Identification or Propagation? Shock-Based Weak-IV Diagnostics and Bootstrap Inference for Instrumented VARs*

Daniele Colombo†

This version: November 28, 2025.

Preliminary draft, newest version here.

Abstract

This paper documents an inconsistency in prevailing practice: instrument relevance in structural VARs is often assessed using an anchor equation's reduced-form residual rather than the identified structural shock. As a result, the test is anchor-dependent and does not directly target the moment condition underlying identification. I propose a shock-based, anchor-independent weak instrument test based on the usual F-statistic—together with a bootstrap procedure that normalizes sign on the recovered shock—that yields invariant and stronger diagnostics, as well as typically tighter inference. I also show that when SVARs produce "structural" IRFs that are statistically indistinguishable from reduced-form dynamic multipliers, the instrument is likely weak, mis-specified, or simply superfluous.

JEL classification: C32, C36, C52, E32, E52.

Keywords: Instrumental variables, instrument relevance, reduced-form, structural IRFs.

*I am grateful to Raffaella Giacomini, James D. Hamilton, Alexey Onatski, James H. Stock, and Mark W. Watson for helpful discussions that inspired this research.

†Department of Economics, London Business School, UK. Email: dcolombo@london.edu

1 Introduction

In recent years, instrumental variables have become a widely used tool for identifying structural shocks in macroeconomic time-series models in a variety of settings (e.g. Stock and Watson, 2012; Mertens and Ravn, 2013; Gertler and Karadi, 2015; Känzig, 2021). The key idea is to use an observed instrument z_t that is informative about a particular unobserved structural shock $w_{k,t}$ of interest, while being otherwise orthogonal to the remaining shocks in the system. Formally, identification relies on two moment restrictions: a relevance condition, $\mathbb{E}[z_t w_{k,t}] \neq 0$, and an exclusion (or exogeneity) condition, $\mathbb{E}[z_t w_{\ell,t}] = 0$ for all $\ell \neq k$. Under these restrictions, the instrument identifies the shock of interest, $w_{k,t}$, up to normalization.

However, since the structural shock $w_{k,t}$ is typically unobserved, in practice, instrument strength is usually evaluated by regressing a particular equation's residual $u_{j,t}$ (the "anchor" or "target" variable) on z_t and reporting the first-stage F-statistic (e.g. Stock and Watson, 2012; Gertler and Karadi, 2015; Känzig, 2021; Montiel Olea et al., 2021). This procedure tests $\mathbb{E}[z_t u_{j,t}] \neq 0$ rather than the identification-relevant condition $\mathbb{E}[z_t w_{k,t}] \neq 0$, though the two are related. When instrument strength is assessed in this way, the reported first-stage F-statistic depends on the choice of anchor equation. Alternative, equally reasonable, anchors can yield different F-statistics even when the identified impulse responses coincide up to normalization. Consequently, rules-of-thumb such as "$F \geq 10$" are not informative in this setting.

Consider, as an example, a researcher who seeks to identify a supply shock using an arguably exogenous supply instrument (or proxy). A common implementation proceeds as follows: (i) regress the price residual on the instrument; (ii) report the associated F-statistic as a measure of instrument relevance; (iii) project the residuals of the remaining variables (e.g., quantity, income) on the instrument; and (iv) normalize the estimated impact coefficients and impulse responses to a specified price increase (e.g., 10%). However, it may be equally reasonable to use the same proxy to instrument quantity, rather than price, in step (i). This would yields a different first-stage F, potentially altering conclusions about instrument weakness, while delivering identical impulse responses up to scale. Moreover, under standard bootstrap implementations, the resulting confidence bands can also depend on the equation chosen as anchor, because the bootstrap typically enforces a sign normalization based on the sign of the correlation between the instrument and the chosen residual.

Building on Stock and Watson (2018), who show that the structural shock series can be recovered even under partial identification, I propose an anchor-invariant diagnostic. The diagnostic computes the F-statistic from a regression of the recovered shock index on the instrument, thereby targeting the moment $\mathbb{E}[z_t w_{k,t}] \neq 0$ directly and aligning with the identification condition. I show that the resulting F-statistic is weakly larger than that obtained from regressing the instrument on any anchor residual, implying that some sources of exogenous variation deemed weak by anchor-based tests may in fact not be. As an illustrative example, I show that the FF4 instrument of Gertler and Karadi (2015) fails the conventional $F \geq 10$ threshold under the anchor-based test on their full 1979M7 to 2012M6 estimation sample but satisfies it when the shock-based F is used. I also propose a bootstrap procedure that normalizes the sign of the impulse responses in each draw using the correlation between the instrument and the recovered structural shock, thereby delivering unique, anchor-independent confidence bands. By fixing the sign through the shock index—which aggregates the instrument's comovement with all reduced-form residuals—this procedure avoids the instability that arises when the chosen anchor's impact response is close to zero and yields more stable inference across bootstrap draws, typically resulting in narrower confidence bands.

A second important question is whether identification with instruments meaningfully changes the identified impulse responses relative to the reduced-form propagation implied by the VAR. Since Stock (2008) popularized the proxy-VAR approach, considerable effort has gone into constructing credibly exogenous instruments to identify macroeconomic shocks. Yet, in practice the resulting impulse responses are often statistically indistinguishable from reduced-form dynamic multipliers, especially in high-frequency settings (see e.g., Känzig, 2021; Alessandri and Gazzani, 2025; Känzig, 2025). When this occurs, the value added by using an instrument is questionable: the recursive (Cholesky) restrictions—that non-target variables adjust only after the shocked variable—are satisfied, and the resulting responses are qualitatively similar. Although the researcher should not choose the identification scheme based on realized responses, interpretation even with a credibly exogenous instrument ultimately rests on economic restrictions. Thus, when theory already implies limited contemporaneous adjustment outside the target variable, constructing an instrument could be superfluous.

Otherwise, the resemblance of the "structural" impulse responses to the reduced form signals potential problems with the instrument or specification: (i) the proxy may not be fully exogenous and could be capturing a reduced-form relation; (ii) the proxy may be weak because it loads primarily on a single equation's residual; or (iii) a relevant "fast-moving" variable may be omitted, preventing identification from departing from the reduced form. As a matter of practice, I argue that it is always useful to report the reduced-form dynamic multipliers—and, where appropriate, impulse responses from a recursive ordering—alongside the main SVAR impulse-responses. Juxtaposing these objects allows to assess whether the instrument contributes information beyond VAR propagation and to gauge how much the findings hinge on the instrument rather than on the model's dynamics.

Given the resource costs of building credible instruments, their use should be justified by informational gains. Credibility is enhanced when the instrument produces impulse responses that are not fully driven by reduced-form propagation—i.e., when variables beyond the shocked equation display economically meaningful impact responses consistent with the identification.

Layout. The remainder of the paper is structured as follows. Section 2 reviews the proxy-VAR framework, which serves as the primary model for exposition. Section 3 discusses the use of anchor variables for normalization and the computation of the first-stage F-statistic to assess instrument relevance, and an anchor-independent bootstrap procedure for inference. Section 4 discusses cases in which external-instrument identification yields impulse responses that are statistically indistinguishable from their reduced-form counterparts. Section 5 argues that these issues are not unique to the proxy-VAR framework and can also arise in internal-instrument VARs and LP-IV settings. Finally, Section 6 concludes.

2 The standard proxy-VAR setting

Consider the following structural VAR(p) model:

$$\mathbf{B}_0\mathbf{y}_t = \mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \mathbf{w}_t, \tag{1}$$

where $\mathbf{y}_t$ is a $K \times 1$ vector of endogenous variables, assumed to have zero mean without loss of generality. The vector $\mathbf{w}_t$ is a $K \times 1$ vector of structural shocks. As in most applications, we assume that the structural shocks $\mathbf{w}_t$ are mutually uncorrelated, i.e. $\mathbb{E}(\mathbf{w}_t\mathbf{w}_t') = \mathbf{\Sigma}_w$ is diagonal. Since the matrices $\mathbf{B}_0$ and $\mathbf{w}_t$ are generally unobserved, we rely on the reduced-form representation to estimate the model:

$$\begin{aligned}\mathbf{y}_t &= \mathbf{B}_0^{-1}\mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_0^{-1}\mathbf{B}_p\mathbf{y}_{t-p} + \mathbf{B}_0^{-1}\mathbf{w}_t \\ &= \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t,\end{aligned} \tag{2}$$

where the reduced-form parameters $\mathbf{A}_1, \ldots, \mathbf{A}_p, \mathbf{\Sigma}_u$, and the reduced-form residuals $\mathbf{u}_t$ are generally estimated by OLS or using Bayesian methods. The key equation linking the reduced-form innovations to the structural shocks is:

$$\mathbf{u}_t = \mathbf{B}_0^{-1}\mathbf{w}_t. \tag{3}$$

The objective is to identify the impact of a structural shock on the system. This corresponds to identifying one column of the impact matrix $\mathbf{B}_0^{-1}$.

