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

Francesco Amodeo*

UC San Diego

July 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. A simple model shows that the fix is order-dependent: orthogonalization and time aggregation do not commute, so an instrument that sums surprises first and purges the total afterwards retains a predictable component. Instead, purging each surprise against its own pre-announcement information, and aggregating afterwards, removes it. Applied to the orthogonalized series of Bauer and Swanson (2023a), the corrected ordering collapses the instrument's strength, leaving no statistically detectable macroeconomic response. The data confirm the mechanism: almost all of the original instrument's power comes from the 8 percent of months containing more than one announcement, exactly where the two orderings diverge.

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

---

*famodeo@ucsd.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 do 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) summoned an unscheduled meeting on October 8 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. 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 cannot use the predictors observed before the October 29 decision, leaving their predictive power unscathed. 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 8.0 basis points, compared with a median absolute surprise of 2.4 basis points in the sample, 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. Separately, the model decomposes the observed proxy gap into a direct later-announcement fitted-value component and a projection-coefficient component. 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 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

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

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 10.19 for the raw surprise, 3.62 for the sum-then-purge series, and 0.15 for purge-then-sum. Analogous reductions obtain for the federal funds rate and the one-year Treasury yield. With an F-statistic of 0.15, 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 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 385 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 nonrepresentative 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, as the Global Financial Crisis, the onset of the Covid pandemic 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 0.98 to 0.80, and the mean absolute gap rising from 0.89 to 3.82 basis points. Second, later-announcement predictors remain jointly significant for the orthogonalized sum-then-purge series in those multi-announcement months, validating them as the origin of the divergence. Third, the model's analytical decomposition shows that most of the gap is later-announcement information that a first-announcement purge cannot remove,

[2] Data available at the Federal Reserve Bank of San Francisco's monetary policy surprises data page, linked here .

with a smaller part coming from the different projection coefficients.

These findings indicate that multi-announcement months are the origin of both 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 a (orthogonalized) one basis point surprise is 0.02 in single-announcement months and 1.96 in multi-announcement months, compared with a pooled response of 0.56. Multi-announcement months receive a pooled-regression weight of 28 percent despite representing only 8.1 percent of the sample; because their estimated slope is much larger, their weighted contribution accounts for nearly all of the 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 two to five 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 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 rather than information effects, and Cieslak and Schrimpf (2019) show more broadly that central-bank communication often conveys non-monetary news about growth and risk premia. 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. 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 Swanson (2023a) (*BS* henceforth), comprehensively reassessing the impulse responses and discussing the lessons that emerge. Section V concludes.[3]

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

[3]Throughout, for ease of comparison, I use Arabic numerals for BS sections/displays, and Roman numerals for this paper.

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.

Let months be indexed by $m = 1, \dots, M$ and within-month announcements by $t = 1, \dots, T_m$, where T_m denotes the number of monetary policy announcements in month m.[4] Let $\boldsymbol{X}_{t^- m} \in \mathbb{R}^k$ be the vector of news variables observed just before announcement (t, m), and define the contemporaneous information set $\Omega_{tm} = \{\boldsymbol{X}_{t^- m}, \boldsymbol{X}_{1m-1}, \dots\}$.

Consider a simple scenario of a month with two monetary events. An early announcement, observed under information set X_{1m}, and a later announcement, observed under information set 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 Canonical case: October 2008

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

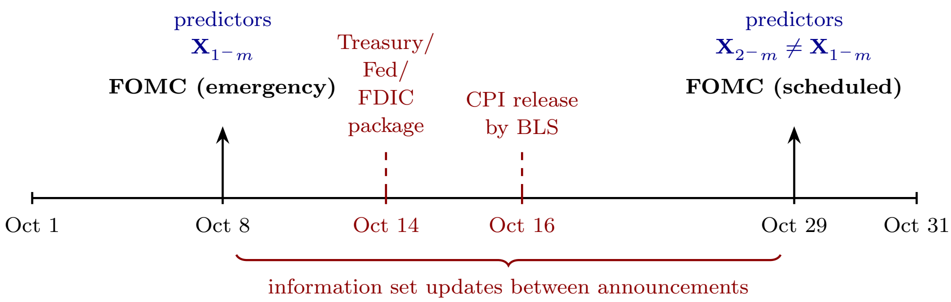


Figure I: October 2008, stylized timeline.

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

The Fed held an unscheduled emergency meeting on October 8 following the developments on the unfolding financial crisis, and reconvened at the scheduled meeting on October 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 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).[5]

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 (≈ 3.0 versus ≈ -3.7 basis points). Appendix A shows that the largest gaps between the two ordering procedures line up exactly with this type of month.

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 start of this section.[6] Table I lists all 31 multi-announcement months compactly, providing some historical context.

[5] For scale, the median surprise in the sample is 2.4 basis points.

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

Table I: Multi-announcement months in the baseline sample

Month	# Ann.	# Unsched.	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; a borrowing-target reset lifts funds to 8.375%.
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. After that, the Fed began announcing policy changes and subsequently expanded this into a statement after every FOMC meeting.

These episodes do not constitute a representative sample of monetary policy conduct. 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 as the Global Financial Crisis, the onset of the Covid pandemic 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. 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 to the Bauer and Swanson (2023a) results.

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.[7] Denote the raw high-frequency surprise around announcement (t, m) by $\texttt{MPS}_{tm} \in \mathbb{R}$. Let $\boldsymbol{X}_{t^-m} \in \mathbb{R}^k$ be the vector of news variables observed just before that announcement and define the contemporaneous information set as $\Omega_{tm} = \{\boldsymbol{X}_{t^-m}, \boldsymbol{X}_{1m-1}, \ldots\}$.

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.[8] BS present a compelling case that monetary surprises 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 exactly in Section IV.II), they recommend orthogonalization of monetary surprises.

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

[8] For simplicity of exposition, I adopt BS notation, but clarify that x_{tm} and $\boldsymbol{X}_{t^-m}$ 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").

Then, I specify

$$\texttt{MPS}_{tm} = \boldsymbol{\beta}'\boldsymbol{X}_{t^-m} + u_{tm}, \tag{1}$$

where u_{tm} denotes the residual from the event-level predictability regression.[9] Let us see what the two orthogonalization protocols imply. The purge-then-sum uses:

$$\tilde{z}_m := \sum_{t=1}^{T_m} u_{tm}, \qquad \mathbb{E}[\tilde{z}_m \mid \Omega_{m-1}] = 0, \tag{2}$$

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

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

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

$$z_m = \boldsymbol{\delta}'\boldsymbol{X}_{1^-m} \ + \ \boldsymbol{\beta}'\sum_{t=2}^{T_m}\boldsymbol{X}_{t^-m} \ + \ \tilde{z}_m \tag{4}$$

where (4) embeds $\tilde{z}_m$ and some combination of the predictable relationship between $\texttt{MPS}_{tm}$ and the news variables.

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 ($\boldsymbol{X}_{2^-m}, \dots \boldsymbol{X}_{T_m^-m}$) and potentially also contaminating the purging of the first announcement. In fact, the two approaches coincide only if the first announcement news vector fully spans every later-announcement information, leaving no extra predictable component to purge.

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:

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

Purge-then-sum delivers

$$\tilde{z}_m = (\texttt{MPS}_{1m} - \beta\, X_{1^-m}) + (\texttt{MPS}_{2m} - \beta\, X_{2^-m}) = u_{1m} + u_{2m}, \quad \mathbb{E}[\tilde{z}_m \mid \Omega_{m-1}] = 0 \tag{6}$$

Sum-then-purge delivers

$$z_m = \texttt{MPS}_m - \gamma\, X_{1^-m} \qquad \mathbb{E}[z_m \mid \Omega_{m-1}] = 0 \tag{7}$$

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

$$\text{with} \quad \texttt{MPS}_m = \beta\, X_{1-m} + \beta\, X_{2-m} + (u_{1m} + u_{2m}),$$

where OLS estimates

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

Note that for either approach $\mathbb{E}[z_m, \tilde{z}_m \mid \Omega_{m-1}] = 0$, making them both *exogenous* monthly instruments. However, plugging γ in z_m yields

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

so z_m retains a predictable component unless $\beta = 0$ (i.e., no predictability to start with) or $X_{2-m} \equiv \rho\, X_{1-m}$ (i.e., the second-announcement information is fully spanned by first-announcement), both of which the data reject.[10,11]

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:[12]

$$Y_{m+h} = \Theta_h\, \varepsilon_m + \Psi_{1h}\, X_{1-m} + \Psi_{2h}\, X_{2-m} + \xi_{m+h}, \qquad \mathbb{E}[\xi_{m+h} \mid \Omega_{m-1}] = 0 \tag{9}$$

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. Both news variables may influence the outcome. Because the contamination operates at the level of the instrument rather than the second-stage specification, 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:

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

[10] For the application: see Section IV.II for the rejection of $\boldsymbol{\beta} = 0$, and Section IV.I for the rejection of the spanning condition.

[11] Full derivation in Appendix H.

[12] 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}.

Equation (10) says that the sum-then-purge IV estimator recovers the true impulse response Θ_h plus a contamination term.[13] The bias originates from the extra term $\beta\,(X_{2-m} - \rho\,X_{1-m})$ 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.

When β and Ψ_{2h} share the same sign, the bias term is positive: where good economic news both predicts tighter policy ($\beta > 0$) and is associated with stronger output ($\Psi_{2h} > 0$), this 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 ε_m: when later-announcement news is largely independent of first-announcement news ($\mathrm{Var}[X_{2-m} - \rho\,X_{1-m}]$ is sizable relative to $\mathrm{Cov}(\varepsilon_m, z_m)$), the contamination is large; in the knife-edge case when X_{2-m} contains no new information beyond X_{1-m}, the bias vanishes.

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 meetings. This provides a simple explanation for the empirical results next in Section IV: once $\tilde{z}_m$ is used, the bias term vanishes and the resulting impulse response functions are significantly impacted, flattening to statistical insignificance.

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 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.

[13] Full derivation in Appendix H.III.

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 emphasizes the importance of orthogonalizing the surprises. For instance, on page 104: *"[...] 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.

This section asks 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.[14] 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 BS 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 disproportionally 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. Define the partition

$$z_m^S = z_m \cdot \mathbf{1}\{T_m = 1\}, \tag{11}$$

[14] The same logic propagates to any of the other empirical techniques considered in BS.

$$z_m^M = z_m \cdot \mathbf{1}\{T_m \geq 2\}, \tag{12}$$

where T_m is the number of FOMC-related announcements in month m, so that 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).

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

$$y_{m+h} = \alpha_h + \beta_{R,h}\, z_m + \mathbf{X}_m' \gamma_h + u_{m+h}, \tag{13}$$

and then the partitioned specification

$$y_{m+h} = \alpha_h + \beta_{1,h}\, z_m^S + \beta_{2,h}\, z_m^M + \mathbf{X}_m' \gamma_h + u_{m+h}. \tag{14}$$

The same controls $\mathbf{X}_m$ appear in both 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,

$$\hat{\beta}_{R,h} = \hat{\omega}_{1,h}\hat{\beta}_{1,h} + \hat{\omega}_{2,h}\hat{\beta}_{2,h}, \qquad \hat{\omega}_{1,h} + \hat{\omega}_{2,h} = 1, \tag{15}$$

where the sample weights $\hat{\omega}_{1,h}$ and $\hat{\omega}_{2,h}$ 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 $(\hat{\beta}_{1,h}, \hat{\beta}_{2,h})$ asks how much the outcome moves per basis-point surprise in single- and multi-announcement months; second, the products $\hat{\omega}_{1,h}\hat{\beta}_{1,h}$ and $\hat{\omega}_{2,h}\hat{\beta}_{2,h}$ ask how much each group contributes to the pooled BS response.[15]

I implement equation (14) over the same local projection window as BS' specification, February 1988 through February 2020. Of the 385 months in this window, 238 are single-announcement, 31 are multi-announcement, and 116 contain no FOMC-related announcement.[16] Outcomes, controls, lag length, and confidence bands closely follow the local projection specification in BS; the main outcome of interest is the two-year Treasury yield.

