IDENTIFICATION AND INFERENCE IN PROXY-SVARS WITH NON-GAUSSIAN SHOCKS

PARITOSH SHANKARRAO JUNARE

(Click here for the latest version: November 2026)

Abstract. Two frequent approaches for identifying structural VARs are external instruments, which carry economic content but are often weak, and non-Gaussianity of the shocks, which provides statistical identification but carries no economic meaning. We combine the two strategies in a single generalized method of moments framework. This hybrid approach point-identifies the target shocks while also identifying the non-target shocks up to sign and ordering. Under suitable rank conditions, the higher-order moments of the structural shocks anchor the identification uniformly over the instrument strength. Consequently, under local-to-zero proxy relevance, estimators of the dynamic causal effects remain consistent, and standard asymptotic inference remains valid. Moreover, the Anderson-Rubin confidence sets are substantially narrower than their instrument-only counterparts. The hybrid estimator is also more efficient than either source of identification used in isolation: at any fixed proxy relevance, even a weak instrument increases the efficiency of the estimator through its covariance with the non-Gaussian moment block. Under local deviations from proxy exogeneity, we provide asymptotic bias bounds and show that stronger non-Gaussianity of the shocks compresses the bias. Finally, the over-identified structure yields two mutually orthogonal specification tests for proxy exogeneity and the validity of the higher-order moment conditions. We derive their limiting distributions and provide a bootstrap procedure for finite-sample critical values. Monte Carlo simulations and two applications with identification of an oil-news shock and a Euro-area monetary policy shock demonstrate the potential of our framework.

1. Introduction

The identification and estimation of dynamic causal effects is crucial to understand the transmission mechanisms of structural shocks in empirical macroeconomics. Identification through structural vector autoregressions (SVARs) generally requires restrictions on the structural impact matrix, denoted by B. It captures the instantaneous effects of the structural

JEL Classification. C32, C38, C52, Q43, E32.

Keywords. Identification, Proxy-SVARs, non-Gaussianity, Weak instruments.

Correspondence: Department of Economics, University of Bologna, Italy. Email: paritosh.junare@unibo.it.

I am immensely grateful to Luca Fanelli, Giuseppe Cavaliere and Giovanni Angelini for their guidance and support throughout this work. I am also thankful to Markku Lanne, Jani Luoto, Mika Meitz, Savi Virolainen and Filippo Ferroni for their insightful comments. Discussions with Raffaella Giacomini and Lutz Kilian at Nordic Econometric Meeting 2026; Silvia Sarpietro and Kirill Ponomarev at CEPR Job Market Bootcamp 2026, Tuscany; Giovanni Caggiano as a discussant at SIdE WEEE 2026, and other participants at these conferences are gratefully acknowledged.

shocks on variables, and hence it is essential for the estimation of impulse response functions (IRFs). Two identification strategies have become increasingly prevalent. The first strategy employs economic theory-based external instruments or proxies: variables which are "external" to the VAR estimation, and are correlated with the target shocks (relevance condition) and orthogonal to other structural shocks (exogeneity condition), see Mertens and Ravn (2013); Stock and Watson (2018). The second strategy exploits the non-Gaussianity of the structural shocks to obtain statistical identification: with at most one Gaussian shock, certain higher-order moment conditions recover B up to sign and column permutations without additional restrictions, see Comon (1994); Lanne and Luoto (2021); Keweloh (2021).

On one hand, in proxy-SVARs, an external instrument is only as good as its correlation with the shock it is meant to isolate. Under local-to-zero relevance, Staiger and Stock (1997) show that the estimates are inconsistent, and standard asymptotic inference is unreliable, also detailed in Montiel Olea et al. (2021). Ramey (2016) notes that weak instruments are pervasive in the literature, and Lewis and Mertens (2026) show that conventional critical-value thresholds are ill-suited for assessing weak-instrument bias in IRFs. Weak instruments are therefore empirically typical, and routinely under-diagnosed; Appendix K (Table K.1) lists prominent applications in which this sensitivity is visible. On the other hand, a large body of evidence in macro-financial literature finds that the structural shocks are pronouncedly non-Gaussian. The recovered structural shocks reject Gaussianity in favor of heavy-tailed distributions as shown in Gouriéroux et al. (2017), Ludvigson et al. (2021); see Appendix K (Table K.2). However, as noted earlier, this is a source of statistical identification where we still rely on labeling conventions to obtain meaningful economic interpretation of the shocks. Both of these identification strategies therefore face an obstacle that the other can overcome.

We propose a "hybrid" identification[1] scheme where we combine the two channels within a single generalized method of moments (HGMM) framework that stacks proxy exclusion restrictions with higher-order moment conditions into a single over-identified system. With k instruments, under standard sign normalizations, we show global point identification of the target k columns of B, and the non-target columns are identified up to signed permutations. The identification is global in the sense of Komunjer (2012): non-Gaussianity eliminates the continuous rotational manifold of B, reducing identification to a finite discrete set $\mathcal{B}$; and then the valid proxies select a unique element from $\mathcal{B}$, delivering global point identification of the target columns.

We provide three main results that directly motivate the HGMM estimator as a preferred alternative to proxy-only approach for researchers.

First, the HGMM estimator remains robust to weak proxies. Under local-to-zero relevance, the proxy-only estimator is inconsistent, whereas under certain rank conditions, the HGMM estimator remains consistent with uniformly valid standard asymptotic inference. This result is explained by the fact that the non-Gaussian (NG) channel acts as an identification anchor.

[1]We denote the identification as *hybrid* because the two merged identification channels are of different kinds: one economic (external instruments), and one statistical (the non-Gaussianity of the shocks).

Since the Jacobian matrix pertaining to the NG moment block can remain full rank regardless of the proxy strength, the joint information matrix remains bounded away from zero even as the proxy relevance degenerates. With sufficient non-Gaussianity in the shocks, this allows practitioners to rely on standard inference methods even in the presence of weak instruments where proxy-SVAR inference breaks down.

Second, the joint estimator is asymptotically at least as efficient as the proxy-only estimator: the difference in their asymptotic variances is positive semidefinite. The efficiency gain has a closed form. The joint information matrix decomposes additively into two sources: the information obtained from the proxy exclusion conditions, and the information added by the higher-order moments after partialing out the identifying directions already spanned by the valid proxies. The gain is strict when the NG moment block identify parameters along directions that the proxy block does not capture. Interestingly, one might expect the hybrid estimator to collapse onto the non-Gaussian GMM estimator if the proxy is very weak. It does not. The covariance between the two moment blocks continues to sharpen the weighting matrix, so at any fixed proxy relevance the HGMM estimator has smaller variance than that of the non-Gaussian estimator alone. This generates an efficiency gain that a researcher who entirely discards the proxies would forego.

Third, when the instrument locally violates the exogeneity condition, the resulting first-order bias is inversely proportional to the minimum eigenvalue of the NG information matrix. Stronger deviations from Gaussianity therefore compress the upper bound on the bias directly, delivering a partial mis-specification robustness that proxy-only approaches do not have. On all three margins, therefore, augmenting a proxy-SVAR with non-Gaussian identification weakly dominates the standalone alternatives: precision improves under correct specification, consistency and standard inference survive a weak proxy, and the bias is bounded when the proxy is locally invalid.

Another contribution is of separate, valid specification tests for each identification channel. The standard J-statistic is invalid in this setting: restricting the criterion to the NG block alone admits multiple observationally equivalent solutions. This implies that the rotation parameters would lack a unique probability limit and the χ^2 limiting distribution of the test statistic is not restored without an anchor. Hence, for both identification channels, we construct mutually orthogonal test statistics from the *joint* GMM estimate. By orthogonalizing each block's sample moments against the other, we obtain separate statistics for each identification channel, J_{prx} and J_{NG}, such that neither test is distorted by the other moment block (Newey, 1985). We also correct for the shared parameter estimation uncertainty in their asymptotic variances. This implies that, under the joint null of valid identification, the limiting distributions of the test statistics follow χ^2 with degrees of freedom equal to the number of moment conditions in the relevant block.

However, we note that asymptotic χ^2 critical values can be unreliable because higher-order cumulant estimation in the optimal weighting matrix may induce substantial finite-sample distortion. Hence, we adopt the residual-based moving block bootstrap of Brüggemann et al.

(2016) that jointly resamples blocks of estimated residuals-proxy pairs, preserving the cross-dependence structure exploited by the proxy identification. It also allows for conditional heteroskedasticity in the structural shocks. Hall and Horowitz (1996) recentering ensures that the bootstrap data-generating process satisfies the over-identifying restrictions exactly at the estimated value, so that bootstrap critical values adapt to the effective degrees of freedom.