Identification via external instruments. Let $\mathbf{b}_k$ denote the k-th column of the matrix $\mathbf{B}_0^{-1}$, representing the effect of a structural shock of interest, $w_{k,t}$, on all K variables in the system, and let $\mathbf{B}_{-k}$ collect all the other columns, with $k \in \{1, \ldots, K\}$. For expository purposes, we here set $k = 1$ without loss of generality. Further, let $\mathbf{z}_t \in \mathbb{R}^m$ denote a vector of instruments, with m the number of instruments, which satisfies the relevance and exogeneity conditions:

$$\mathbb{E}[\mathbf{z}_t w_{1,t}] \neq \mathbf{0} \tag{4}$$

$$\mathbb{E}[\mathbf{z}_t \mathbf{w}'_{2:K,t}] = \mathbf{0} \tag{5}$$

with $\mathbf{w}_t = (w_{1,t}, \mathbf{w}'_{2:K,t})'$. We can partition (3) as $\mathbf{u}_t = \mathbf{b}_1 w_{1,t} + \mathbf{B}_{2:K}\mathbf{w}_{2:K,t}$, and, transposing, premultiplying by the instrument vector and taking expectations, we get

$$\mathbb{E}[\mathbf{z}_t \mathbf{u}'_t] = \mathbb{E}[\mathbf{z}_t w_{1,t}]\mathbf{b}'_1 + \mathbb{E}[\mathbf{z}_t \mathbf{w}'_{2:K,t}]\mathbf{B}'_{2:K} = \mathbb{E}[\mathbf{z}_t w_{1,t}]\mathbf{b}'_1, \tag{6}$$

where the second equality follows from the exogeneity condition. Hence, for any index $j \in \{1, \ldots, K\}$ we have

$$\mathbb{E}[\mathbf{z}_t \mathbf{u}'_{-j,t}] = \mathbb{E}[\mathbf{z}_t u_{j,t}]\frac{\mathbf{b}'_{-j,1}}{b_{j,1}}, \tag{7}$$

where $b_{j,1}$ fixes the sign and scale of the shock, and $\mathbf{b}_{-j,1}$ contains the impact coefficients.

A commonly used consistent estimator for $\boldsymbol{\theta} := \dfrac{\mathbf{b}_{-j,1}}{b_{j,1}}$ is the 2SLS/IV estimator

$$\widehat{\boldsymbol{\theta}} = \left(U'_j P_Z U_j\right)^{-1} U'_j P_Z U_{-j}, \tag{8}$$

with $P_Z = Z(Z'Z)^{-1}Z'$.[1][2] Note that in the case of a single instrument ($m = 1$), after appropriate rescaling, the in-sample estimate $\widehat{\boldsymbol{\theta}}$ is invariant to the choice of j.[3] When $m > 1$,

[1]The relevance condition implies $\mathbb{E}[\mathbf{z}_t u_{j,t}] = \mathbb{E}[\mathbf{z}_t w_{1,t}]\, b_{j,1} \neq \mathbf{0}$ for any j with $b_{j,1} \neq 0$, and there must be at least one such j, unless the shock has no impact at all. We also require $\mathbb{E}[\mathbf{z}_t \mathbf{z}'_t] < \infty$.

[2]We define the data matrices $U_j := (u_{j,1}, \ldots, u_{j,T})'$, $U_{-j} := [\,\mathbf{u}_{-j,1}, \ldots, \mathbf{u}_{-j,T}\,]'$, and $Z := [\,\mathbf{z}_1, \ldots, \mathbf{z}_T\,]'$.

[3]In this case, z_t is scalar and the 2SLS formula simplifies to

$$\widehat{\theta}_k^{(j)} = \frac{U'_j P_Z U_k}{U'_j P_Z U_j} = \frac{Z' U_k}{Z' U_j} = \frac{s_k}{s_j}, \qquad k \neq j,$$

with $s_k := \frac{1}{T}\sum_{t=1}^{T} z_t u_{k,t}$. Hence $\widehat{\mathbf{b}}_1^{(j)} := \begin{bmatrix} 1 \\ \widehat{\boldsymbol{\theta}}^{(j)} \end{bmatrix} = \begin{bmatrix} 1 \\ s_{-j}/s_j \end{bmatrix}$ is the same vector for any j with $s_j \neq 0$. Another way to see this is that in the case of a single instrument, the fitted value from the first stage is just the instrument itself scaled by a constant. As a result, the second stage boils down to comparing how the invariance to the choice of j only holds asymptotically.

Normalization. The scale $b_{j,1}$, which represents the contemporaneous impact on $y_{j,t}$ of a one–unit change in $w_{1,t}$, can be set via any normalization subject to

$$\boldsymbol{\Sigma}_{\boldsymbol{u}} := \mathbb{E}(\mathbf{u}_t\mathbf{u}_t') = \mathbb{E}(\mathbf{B}_0^{-1}\mathbf{w}_t\mathbf{w}_t'(\mathbf{B}_0^{-1})') = \mathbf{B}_0^{-1}\mathbb{E}(\mathbf{w}_t\mathbf{w}_t')(\mathbf{B}_0^{-1})' = \mathbf{B}_0^{-1}\boldsymbol{\Sigma}_{\mathbf{w}}(\mathbf{B}_0^{-1})'. \tag{9}$$

Because $\mathbf{B}_0^{-1}$ and $\boldsymbol{\Sigma}_w$ are unobserved, the scale of column k of $\mathbf{B}_0^{-1}$ and the variance of the corresponding structural shock $\sigma^2_{w_k}$ are not separately identified. A common normalization choice is to rescale the structural shocks to have unit variance: $\boldsymbol{\Sigma}_{\tilde{w}} = \mathbf{I}_K$, under which a unit realization of $\tilde{w}_{k,t}$ equals a one–standard–deviation structural shock. This leads to the unit-variance representation of the model in (2) and (3):

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \tilde{\mathbf{B}}_0^{-1}\tilde{\mathbf{w}}_t \tag{10}$$

$$\tilde{\mathbf{w}}_t = \boldsymbol{\Sigma}_w^{-1/2}\mathbf{w}_t \tag{11}$$

$$\tilde{\mathbf{B}}_0^{-1} = \mathbf{B}_0^{-1}\boldsymbol{\Sigma}_w^{1/2} \tag{12}$$

Alternatively, one may generally normalize by setting $b_{j,1} = c$, which implies that a unit value of $w_{1,t}$ has a positive effect of magnitude c on $y_{j,t}$, and the structural shocks will have some variance different from one: $\boldsymbol{\Sigma}_{\boldsymbol{w}} = \mathrm{diag}(\sigma^2_{w_1}, \ldots, \sigma^2_{w_K})$.

Recovering the structural shocks. In the partially identified case, where not all columns of $\mathbf{B}_0^{-1}$ are recovered, the matrix cannot be directly inverted to obtain the structural shocks via equation (3). Nevertheless, Stock and Watson (2018) demonstrated that the corresponding structural shock can still be recovered using the following identities:

$$\begin{aligned}\tilde{\mathbf{b}}_k'\boldsymbol{\Sigma}_u^{-1}\mathbf{u}_t = \tilde{\mathbf{b}}_k'(\tilde{\mathbf{B}}_0)'\tilde{\mathbf{B}}_0\mathbf{u}_t = \mathbf{e}_j'\tilde{\mathbf{B}}_0\mathbf{u}_t = \mathbf{e}_j'\tilde{\mathbf{B}}_0\mathbf{B}_0^{-1}\mathbf{w}_t \\ = \mathbf{e}_j'\tilde{\mathbf{B}}_0\tilde{\mathbf{B}}_0^{-1}\boldsymbol{\Sigma}_w^{-1/2}\boldsymbol{\Sigma}_w^{-1/2}\tilde{\mathbf{w}}_t = \mathbf{e}_j'\tilde{\mathbf{w}}_t = \tilde{w}_{k,t}.\end{aligned} \tag{13}$$

where $\tilde{\mathbf{b}}_k$ denotes the k-th column of the $\tilde{\mathbf{B}}_0^{-1}$ matrix, $\mathbf{e}_j$ denotes the k-th standard basis vector and, without loss of generality, we adopted the unit-variance representation of the model given by equations (10), (11), and (12).

each residual co-moves with that same instrument. Hence, each entry of the impact vector is simply the projection of its corresponding residual onto the instrument (all defined up to a common scale and sign).

3 On the use of an "anchor variable" and the first-stage F-statistic.

Most proxy-VAR studies select an "anchor variable" that serves both to rescale impulse responses and to compute the first-stage F-statistic (e.g., the one-year rate in Gertler and Karadi, 2015 or the real price of crude oil in Känzig, 2021). While picking specifically these variables for rescaling is one of many possible normalizing options, it's often not recognized that they're not special for identification, they're merely a normalization/interpretation choice, not an identifying one, as should be clear from the discussion in Section 2.[4] The reliance on anchor-based normalizations has even led some studies to claim that, when the chosen anchor's loading is near zero, the IV estimator is undefined (Kilian et al., 2023).[5] Such claims reflect a fragile normalization, not a failure of identification, as normalization is arbitrary: one can normalize on any nonzero loading or on the magnitude of any impulse response at horizon $h \geq 0$.

Consistent with this fact, there is no compelling reason to compute the first-stage F-statistic from the correlation between the instrument and a particular anchor-variable residual (as is common also in methodological contributions like Stock and Watson, 2012; Montiel Olea et al., 2021; Jentsch and Lunsford, 2022). The relevance condition (4) is formulated for the structural shock, not for any single reduced-form residual. Accordingly, the diagnostic should be based on the shock index implied by (13), rather than on a specific equation's residual. The IRFs are invariant—up to a normalization—to the choice of anchor variable, since they are a linear combination of the dynamic multipliers and the impact vector $\mathbf{b}_k$, which—as discussed—is itself identified only up to scale (see equation 23 in the Appendix).

Example: natural gas demand and supply shock. Colombo and Toni (2025) identify a natural gas demand shock using exogenous temperature variation as an external instrument. This shock moves the real price of natural gas and the quantity demanded in the same direction, so either variable provides a reasonable anchor choice. Under the common practice of computing the first-stage F-statistic from a regression of the anchor equation's reduced-

[4] Because the impact vector is identified only up to sign, conventional bootstrap inference (see e.g. Jentsch and Lunsford, 2019) requires a sign normalization. If the anchor variable's impact is close to zero, sampling variation can flip its sign across bootstrap draws, yielding misaligned ("flipped") impulse responses and artificially wide intervals. Choosing an anchor with a sizable contemporaneous loading reduces such sign flips, but the resulting normalization remains arbitrary.