Table II and Figure II report the results. At horizon $h = 0$ for the two-year Treasury yield, the pooled slope is $\hat{\beta}_R = 0.56\,(0.30)$, the single-announcement slope is $\hat{\beta}_1 = 0.02\,(0.31)$, and the multi-announcement slope is $\hat{\beta}_2 = 1.96\,(0.47)$; the null $\beta_1 = \beta_2$ is strongly rejected. Across horizons, $|\hat{\beta}_2| > |\hat{\beta}_1|$ at 47 of 49 horizons at the 5% confidence level, and the same pattern appears for the federal funds rate and the one-year Treasury yield, as displayed in the next exhibit. In summary, single-announcement months contribute essentially none of the interest rate response that gives the monthly BS proxy its instrumental relevance.

[15] The full proof is detailed in Appendix H.I.

[16] Multi-announcement months are 8.1% of all months and 11.5% 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.

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

h	$\hat{\beta}_R$	$\hat{\beta}_1$	$\hat{\beta}_2$	$p_{H_0:\beta_1=\beta_2}$	$\hat{\omega}_1$	$\hat{\omega}_2$	N
0	0.56 (0.30)	0.02 (0.31)	1.96 (0.47)	0.001	0.720	0.280	384
3	−0.08 (0.71)	−1.23 (0.78)	2.89 (1.29)	0.007	0.720	0.280	381
6	1.15 (0.68)	0.07 (0.75)	3.96 (1.41)	0.021	0.724	0.276	378
12	0.47 (0.77)	−0.91 (0.89)	4.10 (1.14)	0.000	0.723	0.277	372
18	1.50 (0.75)	−0.15 (0.72)	5.78 (1.45)	0.001	0.722	0.278	366
24	1.17 (0.69)	0.19 (0.84)	3.70 (1.33)	0.063	0.721	0.279	360
36	−0.75 (0.88)	0.33 (0.91)	−3.53 (1.47)	0.026	0.720	0.280	348
48	−1.60 (1.06)	0.09 (1.05)	−5.91 (2.82)	0.075	0.717	0.283	336

Notes: Coefficients from the unrestricted local projections with controls as in the public-data local-projection specification reported in Appendix F. Newey–West standard errors with bandwidth in parentheses. p_{H_0} is the Wald test of $\beta_1 = \beta_2$. Sample: 1988:02–2020:02, excluding 2001:09.

The weighting arithmetic makes the same point from the pooled-regression perspective. The estimates satisfy $\hat{\beta}_{R,h} = \hat{\omega}_{1,h}\hat{\beta}_{1,h} + \hat{\omega}_{2,h}\hat{\beta}_{2,h}$ horizon by horizon, with $(\hat{\omega}_1, \hat{\omega}_2) \approx (0.72, 0.28)$ 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 28% of the residualized variation in the pooled proxy despite representing only about 8% of the sample. Because $\hat{\beta}_1$ is near zero, the pooled BS response is in practice close to a rescaled version of the multi-announcement slope.[17]

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.[18]

In the baseline sample, there are only 31 months with more than one announcement.[19] 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 later-announcement information can enter the monthly orthogonalization even though it was not available before the first announcement. 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

[17] At $h = 0$, $0.72 \times 0.02 + 0.28 \times 1.96 \approx 0.56$, which reproduces the pooled coefficient exactly.

[18] Appendix I studies whether irregular announcement timing overweight multi-announcement months in the event-level regression. I test this and find it empirically moot. I thank Domenico Giannone for encouraging me to explore this dimension.

[19] Table I lists all 31, together with some narrative context.

months.[20]

The unrestricted regression in (14) shows that multi-announcement months are therefore 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 both 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.

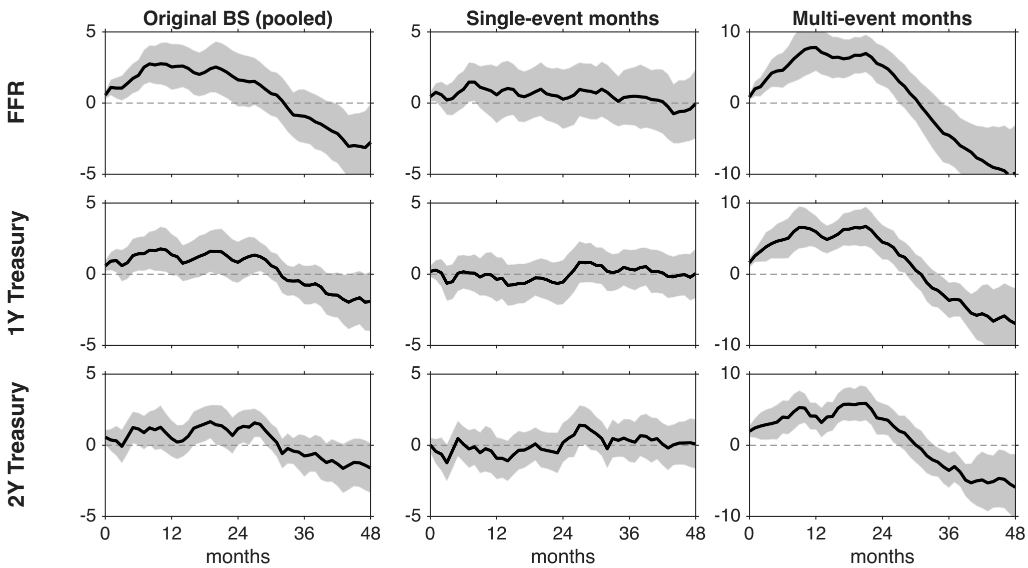


Figure II: Local projection responses to a unit monetary surprise, by event-month type. Rows: federal funds rate (top), 1-year Treasury yield (middle), 2-year Treasury yield (bottom). Columns: pooled BS response $\hat{\beta}_{R,h}$ from the pooled regression on z_m (left), single-announcement months $\hat{\beta}_{1,h}$ (center), multi-announcement months $\hat{\beta}_{2,h}$ (right). Sample, controls, lag length, and 90% bands follow the public-data local-projection specification reported in Appendix F.

The role of multi-announcement months. This interpretation is directly supported by the data. The contamination mechanism makes a direct testable prediction. Restricting our attention to multi-announcement months, later-announcement predictors should remain jointly significant in regressions of the monthly BS series on the information observed before later announcements within the same month. If the BS procedure fully removed same-month predictability, this block should be jointly insignificant.

The data reject this implication. Later-announcement predictors are jointly significant for the monthly BS series: unconditionally $F = 3.36$ ($p = 0.012$), and

[20]The correlation falls from 0.978 to 0.805 (Table A.1, Figure A.1), the direct later-announcement component explains 77.4% of the covariance shortfall (Table A.2), and later-announcement predictors remain jointly significant in the BS monthly series (Table B.1).

conditional on the first-announcement predictor vector $F = 7.00$ ($p < 0.001$) (Appendix Table B.1). The BS procedure retains same-month predictable variation that arrives after the first announcement. That leftover variation improves the first stage, but it does so by contaminating the instrument. Purging at the announcement level removes this predictable component. The resulting series is therefore cleaner on timing and exogeneity grounds, but — as it has been argued above — also less relevant in the first stage.

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. Therefore, the premise of my findings above generalizes. Using the BS predictor set, I show that the leading public surprise series are predictable at the announcement level (Appendix C), exposing the same concerns on instrumental validity. Although a systematic analysis of the properties of each series is beyond the scope of the illustration (and of this paper), Figure III reports the two orderings' first stage for the two-year Treasury yield, in the baseline proxy-SVAR and with the BS predictor set held fixed throughout.

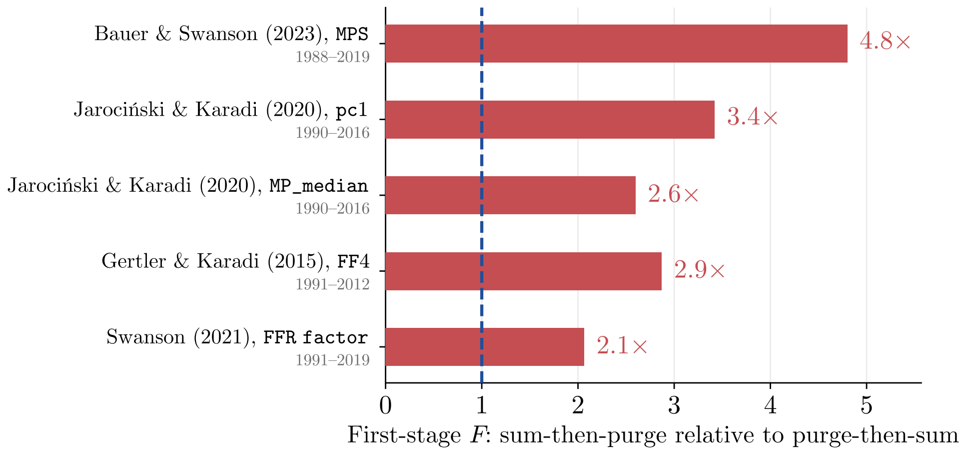


Figure III: Ratio of the first-stage F under sum-then-purge to that under purge-then-sum, two-year Treasury yield, in the baseline proxy-SVAR above with the BS predictor set.

In every series the sum-then-purge instrument is the more relevant of the two, by a factor 2 to 5. The pattern throughout this section, i.e. relevance supported by variation that an announcement-level purge removes, is therefore common to instruments built by different authors from different underlying data.[21] The ex-

[21]Two qualification apply, developed in Appendix C: (i) even an incorrect orthogonalization weakens the series under analysis below acceptable level for inference; (ii) I impose one predictor set on every series, so the figure asks what a monthly-aggregate orthogonalization would do to each surprise measure; no row reproduces its own paper's construction.

ercise therefore to the source of instrumental relevance; it says nothing about the absolute strength of any published construction. The direction is what generalizes: 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 of the instrument.

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.[22] 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 BS 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. Table 1 in BS summarizes these regressions, which I replicate exactly in Table III, confirming the first step in the construction of `MPS_ORTH`.