Monte Carlo evidence from a three-variable VAR model validates the theoretical results. The estimates are consistent and standard inference remains valid irrespective of the strength of the proxies. For identification robust inference, we also show the HGMM estimator provides narrower Anderson-Rubin (AR) confidence sets relative to the inference based only on proxy exclusion conditions. The joint criterion eliminates the failure of proxy identification without sacrificing the validity of the identification-robust coverage. For the specification tests, under the null hypothesis of valid identification, all three test statistics J_{joint}, J_{prx}, and J_{NG} exhibit close to nominal size. Under failure of proxy exogeneity, the power ordering $J_{\text{prx}} \gg J_{\text{NG}}$ holds uniformly across sample sizes: the dominance of J_{prx} is consistent with violation of proxy exogeneity as the primary source of mis-specification, with J_{NG} inheriting relatively lower, albeit false, power only through contamination via the shared estimate. Also, J_{prx} provides a valid formal test of instrument orthogonality even when the instrument is weak.

We demonstrate the framework by revisiting identification of an oil-news shock and a Euro-area monetary policy shock. For the oil news shock of Känzig (2021), estimated under the instrument correction of Kilian (2024), the hybrid criterion delivers substantially narrower Anderson–Rubin confidence sets than the proxy alone. The median reduction in set length is 71.6%, and are sharp enough to characterize the shock as predominantly oil demand news shock. Notably for the real oil price, the confidence set for the on-impact estimate shrinks by a factor of nearly ten once the higher-order moment conditions are added. For a Euro-area monetary policy shock identified from the high-frequency surprises of Altavilla et al. (2019), the gains are more modest with median reduction in set length of 26%. Together the illustrations show that the non-Gaussianity of the structural shocks sharpen the inference when the higher-order moment conditions are informative enough to offset the cost of the additional degrees of freedom.

The remainder of the paper is organized as follows. Section 2 presents the hybrid GMM identification, asymptotic properties of the estimator and derives joint limiting distributions of the structural parameters; Section 3 proves consistency with standard asymptotic inference under weak proxies, and asymptotic bias bounds under local violations of proxy exogeneity; Section 4 develops the mutually orthogonal specification tests and the required bootstrap procedure for finite-sample critical values; Section 5 reports Monte Carlo evidence; Section 6 presents the empirical illustrations and Section 7 concludes. Appendix A consolidates preliminary notations and conventions used in the paper, followed by proofs of the propositions, additional results and robustness checks. A brief review of related literature follows next.

1.1. **Related Literature.** A growing literature combines different sources of identifying information in SVARs, and it largely falls into two groups. The first works within a Bayesian framework and augments a primary identification scheme with sign, zero or narrative restrictions: Baumeister and Hamilton (2015) combine sign restrictions with prior information on structural parameters, Arias et al. (2021) embed proxies in traditional zero and sign restrictions, and Giacomini et al. (2022) develop robust inference for proxy-SVARs with narrative restrictions. Since these restrictions are typically inequalities, the resulting models are generally set-identified, and inference relies on posterior or robust-Bayesian bounds. The second group obtains additional identifying information from time variation in second moments: Carriero et al. (2024) blend heteroskedasticity with sign and narrative restrictions and instrumental variables, and Angelini et al. (2024b) combine external instruments with changes in volatility regimes. Here the gain in identification depends on the existence, timing and size of the volatility shifts. Moreover, our approach differs from both groups. The additional information comes from the higher-order moments of the structural shocks, which are available in the sample without specifying volatility regimes. Also, these moment conditions enter a single frequentist GMM criterion that point-identifies the target shocks.

Most closely related studies are the over-identified GMM proxy-SVAR of Gregory et al. (2024), and its generalization by Bruns et al. (2025) which relaxes the proxy count by focusing the criterion on the proxy-targeted columns. Both, however, draw their over-identifying content solely from *second-moment* uncorrelatedness, which vanishes when only a single shock is targeted and offers no identifying power for the non-targeted columns. We instead draw identifying information from higher-order moments under non-Gaussianity. This implies that even with a single proxy we obtain a test for proxy validity as the non-Gaussian moment block supplies over-identifying information. We further show robustness to weak proxy relevance where uncorrelatedness-based approaches lose identifying content entirely. Relative to the Bayesian non-Gaussian SVAR of Keweloh et al. (2025), which combines statistical identification with potentially endogenous proxies, ours is a frequentist approach which provides explicit weak instrument robustness, efficiency results, formal specification tests, and an explicit bias bound under local exogeneity violations.

When instruments are weak, Montiel Olea et al. (2021) document severe under-coverage of standard proxy-SVAR confidence sets and propose Anderson–Rubin (AR) test inversion for robust inference, Angelini et al. (2024a) restore standard asymptotics for weakly instrumented target shocks by requiring strong auxiliary instruments for the non-target shocks. The HGMM framework remains robust under local-to-zero proxy relevance through a different mechanism: the non-Gaussian Jacobian can retain full column rank as the proxy relevance collapses, keeping the GMM information matrix bounded away from zero in the limit. We consequently obtain substantially narrower AR confidence sets than the proxy-only criterion, comparable in length to the standard HGMM intervals. An applied researcher therefore needs neither additional instruments nor the heavy computation of AR sets to retain valid inference.

In statistical identification, Comon (1994) establishes that non-singular linear mixing matrices are identified up to sign and column permutation when the sources are mutually independent with at most one Gaussian. Keweloh (2021) provides a unified co-skewness and co-kurtosis moment structure, and Lanne and Luoto (2021) construct a GMM estimator based on fourth-order co-kurtosis conditions; Lanne et al. (2023) show that $n(n-1)/2$ symmetric co-kurtosis conditions globally identify the impact matrix when at least $n-1$ shocks have excess kurtosis of a common sign; the condition we adopt for global identification. In our moment set the co-skewness conditions deliver local identification and the common-sign co-kurtosis conditions deliver global identification. We combine them with proxy exclusion restrictions into a joint criterion and develop Neyman-orthogonal specification tests. Lewis (2024) provides a comprehensive review of statistical identification.

2. Hybrid GMM methodology

This section develops the hybrid GMM (HGMM) framework. We parameterize the impact matrix B as a function of Givens rotation angles, ϕ, as in Matteson and Tsay (2011), and estimate the structural parameters by minimizing an efficiently weighted moment criterion using residuals from a first-stage OLS estimation. Combining the two channels delivers global point identification of the target columns of B (Proposition 2.1). The joint asymptotic distribution of the full parameter vector, including autoregressive coefficients, is derived. We characterize the efficiency gains (Proposition 2.2) of the hybrid estimator relative to the proxy-only estimator. The asymptotic variance of $\hat{\phi}$ is corrected for the generated-regressor distortion from first-stage estimation error as noted in Murphy and Topel (1985), and the standard errors for all elements of B and the structural impulse-response functions follow with standard derivations from Lütkepohl (2005).

2.1. SVAR setup.

We consider a structural VAR(p) model:

$$y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + B\varepsilon_t, \qquad t = 1, \ldots, T, \tag{2.1}$$

where y_t is an $n \times 1$ vector of endogenous variables, ν is a vector of intercepts and A_j are $n \times n$ autoregressive matrices. The reduced-form innovations, u_t and structural shocks are linearly related through the $n \times n$ non-singular impact matrix B:

$$u_t = B\varepsilon_t \tag{2.2}$$

where u_t is a sequence of serially uncorrelated reduced-form disturbances with $\mathbb{E}[u_t] = 0$ and covariance matrix $\Sigma_u := \mathbb{E}(u_t u_t')$. We further assume y_t is stable, i.e.

Assumption 2.1. $\det A(z) = \det(I_n - A_1 z - \cdots - A_p z^p) \neq 0, \quad |z| \leq 1$

To span the space of observationally equivalent impact matrices, let H denote the lower-triangular Cholesky factor satisfying $HH' = \Sigma_u$. We adapt the notation of Matteson and Tsay (2011) and parameterize $B = HQ(\phi)$, where $\phi \in \Phi \equiv (-\pi, \pi]^{p_\phi}$, $Q(\phi) \in SO(n)$ is an orthogonal rotation matrix governed by the $p_\phi \times 1$ ($p_\phi = n(n-1)/2$) vector of rotation

angles ϕ. $SO(n)$ is a subgroup within the group of all $n \times n$ orthogonal matrices $O(n)$, with determinant equal to $+1$. There exists a unique mapping of $Q \in SO(n)$ into $\phi \in (-\pi, \pi]^{p_\phi}$, continuous if either all elements on the main diagonal of Q are positive or all elements of Q are non-zero. For a detailed exposition, see Matteson and Tsay (2011, 2013). Hence, the structural shocks are recovered as

$$\varepsilon_t(\phi) = Q(\phi)' H^{-1} u_t = B^{-1} u_t. \tag{2.3}$$

2.2. **Proxy identification.** Consider m_t as a $k \times 1$ vector of external instruments. Identification through instruments requires that each proxy $m_{l,t}$ satisfies a relevance condition, i.e. nonzero correlation with its target shock $\varepsilon_{j,t}$ and an exogeneity condition to all remaining shocks (Stock and Watson, 2018):

$$\mathbb{E}[m_{l,t}\, \varepsilon_{j,t}] = \psi_l \neq 0, \qquad \mathbb{E}[m_{l,t}\, \varepsilon_{i,t}] = 0 \quad \text{for } i \neq j. \tag{2.4}$$