[5] Kilian et al. (2023) erroneously claim that *"when constructing the IV estimator one needs to normalize the responses by scaling the response of interest by the impact response of the variable being instrumented. In our case, one would instrument the TFP residual by the TFP news instrument. [...] one cannot rule out that the impact response of TFP to news is zero in population, resulting in responses that are infinite. This means that the IV estimator of the VAR responses is not well defined in general."*

form residual on the instrument, the reported strength depends on the anchor choice. In this application, using quantity (gas consumption) yields $F = 120.11$ (heteroskedasticity–robust $F = 99.06$), whereas using price yields $F = 26.56$ (robust $F = 14.07$). Which of the two is the correct F-statistic? The top panel of Figure 1 shows that the identified impulse-response point estimates are identical regardless of the chosen anchor variable. The same issue arises for inference: which set of confidence bands should be reported (green, using quantity as the anchor; red, using price as the anchor)? The IRFs are rescaled to a 10% point estimate increase in the natural gas price on impact.[6]

Colombo and Toni (2025) also identify a natural gas supply shock using a news-based high-frequency strategy. In this case, using quantity (gas availability, defined as domestic production plus net imports) yields $F = 5.98$ (robust $F = 12.92$), whereas using price yields $F = 1.38$ (robust $F = 1.32$). The anchor dependence is more pronounced because the supply instrument is generally less correlated with all the reduced-form residuals. While both anchors would lead to the same qualitative conclusion about instrument weakness under conventional first-stage thresholds, the implications for inference differ sharply: with quantity as the anchor, the resulting confidence bands render the responses statistically indistinguishable from zero at all horizons, whereas price-based normalization does not.

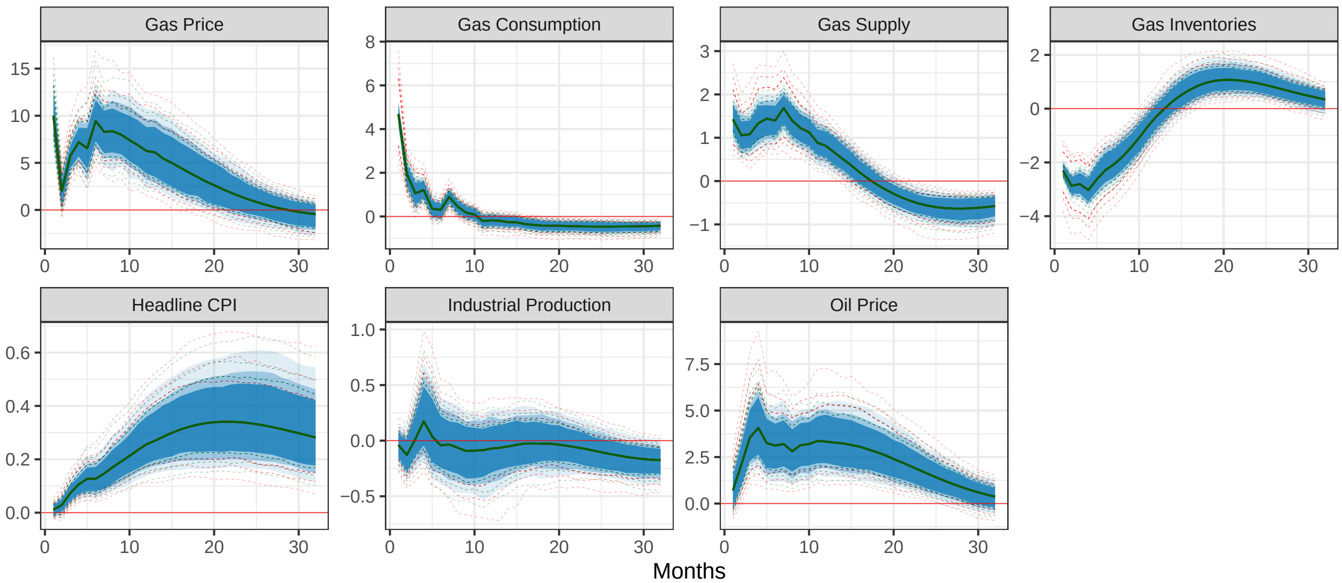
Anchor-independent diagnostics. Using equation (13), we can rewrite the relevance condition (4) for the k^{th} shock of interest as a function of observable objects:

$$\mathbb{E}[\mathbf{z}_t \tilde{w}_{k,t}] = \mathbb{E}[\mathbf{z}_t \tilde{\mathbf{b}}_k' \boldsymbol{\Sigma}_u^{-1} \mathbf{u}_t] \neq \mathbf{0}. \tag{14}$$

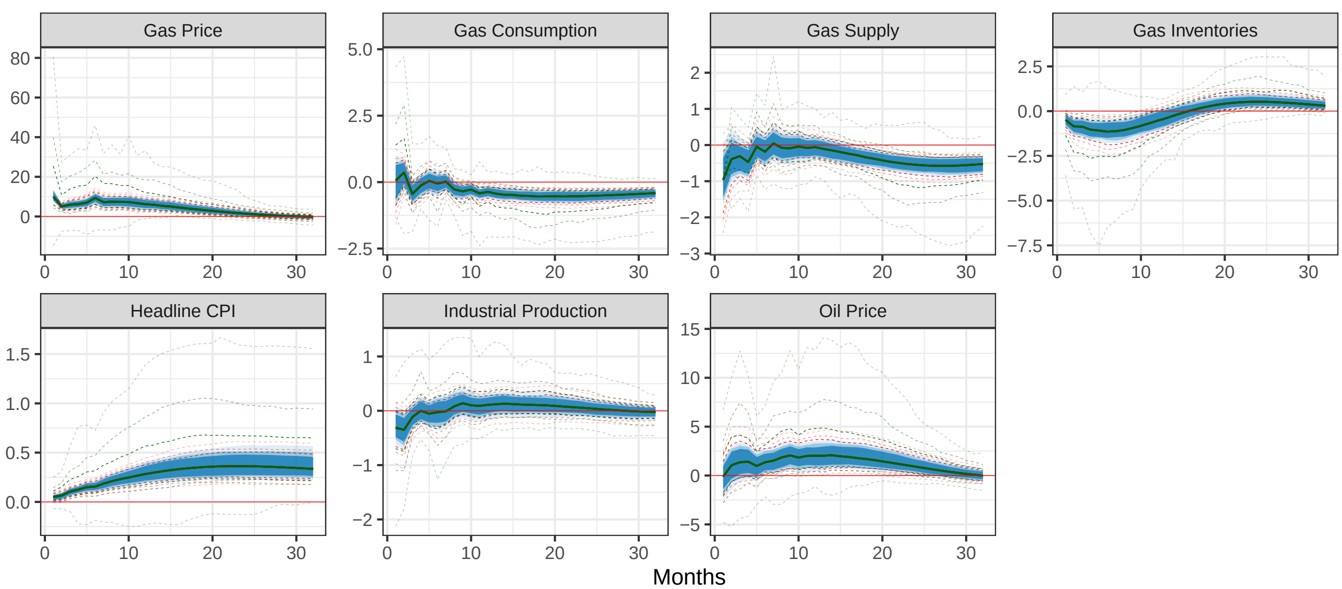
Therefore, the appropriate first-stage F-statistic can be computed as follows. First, estimate the reduced form VAR and obtain $\hat{\mathbf{u}}_t$ and $\hat{\boldsymbol{\Sigma}}_u$. Second, use the proxy-VAR/IV step to estimate the k-th impact vector up to scale, $\hat{\mathbf{b}}_k'$. Third, construct the shock index $\hat{w}_{k,t} = \hat{\mathbf{b}}_k' \hat{\boldsymbol{\Sigma}}_u^{-1} \hat{\mathbf{u}}_t$. Finally, regress $\hat{w}_{k,t}$ on $\mathbf{z}_t$ and compute the associated first-stage F-statistic. This F-statistic is unique and does not depend on the specific anchor variable that is chosen. In the natural gas application, it yields $F = 181.96$ (robust $F = 131.71$) for the demand instrument and $F = 11.03$ (robust $F = 7.99$) for the supply instrument. Notably, the supply instrument satisfies the conventional $F \geq 10$ threshold under the shock-based diagnostic, but fails to do so under any of the anchor-based measures.

This approach is consistent with Bayesian practice, in which instrument strength is assessed by the tightness of the posterior for the full set of impact coefficients rather than for a single, chosen coefficient (see, e.g., Caldara et al., 2019; Giacomini et al., 2022).

[6] To construct confidence bands under external-instrument identification, I use the moving block bootstrap proposed by Jentsch and Lunsford (2019).



Natural gas demand shock



Natural gas supply shock

Figure 1: *Baseline specification for natural gas demand (top panel) and supply (bottom panel) shocks in Colombo and Toni (2025). Green lines use quantity (gas consumption for demand; gas availability for supply) as the anchor, while red lines use price as the anchor. Solid lines show point estimates; dashed lines indicate 68%, 80%, and 90% confidence bands. Blue shaded bands report confidence intervals obtained under shock-based sign normalization.*

Notes: The original paper estimates the model using Bayesian methods. I adopt a frequentist approach for expositional clarity; the substantive implications are unchanged and do not hinge on the estimation method.

