the knife-edge case when X_{2m} contains no new information beyond X_{1m}, the bias vanishes.

The news raises the covariance of z_m with v_m by $\Psi_2^v\, \beta\, \mathrm{Var}(X_{2m} - \rho\, X_{1m})$ (equation (11)) and the variance of z_m by $\beta^2\, \mathrm{Var}(X_{2m} - \rho\, X_{1m})$ (equation (8)). If news left the yield unchanged ($\Psi_2^v = 0$), only the variance would rise, making z_m the weaker instrument. When news moves the yield enough, as in the data, the covariance gain dominates and z_m becomes the stronger instrument.[12] Both instruments have the same covariance with the shock, $\mathrm{Cov}(\varepsilon_m, z_m) = \mathrm{Cov}(\varepsilon_m, \tilde{z}_m)$, so all of the extra strength of z_m is covariance with news. The news term thus inflates the first-stage covariance and attenuates the response. With $\Psi_2^v \beta > 0$ and $\omega := \mathrm{Cov}(v_m, \tilde{z}_m)/\mathrm{Cov}(v_m, z_m)$, equation (10) reads

$$\widehat{\Theta}_h^{(z)} = \omega\, \Theta_h + (1 - \omega)\, \frac{\Psi_{2h}}{\Psi_2^v}, \tag{12}$$

an average of the true response and the response to news, weighted by each component's share of $\mathrm{Cov}(v_m, z_m)$.[13]

Equation (12) says that sum-then-purge estimates a blend of two responses, the target Θ_h, which $\tilde{z}_m$ recovers, and the response to the news that the monthly purge leaves in, Ψ_{2h}/Ψ_2^v. For instance, two years after a shock that raises the yield by 25 basis points, purge-then-sum puts consumer prices 3.0 percent lower and the gap $z_m - \tilde{z}_m$ alone puts them 0.3 percent higher. The gap supplies two thirds of the covariance of z_m with v_m, so $\omega \approx 1/3$ and equation (12) gives $(1/3)(-3.0) + (2/3)(0.3) \approx -0.8$ percent, the sum-then-purge estimate.

In contrast, the fully purged instrument $\tilde{z}_m$ identifies Θ_h, because removing predictability at the announcement level eliminates the contaminant at the source: no residual predictability is carried over across announcements. This provides a simple explanation for the empirical results next in Section IV: once $\tilde{z}_m$ is used, the bias term vanishes together with the first-stage covariance that news supplied, and what remains identifies Θ_h with a much weaker first stage.

IV. Empirical Consequences

The present section presents an empirical application of the lessons emerged in the model using Bauer and Swanson (2023a) as a case study of a paper adopting a sum-then-purge approach, holding everything but the timing of the orthogonalization

[12] Exactly, when $\beta > 0$ and $\mathrm{Cov}(v_m, z_m) > 0$, the correlation of z_m with v_m exceeds that of $\tilde{z}_m$ if and only if $\Psi_2^v > \beta/(1 + \sqrt{1+q})$, with $q := \beta^2\, \mathrm{Var}(X_{2m} - \rho X_{1m})/\mathrm{Var}(\varepsilon_m)$ (Appendix F.III).

[13] In the illustration the extra strength comes from news. With a varying number of announcements the gap also contains the carried-forward prediction of equation (4), which supplies first-stage strength only if the yield moves with it (Appendix F.IV).

fixed and tracing its consequences on instrumental strength and macroeconomic inference.

IV.I Instrument Strength

BS estimate surprises around FOMC announcements using a high-frequency event-study approach (Kuttner (2001), Cook and Hahn (1989)). The financial assets studied in the 30-minute window are one- to four-quarter-ahead Eurodollar futures (ED1–ED4) (Gurkaynak et al. (2005), Nakamura and Steinsson (2018)). Then, they calculate intradaily interest rate changes in ED1-ED4 in a 30-minute window around the FOMC announcement, extracting the first principal component of these changes and interpreting it as monetary policy surprises.

BS argue that incomplete information about the Fed's reaction function can give rise to correlations between monetary policy surprises and economic news. When using these surprises as instruments for policy shocks, such predictability constitutes a confounder for the causal analysis of monetary effects. Thus, they rightly emphasize the importance of orthogonalizing the surprises. For instance, on page 102: *"[...] we recommend orthogonalizing mps_t with respect to the macroeconomic and financial variables that are observed before the FOMC announcement [...]"*. However, the aggregation matter discussed in the previous sections emerges when implementing their procedure, since multiple FOMC announcements can occur within the same month. BS adopt the sum-then-purge approach, and they choose as orthogonalizing information set the value of the news variables available before *the first* of the FOMC announcements within each month (p. 121, endnote 45).[14]

I measure instrument strength by the first stage of BS's proxy SVAR on FOMC announcements (their Figure 3), the regression behind the responses of Section IV.II. As BS explain (p. 122), the proxy SVAR has two steps, estimated on different samples. First, as in BS, a reduced-form VAR in the two-year Treasury yield at the end of the month, the logs of the CPI and of industrial production, and the excess bond premium, with 12 lags and a constant, is estimated on all months from 1973:01 to 2020:02.[15] Let v_m be the residual of its two-year-yield equation. Second, the first stage regresses v_m on the instrument over the 383 months that the instrument covers.[16] For sum-then-purge it is

$$v_m = \alpha + \pi\, z_m + u_m, \qquad m = 1988{:}02, \ldots, 2019{:}12, \tag{13}$$

[14] Summed by month, BS' orthogonalized surprises give indeed the purge-then-sum series.

[15] From the zero-coupon yield curve of Gürkaynak et al. (2007). BS use end-of-month values because they correspond naturally to the surprises (their endnote 37). With the month-average of the same yields, the first stage is 8.20, 1.41 and 0.03 for raw, sum-then-purge and purge-then-sum.

[16] The instruments are BS's posted monthly series and, for purge-then-sum, the monthly sum of their announcement-level series. As in BS's monthly file, they are zero in the 117 months of the window without a recorded surprise, and these months enter neither purge. BS record the six news variables before each announcement, so the later-announcement predictors are BS's own.

and the first stages of purge-then-sum and of the raw surprise replace z_m with $\tilde{z}_m$ and with the raw surprise. The F-statistic is the squared t-statistic of $\hat{\pi}$, with homoskedastic standard errors.[17]

With the two-year yield, purge-then-sum keeps 12% of the first stage: the F-statistic falls from 2.47 to 0.29 (Table II). The sum-then-purge instrument is weak to begin with, and most of the strength it has comes from the gap, the part of the instrument that the announcement-level regression classifies as predictable and sum-then-purge leaves in. The gap carries two thirds of the instrument's covariance with the yield, and the covariance differs statistically from zero (Section IV.I).

Table II: First stage of the three instruments

	Raw `MPS`	Sum-then-purge z_m	Purge-then-sum $\tilde{z}_m$
Two-year yield	9.65	2.47	0.29
One-year yield	17.76	6.34	2.25
Federal funds rate	13.80	6.35	3.34

Notes: F-statistics of the proxy-SVAR first stage, 1988:02–2019:12 (383 months), homoskedastic standard errors. The second and third rows replace the two-year yield in the VAR with the end-of-month one-year yield of the Gürkaynak et al. (2007) curve, as in BS's Figure A1, and with the monthly average of the effective federal funds rate, spliced with the shadow rate of Wu and Xia (2016) at the zero lower bound. Table A.3 adds the 2SLS first stage of BS's endnote 52 and further specifications.

Table II raises a simple question: where does the monthly BS proxy obtain its predictive content for interest rates? And more specifically: what do the multi-announcement months contribute to the strength of their instrument?

To answer, I decompose the monthly orthogonalized BS surprise into the component realized in single-announcement months and the component realized in multi-announcement months, and I re-estimate the regressions allowing the two components to enter separately. The exercise is informative in two ways. First, the unrestricted coefficients compare the response to a one-basis-point surprise across the two types of months. Second, because the pooled sum-then-purge instrument is the sum of the two components, the same decomposition shows how much each group contributes to the pooled coefficient, establishing a direct analytical relationship between the restricted coefficient and the unrestricted coefficients estimated through the partitioned regression.

I show that the restricted pooled coefficient is a weighted average of the single- and the multi-event shock proxy. Crucially, the latter component disproportion-

[17] With the two-year yield, robust standard errors give smaller statistics, 7.71, 2.03 and 0.27. BS estimate the second step by 2SLS (their eq. 22) and report "the first-stage F-statistic in the two-stage least squares regression" (p. 126), which adds the VAR's lags as controls and uses robust standard errors (endnote 52). This 2SLS first stage gives 8.19, 1.98 and 0.25 (Appendix A.II).

ally drives the bulk of the instrumental power of the restricted specification. This finding is highly consequential for the interpretation of the monetary instrument. It establishes that the relevance of the instrument is concentrated in a small minority of non-representative observations – the multi-announcement months –, casting doubt on the robustness of the instrument's validity; relatedly, it explains why the monthly orthogonalized proxy retains any relevance at all, since what survives a monthly purge is concentrated in the months where the interest rate response is largest. Moreover, it emphasizes the salience of the purge-then-sum timing protocol, as the strength of the instrument completely disappears when the orthogonalization correctly accounts for later-announcement information.

Let z_m denote the monthly orthogonalized BS surprise, the sample version of equation (3). Define the partition

$$z_m^S = z_m \cdot \mathbf{1}\{T_m = 1\}, \tag{14}$$

$$z_m^M = z_m \cdot \mathbf{1}\{T_m \geq 2\}, \tag{15}$$

where S and M denote *single-* and *multi-* announcement surprises, respectively. By construction, $z_m = z_m^S + z_m^M$ and $z_m^S \cdot z_m^M = 0$ (non-overlapping support). As in BS's monthly file, $z_m = z_m^S = z_m^M = 0$ in the months without a recorded surprise, and these months stay in the regressions below.

For each horizon h, I estimate the local projection of the interest rate y_{m+h} on the sum-then-purge instrument z_m, effectively measuring its first-stage relevance:

$$y_{m+h} = \alpha_h + \pi_{R,h}\, z_m + \text{CONTROLS}_\text{m} + u_{m+h}, \tag{16}$$

the same regression with $\tilde{z}_m$ in place of z_m for purge-then-sum, and then the partitioned specification for sum-then-purge

$$y_{m+h} = \alpha_h + \pi_{1,h}\, z_m^S + \pi_{2,h}\, z_m^M + \text{CONTROLS}_\text{m} + u_{m+h}. \tag{17}$$

The same controls appear in all three regressions. By Frisch–Waugh–Lovell, after partialling out the common controls and the constant, the pooled coefficient is an exact affine combination of the two unrestricted slopes,

$$\hat{\pi}_{R,h} = \hat{\kappa}_{1,h}\hat{\pi}_{1,h} + \hat{\kappa}_{2,h}\hat{\pi}_{2,h}, \qquad \hat{\kappa}_{1,h} + \hat{\kappa}_{2,h} = 1, \tag{18}$$

where the sample weights $\hat{\kappa}_{1,h}$ and $\hat{\kappa}_{2,h}$ depend on the residualized variation of z_m^S and z_m^M. This identity separates the two above-mentioned notions cleanly: first, the pair $(\hat{\pi}_{1,h}, \hat{\pi}_{2,h})$ asks how much the outcome moves per basis-point surprise in single- and multi-announcement months; second, the products $\hat{\kappa}_{1,h}\hat{\pi}_{1,h}$ and $\hat{\kappa}_{2,h}\hat{\pi}_{2,h}$ ask how much each group contributes to the pooled BS response.[18]

I implement equation (17) over the months covered by BS's published surprises, February 1988 through December 2019. Of the 383 months in this window, 235

[18]The full proof is detailed in Appendix F.I.

are single-announcement, 31 are multi-announcement, and 117 contain no FOMC-related announcement with a recorded surprise.[19] Controls, lag length, and confidence bands closely follow the local projection specification in BS (Table G.1); the main outcome of interest is the two-year Treasury yield.

Table III and Figure II report the results. At horizon $h = 0$ for the two-year Treasury yield, the pooled slope is $\hat{\pi}_R = 0.40\,(0.31)$, against $0.07\,(0.30)$ for purge-then-sum (first stages in Table A.3), the single-announcement slope is $\hat{\pi}_1 = -0.21\,(0.31)$, and the multi-announcement slope is $\hat{\pi}_2 = 2.02\,(0.51)$; the null $\pi_1 = \pi_2$ is strongly rejected. Across horizons, $|\hat{\pi}_2| > |\hat{\pi}_1|$ at 47 of 49 horizons, and the same pattern appears for the federal funds rate and the one-year Treasury yield. In summary, single-announcement months contribute essentially none of the interest rate response that gives the monthly BS proxy its instrumental relevance.

Table III: LP decomposition by event-month type — 2-year Treasury yield

h	$\hat{\pi}_R$	$\hat{\pi}_1$	$\hat{\pi}_2$	$p_{H_0:\pi_1=\pi_2}$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	N
0	0.40 (0.31)	−0.21 (0.31)	2.02 (0.51)	0.000	0.727	0.273	383
3	−0.03 (0.71)	−1.17 (0.76)	2.99 (1.24)	0.005	0.727	0.273	382
6	1.06 (0.69)	−0.03 (0.78)	4.02 (1.33)	0.013	0.730	0.270	379
12	0.38 (0.73)	−0.88 (0.92)	3.76 (1.05)	0.001	0.728	0.272	373
18	1.20 (0.67)	−0.37 (0.78)	5.38 (1.41)	0.001	0.727	0.273	367
24	0.67 (0.67)	−0.30 (0.89)	3.27 (1.22)	0.054	0.727	0.273	361
36	−0.69 (0.82)	0.39 (0.88)	−3.54 (1.46)	0.028	0.724	0.276	349
48	−1.52 (0.98)	−0.04 (1.00)	−5.35 (2.67)	0.095	0.722	0.278	337

Notes: Coefficients from the unrestricted local projections with controls as in the public-data local-projection specification reported in Appendix E. Newey–West standard errors with h lags in parentheses. p_{H_0} is the Wald test of $\pi_1 = \pi_2$. Sample: 1988:02–2019:12.

The weighting arithmetic makes the same point from the pooled-regression perspective. The estimates satisfy $\hat{\pi}_{R,h} = \hat{\kappa}_{1,h}\hat{\pi}_{1,h} + \hat{\kappa}_{2,h}\hat{\pi}_{2,h}$ horizon by horizon, with $(\hat{\kappa}_1, \hat{\kappa}_2) \approx (0.73, 0.27)$ throughout the grid. Least squares weights each group by how much of the instrument's variation survives the controls, so a month type counts for as much as it moves the regressor. Multi-announcement months therefore account for about 27% of the residualized variation in the pooled proxy despite representing only about 8% of the sample. Because $\hat{\pi}_1$ is near zero, the pooled BS response is in practice close to a rescaled version of the multi-announcement slope.[20]

[19] Multi-announcement months are 8.1% of all months and 11.7% of event months in the baseline window. Of the 31 multi-announcement months, 15 contain two announcements, 11 contain three, 3 contain four, 1 contains five, and 1 contains seven.

[20] At $h = 0$, $0.727 \times (-0.21) + 0.273 \times 2.02 \approx 0.40$, reproducing the pooled coefficient exactly.

Figure II shows the full impulse responses for all conventionally used interest rates. The visual pattern is unambiguous: the pooled BS response tracks the multi-announcement series closely, while the single-announcement series remains near zero at essentially all horizons. The decomposition therefore identifies a small but economically distinctive set of observations that is decisive for the instrument's strength.[21]

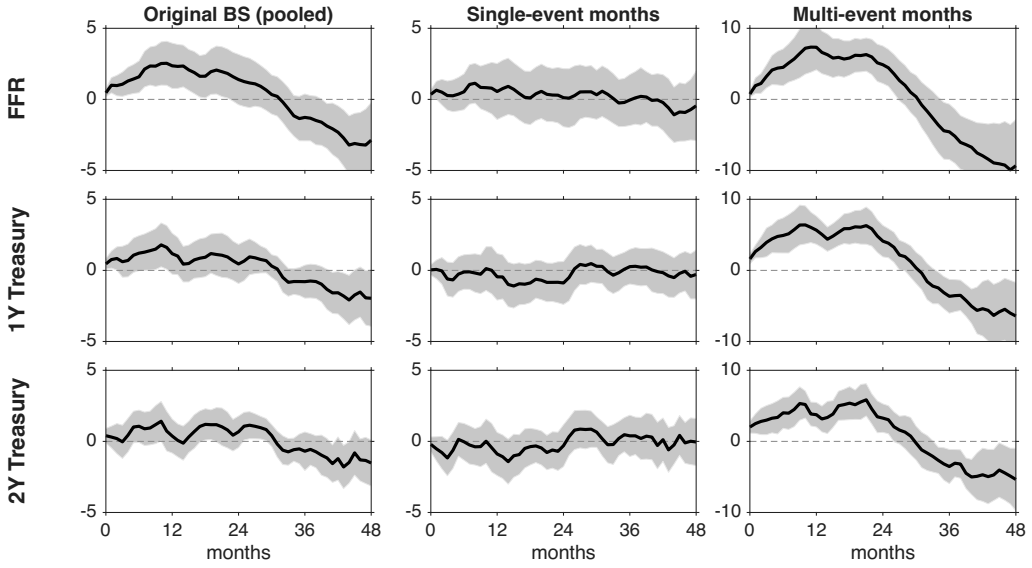


Figure II: Local projection responses to a unit monetary surprise, by event-month type. Rows: federal funds rate, spliced with the shadow rate of Wu and Xia (2016) at the zero lower bound (top), 1-year Treasury yield, FRED's monthly average (middle), 2-year Treasury yield (bottom). Columns: pooled BS response $\hat{\pi}_{R,h}$ from the pooled regression on z_m (left), single-announcement months $\hat{\pi}_{1,h}$ (center), multi-announcement months $\hat{\pi}_{2,h}$ (right). Sample: BS's published series, 1988:02–2019:12. 90% Newey–West bands with h lags.