Hence, we formally assume:

Assumption 2.2. *There exists a $k \times 1$ instrument vector m_t, with $k \geq 1$, satisfying*

$$\mathbb{E}[m_t \varepsilon_t'] = \big(\Psi \ , \ 0_{k \times (n-k)} \big),$$

with targeted shocks indexed $1, \ldots, k$ and $\Psi = \operatorname{diag}(\psi_1, \ldots, \psi_k)$ non-singular

These conditions yield $q_{\text{prx}} = k(n-1)$ (strict) exclusion restrictions. The corresponding sample moment block is

$$g_{P,t}(\phi) = \begin{pmatrix} m_{1,t} \cdot \varepsilon_{i,t}(\phi) \\ \cdots \\ m_{k,t} \cdot \varepsilon_{i,t}(\phi) \end{pmatrix}, \qquad l \in \{1, \ldots, k\}, \quad i \neq \text{target}(l). \tag{2.5}$$

2.3. **Identification through non-Gaussianity.** The statistical identification of SVAR models via non-Gaussianity exploits the information contained in the higher-order moments of the structural shocks. This approach yields identification of B up to sign and column permutation, as established by Comon (1994). For a detailed review, we refer the reader to Hyvärinen (2013) and Lewis (2024). Usually, the key maintained assumption is that the structural shocks, ε_t are mutually independent, but recent literature has shown that it is not a necessary condition for identification, see Mesters and Zwiernik (2024); Lanne and Luoto (2021); Guay (2021). The identifying restrictions require only that the structural shocks be mutually uncorrelated with vanishing higher-order co-moments, together with a non-Gaussianity condition on their marginals; full statistical independence is not needed[2]. Formally, we assume:

Assumption 2.3. *ε_t is strictly stationary, with components of mean zero, unit variance, mutual uncorrelatedness $\mathbb{E}[\varepsilon_{i,t}\varepsilon_{j,t}] = 0$ for $i \neq j$, finite fourth moments, and diagonal third-*

[2]Mutual independence would additionally force all cross-cumulants of every order to vanish, whereas only the third- and fourth-order co-moments enter the moment conditions, so these are imposed directly. This is strictly weaker than independence and does not require recovering independent components from the reduced-form innovations (Kilian and Lütkepohl, 2017); see Remark 2.1.

and fourth-order cumulant tensors,

$$\operatorname{cum}(\varepsilon_{i,t}, \varepsilon_{j,t}, \varepsilon_{k,t}) = 0 \quad \text{unless } i = j = k, \tag{2.6}$$

$$\operatorname{cum}(\varepsilon_{i,t}, \varepsilon_{j,t}, \varepsilon_{k,t}, \varepsilon_{l,t}) = 0 \quad \text{unless } i = j = k = l, \tag{2.7}$$

with diagonal elements $\kappa_{3,k} = \mathbb{E}[\varepsilon_{k,t}^3]$ *and* $\kappa_{4,k} = \mathbb{E}[\varepsilon_{k,t}^4] - 3$.

Assumption 2.4. *The structural shocks satisfy:*
(a) *at most one has zero skewness, i.e.* $\kappa_{3,m} = \mathbb{E}[\varepsilon_{m,t}^3] \neq 0$ *for at least* $n-1$ *shocks; and*
(b) *at least* $n-1$ *have nonzero excess kurtosis of a common sign, i.e.* $\kappa_{4,m} = \mathbb{E}[\varepsilon_{m,t}^4] - 3$ *is of equal sign across the non-mesokurtic shocks.*

Both parts of Assumption 2.4 can be read as deviations from the Stein identity function: for a standard Gaussian X, $\mathbb{E}[Xh(X)] - \mathbb{E}[h'(X)] = 0$ for every continuously differentiable h, the quantities $\kappa_{3,k}$ and $\kappa_{4,k}$ in Assumption 2.3 are exactly the second and third-order deviations. Part (i) of the Assumption 2.4 is standard, Part (ii) is slightly stronger but mild in practice. Apart from the listed studies in Table K.2 and as noted by Lanne et al. (2023), macro-financial data are mostly leptokurtic with few exceptions. Excess kurtosis is positive for asset returns, exchange rates and high-frequency policy surprises, as well as for output and commodity prices at monthly and quarterly frequency, over essentially any sample containing a disruption to the global economy, such as Global Financial Crisis 2007-08, COVID-19. Also, with conditionally heteroskedastic shocks, positive excess kurtosis follows mechanically[3]. Negative excess kurtosis instead requires short-tailed or bounded shocks, and part (b) fails only if at least two of these occur in the same system, since $n-1$ shocks need only share a sign.

We operationalize these assumptions by mapping the relevant set of higher-order moment conditions into vector $g_{NG,t}(\phi)$. For an n-variable VAR, we adopt a baseline parameterization with $q_{\text{NG}} = n(n-1) + n(n-1)/2$ moment conditions:

- *Co-skewness (Group 1):* The $n(n-1)$ third-order cross-moment conditions $\mathbb{E}[\varepsilon_{i,t}^2\, \varepsilon_{j,t}] = 0$ for all ordered pairs (i, j) with $i \neq j$.
- *Symmetric co-kurtosis (Group 2):* The $n(n-1)/2$ fourth-order conditions $\mathbb{E}[\varepsilon_{i,t}^2\, \varepsilon_{j,t}^2] = 1$ for all $i < j$.

Two properties of these moment conditions are worth making explicit. First, they are unconditional moment restrictions that hold under Assumption 2.3 without independence: uncorrelatedness and the diagonal third- and fourth-order cumulant tensors give $\mathbb{E}[\varepsilon_{it}^2 \varepsilon_{jt}] = 0$ and $\mathbb{E}[\varepsilon_{it}^2 \varepsilon_{jt}^2] = 1$ directly. Under the dependence conditions of Section 2.6.1, we allow for idiosyncratic conditional heteroskedasticity in the shocks (ARCH/GARCH, stochastic volatility): each $\varepsilon_{i,t}$ is a martingale difference sequence, serially uncorrelated, while its conditional variance is serially dependent. Second, the central limit theorem for the quartic conditions

[3]For $\varepsilon_{m,t} = \sigma_{m,t} z_{m,t}$ with $z_{m,t}$ independent of the scale, $\kappa_{4,m} \geq \kappa_{4,z_m}$ with strict inequality whenever $\sigma_{m,t}$ is non-degenerate, and $\kappa_{4,m} > 0$ when $z_{m,t}$ is Gaussian. The variations in shocks' volatility therefore shift excess kurtosis upward for every shock and cannot by itself generate a mixture of signs.

requires $\mathbb{E}\|\varepsilon_t\|^{8+4\delta} < \infty$; under cross-sectional independence the relevant cross-moments factorize and $4 + 2\delta$ marginal moments suffice. The general requirement is satisfied by the non-Gaussian parametric families considered (normal-inverse Gaussian, χ^2, and Student-t with degrees of freedom exceeding $8 + 4\delta$).

Remark 2.1 (Scope of the non-Gaussian moment block)**.** Assumption 2.3 already relaxes full statistical independence: it requires only mutual uncorrelatedness and diagonal third- and fourth-order cumulant tensors. There are other sets of moment conditions in the literature, resting on different maintained assumptions: Lanne and Luoto (2021) use co-kurtosis conditions under mutual orthogonality, and Keweloh (2021) uses all third- and fourth-order moments under independence.

We do not rank these strategies by identification strength. The central claim is that *any* non-Gaussian identified SVAR, once expressed as sufficiently identifying moment conditions under appropriate assumptions, admits augmentation to proxy exclusion conditions within a unified GMM criterion. To illustrate, consider the assumptions of Liu (2025): (i) all structural shocks have zero co-skewness and zero co-kurtosis, except that symmetric co-kurtoses may be nonzero and no two shocks share identical fourth-moment relationships, i.e. $\sum_{i=1}^n \mathbb{E}(\varepsilon_{it}^2\varepsilon_{jt}^2) \neq \sum_{i=1}^n \mathbb{E}(\varepsilon_{it}^2\varepsilon_{kt}^2)$ for all $j \neq k$; and (ii) at most one structural shock has both zero skewness and zero excess kurtosis. Under these assumptions, our symmetric co-kurtosis conditions can be replaced by asymmetric co-kurtosis conditions. The resulting moment conditions identify B up to sign and column permutation, see Liu (2025); Mesters and Zwiernik (2024). These assumptions are strictly weaker than full independence and also allow for common volatility processes where the structural shocks exhibit correlated volatilities.

The analysis of the next section therefore takes $g_{NG,t}(\phi)$ as given under Assumptions 2.3 and 2.4, and characterizes its incremental identifying information relative to the proxy exclusion conditions.

2.4. **Global identification of B.** By combining the proxy exclusion and sufficient higher-order moment conditions within the GMM objective, we achieve global point identification of the target columns of B, as established in Proposition 2.1.