Relation to anchor-based F. The natural gas example suggests that the shock-based F-statistic can exceed its anchor-based counterparts. I now show that the F-statistic from regressing the recovered structural shock on the instrument is always weakly larger than any anchor-based F-statistic. From (6) we have $\mathbb{E}[\mathbf{z}_t\mathbf{u}_t'] = \mathbb{E}[\mathbf{z}_t\tilde{w}_{k,t}]\tilde{\mathbf{b}}_k'$, where the expression has been rewritten for a generic shock k and using the unit-variance representation without loss of generality. Hence we can write

$$\mathbb{E}[\mathbf{z}_t\tilde{w}_{k,t}]\tilde{\mathbf{b}}_k'\boldsymbol{\Sigma_u}^{-1}\tilde{\mathbf{b}}_k\mathbb{E}[\mathbf{z}_t\tilde{w}_{k,t}]' = \mathbb{E}[\mathbf{z}_t\mathbf{u}_t']\boldsymbol{\Sigma_u}^{-1}\mathbb{E}[\mathbf{z}_t'\mathbf{u}_t]$$

and, under unit-variance normalization, the expression simplifies to:[7]

$$\mathbb{E}[\mathbf{z}_t\tilde{w}_{k,t}]\mathbb{E}[\mathbf{z}_t\tilde{w}_{k,t}]' = \mathbb{E}[\mathbf{z}_t\mathbf{u}_t']\boldsymbol{\Sigma_u}^{-1}\mathbb{E}[\mathbf{z}_t'\mathbf{u}_t]$$

which, in the case of a single instrument, yields

$$\mathbb{E}[z_t\tilde{w}_{k,t}] = \sqrt{\mathbb{E}[z_t\mathbf{u}_t']\boldsymbol{\Sigma_u}^{-1}\mathbb{E}[\mathbf{u}_t'z_t]} \tag{15}$$

How does this compare to the commonly-used anchor moment $\mathbb{E}[z_t u_{k,t}]$? Using the Cauchy-Schwartz inequality we have[8]

$$\big|\mathbb{E}[z_t\tilde{w}_{k,t}]\big| \geq \frac{\big|\mathbb{E}[z_t u_{k,t}]\big|}{\sqrt{\Sigma_{u,kk}}}.$$

which we can rewrite in correlation terms (recalling that $\mathrm{Var}(\tilde{w}_{k,t}) = 1$):

$$\big|\rho(z_t,\tilde{w}_{k,t})\big| = \frac{\big|\mathbb{E}[z_t\tilde{w}_{k,t}]\big|}{\sigma_z} \geq \frac{\big|\mathbb{E}[z_t u_{k,t}]\big|}{\sigma_z\sqrt{\Sigma_{u,kk}}} = \big|\rho(z_t,u_{k,t})\big|. \tag{16}$$

Therefore, for a single instrument the *shock-based* relevance (and thus the associated first-stage F) is weakly larger than any anchor-based counterpart, for any k. Another way to see this is that the shock index $\tilde{w}_{k,t} = \tilde{\mathbf{b}}_k'\boldsymbol{\Sigma}_u^{-1}\mathbf{u}_t$ is an optimal linear combination of residuals,

[7] $\tilde{\mathbf{b}}_k'\boldsymbol{\Sigma}_u^{-1}\tilde{\mathbf{b}}_k = \tilde{\mathbf{b}}_k'\Big(\tilde{\mathbf{B}}_0'\tilde{\mathbf{B}}_0\Big)\tilde{\mathbf{b}}_k = \Big(\tilde{\mathbf{B}}_0\tilde{\mathbf{b}}_k\Big)'\Big(\tilde{\mathbf{B}}_0\tilde{\mathbf{b}}_k\Big) = \mathbf{e}_k'\mathbf{e}_k = 1.$

[8] For any symmetric positive definite matrix M there exists $M^{1/2}$ with $M^{1/2}M^{1/2} = M$ and $M^{-1/2}$. Set $u = M^{1/2}x$ and $v = M^{-1/2}y$. Then

$$x'y \;=\; u'v \;\Rightarrow\; |u'v|^2 \;\leq\; (u'u)(v'v) \;=\; (x'Mx)\,(y'M^{-1}y),$$

which follows from the Cauchy–Schwarz inequality. Taking $x = \mathbf{e}_k$, $y = \mathbf{s} := \mathbb{E}[z_t\mathbf{u}_t]$, and $M = \boldsymbol{\Sigma}_u$ yields

$$(\mathbf{e}_k'\mathbf{s})^2 \;\leq\; (\mathbf{e}_k'\boldsymbol{\Sigma}_u\mathbf{e}_k)\,(\mathbf{s}'\boldsymbol{\Sigma}_u^{-1}\mathbf{s}) \;\Rightarrow\; \frac{\big(\mathbb{E}[z_t u_{k,t}]\big)^2}{\Sigma_{u,kk}} \;\leq\; \mathbf{s}'\boldsymbol{\Sigma}_u^{-1}\mathbf{s} = \mathbb{E}[z_t\tilde{w}_{k,t}],$$

so that $\big|\mathbb{E}[z_t\tilde{w}_{k,t}]\big| \geq \frac{\big|\mathbb{E}[z_t u_{k,t}]\big|}{\sqrt{\Sigma_{u,kk}}}$.

so its covariance with the instrument cannot be smaller than that of any single residual equation.

The fact that the shock-based relevance is weakly larger than any anchor-based counterpart also holds in the case of multiple instruments.

Anchor-independent bootstrap. In line with this approach, I implement bootstrap inference using a structural shock–based sign normalization in each draw, which uses the available information in the instrument more efficiently than anchor-based normalization. Specifically, I normalize the sign of the estimated impact vector so that the recovered structural shock preserves a pre-specified sign of correlation with the instrument in every bootstrap iteration. This normalization is unique and does not depend on the choice of anchor; consequently, the resulting percentile confidence bands are also unique and anchor-independent. Moreover, because the sign is fixed using the shock index—which aggregates information from the instrument's correlation with all reduced-form residuals, rather than with a single chosen residual—the resulting bands are typically narrower at most horizons, as illustrated in Figure 1.

By contrast, conventional bootstrap implementations sign-normalize each draw using the anchor variable's impact response. This construction mechanically implies zero impact uncertainty for the anchor variable, so the confidence bands coincide with the point estimate on impact (e.g., see the red bands on impact for price and the green bands for quantity). Away from impact for the anchor variable, and at all horizons for non-anchor variables, anchor-based normalization is less stable: weaker correlation between the instrument and the chosen anchor residual increases the frequency of sign reversals across bootstrap draws, inflating dispersion and widening the bands. Normalizing on the shock index mitigates these sign reversals and yields tighter confidence sets.

4 On the difference between reduced-form and "structural" impulse responses

Since Stock (2008) popularized the proxy-VAR approach, a lot of effort has been put in constructing instruments that are credibly exogenous and that allow identification of important macro shocks of interest. However, in practice, in several applications, the impulse responses that are obtained are statistically indistinguishable from the reduced-form dynamic multipliers,[9] or to the IRFs obtained imposing a recursive (Cholesky) ordering, particularly

[9] By dynamic multipliers I mean the impulse responses that are purely reduced-form and that are obtained with the coefficients in $\mathbf{A}_1, \ldots, \mathbf{A}_p$ without identifying any coefficient of the impact matrix $\mathbf{B}_0^{-1}$. These are

in high-frequency identification settings (see e.g. Känzig, 2023; Alessandri and Gazzani, 2025; Känzig, 2025).

An illustration of this is the popular proxy-VAR in Känzig (2025), that identifies an oil news supply shock. We can see that, on impact, almost all coefficients of $\mathbf{b}_{\mathbf{oil}}^{\mathbf{proxy}}$ except the one on the real oil price are statistically zero (see Figure 2, middle panel). This is the case whenever the instrument is correlated significantly only with the residuals of a single endogenous variable, which is typically also chosen as target variable. Mechanically, the IRFs obtained are very similar to the dynamic multipliers (see Figure 4), as these are simply being multiplied by a vector that is almost proportional to the first canonical basis vector (see Appendix A.2 for the construction of the IRFs). The Cholesky impact vector obtained by ordering the shock variable first will be close to a scalar multiple of the first canonical basis vector when the first equation's residual variance dominates its contemporaneous covariances with the other residuals. Moreover, when the shock variable is ordered last, the Cholesky restriction implies that only this variable moves on impact, so the impact vector coincides with that of the dynamic multiplier up to the chosen scale (see Appendix A.3).

$$\phi_{\text{oil}} = \begin{bmatrix} 10.00 \\ \text{(6.56, 11.26)} \\ 0.00 \\ \text{(0.00, 0.00)} \\ 0.00 \\ \text{(0.00, 0.00)} \\ 0.00 \\ \text{(0.00, 0.00)} \\ 0.00 \\ \text{(0.00, 0.00)} \\ 0.00 \\ \text{(0.00, 0.00)} \end{bmatrix} \qquad \mathbf{b}_{\text{oil}}^{\text{proxy}} = \begin{bmatrix} 10.00 \\ \text{(6.56, 11.26)} \\ 0.12 \\ \text{(-0.33, 0.73)} \\ 0.32 \\ \text{(-0.01, 0.70)} \\ 0.00 \\ \text{(-0.23, 0.22)} \\ -0.17 \\ \text{(-0.41, 0.09)} \\ 0.14 \\ \text{(0.05, 0.22)} \end{bmatrix} \qquad \mathbf{b}_{\text{oil}}^{\text{chol}} = \begin{bmatrix} 10.00 \\ \text{(8.91, 10.83)} \\ -0.11 \\ \text{(-0.27, 0.02)} \\ -0.12 \\ \text{(-0.20, -0.05)} \\ 0.07 \\ \text{(0.01, 0.13)} \\ 0.01 \\ \text{(-0.04, 0.06)} \\ 0.11 \\ \text{(0.07, 0.14)} \end{bmatrix}$$

Figure 2: *Impact vectors with 90% confidence intervals in brackets: dynamic multipliers ϕ_{oil}, proxy-VAR $\mathbf{b}_{oil}^{proxy}$, and Cholesky $\mathbf{b}_{oil}^{chol}$ (Real oil price ordered first). Confidence intervals are obtained with a block bootstrap procedure (Jentsch & Lunsford, 2019).*

This pattern is common in high-frequency applications. Because the proxy is typically constructed from intraday or announcement-window movements in a single market price, the covariance of the instrument and the residuals concentrates on that price equation's residual, while cross-equation covariances are negligible on impact. In addition, HF instruments often capture "news" shocks for which real activity and most prices are slow-moving at the sam-

given by the coefficients contained in $\mathbf{\Phi}_i$, using the notation of Appendix A.