In the baseline sample, there are only 31 months with more than one announcement (Table I). The analysis above has shown that these are exactly the months in which the sum-then-purge monthly proxy is most exposed to the timing problem, because the monthly purge removes one fitted value per month whatever the number of announcements, and later-announcement information adds to what it leaves in. Consistent with that mechanism, the sum-then-purge and purge-then-sum instruments are nearly identical in single announcement months but diverge sharply in multi-announcement months.[22]

The unrestricted regression in (17) shows that multi-announcement months are

[21] Appendix A.I tests whether irregular announcement timing gives months with several announcements a different event-level purge. The data do not reject common coefficients ($p = 0.54$), and weighting each month equally leaves purge-then-sum nearly unchanged (correlation 0.998).

[22] The correlation falls from 0.988 to 0.843 (Table A.1, Figure A.1) and net of BS's monthly regressors the carried-forward prediction accounts for 55% of the gap's covariance with the two-year yield residual (Section IV.I).

not just the most salient episodes where the timing issue occurs: they are the observations where the power of the BS instrument *originates*. These are two views of the same fact: the same monetary episodes generate the divergence between the two instruments and account for most of the BS interest rate response. Once the orthogonalization is moved to the announcement level, that source of predictive content largely disappears, which is precisely the loss of instrumental strength formalized in Section III, equation (11).

First-stage diagnostics. The two orderings differ by one series, the gap $g_m = z_m - \tilde{z}_m$, so their first stages differ by one covariance. In fact, because $z_m = \tilde{z}_m + g_m$,

$$\mathrm{Cov}(v_m, z_m) = \mathrm{Cov}(v_m, \tilde{z}_m) + \mathrm{Cov}(v_m, g_m). \tag{19}$$

The gap is what equation (1), BS's own regression at the announcement level, classifies as predictable beyond what their monthly purge removes. It has a tenth of the variance of z_m and carries two thirds of the covariance of z_m with v_m. With one instrument, the first-stage F-statistic depends on the instrument only through the square of its covariance with v_m divided by its variance.[23] As z_m and $\tilde{z}_m$ have almost the same variance and the statistic squares the covariance, the third of the covariance that survives purge-then-sum keeps 12% of the first stage, which falls from 2.47 to 0.29, even if the two series correlate at 0.95. In words, the announcement-level purge removes a predictable part with a tenth of the variance of the BS instrument, and that part carries most of the instrument's comovement with the yield residual.

Regressing v_m on the gap tests whether the two instruments covary equally with v_m, because $\mathrm{Cov}(v_m, z_m) = \mathrm{Cov}(v_m, \tilde{z}_m)$ exactly when $\mathrm{Cov}(v_m, g_m) = 0$. The gap's slope has a t-statistic of 3.39 (4.42 with robust standard errors, Table A.4). Both instruments are residuals of estimated regressions, so I also resample months with replacement, re-estimate both purges in every draw and hold v_m fixed. The gap's covariance with v_m is at or below zero in 0.9% of 4,999 draws, a two-sided p-value of 0.02. Sum-then-purge therefore covaries with v_m more than purge-then-sum does, by more than sampling error allows.

[23] It equals $(n-2)\, r^2/(1-r^2)$, where r is the correlation between the instrument and v_m and n the number of months.

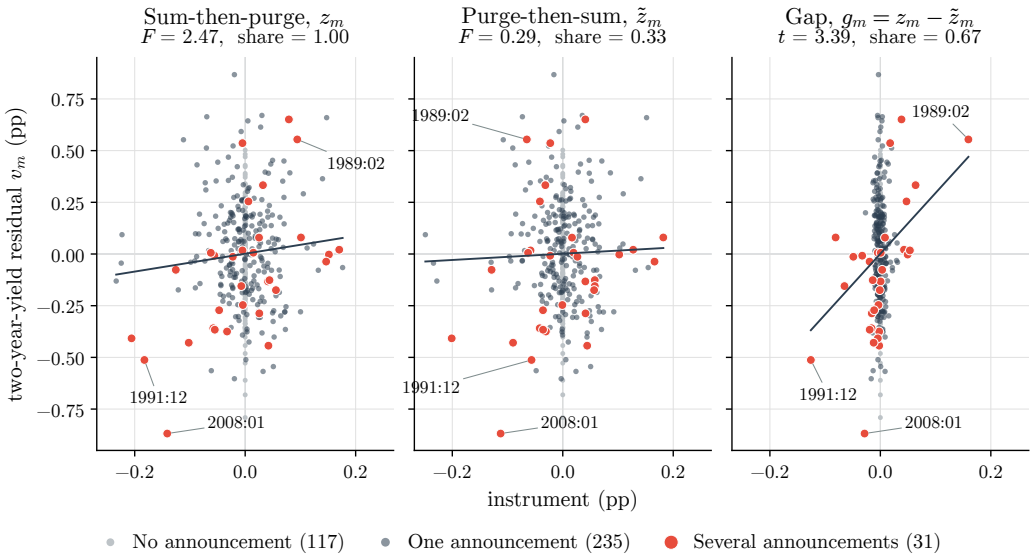


Figure III: Two-year-yield residual against each instrument and the gap, on a common horizontal scale. v_m is the residual of the two-year-yield equation of the VAR (12 lags and a constant, 1973:01–2020:02); the instruments run over 1988:02–2019:12 (383 months) and are zero in the 117 months without a recorded surprise. Red dots mark the 31 months with several announcements, and labels mark the three months named in the text. Above each panel: F, the proxy-SVAR first stage; t, the test of $\text{Cov}(v_m, z_m) = \text{Cov}(v_m, \tilde{z}_m)$; share, the series' covariance with v_m relative to that of z_m.

Figure III shows where the two orderings differ. In the 31 months with several announcements (red dots), sum-then-purge covaries with v_m more than twice as much as purge-then-sum, whereas in the single-announcement months both covary negatively. February 1989 and December 1991, with five and four announcements (most of them unscheduled), supply 59% of the gap's covariance. January 2008, the largest contributor to the covariance of z_m (32%), moves both orderings alike. Both first stages are concentrated in these three months, chosen here by their contributions, so without them neither ordering has a first stage left and the test can no longer tell the two apart ($p = 0.24$). Appendix A.III reports leave-one-out and cumulative deletions for both series. Put simply, the two orderings differ mainly in the months in which several announcements share one monthly purge, the same months that carry BS's response in Table III.[24]

What the gap is made of. The carried-forward prediction accounts for most of the gap's covariance with v_m. Split month by month with equation (4), net of BS's monthly regressors, it accounts for 55% and news for 28% (Table A.2). February 1989 and December 1991 are the two months that combine the largest carried-forward predictions of the sample with large two-year-yield residuals. In

[24]Holding the months fixed isolates the ordering: on BS's series, the first stage of sum-then-purge is 3.72 before February 1994 and 12.23 in the 31 months, against 0.07 and 1.97 for purge-then-sum. Dropping the unscheduled announcements leaves four months with multiple announcements, so the two orderings nearly coincide (Appendix A.IV).

the other months news carries most of the gap's covariance and the carried-forward prediction nets out, because sum-then-purge removes about 1.3 times the first announcement's predictable part in every month, too much for months with one announcement and too little for months with several (Appendix A.I).

Later announcements carry a predictable part of their own. Fit on the first announcement of each month, equation (1) predicts the surprises of the 56 later announcements from their own news variables, and the surprises move 0.84 basis points per basis point of prediction, close to the 1 of a full predictable part and far from zero.[25]

As expected, the response of later news variables to the month's earlier surprises explains at most 2.5% of the gap's covariance with v_m, and the financial variables, the only news that can respond within the month, carry 3.4% of it (Appendix B). The response cannot reach the carried-forward prediction, whose news variables are dated before the month's first announcement, and whether a second announcement follows shows no detectable response to the month's first purged surprise. The same response explains the negative stock-market coefficients of Table B.1, as both orderings contain the earlier surprise, while the gap keeps the positive sign of the announcement-level regression. For the same reason, the six later news variables of Table B.1 predict the month's first purged surprise u_{1m} at least as strongly as either ordering, so these regressions cannot separate the two orderings.[26]

Other surprise series. The timing margin is a property of the instrument construction, and it applies wherever a predictable high-frequency surprise is orthogonalized and aggregated. Using the BS predictor set, I show that the leading public surprise series are predictable at the announcement level, exposing the same concerns on instrumental validity, and that in every one of them the sum-then-purge instrument is the more relevant of the two.[27] Although a systematic analysis of the properties of each series is beyond the scope of the illustration (and of this paper), the direction generalizes across instruments built by different authors from different underlying data: wherever announcements cluster within the aggregation period, purging the aggregate leaves in variation that the announcement-level purge removes, and that variation inflates the first-stage strength. Appendix C reports both exercises in full.

[25] Without February 1989 and December 1991 the slope is 0.61. Purging the carried-forward prediction as well lowers BS's first stage from 2.47 to 1.31; purging a placebo built from other months' news instead leaves 2.42 in the median reshuffle (Appendix A.I).

[26] Two tests that the financial variables' response cannot contaminate, one on the payroll variables and one on purges built only from first-announcement news, still separate the orderings as the carried-forward prediction implies (Appendix B).

[27] A qualification applies, developed in Appendix C: I impose one predictor set on every series, so the exercise asks what a monthly-aggregate orthogonalization would do to each surprise measure, even if the majority of them are not orthogonalized in their original papers.

IV.II Macroeconomic Inference

Sections 3, 4, and 5 of BS examine, respectively, (a) the predictability of monetary policy surprises; (b) its consequences for estimating the effects of monetary policy on asset prices; and (c) its consequences for the macroeconomy. This section revisits all of these empirical results in light of the issues uncovered in the previous sections, and reassesses their interpretation through that lens.

Announcement-level Predictability

BS Section 3 introduces two surprise measures: a conventional one (`MPS`), and its orthogonalized counterpart (`MPS_ORTH`), which removes correlations with predictors available before the announcement.[28] The authors argue that both measures show significant effects on macroeconomic outcomes, although the orthogonalized surprises address endogeneity concerns and amplify such effects significantly (up to a factor of 4). The orthogonalized instruments are constructed by taking the residuals of OLS regressions of `MPS` on a series of predictors, as per their equation (14). Following up on previous studies (Cieslak (2018), Bauer and Swanson (2023b)), the authors focus on six macrofinancial variables known to predict estimated surprises. Their Table 1 summarizes these regressions, which I replicate coefficient by coefficient in Table IV, confirming the first step in the construction of `MPS_ORTH`.

BS Section 4 formalizes the construction of orthogonalized surprises (per their eq. (16)) and compares the effects of unadjusted and adjusted monetary surprises on a set of financial assets — BS equation (15). The results are reported in BS Table 3, which I replicate coefficient by coefficient in Table V. This confirms that the first stage in the construction of the proxy-SVAR is correct.

At the announcement level, the two procedures are indistinguishable: both target the same event-level residual. The divergence between z_m and $\tilde{z}_m$ begins when event-level surprises are aggregated to the monthly frequency. The two monthly series use the same surprises and differ in the timing of orthogonalization; they remain closely aligned overall (correlation 0.95), and especially so in single-announcement months (0.99), but diverge in multi-announcement months (0.84) — precisely where several announcements share one monthly purge (Figure A.1; Table A.1).[29]

[28] The predictability results may depend on how the predictors are constructed. Acosta (2022) shows that the contribution of the pre-announcement S&P 500 return is sensitive to the choice of nearby aggregation window. This is separate from the timing question studied here: holding fixed the BS predictor set, the present note asks whether orthogonalization should be applied after monthly aggregation or at the announcement level before aggregation.

[29] Appendix A provides a detailed comparison.

Table IV: Predictive Regressions Using Macroeconomic and Financial Data

	(1)	(2)	(3)
Nonfarm payrolls	0.094	0.113	0.082
	(2.425)	(1.976)	(1.788)
Employment growth (12m)	0.005	0.004	0.005
	(2.144)	(1.403)	(1.216)
Δ log S&P 500 (3m)	0.084	0.112	0.154
	(1.447)	(1.578)	(1.943)
Δ Slope (3m)	-0.010	-0.010	-0.011
	(-1.393)	(-1.153)	(-1.035)
Δ log Comm. Price (3m)	0.119	0.093	0.224
	(2.380)	(1.460)	(3.488)
Treasury skewness	0.032	0.035	0.050
	(3.016)	(2.917)	(2.126)
Constant	-0.011	-0.009	-0.021
	(-2.345)	(-1.743)	(-2.601)
R^2	0.162	0.173	0.192
Sample	1988:02–2019:12	1994:01–2019:12	1988:02–2007:06
N	322	218	216
Policy surprise	mps	mps	mps

Notes: Regressions $\text{mps}_t = \alpha + \boldsymbol{\beta}'\boldsymbol{X}_t + u_t$ over FOMC announcements t, with BS's monetary policy surprise and the six news variables of their Table 1, observed before each announcement: the payroll entry is the surprise in the latest release, the three-month changes end the day before the announcement, and Treasury skewness is from Bauer and Chernov (2024). HC0 t-statistics in parentheses. As in BS, this table, Table V and Table C.1 report t-statistics, the other coefficient tables standard errors.

Table V: Asset Price Responses to Monetary Policy Surprises

	mps_t			$\text{mps}_t^{\perp}$		
Series	Coef.	t-stat	R^2	Coef.	t-stat	R^2
2–year yield	.73	(18.6)	.783	.74	(16.7)	.689
5–year yield	.63	(14.4)	.626	.64	(13.8)	.550
10–year yield	.41	(9.5)	.436	.41	(9.9)	.363
30–year yield	.25	(6.3)	.206	.25	(6.7)	.173
S&P 500	-5.39	(-7.7)	.304	-5.50	(-6.6)	.266
Observations	322			322		

Notes: Regressions over FOMC announcements t of the change in each Treasury yield or in the log S&P 500 index, in a narrow window around the announcement, on mps_t or on $\text{mps}_t^{\perp}$, its residual on the predictors of Table IV. HC0 t-statistics in parentheses. Sample: 1988:02–2019:12.

Impulse Responses

Finally, section 5 in BS offers an empirical panorama of results that aim to capture the effects of monetary policy on the macroeconomy. In particular, it employs a spectrum of methodologies including External IV SVAR (5.A), LP methods (5.B) internal IV SVAR (5.C), and others, concluding with the summarized best prac-

tices for monetary analysis (5.E). This section focuses on revisiting 5.A and 5.E, comparing BS' proposed orthogonalized monetary surprises with the same object obtained with an alternative orthogonalization scheme; the same issue affects each other empirical approach as well, which I revisit later.

I now turn to one of the main research questions BS aim to answer: how much difference does orthogonalizing the high-frequency surprises make for estimating the effects of monetary policy on the economy? Figure 3 in their paper answers this question by demonstrating that the responses of all variables to `MPS_ORTH` *"are all larger [...] by a factor of about four"* in an External IV SVAR. The alternative approach I propose — which removes the later-announcement predictable component characterized in Section III — fails to replicate these findings.

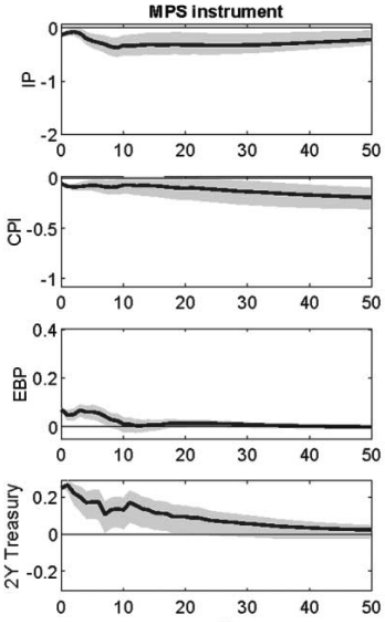
Figure IV and V reproduce the left and the right columns of Figure 3 in BS respectively. Figure IV uses the unadjusted surprise series (`MPS`), and shows exact replication of the point estimates; Figure V uses the purge-then-sum series to address the predictability of the unadjusted series, and by doing so *"[...] provide*[s] *better estimates of monetary policy's true effects on macroeconomic variables"*.[30] No statistically detectable responses emerge for the orthogonalized series. Variations of VAR specifications (number of lags, deterministic component, data transformations) yield analogous findings.