BS Section 4 formalizes the construction of orthogonalized surprises (per BS 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 exactly in Table IV. 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 regressors and differ only in the timing of orthogonalization; they remain closely aligned overall (correlation 0.93), and especially so in single-announcement months (0.98), but

[22] 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.

diverge in multi-announcement months (0.81) — precisely where the information differ (Figure A.1; Table A.1).[23]

Table III: Predictive Regressions Using Macroeconomic and Financial Data

	(1)	(2)	(3)
Nonfarm payrolls	0.094	0.113	0.082
	(2.399)	(1.944)	(1.759)
Employment growth (12m)	0.005	0.004	0.005
	(2.121)	(1.380)	(1.196)
Δ log S&P 500 (3m)	0.084	0.112	0.154
	(1.431)	(1.553)	(1.912)
Δ Slope (3m)	-0.010	-0.010	-0.011
	(-1.377)	(-1.134)	(-1.018)
Δ log Comm. Price (3m)	0.119	0.093	0.224
	(2.353)	(1.436)	(3.431)
Treasury skewness	0.032	0.035	0.050
	(2.983)	(2.869)	(2.092)
Constant	-0.011	-0.009	-0.021
	(-2.319)	(-1.714)	(-2.558)
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: Coefficient estimates β from regressions $mps_t = \alpha + \boldsymbol{\beta}'\boldsymbol{X}_t + u_t$, where t indexes FOMC announcements. Columns use BS (2023a)' baseline monetary policy surprise. Predictors are observed prior to the announcement: the surprise component of the most recent nonfarm payrolls release, employment growth over the last year, the log change in the S&P 500 index from 3 months before to the day before the FOMC announcement, the change in the yield curve slope over the same period, the log change in a commodity price index over the same period, and the option-implied skewness of the 10-year Treasury yield from Bauer and Chernov (2024). Heteroskedasticity-consistent t-statistics in parentheses.

Table IV: Asset Price Responses to Monetary Policy Surprises

	mps_t			$\mathrm{mps}_t^{\perp}$		
Series	Coef.	t-stat	R^2	Coef.	t-stat	R^2
2–year yield	.73	(18.5)	.783	.74	(16.6)	.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	-5.44	(-7.7)	.284	-5.57	(-6.6)	.249
Observations	322			322		

Note: Estimated coefficients β and regression R^2 from high-frequency event-study regressions $y_t = \alpha + \beta\,\mathrm{mps}_t + u_t$, where t indexes FOMC announcements, y_t denotes the change in the 2-, 5-, 10-, or 30-year Treasury yield or log S&P (Standard & Poor's) 500 price index in a narrow window of time around each announcement, and the regressor mps_t is either the unadjusted high-frequency monetary policy surprise measure mps_t or its orthogonalized residual $\mathrm{mps}_t^{\perp}$ from regressing mps_t on the predictors in Table 1. Heteroskedasticity-consistent t-statistics are in parentheses. Sample: 1988:02–2019:12.

[23] If the correction merely added noise, divergences would be uniform across months. Appendix A provides a detailed comparison.

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 practices 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.

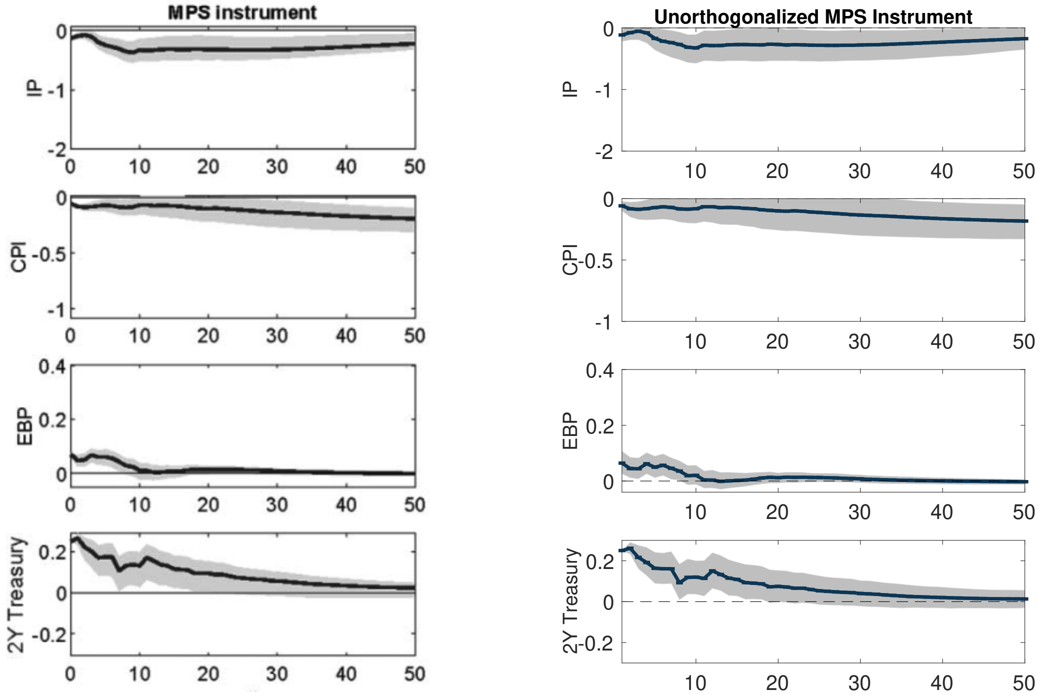
I now turn to one of the main research questions BS aim to answer: how much difference does orthogonalizing the high-frequency surprises make for estimating the effects of monetary policy on the economy? Figure 3 in their paper answers this question by demonstrating that the responses of all variables 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; Figure V uses the announcement-level orthogonalized 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"*. 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 adjusted series' first-stage F-statistic collapses from 10.2 to 0.1, well below any conventional weak-instrument threshold.[24] The corrected measure has essentially no predictive power for the baseline two-year Treasury yield (and similarly for all other interest rates measures), so the BS-style design no longer delivers statistically informative macroeconomic responses.[25]

[24] 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 and show that conventional first-stage thresholds are even too permissive for impulse response applications. 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) are unbounded at every horizon and every variable under the corrected series, so no magnitude is identified. I thank Daniel Lewis for suggesting this exercise.

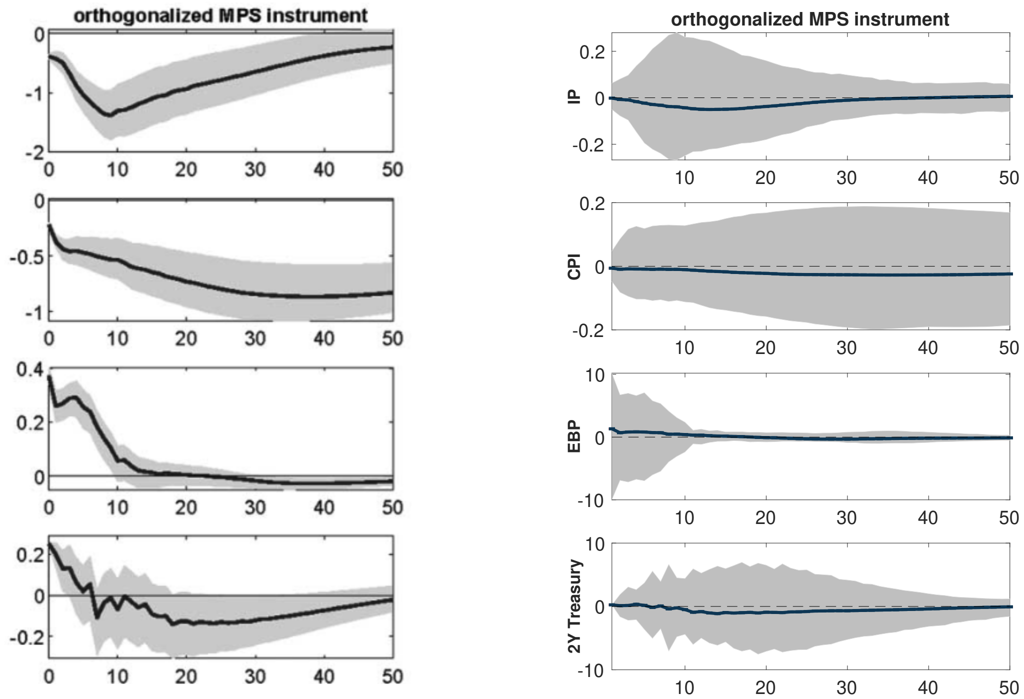
[25] Appendix E reports an extension through 2023. Using the publicly updated BS surprise series and current official updates of the same macro series definitions, the longer sample strengthens all proxies, including the corrected one, but leaves the corrected proxy similarly weaker.



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. Bootstrapped 90% SE bands. EBP = excess bond premium, CPI = Consumer Price Index, IP = ind. production.



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

Revised Method

Figure V: Comparison between Proxy SVAR IRFs to a 25bp policy shock identified using the unadjusted high-frequency `MPS_ORTH` measure. Left: original Bauer and Swanson (2023a) Fig. 3 (Right Panels); Right: Revised Method. Sample: 1973:01–2020:02. Lags: 12. Bootstrapped 90% SE bands. EBP = bond premium, CPI = Consumer Price Index, IP = ind. production.

The mechanism test reported in Section IV.I confirms the origin of this relevance loss: later-announcement predictors are jointly significant for the BS monthly series, and vanish once the orthogonalization is moved to the announcement level.

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 all the other figures in the paper (Fig. 4 — richer external IV SVAR; Fig. 5 — local projections; Fig. 6 — internal IV SVAR; Fig. 8 — "best practices" specification, see below), and confirm the absence of recognizable effects.[26,27]

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 independently reproduces the results. 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 differs from the original only in the absence of those additional events; it therefore represents the closest feasible approximation based on currently available public data.

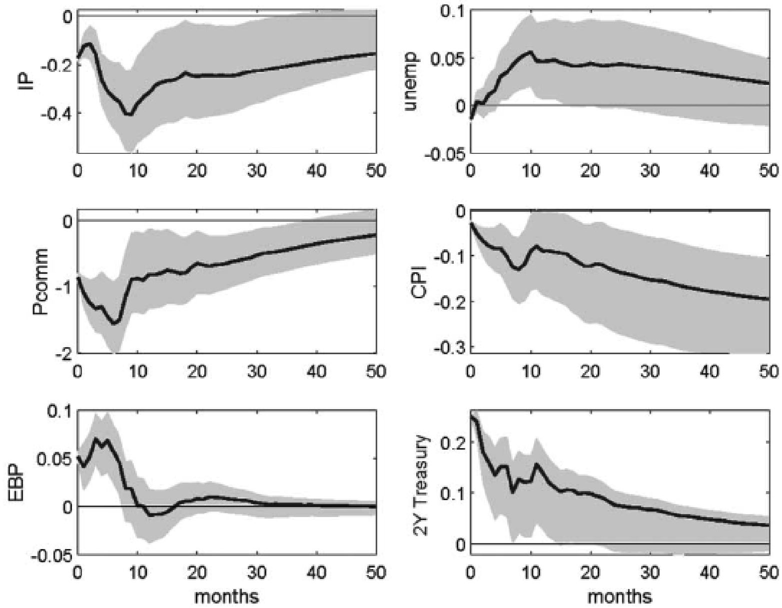
Focusing on Panel (B) of Figure VI, we notice again that the announcement-level orthogonalized surprises show no statistically detectable responses in macroeconomic aggregates under conventional bands, and again see a significant loss in instrumental strength (F-statistic = 0.5).

Under the announcement-level purged instrument, 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.[28] Appendices E and F 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.