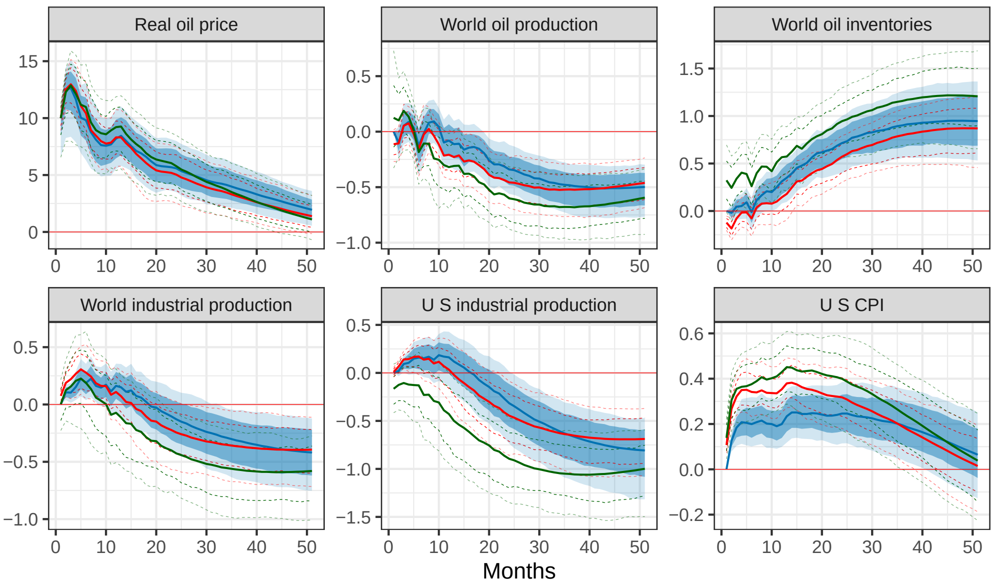


Figure 3: *Baseline specification in Känzig (2021). Blue solid line with shaded confidence bands: dynamic multipliers; green solid line with dashed bands: external-instrument identification; red solid line with dashed bands: recursive identification with the real oil price ordered first. Bands correspond to 68% and 90% confidence intervals.*

Notes: Confidence bands are computed using the moving block bootstrap of Jentsch and Lunsford (2019).

pling frequency, yielding near-zero contemporaneous responses outside the anchor equation. Finally, aggregating daily or intraday surprises to the monthly VAR attenuates same-period comovement with non-anchor variables. This consideration also motivates augmenting the VAR with fast-moving variables (e.g., risk premia or other asset prices) that react on impact, allowing the proxy to load on multiple equations and pulling identification away from a near–basis-vector outcome.

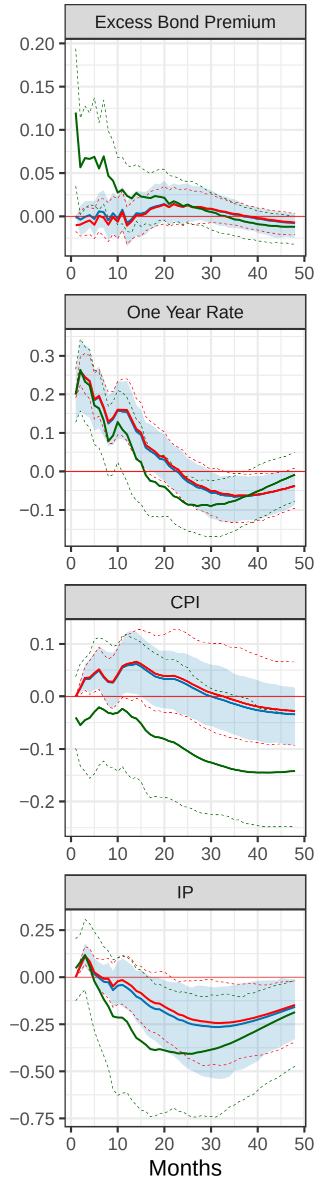
When the instrument is credibly exogenous, validation ultimately rests on economic theory. When the estimated responses accord with theoretical predictions, their similarity to reduced-form dynamic multipliers is not deemed problematic (although, in practice, this comparison is rarely made explicitly because dynamic multipliers are usually not reported). For example, in Känzig (2021), the identified disturbances are interpreted as news shocks, making it plausible that most macroeconomic variables do not react within the same month. At the same time, if theory implies that only one variable adjusts contemporaneously while others are slow-moving, an external instrument is not required for impact identification. In that case, a recursive (Cholesky) scheme that orders the fast-moving variable last imposes precisely those restrictions and will yield responses that are very similar to those from a proxy-VAR and identical (up to scale) to the reduced-form dynamic multipliers; Figure 4 shows that this is the case in Känzig (2021). Given the resource costs of constructing a credible instrument, a strong case for using one should be made.

Irrespectively, it is always useful to report the reduced-form dynamic multipliers-and, where appropriate, impulse responses from a recursive ordering—as a transparency check. Juxtaposing these with the proxy-VAR responses allows to assess whether the instrument contributes information beyond the reduced-form dynamics and to gauge how much the results hinge on the instrument rather than on the VAR's propagation.

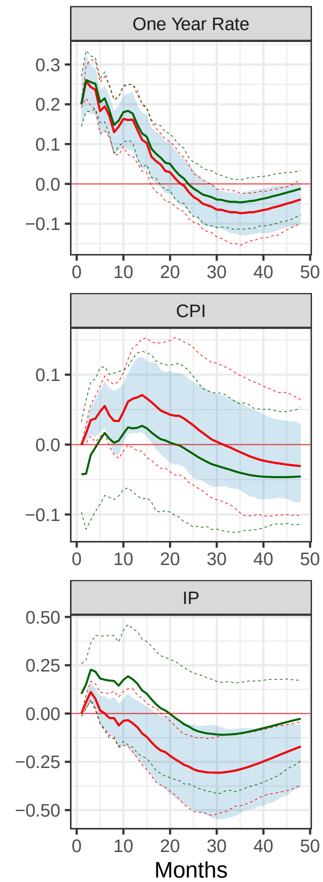
When contemporaneous impacts matter, the case of monetary policy shocks. Consider the classic problem of identifying a monetary policy shock (e.g. Kuttner, 2001; Romer and Romer, 2004). Recursive identification in small VARs produces the price puzzle, at odds with standard theory. This issue can persist even with an external instrument: in canonical trivariate VARs, the impact responses of inflation and output to an interest-rate innovation are typically statistically indistinguishable from zero, so the proxy adds little beyond the reduced form. One approach is to augment the system with a fourth, fast-moving financial variable—the excess bond premium of Gilchrist and Zakrajšek (2012)—which reacts contemporaneously. Doing so shifts the impact vector away from a scalar multiple of the first canonical basis vector and yields impulse responses that depart meaningfully from the reduced form and do not display a price puzzle, with the instrument unchanged relative to the trivariate specification.

Figure 4 illustrates this point in the SVAR of Gertler and Karadi (2015). Moving from the trivariate VAR to the four-variable specification—without changing the instrument—eliminates the price puzzle because the excess bond premium responds on impact. In the trivariate VAR, the anchor-based and shock-based first-stage F-statistics are similar, reflecting the near-zero contemporaneous loadings of the non-policy variables (anchor-based $F = 10.27$, shock-based $F = 11.13$). In the four-variable VAR, by contrast, the discrepancy becomes salient: the anchor-based statistic falls below 10 ($F = 9.02$), while the shock-based statistic is substantially larger ($F = 15.48$). Figure 5 further shows that shock-based sign normalization yields tighter and more stable bootstrap inference than anchor-based normalization. These differences arise because the shock-based approach exploits the instrument's comovement with multiple nonzero impact loadings—here, both the one-year rate and the excess bond premium—rather than relying on a single anchor equation.

Hence, when the purported "structural" impulse responses are statistically close to the dynamic multipliers, one should consider at least three possibilities: (i) the instrument may not be fully exogenous and could be capturing a reduced-form relation; (ii) the instrument



Baseline specification



Trivariate specification

Figure 4: *Baseline four-variable specification of Gertler and Karadi (2015) (left) and the same specification excluding the excess bond premium (right). Blue solid line with shaded confidence bands: dynamic multipliers; green solid line with dashed bands: external-instrument identification using the FF4 instrument; red solid line with dashed bands: recursive identification with the one-year rate ordered after IP and CPI. Bands correspond to 95% confidence intervals.*

Notes: Unlike the original study, which restricts the instrument sample to start in 1991, I use the full sample (1979M7–2012M6) to align with the VAR estimation window. Confidence bands are computed using the moving block bootstrap of Jentsch and Lunsford (2019); the original paper instead employs the wild bootstrap of Mertens and Ravn (2013).