Investigating the stark differences, the conventional estimates become uninformative because of the almost complete loss of statistical power (relevance) of the instrument: the purge-then-sum first-stage F-statistic collapses from 9.65 to 0.29, well below any conventional weak-instrument threshold.[31] The corrected measure has essentially no predictive power for the baseline two-year Treasury yield, so the BS-style design no longer delivers statistically informative macroeconomic responses.[32] Section IV.I traces this loss of relevance to the predictable part of later announcements, which sum-then-purge keeps and purge-then-sum removes.

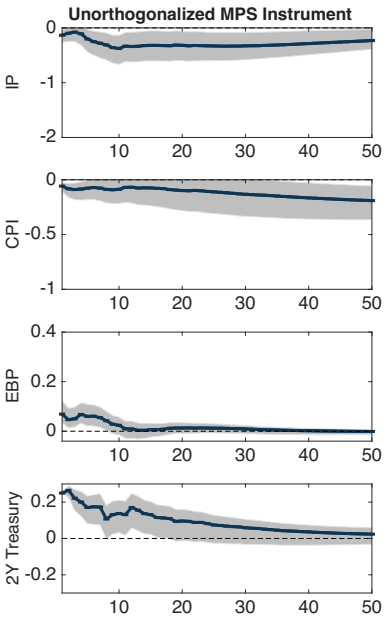
[30] Figure A.4 in Appendix A draws sum-then-purge and purge-then-sum on BS's published data and on a common scale per variable, for the designs of BS's Figures 3 and 8.

[31] For a formal overview of weak-instrument critical values in linear IV, see Olea and Pflueger (2013); Andrews et al. (2019) survey identification-robust inference. Lewis and Mertens (2026) develop IRF-specific diagnostics for weak-instrument bias. I report conventional bands for comparability with BS. Identification-robust inference reaches the same conclusion about magnitudes: confidence sets constructed as in Montiel Olea et al. (2021), with the VAR estimated on 1987:02–2019:12, are unbounded at every horizon and every variable under both orderings, so no magnitude is identified. I thank Daniel Lewis for suggesting this exercise.

[32] Appendix D reports an extension through 2023.

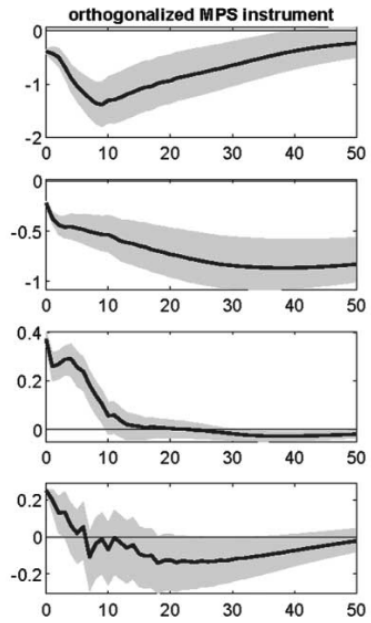


Original Figure 3 (Left Panels)
Bauer and Swanson (2023a)

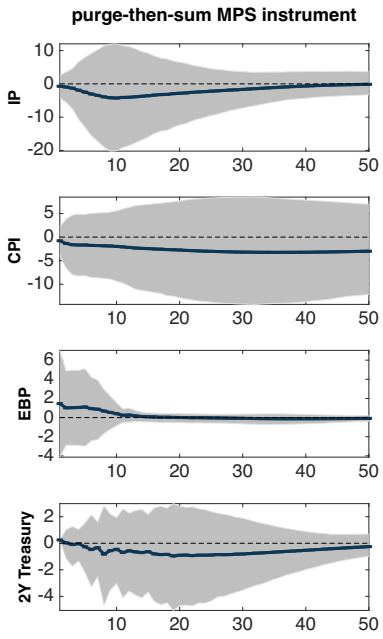


Replication

Figure IV: Comparison between Proxy SVAR IRFs to a 25bp policy shock identified using the unadjusted high-frequency `MPS` measure. Left: original Bauer and Swanson (2023a) Figure 3 (Left Panels); Right: Replication. Sample: 1973:01–2020:02. Lags: 12. 90% bands, BS's bootstrapped standard-error bands (left) and wild-bootstrap percentiles, 10,000 draws (right). EBP = excess bond premium, CPI = Consumer Price Index, IP = ind. production.



Original Figure 3 (Right Panels)
Bauer and Swanson (2023a)



Purge-then-sum

Figure V: Comparison between Proxy SVAR IRFs to a 25bp policy shock identified using BS's orthogonalized `MPS_ORTH` (left) and purge-then-sum (right). Left: original Bauer and Swanson (2023a) Fig. 3 (Right Panels); Right: purge-then-sum. Sample, lags, bands and abbreviations as in Figure IV.

Robustness: LP, internal IV, and additional specifications While some of the results are not directly replicable due to the absence of Fed Chairs speeches from the available replication package referenced above, I reproduce the FOMC-only design of Fig. 8, the "best practices" specification (see below), of Fig. 5 (local projections) and of Fig. 6 (internal IV SVAR) in Appendix E, and confirm the absence of recognizable effects.[33]

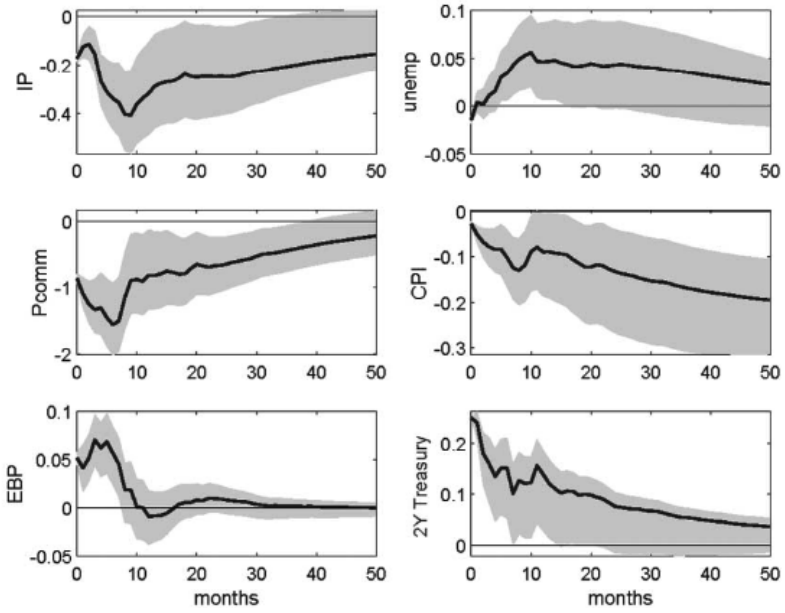
Finally, Figure VI reproduces arguably the most important result in the paper, depicted in Figure 8, i.e., the *"best practice estimates of structural VAR"*. Figure VI.A reports the original Figure 8 in BS verbatim for comparison, while Figure VI.B estimates the same design with purge-then-sum. The main difference with respect to Figure 3 is in the presence of two extra variables in the system: unemployment and an index of commodity prices. Note that this is one instance where the *exact* replication is not possible because the replication data do not include Fed chair speeches among the high-frequency surprise events. The exercise reported here therefore represents the closest feasible approximation based on available data.

Focusing on Panel (B) of Figure VI, we notice again that the purge-then-sum surprises show no statistically detectable responses in macroeconomic aggregates under conventional bands, and again see a significant loss in instrumental strength (F-statistic = 0.67).

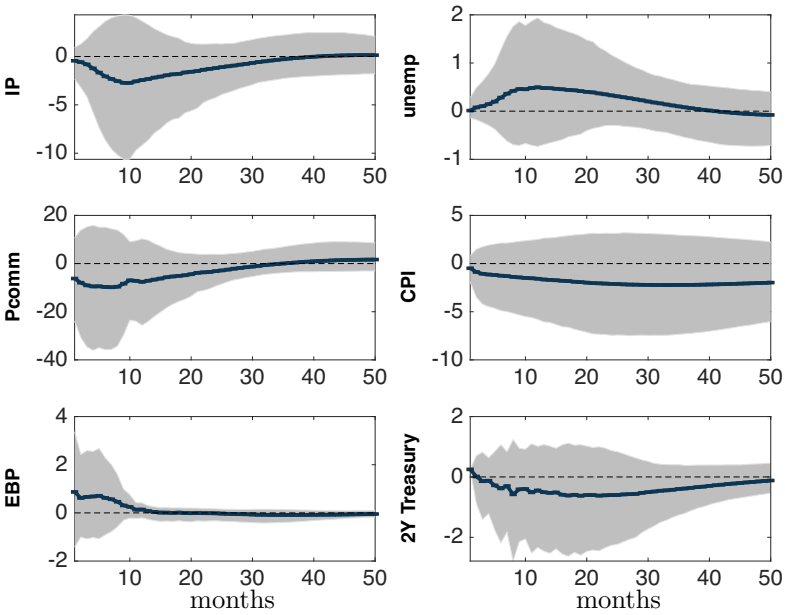
Under purge-then-sum, the orthogonalized monetary surprises yield no statistically significant effects on the macroeconomy through their impact on interest rates, much less the four-times-larger responses on macroeconomic aggregates reported by BS.[34] Appendices D and E show that this pattern is not confined to the baseline Figure 3 exercise. Extending the sample through 2023 strengthens all proxies but leaves the corrected proxy materially weaker; the feasible public-data analogues to the broader BS exercises (local projections, internal-instrument recursive SVAR) yield the same qualitative weakening.

[33] Quoting from the downloadable spreadsheet: *"This spreadsheet contains data only for the FOMC announcements, and excludes data for the Fed Chairs speeches and other types of monetary policy announcements (FOMC minutes releases, speeches by the Vice Chair)."* Appendix E reports the feasible local-projection analogue, closest to Figure C.1 in an older BS draft.

[34] With the end-of-month one-year yield or the federal funds rate as policy variable, the first stage of purge-then-sum rises to 2.25 and 3.34 (Table II), still below conventional thresholds.



(A) Original Bauer and Swanson (2023a) Figure 8



(B) Purge-then-sum

Figure VI: "Best practice" estimates of Proxy SVAR IRFs to a 25bp policy shock identified using BS's orthogonalized `MPS_ORTH` (panel A) and purge-then-sum (panel B). Sample: 1973:01–2020:02. Lags: 12. 90% bands. EBP = bond premium, CPI = Consumer Price, unemp = unemployment, Pcomm = commodity prices, IP = ind. production.

V. Concluding Remarks

Two operations stand between a high-frequency surprise and a monthly instrument: summing surprises within the month, and purging the component predictable from public information. In this paper, I show they do not commute. Sum-then-purge against information available before the month's first announcement cannot remove the part of a later surprise that is predictable from news arriving in between or carried forward by the first announcement's prediction; that variation stays in the instrument. Purging first, each surprise against its own pre-announcement information, removes it by construction. The two procedures coincide when the first announcement's news variables span the later ones and every month has the same number of announcements. Neither condition holds in the data.

That retained component is what the instrument runs on. Moving the orthogonalization to the announcement level removes this predictable variation and, with it, the instrument's statistical relevance. The first-stage F-statistic falls from the 9.65 of the raw surprise and 2.47 of the sum-then-purge instrument to the 0.29 of purge-then-sum, and the corrected instrument is far too weak to size the macroeconomic responses in proxy-SVAR, local-projection and internal-IV specifications. The conventional bands around BS's orthogonalized responses rest on a first stage that owes two thirds of its covariance with the yield to the component purge-then-sum removes. Identification-robust confidence sets are virtually unbounded, not allowing to isolate in BS's baseline proxy SVAR how much industrial production or inflation respond to a monetary shock.

The variation leading to the inflated first stage is concentrated in a small and identifiable set of months. The instrumental power of the monthly orthogonalized Bauer and Swanson (2023a) monetary surprise originates in 31 multi-announcement months — 8% of the baseline sample. Single-announcement months contribute no measurable interest rate response. The predictable part of later-announcement surprises that sum-then-purge retains inflates the first stage and contaminates the impulse responses. The timing flaw and the multi-announcement-month concentration are one finding: the months where the two orthogonalization procedures diverge are exactly the months that drive the instrument's strength.

Three results corroborate the findings: the two monthly series are nearly identical in single-announcement months but diverge sharply in multi-announcement months; later announcements carry a predictable part of their own, which purge-then-sum removes and sum-then-purge largely keeps; and the carried-forward prediction, which the monthly purge leaves in, accounts for most of the gap's covariance with the two-year-yield residual.

For monthly instruments built from high-frequency surprises, orthogonalization

timing is part of the identifying design. The distinction is most consequential in the episodes where announcement clustering is economically salient — crisis interventions, emergency meetings, inter-meeting actions — which are precisely the episodes that supply most of the instrument's statistical power. I recommend purging each announcement on its own news variables before aggregating, and reporting the first stage of the purged series.

References

Acosta, Miguel (2022). "The perceived causes of monetary policy surprises". In: *Published manuscript*.

Andrews, Isaiah, James H Stock, and Liyang Sun (2019). "Weak instruments in instrumental variables regression: Theory and practice". In: *Annual Review of Economics* 11.1, pp. 727–753.

Aruoba, S. Borağan and Thomas Drechsel (2024). *Identifying monetary policy shocks: A natural language approach*. Working Paper 32417. National Bureau of Economic Research.

Bauer, Michael and Mikhail Chernov (2024). "Interest rate skewness and biased beliefs". In: *The Journal of Finance* 79.1, pp. 173–217.

Bauer, Michael D and Eric T Swanson (2023a). "A reassessment of monetary policy surprises and high-frequency identification". In: *NBER Macroeconomics Annual* 37.1, pp. 87–155.

— (2023b). "An alternative explanation for the "fed information effect"". In: *American Economic Review* 113.3, pp. 664–700.

Bu, Chunya, John Rogers, and Wenbin Wu (2021). "A unified measure of Fed monetary policy shocks". In: *Journal of Monetary Economics* 118, pp. 331–349.

Chen, Zhengyang (2026). "Demystifying monetary policy surprises: Fed response to financial conditions and wait and see for new economic data". In: *Journal of Macroeconomics* 87, p. 103736.

Cieslak, Anna (2018). "Short-rate expectations and unexpected returns in treasury bonds". In: *The Review of Financial Studies* 31.9, pp. 3265–3306.

Cieslak, Anna and Andreas Schrimpf (2019). "Non-monetary news in central bank communication". In: *Journal of International Economics* 118, pp. 293–315.

Cook, Timothy and Thomas Hahn (1989). "The effect of changes in the federal funds rate target on market interest rates in the 1970s". In: *Journal of monetary economics* 24.3, pp. 331–351.

Gertler, Mark and Peter Karadi (2015). "Monetary policy surprises, credit costs, and economic activity". In: *American Economic Journal: Macroeconomics* 7.1, pp. 44–76.

Gurkaynak, Refet S, Brian Sack, and Eric T Swanson (2005). "Do actions speak louder than words? The response of asset prices to monetary policy actions and statements". In.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright (2007). "The U.S. Treasury Yield Curve: 1961 to the Present". In: *Journal of Monetary Economics* 54.8, pp. 2291–2304.

Hürtgen, Patrick and Harald Uhlig (Sept. 2026). "Monetary policy shocks: Where do we stand?" Working paper.

Jacobson, Margaret, Christian Matthes, and Todd Walker (2026). "Temporal aggregation bias and monetary policy transmission". In: *Journal of Political Economy Macroeconomics*. Forthcoming. DOI: 10.1086/741548.

Jarociński, Marek and Peter Karadi (2020). "Deconstructing monetary policy surprises—the role of information shocks". In: *American Economic Journal: Macroeconomics* 12.2, pp. 1–43.

Karnaukh, Nina and Petra Vokata (2022). "Growth forecasts and news about monetary policy". In: *Journal of Financial Economics* 146.1, pp. 55–70. DOI: 10.1016/j.jfineco.2022.07.001.

Kuttner, Kenneth N (2001). "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market". In: *Journal of monetary economics* 47.3, pp. 523–544.

Lee, Yeon Jik and Tatevik Sekhposyan (Oct. 2024). "The relevance of temporal aggregation for the propagation of macroeconomic shocks". Working paper.

Lewis, Daniel J. and Karel Mertens (Jan. 2026). *Weak instrument bias in impulse response estimators*. Working Paper 2601. Federal Reserve Bank of Dallas. DOI: 10.24149/wp2601. URL: https://www.dallasfed.org/research/papers/2026/wp2601.

Miranda-Agrippino, Silvia and Giovanni Ricco (2021). "The transmission of monetary policy shocks". In: *American Economic Journal: Macroeconomics* 13.3, pp. 74–107.

Montiel Olea, José Luis, James H. Stock, and Mark W. Watson (2021). "Inference in structural vector autoregressions identified with an external instrument". In: *Journal of Econometrics* 225.1, pp. 74–87.

Nakamura, Emi and Jón Steinsson (2018). "High-frequency identification of monetary non-neutrality: the information effect". In: *The Quarterly Journal of Economics* 133.3, pp. 1283–1330.