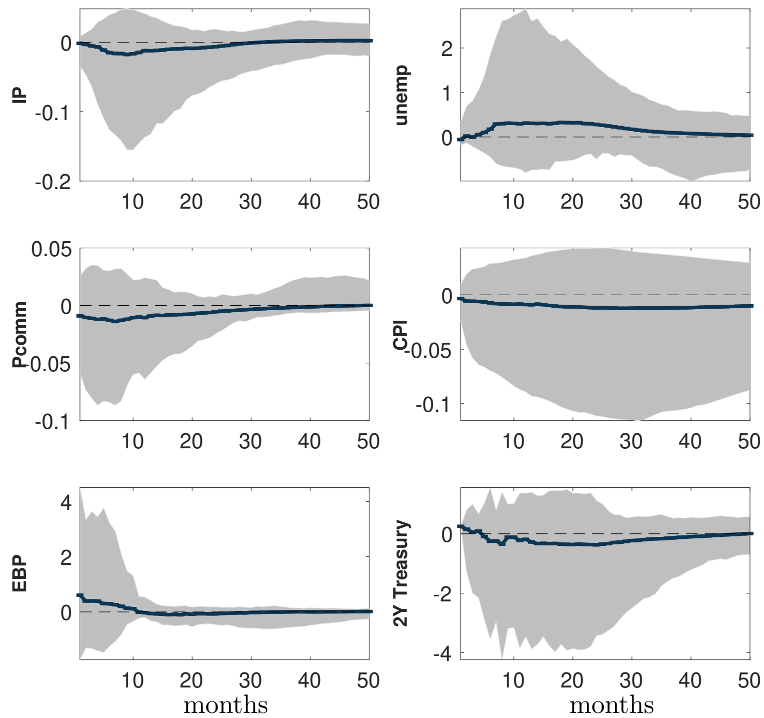
[26] Quoting from the downloadable spreadsheet: *"This spreadsheet contains data only for the FOMC announcements, and excludes data for the Fed Chairs speeches and other types of monetary policy announcements (FOMC minutes releases, speeches by the Vice Chair)."* No public historical Fed Chair speeches surprise dataset is currently available.

[27] Appendix F reports the feasible local-projection analogue, i.e. excluding FOMC speeches (closest to Figure C.1 in an older BS draft).

[28] Qualitatively analogous results hold for the 1-year Treasury and the Fed Funds Rate. For instance, using the 1-year yield (Gertler and Karadi (2015)) raises the corrected instrument's F-statistic from 0.1 to 0.5, still far below conventional relevance thresholds.



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



(B) Revised Method

Figure VI: "Best practice" estimates of Proxy SVAR IRFs to a 25bp policy shock identified using adjusted high-frequency measure (`MPS_ORTH`). Sample: 1973:01–2020:02. Lags: 12. Bootstrapped 90% SE bands. EBP = bond premium, CPI = Consumer Price, unemp = unemployment, Pcomm = commodity prices, IP = ind. production.[29]

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; that variation stays in the instrument. Purging first, each surprise against its own pre-announcement information, removes it by construction. The two procedures coincide only when later information adds nothing to the first — an empirical implausible restriction.

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 3.62 under the BS ordering to 0.15, against 10.19 for the raw surprise, and the four-times-larger macroeconomic responses reported by BS vanish across proxy-SVAR, local projection, and internal-IV specifications.

That variation sits 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 — eight percent of the baseline sample. Single-announcement months contribute no measurable interest rate response. A formal decomposition shows why: by summing within-month surprises before orthogonalizing against first-announcement predictors (*sum-then-purge*), the BS procedure retains a component of later-announcement surprises that is predictable from news arriving between meetings. That leftover predictable component 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 further results corroborate the diagnosis: the two monthly series are nearly identical in single-announcement months but diverge sharply in multi-announcement months; later-announcement predictors remain jointly significant for the monthly BS series; and the direct later-announcement component accounts for most of the covariance gap.

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 the instrument's spurious statistical power.

[29] Uneven widening of the confidence bands is compatible with weak-instrument theory. $\mathrm{Var}(\hat{\Theta}_h) \approx \sigma^2_{Y,h}/(N\pi^2)$ with π the first-stage slope. Heteroskedastic $\sigma^2_{Y,h}$ makes the blow-up uneven across variables.

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.

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.

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.

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.

Appendix

A. Direct comparison of the monthly orthogonalized instruments

To address the concern that the corrected series may depart from the BS procedure only because of additional introduced noise, this exercise directly compares the monthly orthogonalized series of Bauer and Swanson (2023a), `MPS_ORTH`, with the purge-then-sum series. The two measures are constructed from the same predictors set and differ only in the timing of the orthogonalization.

Using the baseline monthly sample 1988:02–2020:02, the overall correlation between the two series is 0.93. In months with a single announcement, the correlation is 0.98 and the mean absolute difference is 0.009. In months with two or more announcements, the correlation falls to 0.80 and the mean absolute difference rises to 0.038. The largest absolute discrepancies occur in 1989:02, 1991:12, and 2008:10, with absolute differences of 0.179, 0.145, and 0.080, respectively. The economically meaningful divergences are therefore concentrated in the months where the relevant information sets differ by construction, rather than being spread uniformly across the sample as one would expect from a purely noisy implementation change.

Table A.1: Comparison of monthly orthogonalized surprise measures

Sample	Observations	Correlation	Mean abs. diff.	Same-sign %
All event months	269	0.932	0.012	90
1 announcement months	238	0.978	0.009	90
2+ announcement months	31	0.805	0.038	87

Notes: The table compares the monthly orthogonalized series of Bauer and Swanson (2023a) (`MPS_ORTH`) with the announcement-level orthogonalized series over 1988:02–2020:02. The same-sign share is the fraction of months in which the two measures have the same sign.

Figure A.1 visualizes the comparison in Table A.1. Panel (A) shows that the two monthly IVs lie close to the 45-degree line in most observations, with the largest departures concentrated in multi-announcement months. Panel (B) shows the same comparison over time and makes clear that the discrepancies (which exist throughout) are mostly evident around specific episodes rather than pervasive. Together, the two panels reinforce the interpretation that the two series are usually close, but diverge systematically precisely when within-month information sets differ.

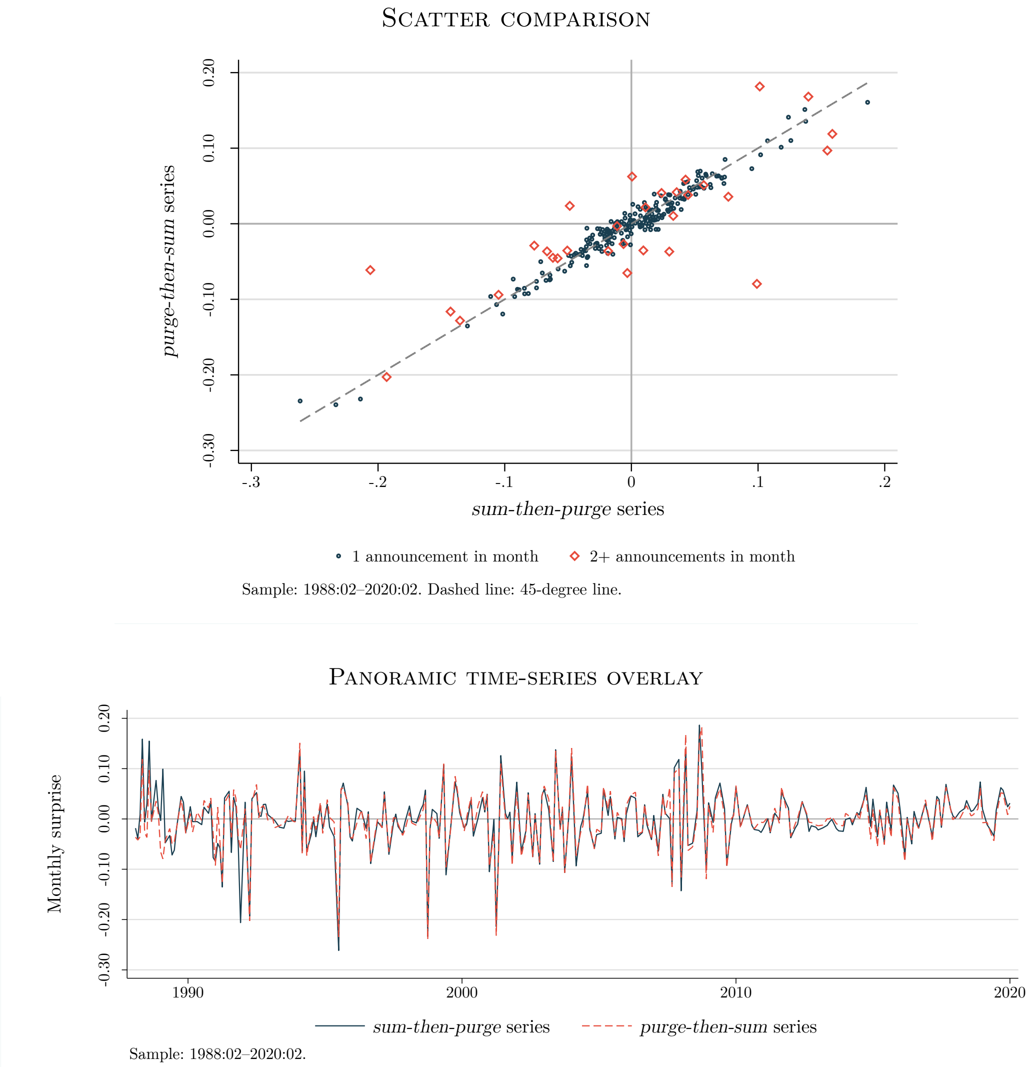


Figure A.1: Direct comparison between the BS orthogonalized series and the purge-then-sum series. Panel (A) plots the two monthly IVs against each other; blue circles denote months with one announcement and red diamonds denote months with two or more announcements. Panel (B) overlays the two monthly series over time. The light gray band in Panel (B) denotes the gap between the two series. Sample: 1988:02–2020:02.

Distribution of announcements per month Figure A.2 offers a visual summary of the number of monetary policy announcements per month in the updated BS announcement file with a single stacked bar, separating months with one announcement from months with at least two announcements.

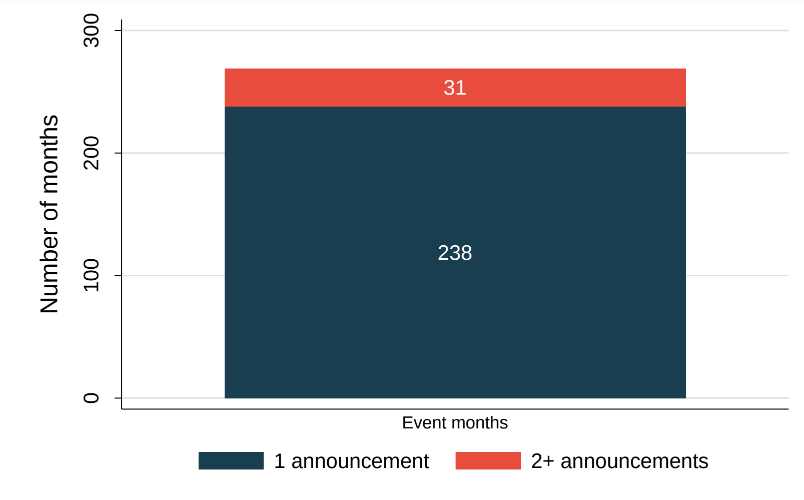


Figure A.2: Stacked-bar summary of monetary policy announcements per month in the updated BS announcement file. The figure counts event months from 1988:02 to 2020:02.