Proposition 2.1 (Point Identification with external instruments and non-Gaussian shocks)**.** *Let $u_t = B_0\varepsilon_t$ with $B_0 \in \mathbb{R}^{n\times n}$ non-singular, the VAR satisfies Assumption 2.1, and let ε_t and the $k\times 1$ proxy vector m_t, $1 \leq k \leq n$, satisfy Assumptions 2.2–2.4. Partition $B_0 = [B_{0,1} \;\; B_{0,2}]$, where $B_{0,1}$ collects the k target columns and $B_{0,2}$ the remaining $n-k$ columns, and impose the sign normalization*

$$[B_0]_{ii} > 0, \qquad i = 1, \ldots, k, \tag{2.8}$$

For $d \geq 1$, let $\mathcal{P}_d$ denote the set of $d\times d$ permutation matrices and $\mathcal{D}_d^{\pm}$ the set of $d\times d$ diagonal matrices with diagonal entries in $\{-1, 1\}$. Then:

(i) ***(Target shocks identification)*** *$B_{0,1}$ is globally point-identified for all $1 \leq k \leq n$.*

(ii) ***(Non-target shocks identification)*** *If $k \geq n-1$, B_0 is globally point-identified. If $1 \leq k < n-1$, $B_{0,2}$ is identified up to the sign and column permutation, with identified*

set

$$\mathcal{E}_{n-k} = \left\{ B_{0,2} P \Lambda : P \in \mathcal{P}_{n-k},\ \Lambda \in \mathcal{D}^{\pm}_{n-k} \right\}, \qquad |\mathcal{E}_{n-k}| = (n-k)! \cdot 2^{n-k}.$$

Proof: See Appendix B.1.

Proposition 2.1 subsumes both "full" and "partial" identification[4] as special cases of a unified framework[5]. When $k < n-1$, the target sub-matrix $B_{0,1}$ is point-identified while the remaining $B_{0,2}$ remains identified only up to a finite signed permutation equivalence class. In both cases, identification of the target columns is achieved through the joint restrictions imposed by proxy exclusion and higher-order moment conditions. As we will see in Section 3.1, this result remains valid under local-to-zero proxy relevance, that is, when the proxy instrument is arbitrarily weak provided that the Jacobian of the non-Gaussian moment block retains full column rank at the true parameter.

This proposition also resolves an interpretive limitation of pure statistical identification. Through non-Gaussianity, we recover the columns of B_0 up to signed permutation indeterminacy, with no intrinsic mapping between statistically identified components and economically meaningful structural shocks. The proxy instrument conditions eliminate this indeterminacy for the target columns. The external instrument, whose economic interpretation is grounded in theoretical prior knowledge, selects the unique permutation and sign normalization consistent with both the higher-order moment conditions and the proxy relevance and orthogonality conditions. Economic identification of the target shocks is therefore achieved as a direct consequence of the moment conditions themselves, rather than through post-estimation labeling conventions applied to anonymous statistical components.

2.5. **GMM estimation.** The unified, over-identified system combines $q = q_{NG} + q_{\mathrm{P}}$ moment equations into the stacked average:

$$\bar{g}_T(\phi) = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} g_{NG,t}(\phi) \\ g_{P,t}(\phi) \end{pmatrix}. \tag{2.9}$$

We can estimate it using standard Hansen (1982) two-step GMM. The first-step estimator minimizes the unweighted criterion: $\hat{\phi}_1 = \arg\min_\phi \ \bar{g}_T(\phi)' \bar{g}_T(\phi)$. The first-step residuals are used to construct the optimal weighting matrix:

$$\hat{S}_T = \frac{1}{T} \sum_{t=1}^{T} \big(g_t(\hat{\phi}_1) - \bar{g}_T(\hat{\phi}_1)\big)\big(g_t(\hat{\phi}_1) - \bar{g}_T(\hat{\phi}_1)\big)'. \tag{2.10}$$

The second-step estimator minimizes the weighted objective function, yielding:

$$\hat{\phi}_T = \arg\min_{\phi} \ \bar{g}_T(\phi)' \, \hat{S}_T^{-1} \, \bar{g}_T(\phi). \tag{2.11}$$

[4]The notion of identification employed is that of global identification in the sense of Komunjer (2012): ϕ_0 is globally identified if the population moment conditions admit ϕ_0 as their unique solution over all of Φ. This is strictly stronger than the local identification condition of Rothenberg (1971), which only rules out a continuum through ϕ_0

[5]Angelini and Fanelli (2019) establish necessary and sufficient conditions for point identification in SVAR models with r instruments identifying $g \leq r$ structural shocks. We do not pursue the general case.

Under the mixing condition of Section 2.6.1, the optimal weighting matrix is the Newey and West (1987) HAC[6] estimator with data-dependent bandwidth of Andrews (1991). Under serial independence of the shocks, the autocovariances vanish and $\hat{S}_T$ reduces to the covariance matrix above. For subsequent derivations, we conformably partition the optimal inverse-weighting matrix as:

$$S = \begin{pmatrix} S_{NN} & S_{NP} \\ S_{PN} & S_{PP} \end{pmatrix}, \qquad S_{NP} = S_{PN}', \tag{2.12}$$

Also, we define the population Jacobian as

$$G(\phi) = \frac{\partial g(\phi)}{\partial \phi'} \in \mathbb{R}^{q \times p_\phi}, \qquad G \equiv G(\phi_0) = \big(G_{NG}', G_P'\big)', \tag{2.13}$$

where $G_P = \partial g_P(\phi_0)/\partial \phi'$ and $G_{NG} = \partial g_{NG}(\phi_0)/\partial \phi'$, and a consistent estimator of Jacobian is $\hat{G}_T(\phi)$.

2.6. **Joint inference with VAR parameters.** With a standard set of regularity conditions, we show consistency and asymptotic normality of the HGMM estimator of the rotation parameters ϕ, and its efficiency gains relative to proxy-only criterion. Also, we provide joint asymptotic distribution of the estimates of autoregressive and structural rotation parameters. Throughout, let $\psi = (\theta', \phi')'$ denote the full parameter vector, with $\theta = \text{vec}(\nu, A_1, \ldots, A_p)$ collecting the reduced-form VAR coefficients and ϕ the structural rotation parameters.

2.6.1. *Regularity conditions.*

(1) The parameter space for the rotation angles $\phi \in (-\pi, \pi]^{p_\phi}$ is compact.
(2) For some $\delta > 0$, with $\mathcal{F}_t = \sigma\big((u_s', m_s')' : s \le t\big)$:
   (a) $\{(u_t', m_t')'\}$ is strictly stationary and $\mathbb{E}[u_t \mid \mathcal{F}_{t-1}] = 0$ almost surely;
   (b) $\{(u_t', m_t')'\}$ is α-mixing with $\sum_{h \ge 1} \alpha(h)^{\delta/(2+\delta)} < \infty$;
   (c) $\mathbb{E}\|u_t\|^{8+4\delta} < \infty$ and $\mathbb{E}\|m_t\|^{4+2\delta} < \infty$.
(3) The long-run variance $S = \sum_h \mathbb{E}\big[g_t(\phi_0) g_{t-h}(\phi_0)'\big]$ is positive definite.
(4) The fourth-order cumulants of $\{g_t(\phi_0)\}$ are absolutely summable, so that the HAC estimator $\hat{S}_T$ is consistent.

Condition 1 guarantees that the extremum estimator is well-defined and that a uniform LLN applies to the criterion, delivering consistency once ϕ_0 is the unique minimizer of the criterion function, as provided by the identification result in Proposition 2.1. Condition 2 governs the temporal dependence of the data. Part (a) implies u_t is a martingale difference sequence, allowing idiosyncratic ARCH/GARCH or stochastic volatility. Parts (b)–(c) deliver a central limit theorem for $T^{-1/2} \sum_t g_t(\phi_0)$. Condition 3 concerns the resulting optimal weighting matrix. Since the moment conditions are serially correlated, the normalized sum has the long-run variance $S = \sum_h \mathbb{E}[g_t(\phi_0) g_{t-h}(\phi_0)']$; positive definiteness of S ensures the efficient weighting S^{-1} and the asymptotic variance $(G' S^{-1} G)^{-1}$ are well-defined (Andrews, 1991).

[6] $\hat{S}_{HAC,T} = \hat{\Gamma}_0 + \sum_{i=1}^{T-1} \omega_{i,T}(\hat{\Gamma}_i + \hat{\Gamma}_i')$, where $\hat{\Gamma}_i$ is a consistent estimator of Γ_i, the ith autocovariance matrix of the moment conditions, $g_t(\phi_0)$, with some kernel weights $\omega_{i,T}$.

Under the regularity conditions in 2.6.1 and Assumptions 2.2 - 2.4, the estimator (2.11) is consistent and asymptotically normal, i.e.

$$\sqrt{T}\,(\hat{\phi}_T - \phi_0) \xrightarrow{d} \mathcal{N}\big(0,\ \mathcal{V}_{joint}\big) \tag{2.14}$$

where, $\mathcal{V}_{joint} = \big(G'S^{-1}G\big)^{-1}$ is the asymptotic covariance matrix[7] of the estimator. See Hall (2005) for details.