Olea, José Luis Montiel and Carolin Pflueger (2013). "A robust test for weak instruments". In: *Journal of Business & Economic Statistics* 31.3, pp. 358–369.

Plagborg-Møller, Mikkel and Christian K Wolf (2021). "Local projections and VARs estimate the same impulse responses". In: *Econometrica* 89.2, pp. 955–980.

Ramey, Valerie A (2011). "Identifying government spending shocks: It's all in the timing". In: *The Quarterly Journal of Economics* 126.1, pp. 1–50.

— (2016). "Macroeconomic shocks and their propagation". In: *Handbook of macroeconomics* 2, pp. 71–162.

Stock, James H and Mark W Watson (2018). "Identification and estimation of dynamic causal effects in macroeconomics using external instruments". In: *The Economic Journal* 128.610, pp. 917–948.

Swanson, Eric T (2021). "Measuring the effects of Federal Reserve forward guidance and asset purchases on financial markets". In: *Journal of Monetary Economics* 118, pp. 32–53.

Wu, Jing Cynthia and Fan Dora Xia (2016). "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound". In: *Journal of Money, Credit and Banking* 48.2-3, pp. 253–291.

Appendix

A. Direct comparison of the orderings

The two orderings nearly coincide in months with one announcement and diverge in months with several. Table A.1 and Figure A.1 compare BS's posted monthly series z_m (`MPS_ORTH`) with purge-then-sum $\tilde{z}_m$, the monthly sum of their posted announcement-level residuals. Both purge the same surprises on the same six news variables, three of which BS's monthly file records differently. Once both purges use the announcement-level records, the ordering alone lowers the first stage from 1.86 to 0.29 (Appendix A.II). With one announcement, they then differ only through their estimated coefficients; with several, sum-then-purge purges the month's sum once, on the first announcement's news variables, and purge-then-sum each announcement on its own. Each posted series is the residual of its regression to within about 0.01 basis points. The three largest differences, 15.9, 12.6 and 8.1 basis points, fall in February 1989, December 1991 and October 2008, all with several announcements. In February 1989 the five residuals sum to −6.5, and BS's monthly series posts +9.4. BS's Figures 3–6, 8 and A2 use sum-then-purge; Figures 7 and A3 compare surprises not purged on forecasts with instruments built as in Miranda-Agrippino and Ricco (2021), which purge each announcement on forecasts before summing.

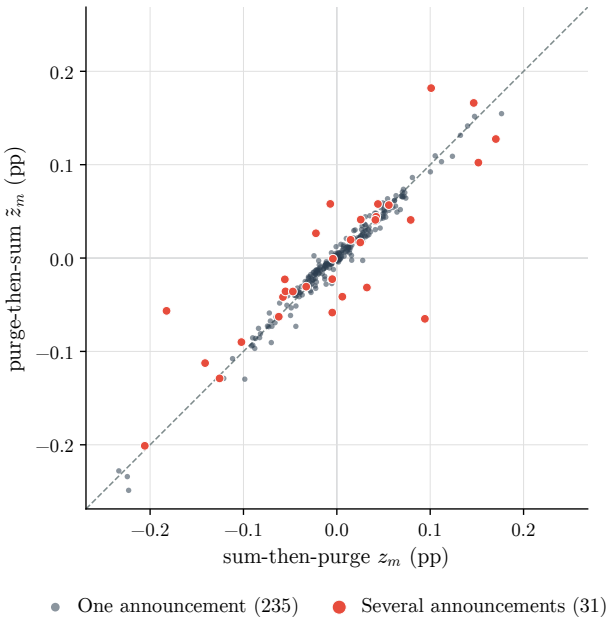


Figure A.1: Sum-then-purge z_m (horizontal axis) against purge-then-sum $\tilde{z}_m$ (vertical axis), in percentage points (0.1 = 10 basis points), in the 266 months with an announcement; red dots mark the 31 months with several announcements, and the dashed line is the 45-degree line. Sample: 1988:02–2019:12.

Table A.1: Sum-then-purge and purge-then-sum, by number of announcements

	Months	Correlation	Mean absolute difference, bp
All months with an announcement	266	0.950	0.91
One announcement	235	0.988	0.60
Several announcements	31	0.843	3.20

Notes: BS's posted monthly series z_m (`MPS_ORTH`) against purge-then-sum $\tilde{z}_m$, the monthly sum of BS's posted announcement-level residuals, over the months with an announcement, 1988:02–2019:12.

A.I What the gap is made of

Net of BS's monthly regressors, the carried-forward prediction accounts for 55% of the gap's covariance with v_m and news for 28% (Table A.2, Panel A). Equation (4) abstracts from two features of BS's implementation, which its sample counterpart includes, so the split is exact for BS's posted series. Both purges have constants, and BS's monthly regression uses the monthly file's records of the news variables, $\boldsymbol{W}_{1m}$, which differ from their announcement-level records in three variables. The largest difference is the yield-curve slope, which $\boldsymbol{W}_{1m}$ measures as the change in the ten-year minus three-month Treasury spread (FRED's T10Y3M) and the announcement-level records, like BS's paper, as the change in the second principal component of the Gürkaynak et al. (2007) yields; the two correlate at 0.85 at the month's first announcement. With $\hat{f}(\boldsymbol{X}) = \hat{\alpha} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}$ from equation (1) and $\hat{c} + \hat{\boldsymbol{\gamma}}'\boldsymbol{W}_{1m}$ the fitted value of BS's monthly regression, the sample counterpart of equation (4) is

$$g_m = \underbrace{(T_m - 1)\,\hat{f}(\boldsymbol{X}_{1m})}_{R_m} + \underbrace{\hat{\boldsymbol{\beta}}' \textstyle\sum_{t=2}^{T_m}(\boldsymbol{X}_{tm} - \boldsymbol{X}_{1m})}_{N_m} + \underbrace{\hat{f}(\boldsymbol{X}_{1m}) - \hat{c} - \hat{\boldsymbol{\gamma}}'\boldsymbol{W}_{1m}}_{C_m} \quad \text{(A.1)}$$

in every month with an announcement, up to the rounding of BS's posted series (about 0.01 basis points). In a month with one announcement the gap is the remainder C_m alone, which is why the two series differ slightly even there.

BS's monthly purge removes about 1.3 times the first announcement's predicted value in every month, whatever the number of announcements. The fitted value of their monthly regression has a slope of 1.30 on $\hat{f}(\boldsymbol{X}_{1m})$ in months with one announcement and 1.34 in months with several. In a month with one announcement it therefore removes about 0.30 predicted values too many, and in a month with several it removes, besides the first announcement's predicted value, only about 0.34 of the $T_m - 1$ further predicted values that the carried-forward prediction R_m adds. Panel A of Table A.2 therefore reports each term net of what BS's regressors span. For any series a_m, $a_m^{\perp}$ denotes its residual on a constant and $\boldsymbol{W}_{1m}$ over the 266 months with an announcement, and the gap splits into $R_m^{\perp} + N_m^{\perp} + C_m^{\perp} - \Pi(\tilde{z}_m)$, where $\Pi(\tilde{z}_m)$ is the small part of $\tilde{z}_m$ that BS's monthly records span; Panel B reports the raw terms. In words, a net term is the part of

a component that BS's monthly purge leaves in its instrument, and the raw term adds the part that BS's regressors span, which is why the two panels differ.

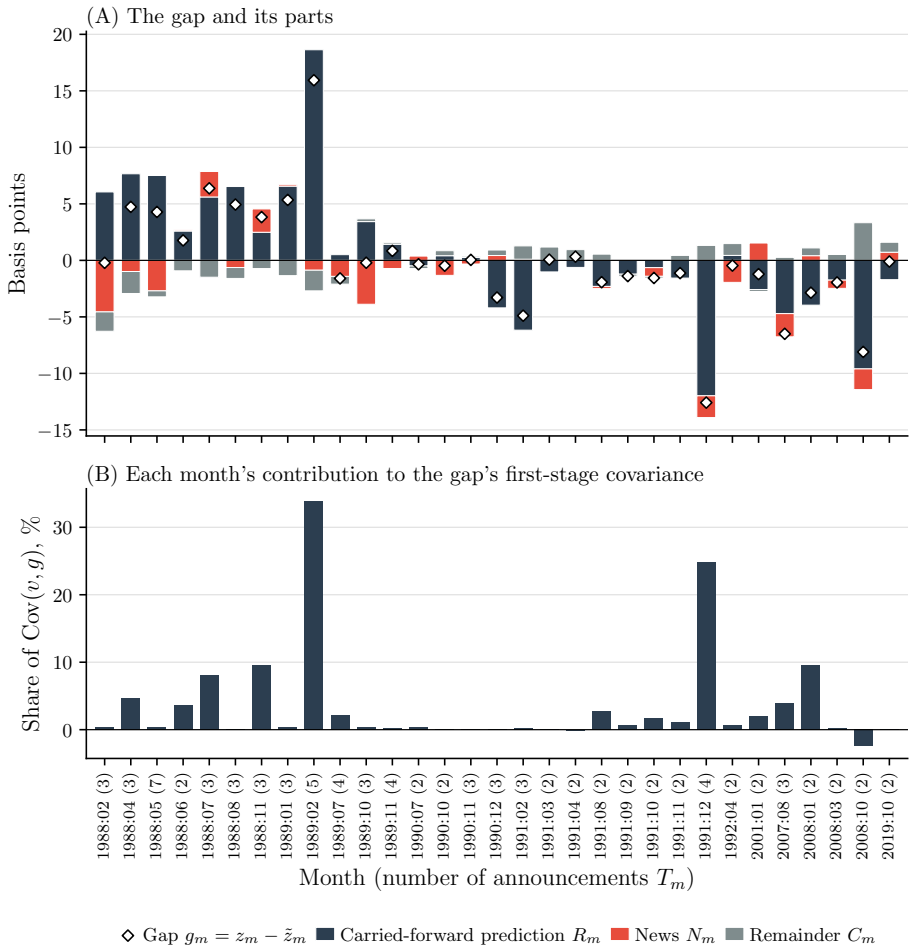


Figure A.2: The gap in the 31 months with several announcements. Panel A: the gap g_m (diamonds) and the terms of equation (A.1), in basis points, positive terms stacked up from zero and negative terms down. Panel B: each month's contribution to $\text{Cov}(v_m, g_m)$, in %; the 31 contributions add up to 109%.

The carried-forward prediction's share comes mostly from two months. In February 1989, R_m assigns the first announcement's predicted surprise to each of the four later announcements, 18.6 basis points in all, more than the whole gap of 15.9 (Figure A.2). Similarly, in December 1991, with four announcements, R_m is nearly the whole gap, and in most months with several announcements it is the largest of the three terms. Both months also have large two-year-yield residuals with the sign of R_m, so they supply more than all of the share of $R_m^{\perp}$ in Table A.2. Over the other months it nets out, as positive contributions in the other months with several announcements offset negative ones in the months with one, where R_m is zero and BS's purge subtracts its fitted value. The other months supply nearly all of the share of $N_m^{\perp}$. The net remainder, which would be zero if BS's

monthly records matched the announcement-level ones, reflects mainly the two records of the yield-curve slope.

Table A.2: What the gap is made of

	Share of $\mathrm{Cov}(v_m, \cdot)$, %	
	of g_m	of z_m
Whole gap, g_m	100	67
A. Net of BS's monthly regressors		
Carried-forward prediction, $R_m^{\perp}$	55	37
News, $N_m^{\perp}$	28	19
Remainder, $C_m^{\perp}$	10	7
Fitted part, $-\Pi(\tilde{z}_m)$	7	5
B. Raw terms		
Carried-forward prediction, R_m	101	67
News, N_m	24	16
Remainder, C_m	−25	−17

Notes: Covariance of each term with the two-year-yield residual v_m, relative to that of the gap g_m and of z_m; within a panel the shares add up to 100 up to rounding. Panel A: residuals on a constant and BS's monthly news variables $\boldsymbol{W}_{1m}$ over the 266 months with an announcement, and the fitted value $\Pi(\tilde{z}_m)$ of $\tilde{z}_m$ on them. Panel B: the terms of equation (A.1). Sample: 1988:02–2019:12, 383 months.

Three checks support reading the carried-forward prediction as predictable variation that carries much of BS's first stage. First, later announcements carry a predictable part of their own, as the carried-forward prediction assumes. Fit on the first announcement of each month, equation (1) predicts the surprises of the 56 later announcements from their own news variables, and the slope of the surprises on these predictions is 0.84 (standard error 0.22, clustered by month). Leaving out any one of the 31 months with several announcements keeps the slope between 0.66 and 0.96, and leaving out February 1989 and December 1991 together gives 0.61. Summed by month, the later surprises move 1.02 (0.21) per unit of the carried-forward prediction built from that fit, holding the news term N_m fixed. In words, each later announcement carries close to a full predicted surprise of its own, which purge-then-sum removes and sum-then-purge largely keeps.

Second, the data do not reject a common announcement-level purge for months with one announcement and months with several. Interacting the constant and the six news variables with an indicator of several announcements gives a joint $F(7, 308)$ of 0.86 ($p = 0.54$, heteroskedasticity-robust standard errors). Weighting each month equally, with weights $1/T_m$, leaves purge-then-sum nearly unchanged (correlation 0.998). Irregular announcement timing thus gives no reason to purge months with several announcements differently.

Third, purging the carried-forward prediction as well lowers the first stage of sum-

then-purge by about half, whereas a placebo built from other months' news variables leaves it almost unchanged. Adding R_m of equation (A.1) to BS's monthly regression lowers it from 2.47 to 1.31. A placebo R_m that keeps each month's number of announcements and takes the news variables of another month with several announcements leaves the first stage at 2.42 in the median reshuffle. So the drop comes from each month's own predicted surprise, which the placebo replaces.

A.II The two first-stage statistics

Purge-then-sum is the weakest of the three instruments under both first-stage statistics and in every specification of Table A.3. BS estimate the second step of the proxy SVAR by two-stage least squares (their eq. 22), and their first-stage statistic (endnote 52) is the squared t-statistic of the instrument in a regression of the two-year yield on a constant, 12 lags of the four VAR variables and the instrument, with heteroskedasticity-robust (HC0) standard errors. Over 1974:01–2020:02, with the instrument set to zero outside 1988:02–2019:12, this 2SLS first stage gives 8.19 for the raw surprise, the number BS print, and 1.98 for their posted monthly series. BS print 1.83, which I reproduce from their April 2022 release.[35]

The 2SLS first stage shares the covariance of the proxy-SVAR first stage, since the yield net of the VAR's lags is v_m, and differs from it in two respects. First, netting the lags out of the instrument removes 7 to 14% of its variance across the three instruments, which raises the slope by 7 to 16%. Second, the 2SLS sample adds 1974:01–1988:01, when the instrument is zero and the residual variance of the two-year yield is 4.5 times that of 1988–2019, so these months hold two thirds of the squared residuals. A robust standard error weights each squared residual by the squared instrument net of the lags, which is small where the instrument is zero, so these months carry only 10 to 18% of the slope's variance. With robust standard errors in both regressions the two statistics are within 7% of each other for every instrument, 1.98 against 2.03 for z_m (Table A.3, Panel A), and the covariance shares of Section IV.I are the same to two decimals in either regression. In words, the two statistics measure the same covariance and agree once both use robust standard errors.

With BS's announcement-level records of the news variables (Appendix A.I) in

[35] BS's initial release of 28 April 2022, kept by the Internet Archive, predates the two corrections in their change log and has no monthly sheet. Summing its surprises by month and purging the sum on the first announcement's news variables, as BS describe their monthly series, gives 1.83; its raw surprise gives 8.13, and purge-then-sum 0.25. Each printed number thus matches one of BS's releases. BS's two-year yield equals the last daily yield of each month on the Gürkaynak et al. (2007) curve, to the four decimals BS post, in all 566 months of the VAR. The VAR keeps September 2001, where the instrument is zero because BS exclude the announcement of 17 September 2001.

both purges, the ordering lowers the first stage from 1.86 to 0.29 (Table A.3, Panel C). On the monthly file's records, as in BS's posted series, sum-then-purge gives 2.47, and 1.90 when only the yield-curve slope takes its announcement-level record. The other two records each move it by at most 0.03: Treasury skewness, which BS's monthly file averages over the 23 trading days before the month's first announcement and the announcement-level file over the 22 before each announcement, and twelve-month payroll growth.