Among the total 269 event months, 31 contain at least two announcements, 36 contain at least one unscheduled event, and the maximum number of announcements in a month is seven (May 1988).

A.I Quantitative decomposition

This appendix quantifies how much of the monthly gap $z_m - \tilde{z}_m$ comes directly from later-announcement information, and how much comes from the change in projection coefficients induced by monthly aggregation. With a constant absorbed into $\boldsymbol{X}_{t-m}$, the exact sample identity is

$$z_m - \tilde{z}_m = \sum_{t=2}^{T_m} \hat{\boldsymbol{\beta}}' \boldsymbol{X}_{t-m} + \left(\hat{\boldsymbol{\beta}}' \boldsymbol{X}_{1-m} - \hat{\gamma}' \boldsymbol{X}_{1-m} \right). \tag{A.1}$$

The first term is the *direct* later-announcement component. The second is the *projection-coefficient* component: the same first-announcement information is fitted differently under the event-level and monthly projections. In single-announcement months, the direct term is zero by construction, but the projection-coefficient component can still generate a gap between the two monthly IVs.

Table A.2 reports the main magnitudes. The mean absolute gap is 1.23 basis points across event months, but this average hides a sharp split: 0.89 basis points in single-announcement months versus 3.82 basis points in months with at least two announcements. In multi-announcement months, the direct term dominates quantitatively: its mean absolute contribution is 3.99 basis points, compared with 0.77 basis points for the projection-coefficient term. The covariance decomposition tells the same story: adding back the direct term closes 77.4% of the covariance gap between the two series, while adding back the projection-coefficient term closes the remaining 22.6%.

Table A.2: Quantitative decomposition of the gap between monthly IVs

Sample	Months	Mean |direct|	Mean |projection|	Mean $\lvert z_m - \tilde{z}_m \rvert$
All months	385	0.0032	0.0061	0.0086
Event months	269	0.0046	0.0088	0.0123
1 announcement month	238	0.0000	0.0089	0.0089
2+ announcement months	31	0.0399	0.0077	0.0382

Notes: The table reports mean absolute values of the two terms in equation (A.1). Units are the same as in the monthly IV series, so 0.01 corresponds to one basis point. The direct term is identically zero in single-announcement months. Let d_m and p_m denote the direct and projection-coefficient terms. Then $\text{Cov}(z, \tilde{z}) = 0.00215$, $\text{Cov}(z, \tilde{z}+d) = 0.00234$, $\text{Cov}(z, \tilde{z}+p) = 0.00221$, and $\text{Var}(z) = 0.00240$. Normalizing each covariance gain by $\text{Var}(z) - \text{Cov}(z, \tilde{z})$ attributes 77.4% of the covariance gap to the direct term and 22.6% to the projection-coefficient term.

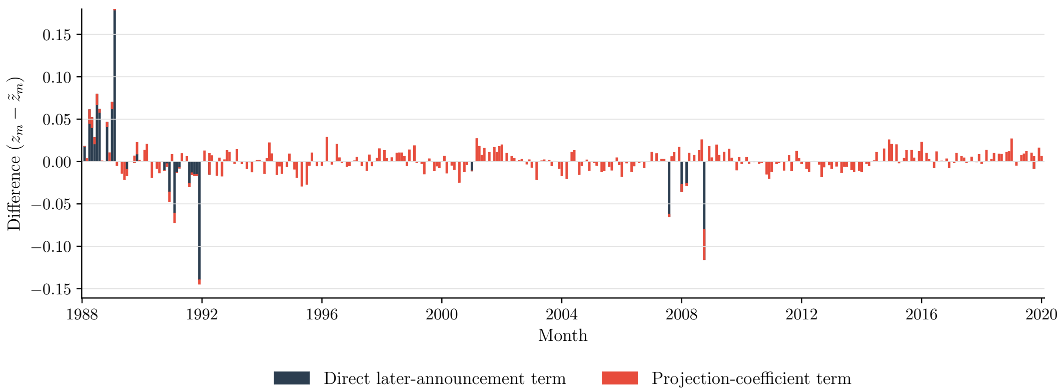


Figure A.3: Chronological decomposition of the monthly gap between `MPS_ORTH` and the announcement-level purge-then-sum IV. Each bar stacks the direct later-announcement term and the projection-coefficient term, so the total height equals $z_m - \tilde{z}_m$. Sample: 1988:02–2020:02 event months.

Figure A.3 visualizes the decomposition. Most bars contain only the red projection-coefficient component because 238 of the 269 event months contain a single announcement, so the direct term is zero by construction. The blue spikes appear only in the small set of multi-announcement months, exactly where later-announcement information can enter the monthly BS measure directly. Overall, the monthly gap is driven mainly by later-announcement information within multi-announcement months, with a smaller but nonzero spillover to all months through the monthly first-announcement projection.

A.II Narrative analysis of large-discrepancy months

The main text already discusses July 1988 and October 2008, two months in which the information set changes sharply between the first and later announcements. Two additional high-gap months show the same mechanism outside those episodes. In 1989:02, the BS file contains five announcements, four of them unscheduled. The payroll predictors are unchanged throughout the month, but several financial predictors move materially between the first and last announcement: the 3-month S&P 500 return rises from 0.078 to 0.092, the 3-month slope change from −0.889 to −0.574, and the 3-month commodity-price change from 0.035 to 0.062. In that month, `MPS_ORTH` equals 9.9 basis points, whereas the announcement-level purge-then-sum series equals −8.0 basis points, so the two measures even differ in sign.

December 1991 provides a similar example without a sign reversal. The month contains four announcements, three unscheduled, including two on December 20. Between the first and last announcement, the 3-month slope change rises from 0.488 to 0.668 and the 3-month commodity-price change falls from −0.010 to −0.068, while the payroll predictors again remain unchanged. In that month, `MPS_ORTH` equals −20.6 basis points, compared with −6.1 basis points for the announcement-level series. Thus, even when the two measures point in the same direction, the stale first-announcement information set can still substantially magnify the monthly orthogonalized series.

Taken together with July 1988 and October 2008 in Section II.I, these episodes cover four of the five largest absolute baseline discrepancies between the two monthly IVs. The common pattern is that the two procedures diverge most in months with multiple announcements and meaningful within-month changes in the predictor vector, which is exactly the timing channel emphasized in this paper.

B. Direct mechanism test using later-announcement information

I restrict attention to months with at least two announcements and estimate

$$\texttt{MPS\_ORTH}_m = \alpha + \Gamma' X_{tm} + u_{tm}, \qquad t = 2, \dots, T_m, \tag{B.1}$$

where X_{tm} is the vector of BS predictors observed immediately before later announcement t in month m. This regression asks whether the monthly BS orthogonalized series remains predictable from information that arrives after the first announcement of the month.[30] If the BS procedure fully removed same-month predictability, the later-announcement predictor block should not be jointly significant. Table B.1 rejects that implication: later-announcement information significantly predicts the monthly BS series, both on its own and conditional on the first-announcement predictor vector X_{1m}.

Table B.1: Direct mechanism test: later-announcement predictability of BS `MPS_ORTH`

	Dependent variable: monthly BS `MPS_ORTH`	
	(1)	(2)
Nonfarm payroll surprise	0.0002	0.0005
	(0.0002)	(0.0002)
12-month payroll growth	0.0290	-0.0491
	(0.0180)	(0.1519)
S&P 500 return (3m)	-0.4841	-1.0355
	(0.1858)	(0.3070)
Yield-curve slope change (3m)	-0.0182	-0.0467
	(0.0504)	(0.0773)
Commodity-price change (3m)	-0.0182	0.8869
	(0.1663)	(0.5057)
Treasury skewness	0.0535	-0.3062
	(0.0623)	(0.1952)
First-announcement predictors X_{1m}	No	Yes
Joint F-test on later-announcement block	3.36	7.00
p-value	0.012	< 0.001
R^2	0.462	0.683
Later-announcement observations	56	56
Underlying multi-announcement months	31	31

Note: Sample ends in 2020:02. The table restricts attention to months with at least two announcements and stacks later-announcement observations within those months. Column (1) regresses the monthly BS orthogonalized series on the later-announcement predictor vector. Column (2) additionally includes the first-announcement predictor vector, whose coefficients are omitted for brevity. Heteroskedasticity-robust standard errors clustered by month are in parentheses.

C. The ordering across public surprise series

This appendix documents the exercise of Section IV.I. 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),

[30] Because the dependent variable is monthly while the predictors vary across later announcements, the monthly observation is repeated for each later announcement and standard errors are clustered by month. Unreported versions that weight each multi-announcement month equally by $1/(T_m - 1)$ are slightly stronger.

Jarociński and Karadi (2020) (surprise factor and monetary component), Swanson (2021) (federal funds rate and forward guidance factors), and BS. Holding the predictor set fixed isolates the ordering; no exercise here replicates any paper's own construction. I also examined Bu et al. (2021) and Nakamura and Steinsson (2018): neither sample contains a month with announcements on more than one date, so the direct later-announcement component is identically zero and the orderings can separate only through the projection coefficients of Appendix D. The Bu et al. (2021) surprises are in addition unpredictable from the BS vector ($R^2 = 0.05$, $p = 0.27$), leaving either ordering nothing to remove. That degeneracy restates the mechanism: with no within-period predictor update, the orderings cannot diverge.

Table C.1 reports the announcement-level predictability regressions that motivate orthogonalizing in the first place, in the format of Table III. Every series in the exercise is predictable from information observed before the announcement, with R^2 between 0.102 and 0.176. The six slopes are jointly significant at the five percent level in four of the five columns; the Swanson federal funds rate factor just misses that level and clears ten percent ($p = 0.053$). The three series left out of the first stage behave as the mechanism implies. The case for removing a predictable component is therefore general across these constructions; what this paper adds is that the order in which that removal and the temporal aggregation are performed is itself part of the design.

Announcements sharing a date are summed to one observation: the six predictors are identical across same-date rows, so the vector cannot update within them. The baseline sample then contains 28 multi-date months, against the 31 multi-announcement months of Section II — three months (1991:04, 1991:09, 1991:11) hold two announcements on a single date, where the ordering cannot bind.[31]

Behind the first-stage gap of Figure III lies the later-announcement information a first-announcement purge cannot remove. Computed with pooled announcement-level coefficients, it amounts to between 0.27 and 0.75 standard deviations of the instrument per multi-date month, and to between 0.02 and 0.08 averaged over all months. It is identically zero in single-date months and under the announcement-level ordering. As in Section IV.I, the concentrated version is what reaches the first stage, because the affected months carry disproportionate regression weight.

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 76 percent of their common variation and the BS

[31]On date-level constructions the two orderings correlate at 0.992 in single-date months and 0.883 in multi-date months, with mean absolute gaps of 0.49 and 3.05 basis points, against 0.978, 0.805, 0.89 and 3.82 in Appendices A and A.I. The post-9/11 announcement of September 17, 2001, for which BS record no surprise value, is excluded throughout, following their convention.

and Jarociński and Karadi (2020) surprises correlating at 0.97. The rows therefore evaluate one mechanism under several standard constructions. What varies is how strongly each series is predictable and how much the predictors move within the affected months.