2.6.2. *Efficiency gains.* The hybrid GMM estimator combines two distinct identification strategies, yielding efficiency gains relative to proxy-based identification alone[8]. The following proposition makes this comparison precise.

Proposition 2.2 (Efficiency of joint GMM)**.** *Under Assumptions 2.2-2.4, and if G_{NG} has full column rank and* $\operatorname{rank}(G_P) = p_\phi$*:*

$$\mathcal{V}_P \ - \ \mathcal{V}_{joint} \ \succeq \ 0. \tag{2.15}$$

where,

$$\mathcal{V}_{joint} \ \equiv \ \mathcal{I}_{joint}^{-1} \ = \ \big(G'S^{-1}G\big)^{-1} \qquad \mathcal{V}_P \equiv \big(G_P'S_{PP}^{-1}G_P\big)^{-1} \tag{2.16}$$

$\mathcal{V}_P$ is the asymptotic covariance matrix of the proxy-only GMM estimator.

The corollary states strict efficiency gains only if the higher-order moment conditions identify ϕ in directions not already identified by the proxy moments.

Corollary 2.1 (Strict efficiency gain)**.** $\mathcal{V}_P - \mathcal{V}_{joint} \succ 0 \iff \operatorname{rank}(\tilde{G}_{NG}) = p_\phi$

where

$$\tilde{G}_{NG} \equiv G_{NG} - S_{NP}S_{PP}^{-1}G_P \tag{2.17}$$

is the "partialed" NG Jacobian. The proof rests on an exact additive decomposition of the joint information matrix, given by:

$$\mathcal{I}_{joint} \ = \ \underbrace{G_P'S_{PP}^{-1}G_P}_{\mathcal{I}_P} \ + \ \underbrace{\tilde{G}_{NG}'\Sigma_{NG}^{-1}\tilde{G}_{NG}}_{\Delta\mathcal{I}_{NG}}, \tag{2.18}$$

where $\Sigma_{NG} \equiv S_{NN} - S_{NP}S_{PP}^{-1}S_{PN} \succ 0$ is the Schur complement of S_{PP} in S, see Newey (1985). Since $\Sigma_{NG}^{-1} \succ 0$, the increment $\Delta\mathcal{I}_{NG} \succeq 0$, and monotonicity of inversion delivers (2.15). A detailed proof can be found in Appendix B.3.

The decomposition (2.18) is a GMM analog of the Frisch-Waugh-Lovell theorem: $\Delta\mathcal{I}_{NG}$ is the information provided by the higher-order moments on ϕ, after projecting out the identifying directions already spanned by the proxy moments. The efficiency gain is thus governed by the extent to which the two channels identify the parameters along distinct directions.

[7]This result assumes the reduced-form innovations (u_t) are known, however in practice the innovations are estimated and hence, we include the generated-regressor correction while deriving the asymptotic distribution of ψ in Section 2.6.3.

[8]Analogous gains hold relative to non-Gaussian identification alone, provided $\operatorname{rank}(\widetilde{G}_P) = p_\phi$. This condition is restrictive in practice, as enough valid proxies to fully identify a moderate-dimensional VAR are rarely available.

2.6.3. *Asymptotic distribution.* Note that the two-step estimation of Section 2.5, requires the autoregressive estimates $\hat{\theta}_T$ which yield the reduced-form innovations $\hat{u}_t$ and the Cholesky factor of the covariance matrix $\hat{\Sigma}_u$, $\hat{H} = \text{chol}(\hat{\Sigma}_u)$. The second stage minimizes the HGMM criterion over ϕ using the generated residuals $\hat{\varepsilon}_t(\phi) = Q(\phi)'\hat{H}^{-1}\hat{u}_t$. Since, the second-stage moment conditions depend on first-stage autoregressive estimates, the asymptotic variance of $\hat{\phi}_T$ requires a generated-regressor correction, as in Murphy and Topel (1985).

Under the regularity conditions in 2.6.1 and Assumptions 2.1 and 2.3,

$$\sqrt{T}\,(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0,\, \mathcal{V}_\theta), \qquad \mathcal{V}_\theta = \mathbb{E}[\zeta_t \zeta_t'], \tag{2.19}$$

where $\zeta_t \equiv (\mathbb{E}[x_t x_t'])^{-1} x_t \otimes u_t$, $\otimes$ denotes the Kronecker product, $x_t = (1, y_{t-1}', \ldots, y_{t-p}')' \in \mathbb{R}^{np+1}$ are VAR regressors (including the intercept). x_t is $\mathcal{F}_{t-1}$-measurable with $\mathcal{F}_{t-1} := \sigma(y_{t-1}, y_{t-2}, \ldots)$ is the natural filtration of $\{y_t\}$. Because ζ_t is MDS under condition 2(a) of Section 2.6.1, $V_\theta = \mathbb{E}[\zeta_t \zeta_t']$ is also its long-run variance; for homoscedastic residuals it simplifies to $V_\theta = \left(\mathbb{E}[x_t x_t']\right)^{-1} \otimes \Sigma_u$.

The joint estimator $\hat{\psi}_T$ solves a stacked moment system identifying ϕ via the proxy exclusion and higher-order moment conditions. Following the two-step estimator theory of Murphy and Topel (1985) and Newey and McFadden (1994), under the regularity conditions of Section 2.6.1 and Assumptions 2.1–2.4,

$$\sqrt{T}\begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{\phi} - \phi_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0,\; \mathcal{V}_\psi \equiv \begin{bmatrix} \mathcal{V}_\theta & \mathcal{V}_{\theta\phi} \\ \mathcal{V}_{\theta\phi}' & \mathcal{V}_\phi \end{bmatrix}\right). \tag{2.20}$$

Since the second-stage moments depend on θ through the generated residuals, first-stage estimation error propagates into $\hat{\phi}_T$. The relevant score is the combined score $g_t + \Gamma_\theta \zeta_t$, with $\Gamma_\theta \equiv \mathbb{E}[\partial g_t(\psi_0)/\partial \theta']$, whose long-run variance we denote Ω_ψ. The generated-regressor corrected variance block is

$$\mathcal{V}_\phi = (G_0' S^{-1} G_0)^{-1} G_0' S^{-1}\, \Omega_\psi\, S^{-1} G_0\, (G_0' S^{-1} G_0)^{-1}, \tag{2.21}$$

estimated by its sample analogue

$$\hat{\mathcal{V}}_\phi = (\hat{G}' \hat{S}_T^{-1} \hat{G})^{-1} \hat{G}' \hat{S}_T^{-1}\, \hat{\Omega}_\psi\, \hat{S}_T^{-1} \hat{G}\, (\hat{G}' \hat{S}_T^{-1} \hat{G})^{-1}. \tag{2.22}$$

Since the higher-order moment conditions of the structural shocks are correlated with the reduced-form innovations that drive the first-stage error, so the cross-covariance between g_t and ζ_t, and with it $\mathcal{V}_{\theta\phi}$, is generically nonzero. Appendix C gives Ω_ψ, $\mathcal{V}_{\theta\phi}$ and their consistent estimators explicitly.

2.6.4. *Inference on the impact matrix.* Since $B = H(\theta)Q(\phi)$ is a smooth differentiable function of ψ, we can apply the delta method to (2.20). Consider the composite Jacobian

$$\mathcal{J}_B = \frac{\partial \,\text{vec}(B)}{\partial \psi'} = \left(\frac{\partial \,\text{vec}(B)}{\partial \theta'},\; \frac{\partial \,\text{vec}(B)}{\partial \phi'}\right) \in \mathbb{R}^{n^2 \times (n^2 p + n + p_\phi)}. \tag{2.23}$$

Both blocks are analytically tractable. For the rotation component:

$$\frac{\partial \operatorname{vec}(B)}{\partial \phi'} = (I_n \otimes H) \frac{\partial \operatorname{vec}(Q(\phi))}{\partial \phi'}, \tag{2.24}$$

where the Givens Jacobian $\partial \operatorname{vec}(Q(\phi))/\partial \phi'$ has explicit closed form as a sum of antisymmetric generator matrices, as shown in Matteson and Tsay (2011). For the Cholesky component:

$$\frac{\partial \operatorname{vec}(B)}{\partial \theta'} = (Q(\phi)' \otimes I_n) \frac{\partial \operatorname{vec}(H)}{\partial \operatorname{vech}(\Sigma_u)'} \cdot \frac{\partial \operatorname{vech}(\Sigma_u)}{\partial \theta'}, \tag{2.25}$$

where the Cholesky derivative uses the implicit differentiation formula of Kleffe (1989) via the duplication matrix D_n, and $\partial \operatorname{vech}(\Sigma_u)/\partial \theta'$ follows from standard theory.