Table A.3: First stage under further specifications

	Raw `MPS`		Sum-then-purge z_m		Purge-then-sum $\tilde{z}_m$	
	VAR	2SLS	VAR	2SLS	VAR	2SLS
A. Standard errors, two-year yield						
Homoskedastic	**9.65**	4.91	**2.47**	1.18	**0.29**	0.14
HC0	7.71	**8.19**	2.03	**1.98**	0.27	**0.25**
HC1	7.67	7.45	2.02	1.80	0.27	0.23
B. Interest rate in the VAR						
Two-year, month average	8.20	6.24	1.41	1.00	0.03	0.02
One-year, end of month	17.76	13.12	6.34	4.68	2.25	1.80
One-year, month average	13.80	9.94	3.47	2.33	0.97	0.62
Federal funds rate	13.80	12.27	6.35	5.20	3.34	2.64
C. Predictors of the monthly purge, two-year yield						
Announcement-level predictors			1.86	1.51	0.29	0.25
D. Local projection on the instrument window						
Equation (16), $h = 0$	5.99		1.69		0.06	

Notes: First-stage F-statistics. VAR: the proxy-SVAR first stage of Section IV.I (383 months); 2SLS: the first stage of BS's endnote 52 (554 months). Bold: the statistic reported for each, used in Panels B and C. HC1 scales HC0 by $n/(n-k)$. Panel B replaces the two-year yield in the VAR and in both first stages; the month-average one-year yield is FRED's GS1, and the federal funds rate is spliced with the shadow rate of Wu and Xia (2016). Panel C: BS's monthly purge on the announcement-level records of the slope, skewness and payroll growth. Panel D: the squared t-statistic of the instrument in equation (16) at $h = 0$, Newey–West standard errors.

A.III Leave-one-out and cumulative deletions

Dropping any single month leaves the first stage of sum-then-purge above every first stage of purge-then-sum, and the four largest contributors to $\mathrm{Cov}(v_m, z_m)$, January 2008, February 1994, December 1991 and April 1992, carry all of that covariance. Single deletions move the proxy-SVAR first stage of z_m between 1.21 and 3.40 and that of $\tilde{z}_m$ between 0.02 and 0.62 (Figure A.3, Panels A and B). Deleted cumulatively, the four largest contributors leave z_m with no first stage ($t = -0.04$, Panel C), and further deletions leave both series with negative slopes. Because the deleted months are chosen by their contributions, the cumulative

deletions measure concentration. The same deletions give the same pattern in the 2SLS first stage (Table A.4).

Table A.4: Where the first-stage covariance sits

	Sum-then-purge z_m		Purge-then-sum $\tilde{z}_m$		Gap g_m	
	VAR	2SLS	VAR	2SLS	VAR	2SLS
A. Full sample, 1988:02–2019:12						
Share of $\text{Cov}(v_m, z_m)$	1.00	1.00	0.33	0.33	0.67	0.67
First-stage F (gap: t)	2.47	1.98	0.29	0.25	3.39	4.51
First-stage F, HC1 (gap: t)	2.02	1.80	0.27	0.23	4.42	4.30
B. Leave one month out, first-stage F (gap: t)						
Minimum	1.21	1.13	0.02	0.01	2.65	2.95
	without 2008:01		without 1994:02		without 1989:02	
Maximum	3.40	3.01	0.62	0.65	3.61	5.25
	without 2003:10		without 2003:10		without 2008:10	
C. Cumulative deletions, largest contributors first; first-stage F (gap: t)						
Without 2008:01	1.21	1.13	0.02	0.03	3.11	4.20
and 1994:02	0.51	0.52	$0.09^{\dagger}$	$0.07^{\dagger}$	3.17	4.27
and 1991:12	0.11	0.14	$0.18^{\dagger}$	$0.16^{\dagger}$	2.55	3.28
and 1992:04	$0.00^{\dagger}$	0.00	$0.65^{\dagger}$	$0.66^{\dagger}$	2.52	3.21
and 1989:02	$0.07^{\dagger}$	$0.06^{\dagger}$	$0.42^{\dagger}$	$0.41^{\dagger}$	1.53	1.70

Notes: VAR and 2SLS: the two first stages of Table A.3, on BS's published series. Panel A, first row: each series' covariance with the yield relative to that of z_m (equation (19)), net of the controls in the 2SLS columns. Columns g_m: the t-statistic, with its sign, of the test of $\text{Cov}(v_m, z_m) = \text{Cov}(v_m, \tilde{z}_m)$. Panels B and C drop months from the first-stage regression; Panel C drops, cumulatively, the largest contributors to $\text{Cov}(v_m, z_m)$. †Negative first-stage slope.

The months that carry $\text{Cov}(v_m, z_m)$ overlap only in part with those that separate the two orderings, as January 2008, February 1994 and April 1992 move both series alike. Without February 1989, December 1991 and January 2008, the proxy-SVAR first stage is 0.26 for z_m and 0.03 for $\tilde{z}_m$, and the p-value of the test of equal covariance rises from 0.02 to 0.24. In units of $\text{Cov}(v_m, z_m)$, the 31 months with several announcements contribute 1.27 for z_m and 0.54 for $\tilde{z}_m$.[36]

[36]Deletions remove a month from the first-stage regression and keep BS's published series. Re-estimating both purges without the deleted month gives similar ranges, with the proxy-SVAR statistic between 1.23 and 3.28 for z_m and between 0.01 and 0.84 for $\tilde{z}_m$.

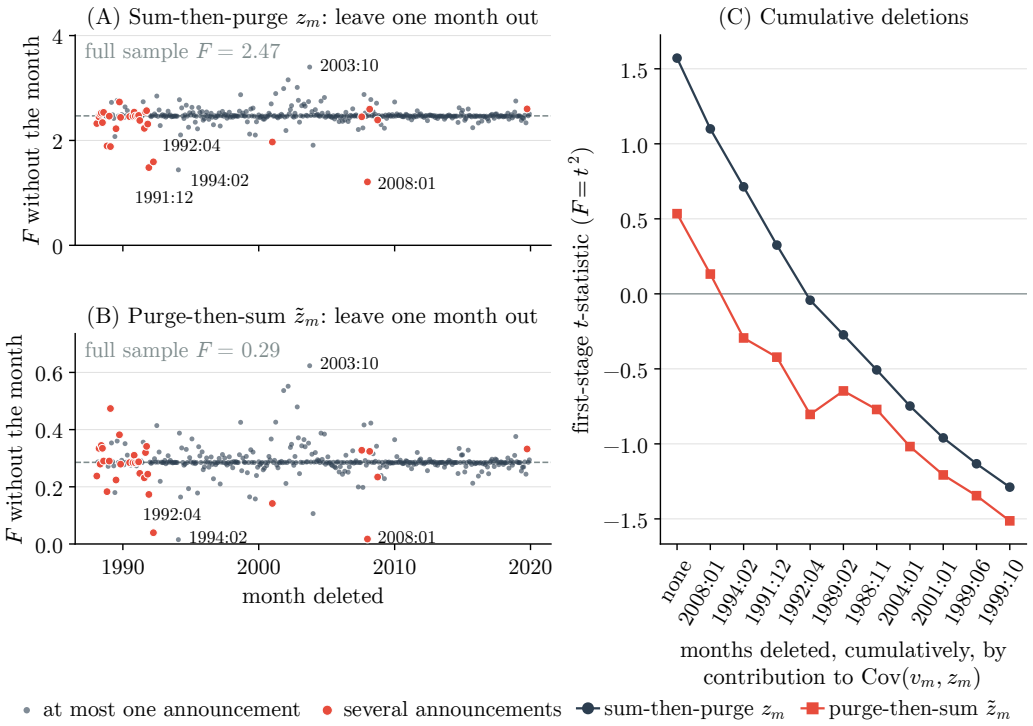
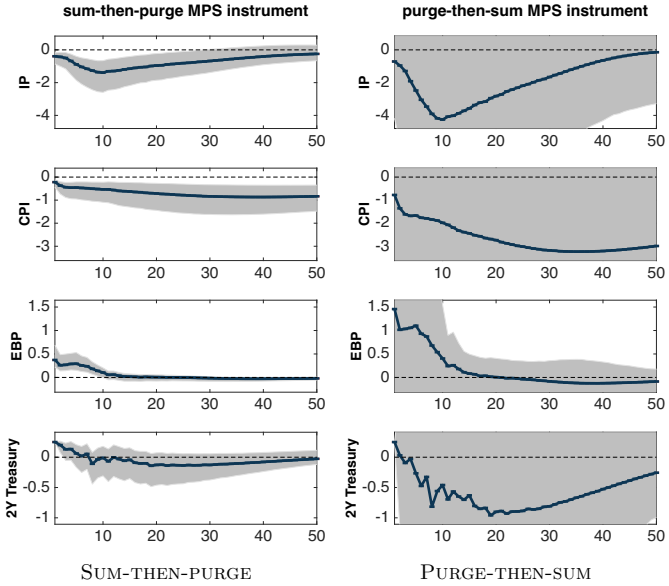


Figure A.3: Leave-one-month-out and cumulative deletions, proxy-SVAR first stage on BS's published series, 1988:02–2019:12. Panels A and B: F after dropping the month on the horizontal axis (dashed line: full sample). Panel C: t-statistic ($F = t^2$) after deleting, cumulatively, the months on the horizontal axis, ordered by their contribution to Cov(v_m, z_m); from the fourth deletion on, the deleted months carry more than all of that covariance, so the slope of z_m is negative by construction.

A.IV Subsamples

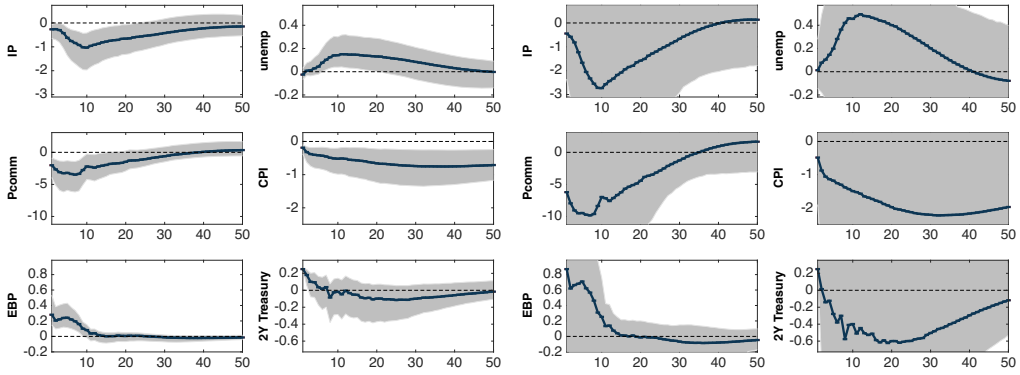
Table A.5 reports the first stages of the raw surprise and of both orderings on subsamples. Without the 31 months with several announcements, or without unscheduled announcements, at most four months hold several announcements, so the carried-forward prediction and the news of equation (4) vanish almost everywhere: the two orderings nearly coincide, and neither they nor the raw surprise has a first stage. From February 1994 six such months remain, the orderings correlate at 0.98, and both purges remove the raw surprise's first stage. With the months held fixed, before February 1994 and in the 31 months, the ordering decides the first stage. Without the 31 months, without unscheduled announcements or from February 1994, the 90% bands of industrial production, consumer prices and the excess bond premium contain zero at every horizon for both orderings, in the proxy SVAR and in the local projections.

(A) Design of BS's Figure 3



Sum-then-purge Purge-then-sum

(B) Design of BS's Figure 8



Sum-then-purge Purge-then-sum

Figure A.4: Same-data comparison on a common scale. Proxy-SVAR responses to a 25bp policy shock identified with sum-then-purge `MPS_ORTH` and purge-then-sum, both on BS's published data; each variable has one scale, which covers both point estimates and the sum-then-purge band, so purge-then-sum's bands extend beyond the axes. Sample: 1973:01–2020:02. Lags: 12. 90% wild-bootstrap percentile bands, 10,000 draws.

BS define an announcement as unscheduled unless it is the first after a scheduled FOMC meeting, and their data also code as scheduled the consecutive announcements right after four meetings, in May 1988, November and December 1990 and November 1991. They keep unscheduled announcements in both their announcement-level and their monthly purge. Of the 31 months, 29 hold an unscheduled announcement, and once these announcements are dropped only those four months keep several. The coding matters only for this subsample and differs across BS's sources, and under all three codings neither ordering has a first stage above 0.25: the flags of BS's 2023 release, used here, code the announcement of

October 19, 1989 as unscheduled; their published flags code it as scheduled, which leaves five such months; and BS's written definition applied literally leaves none.

Table A.5: First stages by subsample

			VAR			2SLS			
	Months	Multi	Raw	z_m	$\tilde{z}_m$	Raw	z_m	$\tilde{z}_m$	Corr.
Full sample	383	31	9.65	2.47	0.29	8.19	1.98	0.25	0.95
Without the 31 months	352	0	0.68	0.28†	0.16†	1.07	0.20†	0.09†	0.99
Without unscheduled	383	4	0.90	0.19†	0.20†	0.84	0.15†	0.16†	1.00
From 1994:02	311	6	3.19	0.30	0.20	2.19	0.17	0.11	0.98
1988:02–1994:01	72	25	7.72	3.72	0.07	11.07	5.17	0.13	0.85
The 31 months alone	31	31	17.38	12.23	1.97	15.25	12.91	3.21	0.84

Notes: First-stage F-statistics; VAR and 2SLS as in Table A.3, the VAR first stage over the row's months and the 2SLS first stage without the other months of 1988:02–2019:12. Multi: months with several announcements. Corr.: correlation of z_m and $\tilde{z}_m$ over the row's months. Every row restricts BS's published series to its months, except the third, which re-estimates both purges on the scheduled announcements under the flags of BS's 2023 release. †Negative first-stage slope.

B. Later news variables and both orderings

The response of later news variables to the month's earlier surprises carries little of the gap, and two tests that the financial variables' response cannot affect both separate the orderings. In Table B.1 the later news variables predict the month's first purged surprise u_{1m} at least as strongly as either ordering, with joint F-statistics of 8.40 and 7.49 against 3.03 to 7.27, and as u_{1m} is realized before them, except at five later announcements on the day of the first, this prediction indicates that they respond to it, so these regressions cannot separate the orderings.

In the 31 months with several announcements, I estimate

$$y_m = \alpha + \boldsymbol{\Gamma}'\boldsymbol{X}_{tm} + \boldsymbol{H}'\boldsymbol{X}_{1m} + e_{tm}, \qquad t = 2, \ldots, T_m, \tag{B.1}$$

where $\boldsymbol{X}_{tm}$ is the vector of news variables observed immediately before later announcement t in month m, $\boldsymbol{X}_{1m}$ the vector at the month's first announcement, entering in columns (2), (4) and (6), and y_m is sum-then-purge z_m, purge-then-sum $\tilde{z}_m$ or the first announcement's purged surprise u_{1m}. All three dependent variables contain the month's first purged surprise, whose estimated effect on later stock returns is negative (Table B.2, Panel A), consistent with the negative coefficient on later stock returns in all six columns. For the gap, (1) minus (3) and (2) minus (4), the coefficient is positive, the sign of the announcement-level regression of Table IV.

Table B.1: Later news variables and both orderings

	Sum-then-purge z_m		Purge-then-sum $\tilde{z}_m$		First announcement u_{1m}	
	(1)	(2)	(3)	(4)	(5)	(6)
Payroll surprise	0.125	0.499	−0.214	0.170	−0.200	−0.111
	(0.240)	(0.173)	(0.173)	(0.157)	(0.095)	(0.097)
Employment growth	0.029	−0.043	0.022	−0.075	0.006	0.107
	(0.019)	(0.151)	(0.015)	(0.133)	(0.007)	(0.073)
S&P 500 return	−0.462	−1.012	−0.517	−1.062	−0.403	−0.670
	(0.189)	(0.310)	(0.143)	(0.268)	(0.094)	(0.139)
Slope change	−0.015	−0.045	−0.002	−0.041	−0.025	−0.001
	(0.053)	(0.077)	(0.042)	(0.067)	(0.026)	(0.043)
Commodity prices	−0.022	0.877	−0.143	0.357	−0.117	−0.506
	(0.169)	(0.498)	(0.127)	(0.424)	(0.069)	(0.260)
Treasury skewness	0.064	−0.316	0.002	−0.259	0.011	−0.020
	(0.066)	(0.209)	(0.049)	(0.157)	(0.028)	(0.088)
$\boldsymbol{X}_{1m}$ included	No	Yes	No	Yes	No	Yes
Joint F	3.03	6.94	7.27	6.75	8.40	7.49
p-value	0.160	0.173	0.125	0.154	0.046	0.031
R^2	0.42	0.67	0.33	0.58	0.38	0.52
Payroll surprise and employment growth only						
Joint F	6.39	0.87	0.89	0.52	2.85	0.90
p-value	0.028	0.635	0.663	0.734	0.220	0.356

Notes: The 56 later announcements of the 31 months with several announcements, 1988:02–2019:12; each month's dependent variable, in percentage points, repeats for each of its later announcements. News variables as in Table IV, the payroll surprise in millions of jobs. Joint F: the six later news variables; bottom rows: the payroll surprise and employment growth only, with their first-announcement values in (2), (4) and (6). Standard errors clustered by month in parentheses; p-values from a wild cluster bootstrap by month, 9,999 draws, as clustered tests over-reject with 31 clusters and up to 13 coefficients. In 1,000 samples with a monthly dependent variable unrelated to the regressors, the clustered F-test rejects at 5% in 24 to 58% of them for the six news variables and in 11 to 23% for the two payroll variables, and the wild cluster bootstrap in 4.2 to 6.8%.