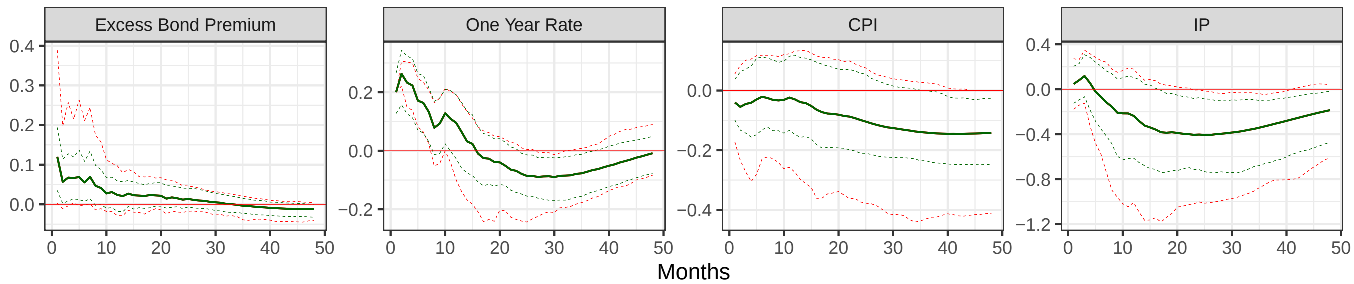


Figure 5: *Green solid line with dashed bands: shock-based sign normalization. Red solid line with dashed bands: anchor-based sign normalization targeting the one-year rate. Bands correspond to 95% confidence intervals.*

may be weak because it loads materially only on one equation's residuals; and (iii) a relevant fast-moving variable may be omitted, preventing identification from deviating from the reduced form.

Identifying multiple shocks that move the dynamic multipliers in different directions. In light of the preceding discussion, designs that identify more than one structural shock and yield impulse responses that are both distinct from any single variable's reduced-form dynamic multipliers and consistent with economic theory are especially convincing. For example, in commodity markets, when both demand and supply shocks are identified using instruments, we expect to observe contemporaneous movements in both prices and quantities, with opposite signs across the two shocks—patterns not mechanically imposed by the identification. Figure 1 illustrates this in Colombo and Toni (2025), where the two shocks move the same dynamic multipliers in significantly different directions: prices rise after either shock, while quantity demanded increases only after a demand shock and quantity supplied decreases only after a supply shock.

A related implication is that specifications in which the identified shock has economically meaningful impact effects on several variables tend to be more robust to misspecification, including the omission of relevant variables that could otherwise serve as anchors. Figures 6 and 7 illustrate this contrast. In the natural gas application of Colombo and Toni (2025), omitting the real natural gas price has little effect on the remaining responses, whereas in Känzig (2021) omitting the real oil price materially alters the estimated dynamics. Furthermore, the shock-based F-statistic drops sharply when the oil price is excluded, indicating that the instrument's relevance is concentrated in that single equation.

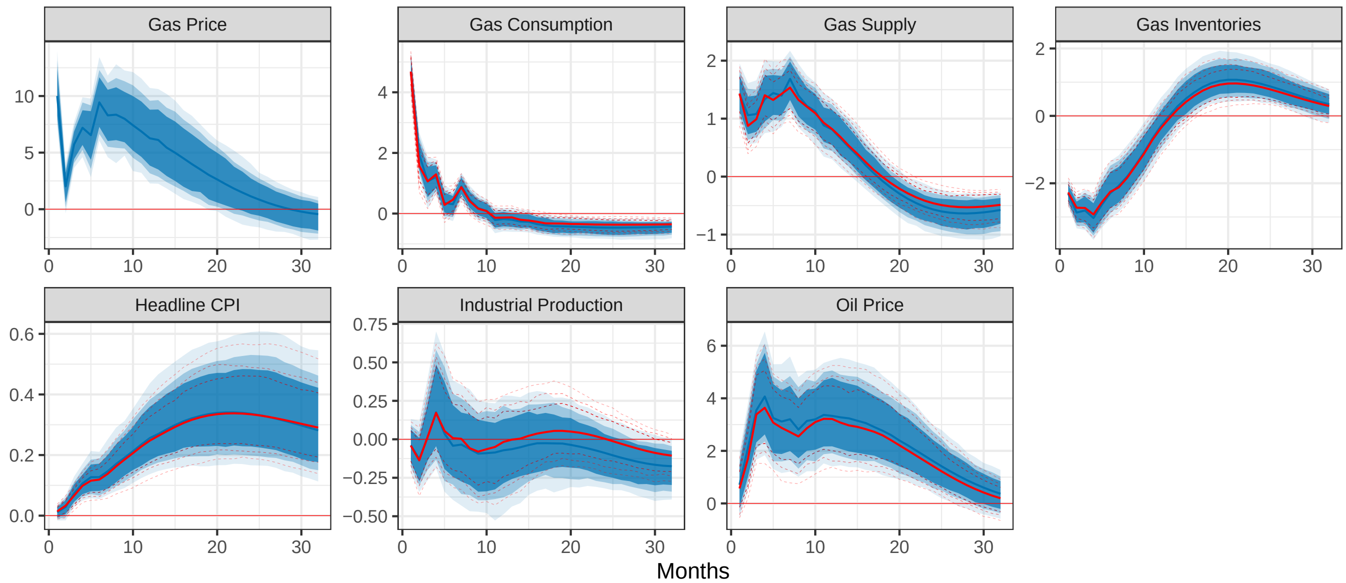


Figure 6: *Blue solid line with shaded confidence bands: baseline specification in Colombo and Toni (2025); red solid line with dashed bands: the same VAR estimated without the real natural gas price. Bands correspond to 68%, 80% and 90% confidence intervals. Shock-based* $F = 181.96$ *(robust* $F = 131.71$*); without price* $F = 175.60$ *(robust* $F = 132.51$*).*

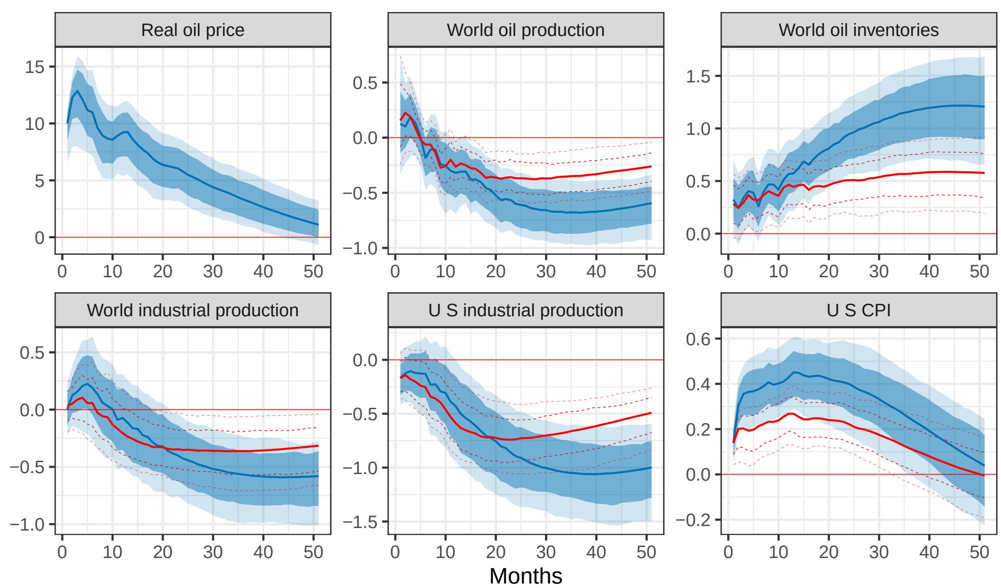


Figure 7: *Blue solid line with shaded confidence bands: baseline specification in Känzig (2021); red solid line with dashed bands: the same VAR estimated without the real oil price. Bands correspond to 68% and 90% confidence intervals. Shock-based* $F = 27.60$ *(robust* $F = 12.86$*); without price* $F = 6.56$ *(robust* $F = 4.47$*).*

5 Beyond External-Instrument VARs

I now argue that the rationale for shock based diagnostics and the concerns that arise when instrument identified impulse responses are largely indistinguishable from reduced form dynamic multipliers carry over to other commonly used estimation and identification frameworks.

Internal instrument approach and LP-IV. An increasingly common alternative to the external-instrument VAR in Section 2 is the "internal instrument" approach (Plagborg-Møller & Wolf, 2021). This method does not rely on invertibility (see Appendix A.4). It incorporates the instrument z_t directly into the VAR and places it first in a recursive identification scheme. In this setting, some applied studies do not report weak-instrument diagnostics and, in some cases, claim that "While the true structural shock is not directly observed, this condition cannot be tested" (Cascaldi-Garcia & Vukotić, 2022). Although, to the best of my knowledge, there is no formal treatment of weak-instrument identification for the internal instrument approach yet, the possibility of weak instruments remains. The approach is therefore not immune to the critique in Section 3, and many applied papers still seek to justify instrument strength.

A prominent example is Ramey (2011), who uses narrative news about military spending to identify a government spending shock and, despite adopting an internal instrument strategy, reports first-stage F-statistics and refers to the conventional threshold $F \geq 10$. In her implementation, the F-statistic is computed outside the VAR by regressing a target variable on the contemporaneous instrument and its lags. She also reports a "marginal F" statistic that additionally controls for lags of variables included in the VAR, thereby approximating the first-stage F-statistic that would arise in an external-instrument VAR. This raises the same conceptual issue as in Section 3: it is unclear why one particular anchor variable should be preferred for the diagnostic. This point is especially clear in this setting, since the internal instrument approach does not involve a distinct first stage. Instead, all endogenous variables are effectively projected onto the instrument. The concerns discussed in Section 4 apply in the same way, since if the instrument mainly loads on a single equation the identified responses largely mirror reduced-form propagation.