The delta method applied to (2.20) with Jacobian $\mathcal{J}_B$ in (2.23) yields

$$\sqrt{T} \operatorname{vec}(\hat{B}_0 - B) \xrightarrow{d} \mathcal{N}\big(0,\ \mathcal{J}_B \mathcal{V}_\psi \mathcal{J}_B'\big). \tag{2.26}$$

We can estimate the standard errors for $[\hat{B}_0]_{ij}$ as the square roots of the corresponding diagonal entries of $\mathcal{J}_B \hat{\mathcal{V}}_\psi \mathcal{J}_B'$. Once we have established the inference for the impact matrix B, the derivation of the asymptotic distribution of the structural impulse response functions (SIRFs) follows directly from Lütkepohl (2005), Chapter 3. We defer it to the Appendix D.

3. Behavior under weak and contaminated proxies

In this section, we provide three main results. First, Proposition 3.1 shows that under local-to-zero weak proxies, the HGMM estimator remains consistent and asymptotically normal, with the non-Gaussian moment block acting as the identifying anchor. The estimator also attains a smaller asymptotic variance than the standalone non-Gaussian GMM estimator, the gain arising from the cross-block long-run covariance between the proxy exclusion restrictions and higher-order moment conditions. Second, Proposition 3.2 shows that when the higher-order moments are sufficiently informative, standard Wald inference is uniformly, rather than only pointwise, valid over the proxy relevance. Third, Proposition 3.3 shows that the estimator incurs a finite asymptotic bias when the proxies show local violations of exogeneity condition through correlation with non-target structural shocks. We derive an explicit bound in terms of the degree of violation of exogeneity and the strength of the non-Gaussian identification.

3.1. Robust identification with weak proxies.

In proxy-SVARs, identification rests on the relevance condition $\mathbb{E}[m_t \varepsilon_{1,t}] = \psi_1 \neq 0$. As $\|\psi_1\|$ shrinks, the Jacobian of the proxy moments loses rank, and a proxy-only estimator inherits the problems of weak identification (Newey and Windmeijer, 2009; Montiel Olea et al., 2021). By augmenting the external instrument with higher-order moments of the structural shocks, we can resolve this rank deficiency. The resulting estimator is consistent, and at least as efficient as the standalone NG-GMM estimator, even when proxy relevance is local-to-zero. The following proposition formalizes this result.

Proposition 3.1 (Behavior under local-to-zero proxy relevance)**.** *Suppose the conditions of Proposition 2.1 hold, but proxy relevance follows the local-to-zero (Pitman) sequence*

$$\mathbb{E}[m_{t,T}\varepsilon_t'] = \frac{1}{\sqrt{T}}\big(\, C \;,\; 0_{k\times(n-k)} \,\big), \qquad C = \mathrm{diag}(c_1,\ldots,c_k), \quad c_i > 0,$$

and let Σ_{NG} be the Schur complement defined in (2.18), suppose $\mathcal{I}^* \equiv G_{NG}'\,\Sigma_{NG}^{-1}\,G_{NG}$ is positive definite. Then:

(i) The population proxy Jacobian satisfies $G_{P,T} = O(T^{-1/2})$, and the joint information matrix:,

$$\mathcal{I}_{\text{joint},T} \;\longrightarrow\; \mathcal{I}^* \tag{3.1}$$

(ii) The hybrid GMM estimator satisfies

$$\hat{\phi} \;\xrightarrow{p}\; \phi_0, \tag{3.2}$$

$$\sqrt{T}\,(\hat{\phi} - \phi_0) \;\xrightarrow{d}\; \mathcal{N}\big(0,\; \mathcal{V}_{NG}^*\big), \tag{3.3}$$

$$\mathcal{V}_{NG}^* = (\mathcal{I}^*)^{-1} \tag{3.4}$$

(iii) $\mathcal{V}_{NG}^* \preceq \mathcal{V}_{NG} \equiv (G_{NG}'\,S_{NN}^{-1}\,G_{NG})^{-1}$, $\mathcal{V}_{NG}$ being the asymptotic variance of the optimal standalone NG-GMM estimator. Equality holds if the cross-block long-run covariance is such that $S_{NP} = 0$.

Proof. See Appendix B.4.

Parts (i) and (ii) show that, even when the relevance of the proxies is local-to-zero, the hybrid estimator remains consistent and asymptotically normal. This result is explained by the fact that identification is anchored by the non-Gaussian moment block, whose Jacobian G_{NG} does not depend on the proxy relevance.

Part (iii) is an interesting result. One might expect that, since the proxy is weak, the hybrid estimator reduces to a standalone NG-GMM estimator and attains asymptotic variance $\mathcal{V}_{NG}$. Part (iii) shows that this conjecture is an understatement. $\mathcal{V}_{NG}^*$ weights the NG moment block by the Schur complement, $\Sigma_{NG} = S_{NN} - S_{NP}S_{PP}^{-1}S_{PN} \preceq S_{NN}$, i.e. the variance after projecting out the proxy moment conditions. Hence, at any fixed proxy relevance, the proxy provides efficiency through its covariance structure with the higher-order moments, even when its direct identifying information is negligible. This correlation need not necessarily vanish in the local-to-zero limit, because the proxy can be weakly correlated with the target shock yet correlated with its higher-power transformations. For example, a proxy of the form $m_t = \psi_{1,T}\,\varepsilon_{1,t} + \kappa\,(\varepsilon_{1,t}^2 - 1) + \eta_t$, with $\kappa \neq 0$, responds to the size of the realized shock regardless of its sign, so that large surprises of either sign move the proxy in the same direction. This implies that $\mathbb{E}[m_t\varepsilon_{1,t}^2] \neq 0$ even as $\psi_{1,T} \to 0$.

The pointwise results above hold for a fixed sequence of data-generating processes. The next subsection shows that under certain conditions where identification through non-Gaussianity is sufficiently strong, standard Wald inference is valid uniformly over a class of data-generating processes in which proxy relevance may be arbitrarily small.

3.2. **Uniform validity of standard inference.** Let $\gamma = (\phi_0, \theta_0, \Psi, F) \in \mathcal{G}$ denote the data-generating process, where F is the joint distribution of the structural shocks and the proxy. All population objects below (G_{NG}, Σ_{NG}, S) are functionals of γ.

Assumption 3.1 (Uniform non-Gaussian identification)**.** Assumptions 2.1–2.4 hold, and $\gamma \in \mathcal{P}$, where

$$\mathcal{P} = \Big\{\gamma \in \mathcal{G} : \ \sigma_{\min}\big(G_{NG}(\phi_0)\big) \geq \underline{g}, \ \operatorname{eig}\big(S(\gamma)\big) \subset [\underline{s}, \bar{s}], \ \min_{l \leq k} |\psi_l| \ \geq \ \underline{\psi}\Big\}$$

for some constants $\underline{g} > 0$, $0 < \underline{s} \leq \bar{s} < \infty$ and $\underline{\psi} > 0$, with $\sigma_{\min}(\cdot)$ the smallest singular value and $\operatorname{eig}(\cdot)$ the set of eigenvalues.

Assumption 3.1 requires the NG moment block to be uniformly strongly identifying. The proxy is relevant[9] but $\underline{\psi} > 0$ can be arbitrarily small. Together, these conditions bound the information available for inference away from zero uniformly over $\mathcal{P}$: the smallest eigenvalue of the joint information matrix $G'S^{-1}G$ is at least $\underline{g}^2/\bar{s}$ for every $\gamma \in \mathcal{P}$. This bound does not involve $\underline{\psi}$, so the asymptotic variance of $\hat{\phi}_T$ stays bounded for arbitrarily weak proxies. Moreover, the conditions of Section 2.6.1 hold for every $\gamma \in \mathcal{P}$, with mixing coefficients, moment bounds and cumulant sums bounded uniformly over $\mathcal{P}$.

Proposition 3.2 (Uniform validity of Wald inference over proxy strength)**.** *Suppose Assumption 3.1 and the regularity conditions of Section 2.6.1 hold. Let R be a fixed $d \times p_\phi$ matrix of full row rank, and let $\hat{\mathcal{V}}_\phi$ be the generated-regressor corrected variance estimator of (2.22). Then, for $H_0 : R\phi_0 = r$, the Wald statistic*

$$W_T \ = \ T\,(R\hat{\phi} - r)'\left(R\,\hat{\mathcal{V}}_\phi\,R'\right)^{-1}(R\hat{\phi} - r)$$

is asymptotically χ^2_d uniformly over proxy strength:

$$\lim_{T\to\infty} \sup_{\gamma\in\mathcal{P}} \sup_{x\in\mathbb{R}} \left|\,\mathbb{P}_\gamma(W_T \leq x) - F_{\chi^2_d}(x)\,\right| \ = \ 0.$$

Accordingly, the Wald confidence set $\mathcal{C}_{1-\alpha} = \{r : W_T(r) \leq \chi^2_{d,1-\alpha}\}$ has correct asymptotic size, $\lim_{T\to\infty} \inf_{\gamma\in\mathcal{P}} \mathbb{P}_\gamma\big(R\phi_0 \in \mathcal{C}_{1-\alpha}\big) = 1 - \alpha$.