Weighted by BS's coefficients, the response of later news variables to the month's earlier purged surprises (Table B.2, Panel A) adds between −1.1% and 2.5% of the gap's covariance with v_m, and removing it moves the first stage of sum-then-purge from 2.47 to between 2.37 and 2.49. Only the four financial variables can respond within the month, and their news term carries 3.4% of that covariance (4.2% net of BS's monthly regressors). As the carried-forward prediction scales with $T_m - 1$, I also test whether a second announcement responds to the month's first purged surprise. Regressing an indicator of a second announcement on that surprise and its absolute value gives a joint p-value of 0.65, and 0.08 before 1994, when 25 of the 31 months with several announcements fall. In words, the later news variables respond too little to the month's earlier surprises to account for the gap.

Table B.2: Later news variables and the month's first announcement

A. Change in later news variables on the month's earlier purged surprises

	Payroll surprise	Empl. growth	S&P 500	Slope	Comm. prices	Skew.	ϕ	All six
Second, on u_{1m}	0.011	0.271	−0.288	−0.349	−0.272	−0.249	−0.059	
	(0.187)	(0.191)	(0.198)	(1.610)	(0.091)	(0.162)	(0.032)	
p-value	0.959	0.149	0.243	0.801	0.090	0.200	0.171	0.789
Later, on $u_{<t,m}$	0.064	0.101	−0.255	−0.257	−0.135	−0.525	−0.046	
	(0.147)	(0.109)	(0.160)	(0.845)	(0.132)	(0.300)	(0.020)	
p-value	0.833	0.341	0.213	0.775	0.421	0.090	0.026	0.207

B. Purges built from the news variables of the month's first announcement: first-stage F

	Sum-then-purge		Purge-then-sum	
	VAR	2SLS	VAR	2SLS
BS's purges, z_m and $\tilde{z}_m$	2.47	1.98	0.29	0.25
(i) Regression on $\boldsymbol{X}_{1m}$	3.59	2.91	0.72	0.65
(ii) $\hat{f}_0$ at $\boldsymbol{X}_{1m}$	3.22	2.58	0.50	0.44
(iii) $\hat{f}$ at $\boldsymbol{X}_{1m}$	3.32	2.66	0.64	0.55

Notes: Panel A: changes since the month's first announcement, at the second (26) and later (51) announcements dated after it, on u_{1m} and on $u_{<t,m}$, the sum of earlier-dated purged surprises; ϕ: slopes weighted by $\boldsymbol{\beta}$ of equation (1); "All six": joint test; otherwise as in Table B.1. Panel B: (i) the 322 surprises regressed on $\boldsymbol{X}_{1m}$; $\hat{f}_0$ and $\hat{f}$: equation (1) on the 266 first announcements and on all 322; purge-then-sum subtracts the fitted value at every announcement, sum-then-purge once per month; F as in Table A.3.

Both tests separate the orderings, and the financial variables' response can affect neither. In the bottom rows of Table B.1, the two payroll variables, which the Bureau of Labor Statistics releases on its own schedule, predict z_m ($F = 6.39$, $p = 0.03$), whereas their p-value for $\tilde{z}_m$ is 0.66, and the prediction of z_m disappears once their first-announcement values enter ($p = 0.64$). With every announcement purged on the news variables of its month's first announcement, which no surprise of the month can move, the orderings differ by the carried-forward prediction alone and give first stages of 3.22 to 3.59 for sum-then-purge and 0.50 to 0.72 for purge-then-sum (Table B.2, Panel B). Put simply, where later news cannot respond to the month's surprises, the orderings still differ as the carried-forward prediction implies.

C. The ordering across public surprise series

In every surprise series of Section IV.I, purging before summing lowers the first stage. For each series I hold the BS predictor vector $\boldsymbol{X}_t$ fixed and build both monthly instruments: announcement-level residuals of the surprise on $[1, \boldsymbol{X}_t]$,

summed within the month; and the within-month sum residualized on the predictors observed before the month's first announcement. The first stage covers the series of Gertler and Karadi (2015), Jarociński and Karadi (2020) (surprise factor and monetary component), Swanson (2021) (federal funds rate factor), and BS. Holding the predictor set fixed isolates the ordering; each series keeps the publication window of its paper, and no exercise here replicates any paper's own construction. In Bu et al. (2021) and Nakamura and Steinsson (2018) no month has announcements on more than one date, so the two orderings coincide; the Bu et al. (2021) surprises are also unpredictable from the BS vector ($R^2 = 0.05$, $p = 0.27$).

Every series is predictable from information observed before the announcement, with R^2 between 0.102 and 0.176 (Table C.1), and the six slopes are jointly significant at the 5% level in four of the five columns and at 10% in the fifth ($p = 0.053$). The case for removing a predictable component is therefore general across these constructions.

Same-date announcements are summed to one observation, as their six predictors are identical; the five announcements at which the decomposition of Jarociński and Karadi (2020) sets both components to zero are left out of both of their series. The baseline sample contains 28 multi-date months, against the 31 multi-announcement months of Section II, as three months (1991:04, 1991:09, 1991:11) hold two announcements on a single date.[37]

The sum-then-purge F exceeds its purge-then-sum counterpart by factors of 2.0 to 4.7 for the two-year Treasury yield (Figure C.1), although both orderings leave the first stage far below conventional relevance thresholds. Behind the difference lie news and the carried-forward prediction, which only the announcement-level ordering removes. Computed with pooled announcement-level coefficients and without the constant, their sum is identically zero in single-date months and averages, in absolute value, between 0.27 and 0.77 standard deviations of the instrument per multi-date month, and between 0.02 and 0.08 over all months with an announcement. As in Section IV.I, this concentrated term reaches the first stage because the affected months carry disproportionate regression weight.

[37] On date-level constructions the two orderings correlate at 0.992 in single-date months and 0.873 in multi-date months, with mean absolute gaps of 0.51 and 3.13 basis points, against 0.988, 0.843, 0.60 and 3.20 in Table A.1. For BS's series the date-level purge-then-sum first stage is 0.39, against 0.29 at the announcement level (Table II), so the factor of 4.7 in Figure C.1 compares 1.86 with 0.39. The post-9/11 announcement of September 17, 2001, for which BS record no surprise value, is excluded throughout, following their convention.

Table C.1: Predictive Regressions Across Public Surprise Series

	(1) BS `MPS`	(2) JK `pc1`	(3) JK `MP`	(4) GK `FF4`	(5) SW `FFR`
Nonfarm payrolls	0.101	0.130	0.104	0.166	1.500
	(2.517)	(2.612)	(2.647)	(3.065)	(2.688)
Employment growth (12m)	0.005	0.006	0.003	0.003	0.003
	(2.244)	(2.272)	(1.363)	(1.474)	(0.096)
Δ log S&P 500 (3m)	0.080	0.081	0.063	0.001	1.543
	(1.355)	(1.210)	(0.857)	(0.020)	(1.531)
Δ Slope (3m)	-0.011	-0.011	-0.007	-0.017	-0.189
	(-1.503)	(-1.194)	(-0.854)	(-1.637)	(-1.416)
Δ log Comm. Price (3m)	0.124	0.157	0.093	0.100	0.382
	(2.380)	(2.614)	(1.594)	(1.591)	(0.492)
Treasury skewness	0.033	0.045	0.028	0.042	0.268
	(2.984)	(2.748)	(1.520)	(1.898)	(1.587)
Constant	-0.012	-0.028	-0.017	-0.021	-0.026
	(-2.373)	(-4.592)	(-2.682)	(-3.484)	(-0.414)
R^2	0.171	0.176	0.102	0.174	0.107
Joint F	5.28	4.54	2.57	3.51	2.11
p-value	0.000	0.000	0.020	0.003	0.053
Sample	1988–19	1990–16	1990–16	1991–12	1991–19
N	315	240	240	192	241

Notes: Regressions $s_t = \alpha + \boldsymbol{\beta}'\boldsymbol{X}_t + u_t$ over announcement dates t of each paper's own high-frequency surprise s_t on the six news variables of Table IV, observed before the announcement and held fixed across columns, with the nonfarm payroll surprise on the same scale. Columns match the series of Figure C.1: BS `MPS`; Jarociński and Karadi (2020) surprise factor (`pc1`) and monetary component (`MP`); Gertler and Karadi (2015) three-month-ahead federal funds futures surprise (`FF4`); Swanson (2021) federal funds rate factor (`FFR`). Column (1) recovers Table IV, column (1) up to the announcement-date convention of this appendix. Joint F is the HC1 Wald statistic for the six slopes, referred to $F(6, N-7)$. HC1 t-statistics in parentheses.

The months in which the orderings can differ are nested: their union is exactly the 28 BS months, and seven — 1991:08, 1991:10, 1991:12, 1992:04, 2001:01, 2007:08, 2008:03 — appear in every series. The series are close substitutes, with the first principal component explaining 87% of their common variation and the BS and Jarociński and Karadi (2020) surprises correlating at 0.97, and they differ in how strongly each is predictable and how much the predictors move within the affected months. Put simply, the rows show one mechanism under several standard constructions, and the order of purging and summing matters in each.

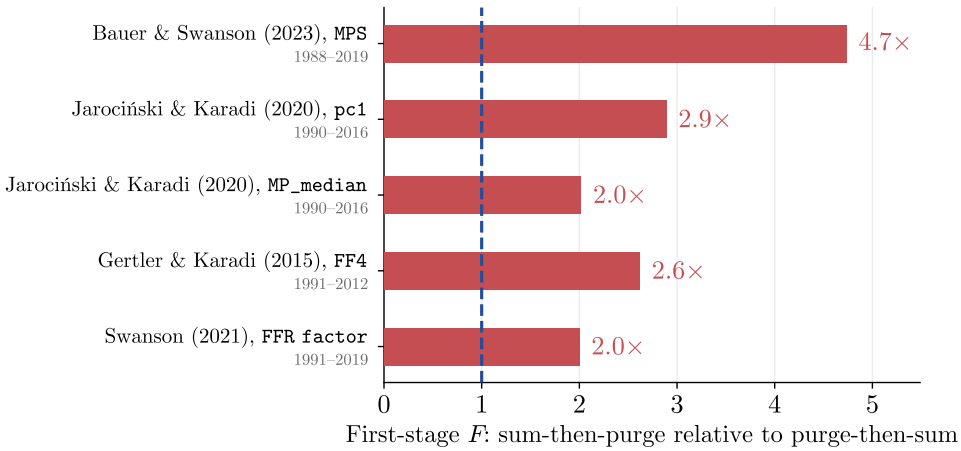


Figure C.1: Ratio of the first-stage F under sum-then-purge to that under purge-then-sum, two-year Treasury yield, in the baseline proxy-SVAR over each series' sample, with BS's six published predictors held fixed.

D. Sample extension to 2023

The longer sample strengthens all three instruments, and purge-then-sum stays the weakest. Extending the sample end from 2020:02 to 2023:12 raises the proxy-SVAR first stage from 9.59 to 18.03 for the raw surprise, from 4.01 to 9.31 for sum-then-purge and from 0.18 to 2.40 for purge-then-sum (Table D.1). The 24 announcements added after 2020 fall one per month, so the extension adds no month in which the carried-forward prediction or news can separate the two orderings. With the announcement-level records in both purges, the first stage is 1.68 for sum-then-purge and 0.18 for purge-then-sum over 1988:02–2020:02, and 5.22 and 2.40 through 2023:12.

Table D.1: First stage through 2023, BS's 2023 release

	1988:02–2020:02			1988:02–2023:12		
	VAR	2SLS	LP	VAR	2SLS	LP
Raw, MPS	9.59	7.98	5.87	18.03	13.95	11.40
Sum-then-purge, z_m	4.01	3.32	3.09	9.31	7.54	7.01
Purge-then-sum, $\tilde{z}_m$	0.18	0.16	0.03	2.40	2.06	1.38
Sum-then-purge, announcement-level records	1.68	1.35	1.14	5.22	4.23	3.66

Notes: First-stage F-statistics with current macro data; the longer sample leaves out March–December 2020 (385 and 421 months in the instrument window). VAR and 2SLS as in Table A.3, LP as in its Panel D. Last row: BS's monthly purge on the announcement-level records of the news variables, which purge-then-sum already uses.

The surprises are those of BS's 2023 release throughout, purged as BS post them, so the shorter sample differs from the paper's baseline by the vintage of the release, by two months and by the macro data; on BS's published series the proxy-SVAR first stages are 9.65, 2.47 and 0.29 (Table II). I leave March–December 2020 out of the VAR and of every first-stage regression, as BS's VAR ends in February 2020. BS leave the twelve announcements of that period out of their announcement-level purge, and their monthly purge keeps the eight months of that period with an announcement. The macro data are current official releases for 1973:01–2023:12: the consumer price index and industrial production from FRED, the excess bond premium of the Federal Reserve Board, and the last daily two-year yield of each month on the Gürkaynak et al. (2007) curve.

In every window and vintage of this appendix, and with either set of records, the raw surprise is the strongest instrument and purge-then-sum the weakest.

E. Local projections and internal instruments

As in the proxy SVAR, no response of industrial production, consumer prices or the excess bond premium to purge-then-sum is statistically distinguishable from zero in the local-projection and internal-instrument analogues of BS's Figures 5 and 6, and the first stage of purge-then-sum is too weak for these bands to carry much information (0.06 at $h = 0$ on BS's published series, Table A.3, Panel D). BS's Figure 5 cannot be replicated exactly, because their posted files exclude the Fed Chair speeches and that data has not been disclosed to the date of writing. Figure E.1 reports the FOMC-only local projection of BS's Appendix Figure A2, the closest analogue the public data allow, with the raw surprise and purge-then-sum as instruments for the two-year yield i_m in

$$y_{m+h} = \alpha_h + \theta_h \, i_m + \text{CONTROLS}_\text{m} + e_{m+h},$$

estimated by two-stage least squares, with the controls 12 lags of the two-year yield, the log of the consumer price index, the log of industrial production and the excess bond premium. Standard errors are Newey–West with h lags. The local projections of Table III and Figure II regress each interest rate directly on the instrument, with the same controls and 12 lags of that rate in place of those of the two-year yield.

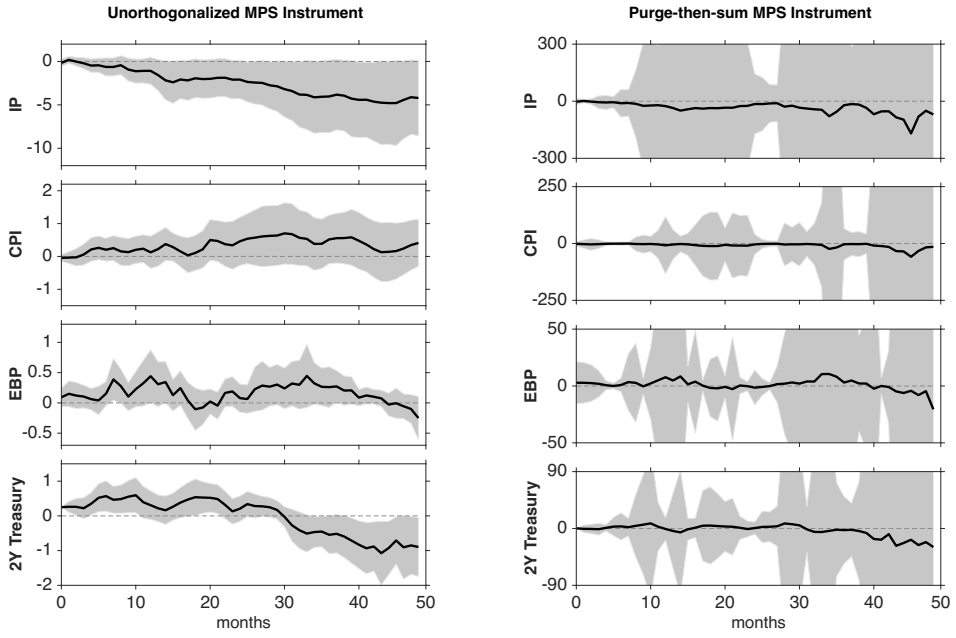


Figure E.1: Local-projection responses to a 25-basis-point rise in the two-year yield, instrumented by the raw surprise (left) and by purge-then-sum (right). Controls and standard errors as in the text. Shaded regions: 90% Newey–West bands. In the right panels the bands extend beyond the axes at some horizons from month 9 and at most horizons after month 28. Sample: 1988:02–2019:12, BS's published series. EBP = excess bond premium, CPI = consumer price index, IP = industrial production.

BS's Figure 6, the internal-instrument SVAR, also uses the speeches, and BS report no version without them. Figure E.2 reports the analogue built from BS's posted FOMC surprises. Under purge-then-sum the responses of industrial production, consumer prices and the excess bond premium are small relative to the bands.

On BS's published data the proxy SVAR reproduces the lines BS print in their Figure 3, for the raw surprise and for sum-then-purge, exactly, to within 0.0003 in the units of every panel. Part of the visible difference between the bands in the two panels of Figures IV and V comes from how the bands are built, as BS's printed 90% bands, which I could not reproduce, are narrower than both the percentile and the ± 1.645-standard-deviation bands of the 10,000-draw wild bootstrap they describe (their endnote 47). With the percentile band, the responses of industrial production and consumer prices to sum-then-purge still exclude zero over the first 29 months, and no response of industrial production, consumer prices or the excess bond premium to purge-then-sum does.