The same conclusion extends to the LP-IV approach. For example, Ramey and Zubairy (2018) use local projection IV to study state dependence in the effects of government spending shocks. As in the previous case, instrument strength is assessed using first-stage F-statistics for a particular anchor variable (government spending), even though the second-stage responses are reported for a broader set of variables.

When can the use of a target variable be justified? The arbitrariness of the anchor choice discussed in Section 3 reflects a common interpretation in macroeconomic identification. Instruments are typically viewed as generating exogenous variation in an unobserved structural shock, rather than as shifting an observed endogenous regressor in the usual 2SLS sense. If the object of interest were instead a policy experiment defined directly in terms of an endogenous variable—for example, the macroeconomic effects of an exogenous 25 basis point increase in the policy rate—then stating relevance with respect to that variable and using the policy-rate equation as the anchor for sign normalization and weak-instrument diagnostics would be appropriate. In most applications, however, the target is a monetary policy shock *that raises the policy rate by 25 basis points* (after normalization). Because such a shock is understood to affect other variables on impact as well, fixing instrument strength and normalization to a single anchor equation is not generally well motivated.

This shock-based interpretation is typically maintained irrespective of the estimation or identification method. For example, Stock and Watson (2018) show that the same instrument can be used in both an external-instrument VAR and an LP-IV implementation. Their unit-effect normalization fixes the scale of the shock so that a 1 percentage point monetary policy shock increases the federal funds rate by 1 percentage point, thereby making explicit the distinction between the shock and the endogenous variable. Likewise, Plagborg-Møller and Wolf (2022) state their identifying assumptions directly in terms of structural shocks, including in internal-instrument and LP-IV settings.

6 Conclusions

In this paper I document an inconsistency in prevailing practice: instrument relevance in proxy-VARs is often tested on an anchor residual rather than on the identified structural shock, which renders the test non-invariant to the choice of anchor residual. I propose a shock-based, anchor-independent weak instrument test based on the usual F-statistic, together with a bootstrap procedure that normalizes sign on the recovered shock, that yields invariant and stronger diagnostics, as well as typically tighter inference. Separately, I argue that when proxy-VARs produce "structural" IRFs that are statistically indistinguishable from reduced-form dynamic multipliers, the instrument is likely weak, mis-specified, or simply superfluous. Crucially, these considerations are not specific to external-instrument VARs and also arise in internal-instrument and LP-IV settings. A natural direction for future work is to develop formal diagnostics that distinguish genuine shock identification from reduced-form propagation.

References

Alessandri, **P.**, & **Gazzani**, **A.** (2025). Natural gas and the macroeconomy: Not all energy shocks are alike. *Journal of Monetary Economics*, 103749.

Caldara, **D.**, **Cavallo**, **M.**, & **Iacoviello**, **M.** (2019). Oil price elasticities and oil price fluctuations. *Journal of Monetary Economics*, *103*, 1–20.

Cascaldi-Garcia, **D.**, & **Vukotić**, **M.** (2022). Patent-based news shocks. *Review of Economics and Statistics*, *104*(1), 51–66.

Colombo, **D.**, & **Toni**, **F.** (2025). Gas prices and the macroeconomy. *Available at SSRN 5237587*.

Gertler, **M.**, & **Karadi**, **P.** (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, *7*(1), 44–76.

Giacomini, **R.**, **Kitagawa**, **T.**, & **Read**, **M.** (2022). Robust bayesian inference in proxy svars. *Journal of Econometrics*, *228*(1), 107–126.

Gilchrist, **S.**, & **Zakrajšek**, **E.** (2012). Credit spreads and business cycle fluctuations. *American economic review*, *102*(4), 1692–1720.

Jentsch, **C.**, & **Lunsford**, **K. G.** (2019). The dynamic effects of personal and corporate income tax changes in the united states: Comment. *American Economic Review*, *109*(7), 2655–2678.

Jentsch, **C.**, & **Lunsford**, **K. G.** (2022). Asymptotically valid bootstrap inference for proxy svars. *Journal of Business & Economic Statistics*, *40*(4), 1876–1891.

Känzig, **D. R.** (2021). The macroeconomic effects of oil supply news: Evidence from opec announcements. *American Economic Review*, *111*(4), 1092–1125.

Känzig, **D. R.** (2023). *The unequal economic consequences of carbon pricing* (NBER Working Paper).

Känzig, **D. R.** (2025). *The macroeconomic effects of supply chain shocks: Evidence from global shipping disruptions* (tech. rep.).

Kilian, **L.**, **Plante**, **M.**, & **Richter**, **A. W.** (2023). Estimating macroeconomic news and surprise shocks.

Kuttner, **K. N.** (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of monetary economics*, *47*(3), 523–544.

Mertens, **K.**, & **Ravn**, **M. O.** (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American economic review*, *103*(4), 1212–1247.

Montiel Olea, **J. L.**, **Stock**, **J. H.**, & **Watson**, **M. W.** (2021). Inference in structural vector autoregressions identified with an external instrument. *Journal of Econometrics*, *225*(1), 74–87.

Plagborg-Møller, **M.**, & **Wolf**, **C. K.** (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, *89*(2), 955–980.

Plagborg-Møller, **M.**, & **Wolf**, **C. K.** (2022). Instrumental variable identification of dynamic variance decompositions. *Journal of Political Economy*, *130*(8), 2164–2202.

Ramey, **V. A.** (2011). Identifying government spending shocks: It's all in the timing. *The quarterly journal of economics*, *126*(1), 1–50.

Ramey, **V. A.**, & **Zubairy**, **S.** (2018). Government spending multipliers in good times and in bad: Evidence from us historical data. *Journal of political economy*, *126*(2), 850–901.

Romer, **C. D.**, & **Romer**, **D. H.** (2004). A new measure of monetary shocks: Derivation and implications. *American economic review*, *94*(4), 1055–1084.

Stock, **J. H.** (2008). *What is new in econometrics: Time series, lecture 7* (tech. rep.).

Stock, **J. H.**, & **Watson**, **M. W.** (2012). *Disentangling the channels of the 2007-2009 recession* (tech. rep.). National Bureau of Economic Research.

Stock, **J. H.**, & **Watson**, **M. W.** (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, *128*(610), 917–948.

Appendix

Identification or Propagation? Shock-Based Weak-IV Diagnostics and Bootstrap Inference for Instrumented VARs

Daniele Colombo[†]

November 28, 2025

Contents

[†]London Business School. Email: dcolombo@london.edu

A SVAR econometrics

A.1 Stability and MA(∞) representation.

Using the lag operator L, (2) can be rewritten as

$$\big(\mathbf{I} - \mathbf{A}_1 L - \cdots - \mathbf{A}_p L^p\big)\mathbf{y}_t = \mathbf{u}_t. \tag{17}$$

If the VAR is stable,[1] it is possible to apply the inverse filter of the lag polynomial, which also admits a convergent power-series expansion:

$$\mathbf{y}_t = \big(\mathbf{I} - \mathbf{A}_1 L - \cdots - \mathbf{A}_p L^p\big)^{-1}\mathbf{u}_t = \left(\sum_{i=0}^{\infty} \mathbf{\Phi}_i L^i\right)\mathbf{u}_t = \sum_{i=0}^{\infty} \mathbf{\Phi}_i \mathbf{u}_{t-i}, \tag{18}$$

with coefficients $\{\mathbf{\Phi}_i\}_{i\geq 0}$ that are absolutely summable. Using (3) and defining $\mathbf{\Theta}_i := \mathbf{\Phi}_i\mathbf{B}_0^{-1}$, we get the one-sided[2] moving-average representation

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \mathbf{\Phi}_i \mathbf{u}_{t-i} = \sum_{i=0}^{\infty} \mathbf{\Phi}_i \mathbf{B}_0^{-1}\mathbf{w}_{t-i} = \sum_{i=0}^{\infty} \mathbf{\Theta}_i \mathbf{w}_{t-i}. \tag{19}$$

Further, by defining $\mathbf{C}(L) := \big(\mathbf{I} - \mathbf{A}_1 L - \cdots - \mathbf{A}_p L^p\big)^{-1}\mathbf{B}_0^{-1}$ this can also be written as

$$\mathbf{y}_t = \mathbf{C}(L)\mathbf{w}_t, \tag{20}$$

which is the structural vector moving-average (SVMA) formulation for $\mathbf{y}_t$.

A.2 Structural impulse response functions

The MA coefficients contained in $\mathbf{\Theta}_i$ give the impulse response functions (IRFs), that is, the responses of each element of $\mathbf{y}_t = (y_{1t}, \ldots, y_{Kt})'$ to a one-time impulse in each element of $\mathbf{w}_t = (w_{1t}, \ldots, w_{Kt})'$:

$$\frac{\partial \mathbf{y}_{t+i}}{\partial \mathbf{w}_t'} = \mathbf{\Theta}_i, \quad i = 0, 1, 2, \ldots, H \tag{21}$$

At every horizon i, this is a $(K \times K)$ matrix whose elements are given by

$$\theta_{jk,i} = \frac{\partial y_{j,t+i}}{\partial w_{kt}}.$$

[1] Stability means that shocks do not generate explosive dynamics and, formally, is equivalent to all roots of the characteristic polynomial lying outside the unit circle: $\det\big(\mathbf{I} - \mathbf{A}_1 z - \cdots - \mathbf{A}_p z^p\big) \neq 0$ for all $|z| \leq 1$. Equivalently, all eigenvalues of the companion matrix associated with $(\mathbf{A}_1, \ldots, \mathbf{A}_p)$ have modulus strictly less than one.

[2] One-sided means the system can be expressed as a linear combination of current and past shocks only.