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	0.093	0.129	0.104	0.166	1.497
	(2.381)	(2.610)	(2.645)	(3.062)	(2.688)
Employment growth (12m)	0.005	0.006	0.003	0.003	0.003
	(2.165)	(2.273)	(1.363)	(1.471)	(0.094)
Δ log S&P 500 (3m)	0.079	0.080	0.063	0.001	1.541
	(1.354)	(1.209)	(0.857)	(0.018)	(1.529)
Δ Slope (3m)	-0.011	-0.011	-0.007	-0.017	-0.189
	(-1.539)	(-1.191)	(-0.851)	(-1.631)	(-1.412)
Δ log Comm. Price (3m)	0.122	0.157	0.093	0.100	0.381
	(2.354)	(2.612)	(1.592)	(1.586)	(0.490)
Treasury skewness	0.032	0.045	0.028	0.042	0.267
	(2.916)	(2.746)	(1.519)	(1.888)	(1.585)
Constant	-0.012	-0.028	-0.017	-0.021	-0.026
	(-2.518)	(-4.593)	(-2.683)	(-3.473)	(-0.410)
R^2	0.165	0.176	0.102	0.174	0.107
Joint F	5.22	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	316	240	240	192	241

Notes: Coefficient estimates β from regressions $s_t = \alpha + \boldsymbol{\beta}'\boldsymbol{X}_{t-} + u_t$, where t indexes announcement dates and s_t is each paper's own high-frequency surprise. Predictors are the six variables of Table III, observed before the announcement and held fixed across columns, with the nonfarm payroll surprise on the same scale. Columns match the series of Figure III: 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 III, column (1) up to the announcement-date convention of this appendix and the updated release of the BS spreadsheet. Joint F is the heteroskedasticity-robust Wald statistic for the six slopes, referred to $F(6, N-7)$. Heteroskedasticity-consistent t-statistics in parentheses.

Caveats. I impose one predictor set on every series and use each paper's own publication window. The first-stage specification is the one this paper applies to BS throughout, carried over unchanged; it belongs to no other paper, and the within-series comparison is the object of interest.

D. Why the IVs Differ in Single-Announcement Months

This paper clarifies a point that may appear puzzling at first glance. In a month with exactly one announcement, the raw surprise is the same object in both constructions, and the predictor vector observed before that announcement is also the same. It is therefore natural to wonder why the two orthogonalized monthly series need not coincide in such months.

The key is that an OLS residual is not determined only by the current observation. It is determined by the current observation together with the coefficient vector estimated from the whole regression. For a single-announcement month m, let the raw surprise be $\texttt{MPS}_{1m}$ and the pre-announcement predictor vector be X_{1-m}. Then the announcement-level purge-then-sum series is

$$\tilde{z}_m = \texttt{MPS}_{1m} - \hat{\beta}' X_{1-m},$$

while the Bauer–Swanson monthly series is

$$z_m = \texttt{MPS}_{1m} - \hat{\gamma}' X_{1-m}.$$

Hence, in a single-announcement month,

$$z_m - \tilde{z}_m = (\hat{\beta} - \hat{\gamma})' X_{1-m}.$$

So equality in that month requires not only the same raw surprise and the same predictor vector, but also the same fitted coefficient vector.

This is precisely where the full-sample difference enters. The event-level coefficients $\hat{\beta}$ are estimated from a regression with one observation per announcement. By contrast, the monthly coefficients $\hat{\gamma}$ are estimated from a regression with one observation per month, where multi-announcement months are collapsed into "sum of surprises on the left-hand side, first-announcement predictors on the right-hand side." Those multi-announcement months affect the fitted monthly projection and therefore shift $\hat{\gamma}$. That shift then spills over to all months, including months with only one announcement. Put differently, the same point can have different residuals when evaluated against two different fitted lines.

The common intuition can therefore be restated as follows. In a single-announcement month, the left-hand side is indeed the same raw surprise, and the right-hand side is indeed the same predictor vector. If the coefficient vector were held fixed, or if both regressions were re-estimated on a sample containing only single-announcement months, the residuals would coincide. But that is not the comparison implemented in the paper. The paper compares residuals from two different regressions, estimated on two different samples and at two different levels of aggregation. For that reason, the same left-hand side and the same right-hand side in a given month do not imply the same residual.

E. Sample extension to 2023

Table E.1 summarizes how first-stage strength changes when I extend the sample from the baseline period to December 2023. I stop at 2023 because that is where the currently posted BS-style public surprise spreadsheet ends; going further would require switching to a different FRBSF data product rather than extending the same public object. Since the exact macro workbook underlying the baseline VAR and local-projection exercises is not publicly posted in a matched updated form, I leave the baseline 1973:01–2020:02 sample unchanged and treat the longer sample separately, using the same public series definitions and current official updates of the macro variables. The longer sample strengthens all proxies, including the corrected one, but the ranking is unchanged. In the proxy-SVAR, the corrected proxy's first-stage F-statistic rises from 0.15 to 2.00, still well below 12.58 for `MPS_ORTH` and 24.26 for `MPS`. In the local-projection exercise, the corresponding weak-IV statistic rises from 0.04 to 2.20 for the corrected proxy, compared with 11.62 for `MPS_ORTH`. I therefore treat the 2023 extension as a useful appendix update rather than as a change in the paper's main message.

Table E.1: Extension of first-stage strength (F-stats) through 2023

Exercise	Proxy	Baseline sample	Extended to 2023
Proxy-SVAR	`MPS`	10.19	24.26
Proxy-SVAR	`MPS_ORTH`	3.62	12.58
Proxy-SVAR	`MPS_ORT`	0.15	2.00
Local projections	`MPS_ORTH`	3.40	11.62
Local projections	`MPS_ORT`	0.04	2.20

Notes: The baseline sample is 1973:01–2020:02. `MPS` denotes the raw unadjusted series, `MPS_ORTH` denotes the BS monthly orthogonalized proxy, and `MPS_ORT` denotes the corrected proxy. The extension uses the BS surprise series through 2023:12 together with current official updates of the same macro series definitions. It adds 46 months and 32 event months beyond 2020:02. The longer sample strengthens all proxies, including the corrected one, but the corrected proxy remains materially weaker than the BS monthly orthogonalized proxy in both exercises.

F. Additional Replicated Results

The exact replication of BS Figure 5 is not feasible because the publicly posted replication files exclude Fed Chair speeches. As a feasible substitute, Figure F.1 reports the currently available public-data FOMC-only local-projection exercise, which is the closest public analogue to BS Figure 5 and corresponds most closely to BS Appendix Figure C.1. The left panel uses the unadjusted surprise series `MPS`; the right panel replaces it with the announcement-level purge-then-sum series `MPS_ORT`. Consistent with the proxy-SVAR results in the main text, the corrected orthogonalized LP responses display no recognizable effects: the impulse responses are essentially flat relative to the confidence bands, and statistical significance disappears throughout.

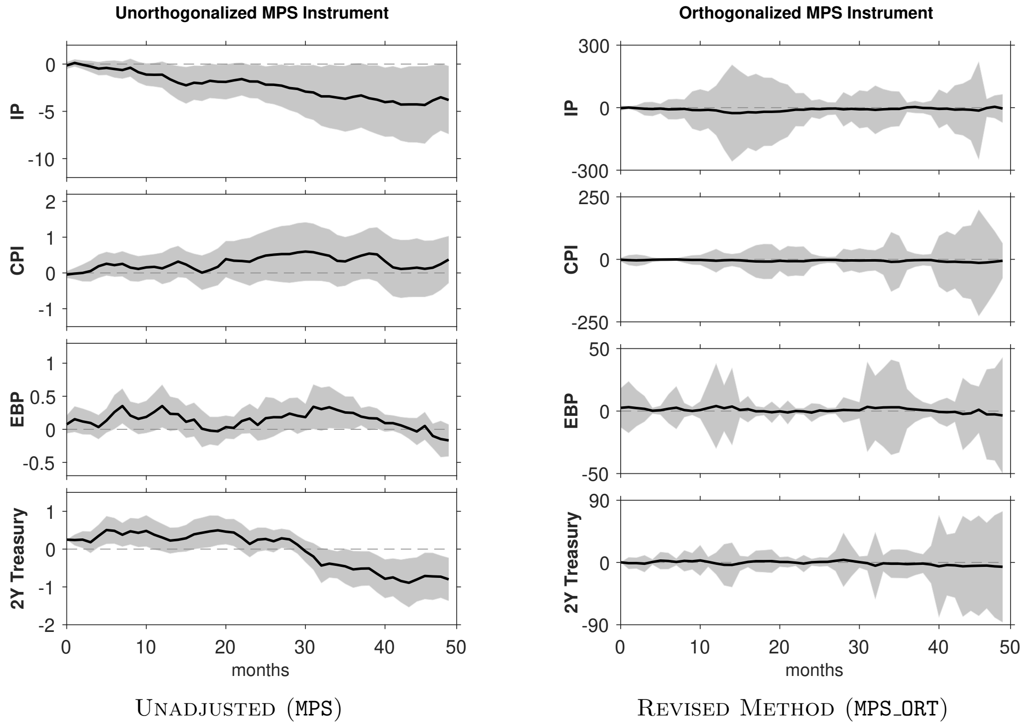


Figure F.1: Local-projection impulse responses to a 25bp policy shock. Left: unadjusted high-frequency `MPS` measure (public-data FOMC-only benchmark, closest analogue to BS Figure 5 / Appendix Figure C.1). Right: announcement-level orthogonalized series `MPS_ORT`. Sample: 1988:02–2020:02. Lags: 12. Shaded regions report 90% Newey–West standard-error bands. EBP = excess bond premium, CPI = Consumer Price Index, IP = ind. production.

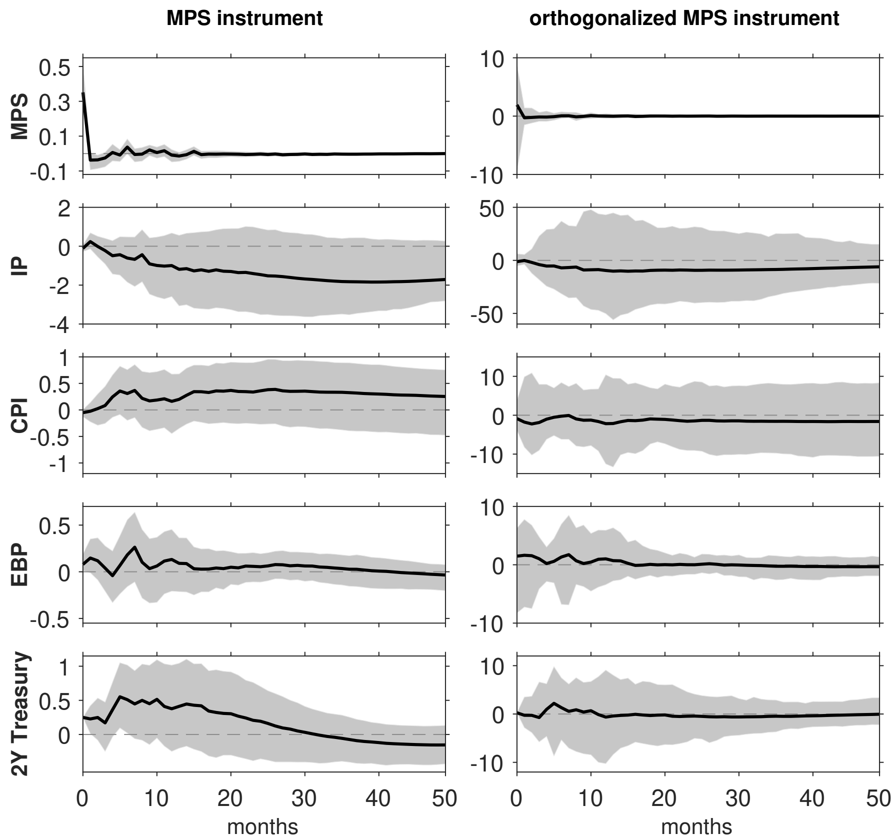


Figure F.2: Recursive SVAR with internal instrument: public-data analogue to BS Figure 6. Left: unadjusted high-frequency `MPS` measure. Right: corrected announcement-level orthogonalized series `MPS_ORT`. The instrument is included in the VAR and ordered first, following the internal-instrument approach of Plagborg-Møller and Wolf (2021b). Public-data sample: 1988:02–2019:12. The first non-missing monthly surprise in the posted workbook is 1988:02. Lags: 12. Shaded regions report bootstrapped 90% standard-error bands. EBP = excess bond premium, CPI = Consumer Price Index, IP = ind. production.