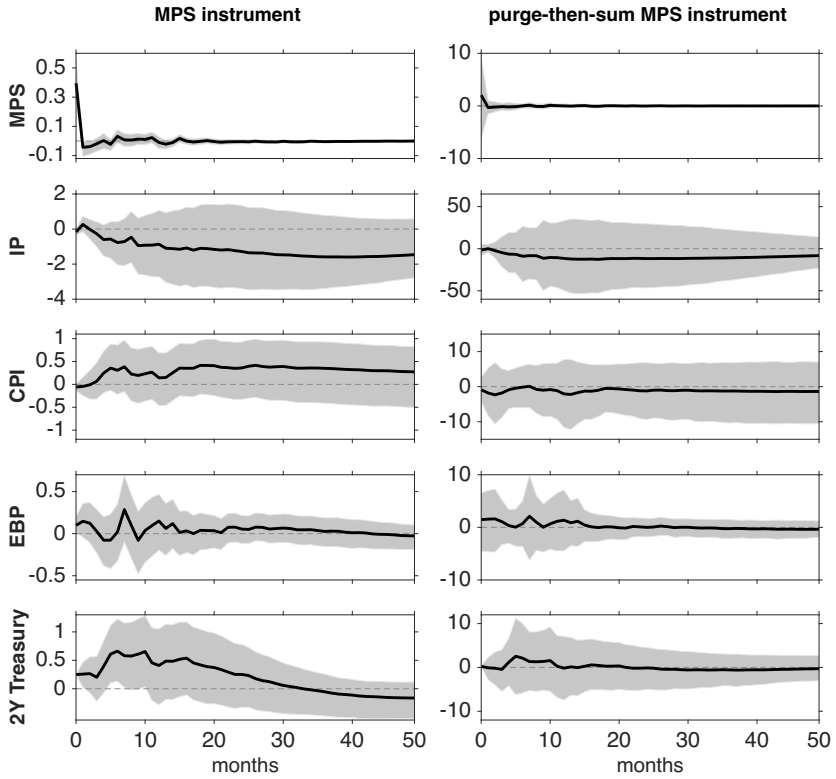


Figure E.2: Internal-instrument SVAR, the public-data analogue of BS's Figure 6, with the raw surprise (left) and purge-then-sum (right) included in the VAR and ordered first, following Plagborg-Møller and Wolf (2021). Lags: 12. Shaded regions: 90% wild-bootstrap percentile bands, 10,000 draws. Sample: 1988:02–2019:12. EBP = excess bond premium, CPI = consumer price index, IP = industrial production.

F. Proofs & Derivations

This appendix collects the algebra behind Sections III and IV. Appendix F.I proves the weighted-average identity (18). Appendices F.II and F.III take a month with two announcements, derive the gap between the two orderings, equation (8), and show how the news in that gap biases the impulse response and moves the first stage, equations (10) and (11). Appendix F.IV lets the number of announcements vary across months, as in BS's sample, so that the gap also contains the carried-forward prediction, and shows when this part moves the first stage.

F.I Proof of equation (18)

Let $\mathbf{M}$ annihilate the constant and the controls, and write $\ddot{\mathbf{y}} = \mathbf{M}\mathbf{y}$, $\ddot{\mathbf{z}}_S = \mathbf{M}\mathbf{z}^S$, $\ddot{\mathbf{z}}_M = \mathbf{M}\mathbf{z}^M$ and $\ddot{\mathbf{z}} = \ddot{\mathbf{z}}_S + \ddot{\mathbf{z}}_M$. Define the cross-products $V_S = \ddot{\mathbf{z}}_S'\ddot{\mathbf{z}}_S$, $V_M = \ddot{\mathbf{z}}_M'\ddot{\mathbf{z}}_M$, $C_{SM} = \ddot{\mathbf{z}}_S'\ddot{\mathbf{z}}_M$, $C_{Sy} = \ddot{\mathbf{z}}_S'\ddot{\mathbf{y}}$ and $C_{My} = \ddot{\mathbf{z}}_M'\ddot{\mathbf{y}}$, and let $V = V_S + 2C_{SM} + V_M$ and $\Delta = V_S V_M - C_{SM}^2$, positive unless $\ddot{\mathbf{z}}_S$ and $\ddot{\mathbf{z}}_M$ are collinear.

Restricted regression (16). By FWL, $\hat{\pi}_{R,h}$ is the OLS coefficient of $\ddot{\mathbf{y}}$ on $\ddot{\mathbf{z}}$:

$$\hat{\pi}_{R,h} \;=\; \frac{\ddot{\mathbf{z}}'\ddot{\mathbf{y}}}{\ddot{\mathbf{z}}'\ddot{\mathbf{z}}} \;=\; \frac{C_{Sy} + C_{My}}{V}. \tag{F.1}$$

Partitioned regression (17). By FWL, $(\hat{\pi}_{1,h}, \hat{\pi}_{2,h})$ are the OLS coefficients of $\ddot{\mathbf{y}}$ on $(\ddot{\mathbf{z}}_S, \ddot{\mathbf{z}}_M)$. They solve the normal equations

$$\begin{bmatrix} V_S & C_{SM} \\ C_{SM} & V_M \end{bmatrix} \begin{bmatrix} \hat{\pi}_{1,h} \\ \hat{\pi}_{2,h} \end{bmatrix} \;=\; \begin{bmatrix} C_{Sy} \\ C_{My} \end{bmatrix},$$

so that

$$\hat{\pi}_{1,h} \;=\; \frac{V_M C_{Sy} - C_{SM} C_{My}}{\Delta}, \qquad \hat{\pi}_{2,h} \;=\; \frac{V_S C_{My} - C_{SM} C_{Sy}}{\Delta}. \tag{F.2}$$

The weights. Let $\hat{\kappa}_{1,h} = (V_S + C_{SM})/V$ and $\hat{\kappa}_{2,h} = (C_{SM} + V_M)/V$, so that $\hat{\kappa}_{1,h} + \hat{\kappa}_{2,h} = 1$. By (F.2),

$$\begin{aligned} V\Delta\,(\hat{\kappa}_{1,h}\hat{\pi}_{1,h} + \hat{\kappa}_{2,h}\hat{\pi}_{2,h}) &= (V_S + C_{SM})(V_M C_{Sy} - C_{SM} C_{My}) \\ &\quad + (C_{SM} + V_M)(V_S C_{My} - C_{SM} C_{Sy}) \\ &= \Delta\,(C_{Sy} + C_{My}), \end{aligned}$$

since the coefficient of C_{Sy}, $(V_S + C_{SM})V_M - (C_{SM} + V_M)C_{SM}$, and that of C_{My}, $(C_{SM} + V_M)V_S - (V_S + C_{SM})C_{SM}$, both equal $V_S V_M - C_{SM}^2 = \Delta$. Dividing by $V\Delta$ and using (F.1),

$$\hat{\kappa}_{1,h}\hat{\pi}_{1,h} + \hat{\kappa}_{2,h}\hat{\pi}_{2,h} \;=\; \frac{C_{Sy} + C_{My}}{V} \;=\; \hat{\pi}_{R,h}. \qquad \blacksquare$$

F.II Derivation of equation (8)

Fix a month m with exactly two announcements. Work in the scalar case $k = 1$, with the news variables measured as deviations from their means. Assume the event-level predictive equation

$$\texttt{MPS}_{tm} = \beta X_{tm} + u_{tm}, \qquad t \in \{1, 2\}, \tag{F.3}$$

with the orthogonality conditions

$$\mathbb{E}[u_{tm} \mid X_{tm}] = 0, \qquad \mathrm{Cov}(u_{tm}, X_{1m}) = 0 \text{ for } t = 1, 2. \tag{F.4}$$

Both follow from equation (1), and the second is what the monthly OLS coefficient below needs. Define the purge-then-sum instrument

$$\tilde{z}_m := (\texttt{MPS}_{1m} - \beta X_{1m}) + (\texttt{MPS}_{2m} - \beta X_{2m}). \tag{F.5}$$

Define the sum-then-purge monthly surprise

$$\texttt{MPS}_m := \texttt{MPS}_{1m} + \texttt{MPS}_{2m},$$

and let z_m be the residual from the population linear projection of $\texttt{MPS}_m$ on X_{1m}:

$$\texttt{MPS}_m = \gamma X_{1m} + z_m, \qquad \gamma = \frac{\mathrm{Cov}(\texttt{MPS}_m, X_{1m})}{\mathrm{Var}(X_{1m})}, \tag{F.6}$$

assuming $\mathrm{Var}(X_{1m}) > 0$.

Step 1: purge-then-sum. Substituting (F.3) into (F.5),

$$\begin{aligned} \tilde{z}_m &= (\beta X_{1m} + u_{1m} - \beta X_{1m}) + (\beta X_{2m} + u_{2m} - \beta X_{2m}) \\ &= u_{1m} + u_{2m}. \end{aligned} \tag{F.7}$$

Hence purge-then-sum removes the predictable component event by event.

Step 2: sum-then-purge. Using (F.3),

$$\begin{aligned} \texttt{MPS}_m &= \texttt{MPS}_{1m} + \texttt{MPS}_{2m} \\ &= (\beta X_{1m} + u_{1m}) + (\beta X_{2m} + u_{2m}) \\ &= \beta X_{1m} + \beta X_{2m} + (u_{1m} + u_{2m}). \end{aligned} \tag{F.8}$$

Compute the OLS coefficient γ from (F.6),

$$\gamma = \frac{\mathrm{Cov}(\texttt{MPS}_m, X_{1m})}{\mathrm{Var}(X_{1m})}.$$

Insert (F.8):

$$\mathrm{Cov}(\texttt{MPS}_m, X_{1m}) = \mathrm{Cov}(\beta X_{1m} + \beta X_{2m} + (u_{1m} + u_{2m}), X_{1m})$$

By (F.4), $\mathrm{Cov}(u_{1m} + u_{2m}, X_{1m}) = 0$, so

$$\gamma = \beta + \beta \frac{\mathrm{Cov}(X_{2m}, X_{1m})}{\mathrm{Var}(X_{1m})}.$$

Define

$$\rho := \frac{\mathrm{Cov}(X_{2m}, X_{1m})}{\mathrm{Var}(X_{1m})}.$$

Then

$$\gamma = \beta + \beta\rho = \beta(1 + \rho).$$

Interpretation of ρ. The quantity ρ is the slope coefficient in the population linear projection of X_{2m} on X_{1m}:

$$\Pi(X_{2m} \mid X_{1m}) = \rho X_{1m}.$$

Therefore $X_{2m} - \rho X_{1m}$ is the projection residual, i.e. the component of second-announcement information not linearly spanned by first-announcement information.

Step 3: substitute γ back into z_m. From (F.6) and (F.8),

$$\begin{aligned} z_m &= \texttt{MPS}_m - \gamma X_{1m} \\ &= [\beta X_{1m} + \beta X_{2m} + (u_{1m} + u_{2m})] - (\beta + \beta\rho) X_{1m}. \end{aligned}$$

Now distribute and cancel the common term βX_{1m}:

$$\begin{aligned} z_m &= (u_{1m} + u_{2m}) + \beta X_{2m} - \beta\rho X_{1m} \\ &= (u_{1m} + u_{2m}) + \beta(X_{2m} - \rho X_{1m}). \end{aligned}$$

Using (F.7),

$$z_m = \tilde{z}_m + \beta(X_{2m} - \rho X_{1m}). \tag{F.9}$$

What exactly remains in z_m? By (F.9):

- if $\beta = 0$, there is no event-level predictability and the two instruments coincide;
- if $X_{2m} = \rho X_{1m}$ almost surely, then second-announcement information is fully spanned by first-announcement information and the two instruments coincide;
- otherwise $z_m \neq \tilde{z}_m$, and z_m retains the unpurged component $\beta(X_{2m} - \rho X_{1m})$.

F.III Derivation of the bias equation

This subsection derives the probability limits in equations (10) and (11) under the convention of BS's endnote 41: ε_m is a latent structural shock, and $z_m, \tilde{z}_m$ are candidate instruments for ε_m. In the model of Section III, $\varepsilon_m = \sum_{t=1}^{T_m} u_{tm}$, so the clean instrument $\tilde{z}_m$ equals the shock. The impulse response is the instrument's covariance with the outcome divided by its covariance with v_m, the residual of the two-year yield.[38]

All objects are defined in Section III.I: $k = 1$ and two announcements per month; here the news variables are measured as deviations from their means. There the corrected purge-then-sum instrument is $\tilde{z}_m = u_{1m} + u_{2m}$, and by (8) the sum-then-purge instrument is $z_m = \tilde{z}_m + \beta\,\eta_m$, where

$$\eta_m \;:=\; X_{2m} - \rho\, X_{1m}$$

is the innovation in the second-announcement news beyond what is linearly spanned by the first.[39] The outcome at horizon h and the yield follow

$$\begin{aligned} Y_{m+h} &= \Theta_h\,\varepsilon_m + \Psi_{1h}\,X_{1m} + \Psi_{2h}\,X_{2m} + \xi_{m+h}, \\ v_m &= \varepsilon_m + \Psi_1^v\,X_{1m} + \Psi_2^v\,X_{2m} + \xi_m^v, \end{aligned} \tag{F.10}$$

with ξ_{m+h} and ξ_m^v uncorrelated with both instruments.

The numerator: $\mathrm{Cov}(Y_{m+h},\, z_m)$. Substituting (F.10):

$$\mathrm{Cov}(Y_{m+h},\, z_m) = \Theta_h\,\mathrm{Cov}(\varepsilon_m, z_m) + \Psi_{1h}\,\mathrm{Cov}(X_{1m}, z_m) + \Psi_{2h}\,\mathrm{Cov}(X_{2m}, z_m).$$

First, $\mathrm{Cov}(X_{1m}, z_m) = 0$, as z_m is the OLS residual from projecting $\sum_t \texttt{MPS}_{tm}$ on X_{1m}. Second, writing $X_{2m} = \rho\, X_{1m} + \eta_m$ and substituting $z_m = \tilde{z}_m + \beta\eta_m$:

$$\mathrm{Cov}(X_{2m}, z_m) = \underbrace{\mathrm{Cov}(X_{2m}, \tilde{z}_m)}_{=0} + \beta\,\underbrace{\mathrm{Cov}(X_{2m}, \eta_m)}_{=\mathrm{Var}(\eta_m)} = \beta\,\mathrm{Var}(\eta_m),$$

where $\mathrm{Cov}(X_{2m}, \eta_m) = \mathrm{Var}(\eta_m)$ because η_m is the projection residual of X_{2m} on X_{1m}. This is the source of the bias: z_m is correlated with later-announcement news. Third, since $\tilde{z}_m = \varepsilon_m$:

$$\mathrm{Cov}(\varepsilon_m, z_m) = \underbrace{\mathrm{Cov}(\varepsilon_m, \tilde{z}_m)}_{=\mathrm{Var}(\varepsilon_m)} + \beta\,\underbrace{\mathrm{Cov}(\varepsilon_m, \eta_m)}_{=0} = \mathrm{Var}(\varepsilon_m).$$

[38] In the proxy-SVAR notation, the structural impact vector is $\widehat{\Sigma}_{u,z}/\widehat{\Sigma}_{u^{2y},z}$, and the ratio $\mathrm{Cov}(Y,z)/\mathrm{Cov}(v,z)$ is its horizon-h generalization; see Bauer and Swanson (2023a), eq. (22) and endnote 46.

[39] Throughout I maintain that each event-level residual u_{tm} is uncorrelated with news realized at other announcements, $\mathrm{Cov}(u_{tm}, X_{sm}) = 0$ for $s \neq t$. Equation (1) implies it for $s < t$. For $s > t$ it requires that the news variables of later announcements do not react to u_{tm}, which is plausible for the payroll variables and strong for asset prices. Appendix B estimates that reaction and the piece of the gap it creates.

Collecting terms:

$$\mathrm{Cov}(Y_{m+h},\, z_m) = \Theta_h\ \mathrm{Var}(\varepsilon_m) + \Psi_{2h}\,\beta\ \mathrm{Var}(\eta_m). \tag{F.11}$$

For the clean instrument $\tilde{z}_m$, the Ψ_{2h} term vanishes: $\mathrm{Cov}(Y_{m+h}, \tilde{z}_m) = \Theta_h\ \mathrm{Var}(\varepsilon_m)$.