Proof. See Appendix B.5.

Proposition 3.2 is the inferential counterpart of Proposition 3.1. Since the information bound holds uniformly over $\mathcal{P}$, the estimator is asymptotically normal and $\hat{\mathcal{V}}_\phi$ is consistent uniformly in proxy strength, so the Wald statistic retains its χ^2 limit. In practice, when the structural shocks are sufficiently non-Gaussian,[10] conventional delta-method standard errors for the structural parameters are valid.

Remark 3.1 (The case of doubly-weak identification)**.** Proposition 3.2 is uniform over proxy strength while the non-Gaussian block is uniformly strongly identifying ($\sigma_{\min}(G_{NG}) \geq \underline{g}$). If

[9] In the Appendix, Corollary B.2 establishes that strict proxy irrelevance ($\Psi = 0$) collapses identification to the non-Gaussian identified set $\mathcal{B}$. So indeed, the proxy is still required to select the target column, but only out of a discrete (not a continuum) number of choices. This allows the proxy relevance to be arbitrarily close to zero and NG moment block anchors the identification with uniformly valid standard inference.

[10] See Appendix K for the empirical evidence.

non-Gaussianity is itself weak ($\sigma_{\min}(G_{NG}) \to 0$, i.e. near-Gaussian shocks), the DGP leaves $\mathcal{P}$, and the limit is non-standard; Wald inference is then not uniformly valid. For that regime we can rely on the identification-robust Anderson–Rubin confidence sets which retain correct coverage uniformly over the strength of both channels (Montiel Olea et al., 2021; Hoesch et al., 2024).

3.3. **Bounded asymptotic bias under local violations of proxy exogeneity.** In applied macroeconomic and financial research, external instruments may be endogenous by narrative mis-classification, or the inadvertent capture of confounding information, as discussed in Bauer and Swanson (2022); Miranda-Agrippino and Ricco (2021). When a proxy co-varies with the non-target shocks, the orthogonality condition fails, i.e. the population expectation $\mathbb{E}[g_P(\phi_0)] = \mu \neq 0$, inducing a non-vanishing asymptotic bias in the structural parameter estimates.

The hybrid GMM framework permits a precise characterization of this bias and its determinants. Because identification operates through two distinct channels, the asymptotic bias is not governed solely by the degree of the violation of proxy exogeneity but sufficiently strong non-Gaussian identification attenuates the bias. The following proposition makes this relationship explicit.

Proposition 3.3 (Asymptotic bias bound under local violations of proxy exogeneity)**.** *Suppose Assumptions 2.3–2.4 hold, so $G_{NG}(\phi_0)$ has full column rank at ϕ_0, and let the estimator adopts the efficient weighting matrix $W = S^{-1}$. The proxy moment block is locally contaminated,*

$$\mathbb{E}[g_P(\phi_0)] = \mu_T = c/\sqrt{T}, \quad \|c\| < \infty, \qquad \mathbb{E}[g_{NG}(\phi_0)] = 0.$$

Let the partialed-out proxy Jacobian be $\tilde{G}_P = G_P - S_{PN}S_{NN}^{-1}G_{NG}$ and the Schur complement $\Sigma_P = S_{PP} - S_{PN}S_{NN}^{-1}S_{NP}$. Also, suppose $\phi_{*,T}$ minimizes $\mathcal{Q}_T(\phi) = \mathbb{E}[g(\phi)]'S^{-1}\mathbb{E}[g(\phi)]$ and denote $\mathbb{B}_T = \sqrt{T}(\phi_{*,T} - \phi_0)$ as the scaled bias. Then

$$\mathbb{B}_T = -(G'S^{-1}G)^{-1}\,\tilde{G}_P'\,\Sigma_P^{-1}\,c + O(T^{-1/2}), \tag{3.5}$$

The Euclidean norm of the scaled bias is bounded by

$$\|\mathbb{B}_T\|_2 \;\leq\; \frac{\left\|\tilde{G}_P'\,\Sigma_P^{-1}\,c\right\|_2}{\lambda_{\min}\!\left(G_{NG}'S_{NN}^{-1}G_{NG}\right)} \;+\; O(T^{-1/2}). \tag{3.6}$$

Proof. See Appendix B.6. In (3.6), the numerator $\left\|\tilde{G}_P'\,\Sigma_P^{-1}\,c\right\|_2$ is the raw contamination, transmitted only through the proxy moment conditions. The denominator $\lambda_{\min}\!\left(G_{NG}'S_{NN}^{-1}G_{NG}\right)$ is the curvature contributed by the non-Gaussian block alone contributes to the population criterion. The bias ceiling is therefore strictly decreasing in the strength of non-Gaussian identification: as $\lambda_{\min}\!\left(G_{NG}'S_{NN}^{-1}G_{NG}\right)$ grows, the higher-order moment conditions anchor the estimator strongly at ϕ_0 and the distortion from the mis-specified proxy shrinks. A proxy-only estimator has no such denominator. There $\mathcal{I}_{NG} = 0$, and the analogous bound is unbounded whenever $c \neq 0$.

4. Specification tests

The asymptotic theory of Section 2.6 assumes valid specification of both the non-Gaussian moment block and the proxy exogeneity moment conditions. In practice, neither can be taken as given. Proxy exogeneity is not directly testable in isolation. The exclusion restriction $\mathbb{E}[m_{l,t}\varepsilon_{i,t}] = 0$ for $i \neq \text{target}(l)$ involves unobserved structural shocks and cannot be evaluated without auxiliary identifying assumptions. This section constructs the mutually orthogonal specification test statistics for both moment blocks, establishes their asymptotic distributions, and describes the bootstrap procedure used to obtain valid finite-sample critical values. Specifically, we construct three test statistics: J_{joint}, J_{prx}, and J_{NG}.

4.1. **The joint null hypothesis.** Let $\bar{g}_T(\phi)$ denote the sample moment vector of (2.9) with $q = q_{NG} + q_P$ conditions. The joint null hypothesis assumes that the proxy instruments satisfy the exogeneity conditions (2.4), and the structural innovations satisfy the higher-order moment conditions (2.3 and 2.4). Under this null hypothesis, the over-identifying test statistic, as in Hansen (1982), is given by:

$$J_{\text{joint}} = T \cdot \bar{g}_T(\hat{\phi})' \, \hat{S}_T^{-1} \, \bar{g}_T(\hat{\phi}) \xrightarrow{d} \chi^2(q - p_\phi), \tag{4.1}$$

where $\hat{S}_T$ is the long-run covariance estimator from (2.22). The degrees of freedom equal $q - p_\phi = (q_{NG} + q_P) - p_\phi$. Rejection of (4.1) signals mis-specification of at least one moment block, but it does not identify its origin. We develop conditional orthogonal tests to address this lack of attribution.

4.2. **Neyman-Orthogonal tests.** Identification through non-Gaussianity recovers the rotation matrix $Q(\phi)$ only up to a signed permutation group $\mathcal{SP}(n)$. This implies that the unrestricted non-Gaussian estimator, which excludes the proxy moment conditions, admits $2^n \cdot n!$ observationally equivalent solutions and does not converge to a unique probability limit. The standard J-statistic requires the unrestricted estimator to be asymptotically normal around a unique $Q(\phi_0)$. Without a column anchor provided by proxy exclusion conditions to select a single element of the $\mathcal{SP}(n)$-orbit, this condition fails and the χ^2 limit distribution of the J-statistic is not restored. Therefore, the standard test statistic is invalid in this setting.

Instead, we propose to construct the test statistics from the same joint GMM estimator under the joint null. The key challenge is that the raw sample moments $\bar{g}_{NG,T}$ and $\bar{g}_{P,T}$ are correlated through the shared sample, identification space and the common estimate $\hat{\phi}_T$. This means that their naive quadratic forms do not have standard chi-squared limits under the null. We address these concerns by constructing mutually orthogonalized score statistics, as in Newey (1985). The notations for the partitioned moment vectors, variance, and Jacobian matrices remain the same as Section 2.5.