The same speech-data boundary applies to BS Figure 6. Its exact published internal-instrument SVAR cannot be replicated from the public files because the posted surprise data exclude Fed Chair speeches, and the paper does not provide a corresponding no-speeches appendix figure. As a feasible substitute, Figure F.2 reports a public-data internal-instrument analogue built from the FOMC-only workbook used throughout this paper. The left panel uses the unadjusted monthly surprise series `MPS`; the right panel replaces it with the corrected announcement-level series `MPS_ORT`. Consistent with the proxy-SVAR and LP results, the corrected orthogonalized internal-instrument responses display no recognizable macro effects: the point estimates are largely flat relative to the confidence bands, and statistical significance disappears throughout.

G. Projection coefficients and policy-rule coefficients

Projection coefficients from surprise regressions need not equal structural policy-rule coefficients. Suppose

$$i_t = \phi x_t + \varepsilon_t, \qquad \mathbb{E}_{t^-}[i_t] = \bar{\phi} x_t. \tag{G.1}$$

Then the high-frequency monetary surprise is

$$mps_t \equiv i_t - \mathbb{E}_{t^-}[i_t] = (\phi - \bar{\phi}) x_t + \varepsilon_t, \tag{G.2}$$

so projecting mps_t on x_t gives

$$mps_t = \beta x_t + u_t, \qquad \beta = \phi - \bar{\phi}. \tag{G.3}$$

Thus β equals the policy-rule coefficient ϕ only in the knife-edge case $\bar{\phi} = 0$; in general it is a reduced-form predictability coefficient. The gap can be wider when omitted signals also matter: if the Fed reacts to w_t and $w_t = \lambda x_t + \eta_t$ with $\mathbb{E}[x_t \eta_t] = 0$, then the surprise-projection coefficient also loads on $(\phi_w - \bar{\phi}_w)\lambda$.

This distinction also clarifies the decomposition of the gap between sum-then-purge and purge-then-sum. In a two-announcement month, let

$$mps_{1m} = \beta_1' X_{1m} + u_{1m}, \qquad mps_{2m} = \beta_2' X_{2m} + u_{2m}, \tag{G.4}$$

and write the linear projection of later-announcement predictors on first-announcement predictors as

$$X_{2m} = \Pi X_{1m} + \eta_{2m}, \qquad \mathbb{E}[X_{1m} \eta_{2m}'] = 0. \tag{G.5}$$

Projecting the monthly sum $mps_{1m} + mps_{2m}$ on X_{1m} gives

$$\gamma = \beta_1 + \Pi' \beta_2. \tag{G.6}$$

Hence

$$z_m - \tilde{z}_m = \beta_2' X_{2m} + (\beta_1 - \gamma)' X_{1m} \tag{G.7}$$

$$= \beta_2'(X_{2m} - \Pi X_{1m}) \tag{G.8}$$
$$= \beta_2'\eta_{2m}. \tag{G.9}$$

So the second term is best read as the part of later-announcement predictable variation absorbed into the monthly projection coefficient γ. It is a projection-coefficient effect, not automatically a structural policy-rule effect.

H. Proofs & Derivations

H.I Proof of equation (15)

Let $\ddot{\mathbf{y}} = M_W\mathbf{y}$, $u \equiv \ddot{\mathbf{z}}_S'\ddot{\mathbf{y}}$, $v \equiv \ddot{\mathbf{z}}_M'\ddot{\mathbf{y}}$, and recall $\ddot{\mathbf{z}} = \ddot{\mathbf{z}}_S + \ddot{\mathbf{z}}_M$, $V_S = \ddot{\mathbf{z}}_S'\ddot{\mathbf{z}}_S$, $C_{SM} = \ddot{\mathbf{z}}_S'\ddot{\mathbf{z}}_M$, $V_M = \ddot{\mathbf{z}}_M'\ddot{\mathbf{z}}_M$.

Restricted regression (13) *(the pooled one-regressor specification).* By FWL, $\hat{\beta}_{R,h}$ equals the OLS coefficient from regressing $\ddot{\mathbf{y}}$ on $\ddot{\mathbf{z}}$:

$$\hat{\beta}_{R,h} \;=\; \frac{\ddot{\mathbf{z}}'\ddot{\mathbf{y}}}{\ddot{\mathbf{z}}'\ddot{\mathbf{z}}} \;=\; \frac{u+v}{V_S + 2C_{SM} + V_M}. \tag{H.1}$$

Partitioned regression (14). By FWL, $(\hat{\beta}_{1,h}, \hat{\beta}_{2,h})$ are the OLS coefficients from regressing $\ddot{\mathbf{y}}$ on $(\ddot{\mathbf{z}}_S, \ddot{\mathbf{z}}_M)$. The normal equations are

$$\begin{bmatrix} V_S & C_{SM} \\ C_{SM} & V_M \end{bmatrix} \begin{bmatrix} \hat{\beta}_{1,h} \\ \hat{\beta}_{2,h} \end{bmatrix} \;=\; \begin{bmatrix} u \\ v \end{bmatrix},$$

with solution (assuming no exact collinearity, $V_SV_M > C_{SM}^2$)

$$\hat{\beta}_{1,h} \;=\; \frac{V_M\,u - C_{SM}\,v}{V_SV_M - C_{SM}^2}, \qquad \hat{\beta}_{2,h} \;=\; \frac{V_S\,v - C_{SM}\,u}{V_SV_M - C_{SM}^2}. \tag{H.2}$$

Substitute (H.2) into $\omega_{1,h}\hat{\beta}_{1,h}+\omega_{2,h}\hat{\beta}_{2,h}$ with $\omega_{1,h} = (V_S+C_{SM})/(V_S+2C_{SM}+V_M)$ and $\omega_{2,h} = (C_{SM}+V_M)/(V_S+2C_{SM}+V_M)$:

$$\omega_{1,h}\hat{\beta}_{1,h}+\omega_{2,h}\hat{\beta}_{2,h} \;=\; \frac{1}{V_S+2C_{SM}+V_M}\cdot\frac{(V_S+C_{SM})(V_M\,u - C_{SM}\,v) + (C_{SM}+V_M)(V_S\,v - C_{SM}\,u)}{V_SV_M - C_{SM}^2}.$$

Expand the numerator:

$$\begin{aligned}(V_S+C_{SM})(V_M\,u - C_{SM}\,v) + (C_{SM}+V_M)(V_S\,v - C_{SM}\,u) &= (V_SV_M - C_{SM}^2)\,u \;+\; (V_SV_M - C_{SM}^2)\,v \\ &= (V_SV_M - C_{SM}^2)\,(u+v).\end{aligned}$$

Therefore

$$\omega_{1,h}\hat{\beta}_{1,h} + \omega_{2,h}\hat{\beta}_{2,h} \;=\; \frac{(V_SV_M - C_{SM}^2)(u+v)}{(V_S+2C_{SM}+V_M)(V_SV_M - C_{SM}^2)}$$

$$= \frac{u+v}{V_S + 2C_{SM} + V_M}$$

$$\overset{(H.1)}{=} \hat{\beta}_{R,h}. \qquad \blacksquare$$

H.II Derivation of Equation (8)

Fix a month m with exactly two announcements. Work in the scalar case $k = 1$ and suppress the month index where harmless. Assume the event-level predictive equation

$$\mathtt{MPS}_{tm} = \beta X_{t\text{-}m} + u_{tm}, \qquad t \in \{1, 2\}, \tag{H.3}$$

with the orthogonality conditions

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

The second condition is what is needed for the monthly OLS coefficient below. Define the purge-then-sum instrument

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

Define the sum-then-purge monthly surprise

$$\mathtt{MPS}_m := \mathtt{MPS}_{1m} + \mathtt{MPS}_{2m}, \tag{H.6}$$

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

$$\mathtt{MPS}_m = \gamma X_{1\text{-}m} + z_m, \qquad \gamma = \frac{\mathrm{Cov}(\mathtt{MPS}_m, X_{1\text{-}m})}{\mathrm{Var}(X_{1\text{-}m})}, \tag{H.7}$$

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

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

$$\begin{aligned} \tilde{z}_m &= (\beta X_{1\text{-}m} + u_{1m} - \beta X_{1\text{-}m}) + (\beta X_{2\text{-}m} + u_{2m} - \beta X_{2\text{-}m}) && \text{(H.8)} \\ &= u_{1m} + u_{2m}. && \text{(H.9)} \end{aligned}$$

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

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

$$\begin{aligned} \mathtt{MPS}_m &= \mathtt{MPS}_{1m} + \mathtt{MPS}_{2m} && \text{(H.10)} \\ &= (\beta X_{1\text{-}m} + u_{1m}) + (\beta X_{2\text{-}m} + u_{2m}) && \text{(H.11)} \\ &= \beta X_{1\text{-}m} + \beta X_{2\text{-}m} + (u_{1m} + u_{2m}). && \text{(H.12)} \end{aligned}$$

Compute the OLS coefficient γ from (H.7),

$$\gamma = \frac{\mathrm{Cov}(\mathtt{MPS}_m, X_{1\text{-}m})}{\mathrm{Var}(X_{1\text{-}m})}. \tag{H.13}$$

Insert (H.12):

$$\mathrm{Cov}(\mathtt{MPS}_m, X_{1\text{-}m}) = \mathrm{Cov}(\beta X_{1\text{-}m} + \beta X_{2\text{-}m} + (u_{1m} + u_{2m}), X_{1\text{-}m}) \tag{H.14}$$

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

$$\gamma = \beta + \beta \frac{\mathrm{Cov}(X_{2-m}, X_{1-m})}{\mathrm{Var}(X_{1-m})}. \tag{H.15}$$

Define

$$\rho := \frac{\mathrm{Cov}(X_{2-m}, X_{1-m})}{\mathrm{Var}(X_{1-m})}. \tag{H.16}$$

Then

$$\gamma = \beta + \beta\rho = \beta(1 + \rho). \tag{H.17}$$

Interpretation of ρ. The quantity ρ is the slope coefficient in the population linear projection of X_{2-m} on X_{1-m} (in the mean-zero representation):

$$\Pi(X_{2-m} \mid X_{1-m}) = \rho X_{1-m}. \tag{H.18}$$

Therefore $X_{2-m} - \rho X_{1-m}$ 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 (H.7) and (H.12),

$$\begin{aligned} z_m &= \mathtt{MPS}_m - \gamma X_{1-m} && \text{(H.19)} \\ &= [\beta X_{1-m} + \beta X_{2-m} + (u_{1m} + u_{2m})] - (\beta + \beta\rho) X_{1-m}. && \text{(H.20)} \end{aligned}$$

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

$$\begin{aligned} z_m &= (u_{1m} + u_{2m}) + \beta X_{2-m} - \beta\rho X_{1-m} && \text{(H.21)} \\ &= (u_{1m} + u_{2m}) + \beta(X_{2-m} - \rho X_{1-m}). && \text{(H.22)} \end{aligned}$$

Using (H.9),

$$z_m = \tilde{z}_m + \beta(X_{2-m} - \rho X_{1-m}). \tag{H.23}$$

which is the desired decomposition.