In order to recover the IRFs, we first resort to the VAR(1) representation of the VAR(p) process:

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{U}_t, \tag{22}$$

with

$$\mathbf{Y}_t \equiv \begin{bmatrix} \mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix} \quad \mathbf{A} \equiv \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_K & \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K & & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_K & \mathbf{0} \end{bmatrix} \quad \mathbf{U}_t \equiv \begin{bmatrix} \mathbf{u}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}.$$

By recursive substitution, it can be shown that the response of the variable $j = 1, \dots, K$ to a unit shock u_{kt}, i periods in the past, for $k = 1, \dots, K$, is given by $\mathbf{\Phi}_i = \begin{bmatrix} \phi_{jk,i} \end{bmatrix} \equiv \mathbf{J}\mathbf{A}^i\mathbf{J}'$, where $\mathbf{J} \equiv \begin{bmatrix} \mathbf{I}_K & \mathbf{0}_{K\times K(p-1)} \end{bmatrix}$ is a selector matrix. These are sometimes called dynamic multipliers or reduced-form impulse responses. From (19) it follows that the structural impulse responses can be obtained simply by post-multiplying the dynamic multipliers $\mathbf{\Phi}_i$ by $\mathbf{B}_0^{-1}$ for every horizon i.

Partially identified case. When only a subset of structural shocks is identified—consider the first without loss of generality—one recovers only the corresponding column $\mathbf{b}_1$ of $\mathbf{B}_0^{-1}$. The impulse responses of all endogenous variables to that shock are then

$$\frac{\partial \mathbf{y}_{t+i}}{\partial w_{1t}} = \mathbf{\Theta}_i \mathbf{e}_1 = \mathbf{\Phi}_i \mathbf{b}_1, \qquad i = 0, 1, \dots, H, \tag{23}$$

where $\mathbf{e}_1$ denotes the first canonical basis vector.

A.3 The recursive identification scheme

Under the unit-variance representation, we can rewrite (9) as

$$\mathbf{\Sigma}_{\mathbf{u}} = \tilde{\mathbf{B}}_0^{-1}(\tilde{\mathbf{B}}_0^{-1})'$$

Let $\hat{\mathbf{\Sigma}}_u$ denote the sample covariance of the reduced-form residuals. Since $\hat{\mathbf{\Sigma}}_u$ is symmetric positive definite, it admits the (unique) Cholesky factorization

$$\hat{\mathbf{\Sigma}}_u = \mathbf{L}\mathbf{L}',$$

with $\mathbf{L}$ lower triangular with positive diagonal entries. Therefore, choosing $\tilde{\mathbf{B}}_0^{-1} = \mathbf{L}$, satisfies the restrictions of the model, given by (9) and by the fact that the implied structural shocks are orthogonal.[3] Hence, under the recursive identification, the impact matrix is estimated by the lower-triangular Cholesky factor of $\hat{\mathbf{\Sigma}}_u$, $\tilde{\mathbf{B}}_0^{-1} = \mathbf{L}$, and the structural shocks are recovered as $\tilde{\mathbf{w}}_t = \mathbf{L}^{-1}\mathbf{u}_t$.

Since using the Cholesky factorization implies the unit-variance normalization, with this identification we have that a one-standard deviation structural shock has an impact equal to the correponding entries of $\mathbf{L}$ on each endogenous variable. The impact of a one–standard–deviation realization of structural shock k produces the contemporaneous impact $\mathbf{L}_{:,k}$ on $\mathbf{y}_t$; in particular, for the first-ordered variable, the magnitude of the impact also corresponds to the in-sample standard deviation of $u_{1,t}$.[4]

Since $\mathbf{L}$ is lower triangular, this identification scheme imposes that the simultaneous relationships between the variables are *acyclic*, that is, that there are no contemporary feedbacks in the system and that there exists a precise causal ordering of the variables.

A.4 Invertibility

A standard SVAR specification is a particular type of SVMA process (see 20) that assumes an *invertible* (or *fundamental*) representation, meaning that the structural shocks can be recovered from current and past (without requiring future) values of the endogenous variables. Formally, invertibility requires that there exists a one-sided filter $\mathbf{D}(L)$ such that

$$\mathbf{w}_t = \mathbf{D}(L)\mathbf{y}_t. \tag{24}$$

In the SVAR(p) model (1), this condition holds by construction, since

$$\mathbf{w}_t = \mathbf{B}_0\mathbf{y}_t - \mathbf{B}_1\mathbf{y}_{t-1} - \cdots - \mathbf{B}_p\mathbf{y}_{t-p} = \big(\mathbf{B}_0 - \mathbf{B}_1 L - \cdots - \mathbf{B}_p L^p\big)\mathbf{y}_t,$$

which satisfies (24) with $\mathbf{D}(L) := \mathbf{B}_0 - \mathbf{B}_1 L - \cdots - \mathbf{B}_p L^p$.[5] Within the class of one-sided *stable* filters, the representation in (24) is unique, and $\mathbf{D}(L)$ is the unique stable one-sided left inverse of $\mathbf{C}(L)$.

In many economically relevant cases, $\mathbf{w}_t$ is a shock to expectations or information whose consequences for standard macro aggregates only become evident with a delay. In such

[3] $\widehat{\mathbb{E}(\tilde{\mathbf{w}}_t\tilde{\mathbf{w}}_t')} = \mathbb{E}(\mathbf{L}^{-1}\hat{\mathbf{u}}_t\hat{\mathbf{u}}_t'\mathbf{L}'^{-1}) = \mathbf{L}^{-1}\hat{\mathbf{\Sigma}}_u\mathbf{L}'^{-1} = \mathbf{L}^{-1}\mathbf{L}\mathbf{L}'\mathbf{L}'^{-1} = \mathbf{I}_k$.

[4] Note that for a positive-definite matrix such as the variance-covariance matrix $\hat{\mathbf{\Sigma}}_u$, it can be shown that the diagonal entries of its Cholesky factorization are equal to the square root of the conditional variances $\mathbf{L}_{ii} = \sqrt{\widehat{\mathbb{V}}(u_{i,t}|u_{1,t},\ldots,u_{i-1,t})}$, which are the square roots of the corresponding conditional variances under the chosen ordering. For the first-ordered variable, the corresponding entry is the in-sample standard deviation of $u_{1,t}$.

[5] The SVAR also assumes that the number of endogenous variables must be at least as large as the number of shocks, since a necessary condition for $\mathbf{D}(L)$ to be a left inverse of $\mathbf{C}(L)$ is for $\mathbf{C}(L)$ to have full column rank.

environments, future realizations of $\mathbf{y}_t$ "reveal" which part of contemporaneous movements was due to the information shock as opposed to other disturbances. As a result, the shock may not be recoverable by a one-sided (or causal) filter of the form (24), with

$$\mathbf{D}(L) = \sum_{i=0}^{\infty} \mathbf{D}_i L^i,$$

so the invertibility assumption embedded in standard SVAR models may fail.[6] Instead, recovering $\mathbf{w}_t$ may require a two-sided (non-causal) filter that uses both lags and leads,

$$\mathbf{w}_t = \sum_{i=-\infty}^{\infty} \mathbf{H}_i \mathbf{y}_{t-i},$$

so that identification relies on information contained in future realizations of $\mathbf{y}_t$.

Two leading examples are *news* shocks and *anticipated policy* shocks. For instance, agents may receive news at time t about higher future productivity, leading forward-looking variables (e.g., investment and asset prices) to respond immediately even though measured productivity moves only later. Likewise, many policy changes are announced in advance and implemented with a lag, so the economically relevant disturbance is the announcement-driven revision in expectations rather than the subsequent movement in realized policy instruments. In both cases, the relevant shock need not coincide with the one-step-ahead innovation in the vector of observables and may therefore be non-invertible.

Internal instrument approach. An alternative identification strategy with instruments that does not require invertibility of $w_{k,t}$ with respect to $\{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots\}$ is to include the proxy z_t directly in the VAR and order it first in a recursive scheme (Ramey, 2011; Plagborg-Møller & Wolf, 2021). Doing so enlarges the information set to $\{\mathbf{y}_\tau, z_\tau\}_{\tau \leq t}$. When the instrument is included in the VAR, the information set becomes the history of the endogenous variables and the instrument, $\{\mathbf{y}_\tau, z_\tau\}_{\tau \leq t}$. Under the usual relevance and exogeneity conditions, z_t provides a direct contemporaneous signal about the structural shock of interest—formally, it loads on $w_{k,t}$ while being orthogonal to the remaining shocks—so augmenting the system rules out missing observables as the source of non-invertibility.

However, if we think of z_t as a noisy measure of the shock observed with independent measurement error

$$z_t = \lambda w_{k,t} + v_t,$$

with v_t independent of $\mathbf{w}_t$ at all leads and lags, $w_{k,t}$ is still not exactly recoverable from

[6] This leads the SVAR to inconsistently estimate the impulse response functions (Plagborg-Møller & Wolf, 2022).

$\{\mathbf{y}_\tau, z_\tau\}_{\tau \leq t}$ as a one-sided filter: the proxy is only an imperfect signal of the shock, so exact spanning fails. This measurement error induces a common attenuation bias in covariances between z_t and future outcomes, so it affects the overall scale but not the shape of estimated impulse responses once a normalization is imposed. Consequently, the internal-instrument approach delivers correct *relative* impulse responses, even when $w_{k,t}$ is non-invertible with respect to $\{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots\}$ (Plagborg-Møller & Wolf, 2021).

References Appendix

Plagborg-Møller, **M.**, & **Wolf**, **C. K.** (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, *89*(2), 955–980.

Plagborg-Møller, **M.**, & **Wolf**, **C. K.** (2022). Instrumental variable identification of dynamic variance decompositions. *Journal of Political Economy*, *130*(8), 2164–2202.

Ramey, **V. A.** (2011). Identifying government spending shocks: It's all in the timing. *The quarterly journal of economics*, *126*(1), 1–50.