The denominator and the main result. Substituting the yield equation in (F.10) and using the same three covariances,

$$\mathrm{Cov}(v_m, \tilde{z}_m) = \mathrm{Var}(\varepsilon_m), \qquad \mathrm{Cov}(v_m, z_m) = \mathrm{Var}(\varepsilon_m) + \Psi_2^v\,\beta\ \mathrm{Var}(\eta_m), \tag{F.12}$$

which is equation (11). Dividing (F.11) by (F.12),

$$\begin{aligned}
\widehat{\Theta}_h^{(z)} &:= \frac{\mathrm{Cov}(Y_{m+h},\, z_m)}{\mathrm{Cov}(v_m,\, z_m)} \\
&= \frac{\Theta_h\,\mathrm{Var}(\varepsilon_m) + \Psi_{2h}\,\beta\,\mathrm{Var}(\eta_m)}{\mathrm{Var}(\varepsilon_m) + \Psi_2^v\,\beta\,\mathrm{Var}(\eta_m)} \\
&= \Theta_h + \beta\,(\Psi_{2h} - \Psi_2^v\Theta_h)\frac{\mathrm{Var}(\eta_m)}{\mathrm{Cov}(v_m, z_m)},
\end{aligned} \tag{F.13}$$

where the last step adds and subtracts $\Psi_2^v\,\beta\,\Theta_h\,\mathrm{Var}(\eta_m)$ in the numerator. This is equation (10). With $\Psi_2^v \neq 0$ and $\omega := \mathrm{Cov}(v_m, \tilde{z}_m)/\mathrm{Cov}(v_m, z_m)$, the middle expression equals $\omega\,\Theta_h + (1-\omega)\,\Psi_{2h}/\Psi_2^v$. For the clean instrument, the same steps give $\widehat{\Theta}_h^{(\tilde{z})} = \Theta_h$. At $\Psi_2^v = 0$ the denominator is $\mathrm{Cov}(\varepsilon_m, z_m)$, and (F.13) reduces to $\Theta_h + \Psi_{2h}\,\beta\ \mathrm{Var}(\eta_m)\,/\mathrm{Var}(\varepsilon_m)$.

First-stage strength. Let r^2 denote the squared correlation between an instrument and v_m. The homoskedastic first-stage statistic, $F = (n-2)\,r^2/(1-r^2)$, increases with r^2. Since $\mathrm{Var}(z_m) = \mathrm{Var}(\varepsilon_m) + \beta^2\,\mathrm{Var}(\eta_m)$ and $\mathrm{Var}(\tilde{z}_m) = \mathrm{Var}(\varepsilon_m)$,

$$\frac{r^2(z_m)}{r^2(\tilde{z}_m)} = \frac{\mathrm{Cov}(v_m, z_m)^2/\mathrm{Var}(z_m)}{\mathrm{Cov}(v_m, \tilde{z}_m)^2/\mathrm{Var}(\tilde{z}_m)} = \frac{(1+s)^2}{1+q}, \tag{F.14}$$

with $s := \Psi_2^v\,\beta\,\mathrm{Var}(\eta_m)/\mathrm{Var}(\varepsilon_m)$ and $q := \beta^2\,\mathrm{Var}(\eta_m)/\mathrm{Var}(\varepsilon_m)$. Hence, when $\mathrm{Cov}(v_m, z_m) > 0$, z_m is the stronger instrument if and only if $s > \sqrt{1+q} - 1$, that is, for $\beta > 0$, $\Psi_2^v > \beta/(1+\sqrt{1+q})$, a threshold below $\beta/2$ that tends to $\beta/2$ as q shrinks. At $\Psi_2^v = 0$ the ratio is $1/(1+q) < 1$. Both instruments have covariance $\mathrm{Var}(\varepsilon_m)$ with the shock, so the gain is covariance with news.

F.IV The gap with any number of announcements

This subsection derives the gap between the two orderings when the number of announcements varies across months and BS's monthly regression uses X_{1m}. Equation (8) is its case with two announcements in every month. Moments are taken

over the months with at least one announcement, the sample of BS's monthly regression.

Example. With $k = 1$, let the news variables keep their value within the month, $X_{tm} = X_{1m}$, so that the news term of (4) is zero, and let the number of announcements be unrelated to them, with BS's mean of 1.21 announcements per month (322 in 266 months). Purge-then-sum removes βX_{1m} once per announcement. The monthly regression of the summed surprise on X_{1m} removes the average, $1.21\,\beta X_{1m}$, in every month. A month with three announcements thus keeps $1.79\,\beta X_{1m}$ in z_m, and a month with one announcement removes $0.21\,\beta X_{1m}$ too much. The gap is the carried-forward prediction net of BS's regressor, $\beta(T_m - 1.21)X_{1m}$. Its factor $T_m - 1.21$ has mean zero and is unrelated to the news, so the gap adds variance to z_m and no covariance with the yield or the outcome, unless their response to X_{1m} varies with T_m.

Step 1: the identity. Summing equation (1) over the announcements of month m gives $\sum_{t=1}^{T_m} \texttt{MPS}_{tm} = \boldsymbol{\beta}' \sum_{t=1}^{T_m} \boldsymbol{X}_{tm} + \tilde{z}_m$. Substituting this sum in equation (3), subtracting $\tilde{z}_m$ and writing each $\boldsymbol{X}_{tm}$ as $\boldsymbol{X}_{1m} + (\boldsymbol{X}_{tm} - \boldsymbol{X}_{1m})$,

$$\begin{aligned} z_m - \tilde{z}_m &= \boldsymbol{\beta}' \sum_{t=1}^{T_m} \boldsymbol{X}_{tm} - \boldsymbol{\gamma}' \boldsymbol{X}_{1m} \\ &= (T_m - 1)\,\boldsymbol{\beta}' \boldsymbol{X}_{1m} + \boldsymbol{\beta}' \sum_{t=2}^{T_m} (\boldsymbol{X}_{tm} - \boldsymbol{X}_{1m}) + \boldsymbol{\delta}' \boldsymbol{X}_{1m}, \end{aligned} \tag{F.15}$$

which is equation (4). The identity uses only the definitions, so it holds month by month for any $\boldsymbol{\gamma}$, the OLS estimate included.

Step 2: the monthly coefficient. From here on $k = 1$ and the news variables have mean zero. The number of announcements T_m is independent of the news variables, the purged surprises and the residuals ξ_{m+h} and ξ_m^v jointly, all defined in every month, including announcements beyond T_m, which go unobserved. For $t \geq 2$, define

$$\rho_t := \frac{\mathrm{Cov}(X_{tm}, X_{1m})}{\mathrm{Var}(X_{1m})}, \qquad \eta_{tm} := X_{tm} - \rho_t\, X_{1m},$$

the slope and the residual of the projection of X_{tm} on X_{1m}, so that $\rho_2 = \rho$ and $\eta_{2m} = \eta_m$. Writing the month's summed surprise as $\sum_{t\geq 1} \mathbf{1}\{T_m \geq t\}\, \texttt{MPS}_{tm}$,

$$\begin{aligned}
\operatorname{Cov}\Big(\sum_{t=1}^{T_m} \texttt{MPS}_{tm},\, X_{1m}\Big) &= \sum_{t\geq 1} \operatorname{Cov}(\mathbf{1}\{T_m \geq t\}\texttt{MPS}_{tm},\, X_{1m}) \\
&= \sum_{t\geq 1} \Pr(T_m \geq t)\operatorname{Cov}(\texttt{MPS}_{tm}, X_{1m}) \\
&= \beta \sum_{t\geq 1} \Pr(T_m \geq t)\operatorname{Cov}(X_{tm}, X_{1m}) \\
&= \beta\,(1+\bar{\varrho})\operatorname{Var}(X_{1m}),
\end{aligned}$$

where $\varrho_m := \rho_2 + \cdots + \rho_{T_m}$, with $\varrho_m = 0$ when $T_m = 1$, and $\bar{\varrho} := \mathbb{E}[\varrho_m] = \sum_{t\geq 2} \Pr(T_m \geq t)\,\rho_t$. The second line uses the independence of T_m from the news and the purged surprises and $\mathbb{E}[X_{1m}] = 0$, the third $\operatorname{Cov}(u_{tm}, X_{1m}) = 0$ for every t by equation (1), and the last $\Pr(T_m \geq 1) = 1$ and $\operatorname{Cov}(X_{tm}, X_{1m}) = \rho_t \operatorname{Var}(X_{1m})$. Hence

$$\gamma = \beta\,(1+\bar{\varrho}), \qquad \delta = -\beta\bar{\varrho}. \tag{F.16}$$

A month with T_m announcements thus loads $\beta(1+\varrho_m)$ on X_{1m}, and BS's monthly regression removes $\beta(1+\bar{\varrho})X_{1m}$ in every month, too much where $\varrho_m < \bar{\varrho}$ and too little where $\varrho_m > \bar{\varrho}$. In plain words, one coefficient serves all months, so the regression removes from each month what X_{1m} predicts for a month with the average number of announcements. With persistent news, $\rho_t > 0$, a month with more announcements than the average keeps part of its predictable surprise in z_m, and a month with fewer has too much removed.

Step 3: the gap. Substituting $X_{tm} = \rho_t X_{1m} + \eta_{tm}$ in the first line of (F.15) and γ from (F.16),

$$\begin{aligned}
z_m - \tilde{z}_m &= \beta\,(1+\varrho_m)X_{1m} + \beta \sum_{t=2}^{T_m} \eta_{tm} - \beta\,(1+\bar{\varrho})X_{1m} \\
&= \underbrace{\beta\,(\varrho_m - \bar{\varrho})\,X_{1m}}_{\rho_t\text{-weighted carried-forward, net}} + \underbrace{\beta \sum_{t=2}^{T_m} \eta_{tm}}_{\text{unpredicted news}}.
\end{aligned} \tag{F.17}$$

The first term is the carried-forward prediction with each later announcement weighted by ρ_t, net of X_{1m}, and the second is the news that X_{1m} does not predict; the remainder $\delta X_{1m} = -\beta\bar{\varrho}X_{1m}$ removes the average loading on X_{1m} of the carried-forward prediction and the news of (4). When the news since the first annoucement is uncorrelated with X_{1m}, $\rho_t = 1$, the first term is the net carried-forward prediction $\beta(D_m - p)X_{1m}$, with $D_m := T_m - 1$ and $p := \mathbb{E}[D_m]$, and the second is the news of (4). The example is this case without news.

Special cases. Two cases connect equation (F.17) to the rest of the paper. The first recovers equation (8), which assumes two announcements in every month, and the second gives the closed forms used in Steps 4 and 5. With two announcements in every month, $\varrho_m = \bar{\varrho} = \rho$, so $\gamma = \beta(1+\rho)$, the first term of (F.17) is zero and the second is $\beta\eta_m$, which is equation (8). The first term is zero whenever every month has the same number of announcements. With one or two announcements, $\varrho_m = \rho D_m$ and $\bar{\varrho} = \rho p$, so $\gamma = \beta(1+p\rho)$ and

$$z_m - \tilde{z}_m = \underbrace{\beta\rho\,(D_m - p)\,X_{1m}}_{\rho\,\times\,\text{carried-forward, net}} + \underbrace{D_m\,\beta\,\eta_m}_{\text{unpredicted news}} .$$

Step 4: covariances. Only the news in the gap moves the covariances of z_m with the yield and the outcome. Let both follow (F.10) in every month, so that X_{2m} enters them also in months with one announcement, with ξ_{m+h} and ξ^v_m uncorrelated with the news variables and the purged surprises. As in Appendix F.III, each u_{tm} is uncorrelated with the news variables of every announcement.

The first term of (F.17) multiplies X_{1m} by $\beta(\varrho_m - \bar{\varrho})$, which has mean zero and depends only on T_m. It is therefore uncorrelated with X_{1m}, X_{2m}, ξ_{m+h} and ξ^v_m, and, as $\mathbb{E}[X_{1m}u_{tm}] = \mathbb{E}[X_{1m}\eta_{tm}] = 0$ for every t, also with $\varepsilon_m = \sum_{t=1}^{T_m} u_{tm}$ and with the news term, so it adds nothing to either covariance.

The news term is uncorrelated with ε_m, X_{1m} and the residuals ξ_{m+h} and ξ^v_m, and has covariance $\beta\Lambda$ with $X_{2m} = \rho X_{1m} + \eta_m$, where

$$\Lambda := \operatorname{Cov}\Big(\eta_m, \sum_{t=2}^{T_m} \eta_{tm}\Big) = \sum_{t\geq 2} \Pr(T_m \geq t)\operatorname{Cov}(\eta_m, \eta_{tm}).$$

It thus adds $\Psi^v_2\beta\Lambda$ to $\operatorname{Cov}(v_m, z_m)$ and $\Psi_{2h}\beta\Lambda$ to $\operatorname{Cov}(Y_{m+h}, z_m)$, so equations (10) and (11) hold with Λ in place of $\operatorname{Var}(\eta_m)$, and $\Lambda = p\operatorname{Var}(\eta_m)$ with one or two announcements.

Step 5: variance and first-stage strength. Under (F.10) the carried-forward prediction adds variance to z_m and nothing to its covariances, so it weakens the instrument. As ε_m and the two terms of (F.17) are uncorrelated (Step 4), the variance of z_m is

$$\operatorname{Var}(z_m) = \operatorname{Var}(\varepsilon_m) + \beta^2 \operatorname{Var}\Big(\sum_{t=2}^{T_m} \eta_{tm}\Big) + \beta^2 \operatorname{Var}(\varrho_m)\operatorname{Var}(X_{1m}), \tag{F.18}$$

which with one or two announcements reads $\operatorname{Var}(\varepsilon_m) + p\beta^2\operatorname{Var}(\eta_m) + p(1-p)\beta^2\rho^2\operatorname{Var}(X_{1m})$. The ratio (F.14) becomes $(1+s')^2/(1+q')$, where s' is s with Λ in place of $\operatorname{Var}(\eta_m)$ and q' is q with the last two terms of (F.18) in place of

$\beta^2 \operatorname{Var}(\eta_m)$. When $\operatorname{Cov}(v_m, z_m) > 0$ and $\beta\Lambda > 0$, z_m is therefore the stronger instrument if and only if $\Psi_2^v > \operatorname{Var}(\varepsilon_m)(\sqrt{1+q'}-1)/(\beta\Lambda)$. The first term of (F.17) enters only q', so it raises this threshold.

When the carried-forward prediction moves the first stage. In BS's sample the net carried-forward prediction supplies 55% of the gap's covariance with v_m, and February 1989 and December 1991 supply more than all of it (Appendix A.I). In population, with T_m independent of the news, the first term of (F.17) covaries with the yield only if the slope of v_m on X_{1m} varies with the number of announcements:

$$\operatorname{Cov}(v_m,\, \beta(\varrho_m - \bar{\varrho})X_{1m}) = \beta \operatorname{Var}(X_{1m}) \operatorname{Cov}(\varrho_m, b_m),$$

where $b_m := \mathbb{E}[v_m X_{1m} \mid T_m]/\operatorname{Var}(X_{1m})$ is the slope of v_m on X_{1m} among months with T_m announcements, and the same holds for Y_{m+h}. Under (F.10), $b_m = \Psi_1^v + \rho\Psi_2^v$ in every month and the covariance is zero. If instead $v_m = \varepsilon_m + \Psi^v \sum_{t=1}^{T_m} X_{tm} + \xi_m^v$, so that the yield responds to the news before every announcement of the month, then $b_m = \Psi^v(1+\varrho_m)$, and the first term adds $\beta\Psi^v \operatorname{Var}(\varrho_m) \operatorname{Var}(X_{1m})$ to $\operatorname{Cov}(v_m, z_m)$, positive when $\beta\Psi^v > 0$. In both cases the term is uncorrelated with ε_m (Step 4), so none of the first-stage strength it supplies in population comes from the shock.

G. Departures from BS's implementation

The surprises, the news variables and the VAR are BS's throughout, and Table G.1 lists each departure from their implementation that changes an exhibit, the exhibits concerned and the reason for it. BS's own statistic on their posted series gives the same first-stage result: 1.98 for sum-then-purge, 0.25 for purge-then-sum.

Table G.1: Departures from BS's implementation

Exercise	BS	This paper	Reason
Monthly instrument, all exhibits	Monthly purge on the monthly file's records of twelve-month payroll growth, the yield-curve slope and Treasury skewness	BS's posted series; purge-then-sum sums their posted announcement-level residuals (Appendix A.I)	With both purges on the same records, the ordering alone lowers the first stage from 1.86 to 0.29 (Table A.3, Panel C)
First stage, Table II and Appendix A	2SLS with HC0 standard errors (endnote 52): 8.19 raw, 1.83 sum-then-purge	Proxy-SVAR first stage, homoskedastic: 9.65, 2.47, 0.29	The regression behind BS's Figure 3 responses; Appendix A.II also reports BS's statistic, whose 1.83 comes from their April 2022 release
Local projections, Table III and Figure II	LP-IV of macro outcomes on the two-year yield (eq. 23), 1988:01–2020:02	Each rate on the instrument and its own lags, 1988:02–2019:12; one-year yield: FRED's monthly average	Relevance at each horizon (equation (16)) on the months of BS's published surprises; Figure E.1 reports BS's FOMC-only version
Responses with speeches, Figure VI and Appendix E	Fed Chair speeches in Figures 4–6 and 8 and Table 3, columns 3–4	FOMC announcements only	BS do not post the speech surprises
Proxy-SVAR bands, Figures IV–VI and A.4	90% standard-error bands from a wild bootstrap, 10,000 draws (endnote 47)	90% percentile bands from the same bootstrap	Percentile bands follow the skew of the responses under a weak first stage. BS's printed bands are narrower than both bands of their bootstrap (Appendix E)

Notes: Exercises BS do not run (the federal funds rate, identification-robust sets, Appendices C and D) are described where they appear.