What exactly remains in z_m? Equation (H.23) shows that sum-then-purge equals purge-then-sum plus the projection residual from second-announcement information. Thus:

- if $\beta = 0$, there is no event-level predictability and the two instruments coincide;
- if $X_{2-m} = \rho X_{1-m}$ 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 $\beta(X_{2-m} - \rho X_{1-m})$.

H.III Derivation of the bias equation

This appendix derives the probability limit of the estimator in equation (10) under the convention of Bauer and Swanson (2023a)[32]: ε_m is a latent structural shock, and $z_m, \tilde{z}_m$ are candidate instruments for ε_m. Throughout, $\kappa := \mathrm{Cov}(\varepsilon_m, z_m)$ denotes the covariance that appears in the bias formula.

All objects are defined in Section III.I: $k = 1$, two announcements per month. From (8), the corrected purge-then-sum instrument is $\tilde{z}_m = u_{1m} + u_{2m}$, and the sum-then-purge instrument is $z_m = \tilde{z}_m + \beta\,\eta_m$, where

$$\eta_m \;:=\; X_{2-m} - \rho\, X_{1-m} \tag{H.24}$$

is the innovation in the second-announcement news beyond what is linearly spanned by the first.[33] The outcome at horizon h follows

$$Y_{m+h} = \Theta_h\,\varepsilon_m + \Psi_{1h}\,X_{1-m} + \Psi_{2h}\,X_{2-m} + \xi_{m+h}, \qquad \mathbb{E}[\xi_{m+h} \mid \Omega_{m-1}] = 0. \tag{H.25}$$

The numerator: $\mathrm{Cov}(Y_{m+h},\, z_m)$**.** Substituting (H.25):

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

I evaluate each remaining covariance in turn. First, $\mathrm{Cov}(X_{1-m}, z_m) = 0$: z_m is the OLS residual from projecting $\sum_t \mathtt{MPS}_{tm}$ on X_{1-m}, so this follows from OLS orthogonality. Second, writing $X_{2-m} = \rho\, X_{1-m} + \eta_m$ and substituting $z_m = \tilde{z}_m + \beta\eta_m$:

$$\mathrm{Cov}(X_{2-m}, z_m) = \underbrace{\mathrm{Cov}(X_{2-m}, \tilde{z}_m)}_{=\,0} + \beta\; \underbrace{\mathrm{Cov}(X_{2-m}, \eta_m)}_{=\,\mathrm{Var}(\eta_m)} = \beta\; \mathrm{Var}(\eta_m), \tag{H.26}$$

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

$$\mathrm{Cov}(\varepsilon_m, z_m) = \mathrm{Cov}(\varepsilon_m, \tilde{z}_m) + \beta\; \underbrace{\mathrm{Cov}(\varepsilon_m, \eta_m)}_{=\,0} = \kappa. \tag{H.27}$$

Collecting terms:

$$\mathrm{Cov}(Y_{m+h},\, z_m) = \Theta_h\,\kappa + \Psi_{2h}\,\beta\; \mathrm{Var}(\eta_m). \tag{H.28}$$

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

[32] Endnote 41 makes this explicit.

[33] Throughout I maintain that each event-level residual u_{tm} is uncorrelated with news realized at other announcements, $\mathrm{Cov}(u_{tm}, X_{s-m}) = 0$ for $s \neq t$, consistent with the BS-framework interpretation of u_{tm} as the structural monetary shock at announcement (t, m). OLS orthogonality only delivers $\mathrm{Cov}(u_{tm}, X_{t-m}) = 0$; the cross-event piece is what makes $\tilde{z}_m$ a valid instrument rather than merely a projection residual.

The denominator and the main result (Estimand B). In BS' proxy-SVAR, z_m instruments for the unobserved structural shock ε_m, and the impulse response is recovered by dividing the reduced-form covariance by the first-stage covariance.[34] The denominator is therefore $\mathrm{Cov}(\varepsilon_m, z_m)$ from (H.27) — crucially, this is *not* $\mathrm{Var}(z_m)$, which equals $\mathrm{Var}(\tilde{z}_m) + \beta^2 \mathrm{Var}(\eta_m) > \mathrm{Var}(\tilde{z}_m)$. Dividing (H.28) by $\mathrm{Cov}(\varepsilon_m, z_m)$:

$$\widehat{\Theta}_h^{(z)} := \frac{\mathrm{Cov}(Y_{m+h},\, z_m)}{\mathrm{Cov}(\varepsilon_m,\, z_m)} = \Theta_h + \Psi_{2h}\, \beta\, \frac{\mathrm{Var}(\eta_m)}{\mathrm{Cov}(\varepsilon_m, z_m)}. \tag{H.29}$$

This is equation (10) in the body: the true impulse response Θ_h plus a contamination term governed by the news-outcome effect Ψ_{2h}, the predictability coefficient β, and the ratio of news-innovation variance to first-stage relevance. For the clean instrument: $\widehat{\Theta}_h^{(\tilde{z})} = \Theta_h$.

Comparison: reduced-form OLS (Estimand A). The simplest alternative — regressing Y_{m+h} directly on z_m by OLS — uses $\mathrm{Var}(z_m)$ as denominator rather than $\mathrm{Cov}(\varepsilon_m, z_m)$:

$$\theta_h^{(z)} := \frac{\mathrm{Cov}(Y_{m+h},\, z_m)}{\mathrm{Var}(z_m)} = \frac{\Theta_h\, \kappa + \Psi_{2h}\, \beta\ \mathrm{Var}(\eta_m)}{\mathrm{Var}(\tilde{z}_m) + \beta^2\ \mathrm{Var}(\eta_m)}. \tag{H.30}$$

This reduced-form coefficient combines Θ_h and Ψ_{2h}/β, with Θ_h scaled by $\kappa/\mathrm{Var}(z_m)$ — the classical measurement-error attenuation from z_m being a noisy proxy for ε_m. The contamination has the same source as in Estimand B (the $\Psi_{2h}\beta\,\mathrm{Var}(\eta_m)$ term in the numerator), but the larger denominator also shrinks the structural coefficient Θ_h toward zero. The standard IV relationship links the two: $\widehat{\Theta}_h^{(z)} = \theta_h^{(z)} / \pi^{(z)}$, where $\pi^{(z)} := \mathrm{Cov}(\varepsilon_m, z_m)/\mathrm{Var}(z_m)$ is the first-stage coefficient. The IV estimand undoes the attenuation but does not undo the news contamination. For $\tilde{z}_m$, the IV estimand equals Θ_h; the reduced-form coefficient coincides with it only under the additional normalization $\mathrm{Var}(\tilde{z}_m) = \mathrm{Cov}(\varepsilon_m, \tilde{z}_m)$.

I. Event clustering and predictor homogeneity

One possible concern is that the event-level purging regression in equation (1) gives disproportionate influence to months with clustered announcements. If multi-announcement months obeyed a different predictor-to-surprise relationship from single-announcement months, the pooled purge could tilt the estimated coefficient vector toward those episodes and mechanically contribute to the decomposition in Section IV.I.

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

This concern matters only if event clustering changes the object being estimated. Under homogeneous event-level coefficients, unequal spacing affects efficiency but not the probability limit of the purge. The relevant question is therefore empirical: do single- and multi-announcement months display different purging slopes, and does equalizing month-level influence materially change the monthly instrument?

I assess this issue in two ways. First, I interact the full predictor vector in equation (1) with an indicator for months with at least two announcements and test the interaction block jointly. Second, I re-estimate the purging regression with weights $1/T_m$, so that each month receives equal total influence, and compare the resulting monthly purge-then-sum instrument $\tilde{z}_m^w$ with the baseline series $\tilde{z}_m$.

Table I.1 reports the coefficient-homogeneity test. The data do not reject common purging coefficients across single- and multi-announcement months: the joint test yields $F(7, 334) = 1.00$ with $p = 0.43$. Table I.2 reports the weighting exercise. Reweighting the event-level purge leaves the monthly instrument essentially unchanged: the correlation between $\tilde{z}_m$ and $\tilde{z}_m^w$ is 0.999 overall, and the largest absolute gap is 0.022, attained in October 2008.

Table I.1: Homogeneity test: single- vs. multi-event months

	Coefficient	Robust SE	t	p
2+ announcements (intercept shift)	−0.010	0.015	−0.71	0.48
2+ announcements × NFP surprise	0.000	0.000	0.18	0.86
2+ announcements × NFP 12-month	0.005	0.006	0.93	0.35
2+ announcements × S&P 500 3-mo.	−0.178	0.106	−1.67	0.09
2+ announcements × Slope 3-mo.	0.019	0.019	1.01	0.31
2+ announcements × Commodity 3-mo.	−0.102	0.089	−1.16	0.25
2+ announcements × Skewness	0.012	0.024	0.51	0.61
Joint $F(7, 334) = 1.00$, $p = 0.43$				
$R^2_{\text{restricted}} = 0.160$, $R^2_{\text{unrestricted}} = 0.178$				

Notes: The unrestricted model fully interacts all six predictors in equation (1) with an indicator for months with at least two announcements. HC1-robust standard errors. Diagnostics use the public event-level sample through 2023, excluding March–December 2020; the same qualitative conclusion holds on the baseline pre-2020 sample.

Table I.2: Sensitivity of the monthly instrument to $1/T_m$ reweighting

	All event-months	1 announcement	2+ announcements
Months	292	261	31
$\text{Corr}(\tilde{z}_m, \tilde{z}_m^w)$	0.999	0.999	0.998
Mean $\lvert\tilde{z}_m - \tilde{z}_m^w\rvert$	0.002	0.002	0.005
Max $\lvert\tilde{z}_m - \tilde{z}_m^w\rvert$	0.022	0.013	0.022

Notes: $\tilde{z}_m^w$ uses residuals from the event-level purging regression estimated with weights $1/T_m$, so that each month receives equal total influence. The largest absolute difference occurs in October 2008. $\text{SD}(\tilde{z}_m) = 0.056$, so the maximum gap corresponds to about 0.39 standard deviations of the baseline series.

Taken together, these diagnostics indicate that the decomposition in Section IV.I is not being driven by a different event-level purging regression in clustered months, nor by the mechanical overweighting of those months in the first-stage purge. Event clustering is a legitimate concern in principle, but in this application it appears to have limited empirical bite.