To construct a test statistic for the validity of the proxy moment conditions that is asymptotically insensitive (under the null) to the higher-order moment conditions, we orthogonalize

proxy moment vector, $\bar{g}_{P,T}$ with respect to the linear span of $\bar{g}_{NG,T}$, i.e.,

$$\tilde{g}_{P,T} = \bar{g}_{P,T} - S_{PN} S_{NN}^{-1} \bar{g}_{NG,T}. \tag{4.2}$$

The coefficient $S_{PN} S_{NN}^{-1}$ is the linear predictor of g_P given g_{NG} under the long-run variance matrix S. The resulting score satisfies the Neyman orthogonality condition:

$$\frac{\partial}{\partial \theta_{NG}} \mathbb{E}[\tilde{g}_{P,T}(\phi_0, \theta_{NG})] = 0, \tag{4.3}$$

where θ_{NG} parameterizes local perturbations of the non-Gaussian moment block. Condition (4.3) ensures that first-order perturbations in the non-Gaussian moment block does not contaminate the asymptotic distribution of the test statistic for the proxy conditions. An analogous construction yields the orthogonalized non-Gaussian moment vector:

$$\tilde{g}_{NG,T} = \bar{g}_{NG,T} - S_{NP} S_{PP}^{-1} \bar{g}_{P,T}. \tag{4.4}$$

The asymptotic behavior of $T^{1/2} \tilde{g}_P$ is governed by two objects introduced in Proposition 3.3: the projected Jacobian $\tilde{G}_P = G_P - S_{PN} S_{NN}^{-1} G_{NG}$ and the Schur complement $\Sigma_P = S_{PP} - S_{PN} S_{NN}^{-1} S_{NP}$, which are the proxy-block counterparts of $\tilde{G}_{NG}$ in (2.17) and Σ_{NG} in (2.18). Since $\tilde{g}_P$ is evaluated at the estimated value, its asymptotic variance is given by:

$$\mathcal{Q}_P = \underbrace{\Sigma_P}_{\text{Schur complement}} - \underbrace{\tilde{G}_P \big(G' S^{-1} G\big)^{-1} \tilde{G}_P'}_{\text{parameter estimation correction}}. \tag{4.5}$$

The second term removes the part of the uncertainty absorbed by the estimation of ϕ. The analogous construction yields $\mathcal{Q}_{NG}$. Formally, the two specification tests are

$$J_{\text{prx}} = T\, \tilde{g}_P'\, \mathcal{Q}_P^{+}\, \tilde{g}_P, \qquad J_{NG} = T\, \tilde{g}_{NG}'\, \mathcal{Q}_{NG}^{+}\, \tilde{g}_{NG}, \tag{4.6}$$

where $(\cdot)^+$ denotes the Moore-Penrose pseudo-inverse.

4.3. **Limiting distributions.** The specification test of validity of the proxy exclusion conditions is conditional on the non-Gaussian moment block identifying ϕ_0 i.e., formally:

$$H_0^{prx} : \mathbb{E}(\tilde{g}_P) = 0 | rank(G_{NG}) = p_\phi \tag{4.7}$$

under the assumptions 2.3 and 2.4. When evaluated at the joint efficient estimator $\hat{\phi}$, the orthogonalized sample moments $T^{1/2} \tilde{g}_P$ converge in distribution to $\mathcal{N}(0, \mathcal{Q}_P)$. The variance matrix $\mathcal{Q}_P$ explicitly accounts for the parameter estimation error through the projected Jacobian $\tilde{G}_P$, has generic rank q_P. Therefore, with analogous reasoning for the non-Gaussianity block, both test statistics satisfy:

$$J_{\text{prx}} \xrightarrow{d} \chi^2(q_P), \qquad J_{\text{NG}} \xrightarrow{d} \chi^2(q_{\text{NG}}). \tag{4.8}$$

The degrees of freedom equal the number of moment conditions in each block: q_P and q_{NG} respectively. The penalty for estimating the p_ϕ parameters is already integrated into the construction of $\mathcal{Q}_P$ and Q_{NG} via the second term in equation (4.5). This allows the generalized inverse to operate on the full q_P-dimensional (or q_{NG}-dimensional) covariance structure. While

the statistics are asymptotically insensitive under the joint null, mis-specification under the alternative (for instance, invalid proxies or dependent structural shocks) in one block will transmit contamination to both statistics through the shared parameter estimate $\hat{\phi}$. This property is utilized as a diagnostic tool in Section 5. We formalize the asymptotic distributions in the following proposition:

Proposition 4.1 (Asymptotic distributions of the test statistics)**.** *Under Assumptions 2.2-2.4 where G_{NG} has full column rank, with $H_0\colon \mathbb{E}[g(\phi_0)] = 0$, and consistent estimators $\hat{\mathcal{Q}}_P \xrightarrow{p} \mathcal{Q}_P$ and $\hat{\mathcal{Q}}_{NG} \xrightarrow{p} \mathcal{Q}_{NG}$:*

(i) $J_{\text{prx}} \equiv T\,\tilde{g}'_{P,T}\hat{\mathcal{Q}}_P^{-1}\tilde{g}_{P,T} \xrightarrow{d} \chi^2(q_P)$.

(ii) *If* $\operatorname{rank}(G_P) = p_\phi$, *then* $J_{NG} \equiv T\,\tilde{g}'_{NG,T}\hat{\mathcal{Q}}_{NG}^{-1}\tilde{g}_{NG,T} \xrightarrow{d} \chi^2(q_{NG})$.

(iii) *In general, let $\hat{\Pi}$ be the orthogonal projector onto the column space of $\hat{\mathcal{Q}}_{NG}$, consistently estimated from the spectral decomposition of $\hat{\mathcal{Q}}_{NG}$ by retaining eigenvectors with eigenvalues above a threshold $\tau_T \to 0$ with $T^{1/2}\tau_T \to \infty$. Then,*

$$J_{NG} \equiv T\,\tilde{g}'_{NG,T}\,\hat{\Pi}\,\hat{\mathcal{Q}}_{NG}^{+}\,\hat{\Pi}\,\tilde{g}_{NG,T} \xrightarrow{d} \chi^2(r_{NG}), \qquad r_{NG} \equiv q_{NG} - (p_\phi - \operatorname{rank}(G_P)). \tag{4.9}$$

Proof: See Appendix B.7.

Remark 4.1 (Compression of degrees of freedom)**.** Under the null, both statistics have chi-squared limits, but their degrees of freedom depend on how much of ϕ_0 the other block identifies. The correction term in (4.5) absorbs the directions of ϕ that the conditioning block leaves unidentified. Hence, if G has full column rank p_ϕ, then the generalized versions of the asymptotic distributions are given by:

$$J_{\text{prx}} \xrightarrow{d} \chi^2\big(q_P - (p_\phi - \operatorname{rank}(G_{NG}))\big), \qquad J_{NG} \xrightarrow{d} \chi^2\big(q_{NG} - (p_\phi - \operatorname{rank}(G_P))\big). \tag{4.10}$$

J_{prx} thus retains all q_P degrees of freedom when the non-Gaussian moments identify ϕ_0 on their own.

This characteristic can be salient in SVAR applications with partial identification through external instruments, where the number of rotation angles p_ϕ grows quadratically with the number of variables, while the available valid instruments are typically small. As a result, in empirical applications, the test statistic J_{NG} becomes conservative, and is not always a reliable standalone test of the non-Gaussian specification. Nonetheless, J_{prx} remains valid when non-Gaussian block retains full column rank (Lemma B.1) and anchors estimation regardless of proxy strength, so the test for proxy exogeneity stays valid even when there is a single instrument in the system.

4.4. **Residual-based moving block bootstrap.** The chi-squared approximations in (4.1) and (4.8) rely on the asymptotic distribution of the standardized sample moments. However, the weighting matrix involves up to fourth-order cumulants. Also, as discussed in Remark 4.1, the asymptotic degrees of freedom for the test statistics also depend on the rank of the Jacobian of each moment block. This may invalidate the use of critical values from asymptotic χ^2 distributions in empirical settings with finite samples.

We adopt the residual-based moving block bootstrap procedure of Brüggemann et al. (2016), which also allows for conditional heteroskedasticity. To preserve the empirical dependence structure between the residuals and proxies $(\hat{u}_t, m_t)$, we resample overlapping blocks of the joint residual-proxy pairs (Jentsch and Lunsford, 2022).

To efficiently obtain the bootstrap analog of the sample moments, we use the k-step iterative estimation procedure of Andrews (2002). Moment recentering through Hall and Horowitz (1996) subtracts $\bar{g}_T(\hat{\phi})$ from each bootstrap moment average, so that the bootstrap data-generating process satisfies the over-identifying restrictions exactly at $\hat{\phi}$. With recentering, the bootstrap correctly mimics the non-standard limit distribution due to the degrees-of-freedom compression. The detailed procedure can be found in Appendix E.

By repeating the bootstrap procedure over N_B replications, we obtain the empirical distributions of $(J^*_{\text{joint}}, J^*_{\text{prx}}, J^*_{NG})$. Hence, we can compute bootstrap p-values as the proportion of bootstrap replications exceeding the corresponding sample statistic.

5. Monte Carlo evidence

We assess the finite-sample properties of the hybrid GMM estimator and the specification tests via a Monte Carlo simulation study.

5.1. **Data-generating process.**

The baseline data generating process is a stable VAR(2) with $n = 3$ variables. The true structural impact matrix is $B_0 = H_0 Q(\phi_0)$, with $H_0 = I_3$ and rotation angles $\phi_0 = (0.80, -0.40, 1.20)'$.

Under the null hypothesis of valid specification, the structural shocks[11] are non-Gaussian, satisfying the moment conditions for identification via co-skewness and co-kurtosis, see Section 2.3. Each shock is normalized to zero mean and unit variance:

- Shock 1 is drawn from a Normal-Inverse Gaussian distribution with parameters $(\alpha, \beta, \delta) = (1.6, 0.8, 1.6)$, yielding skewness $\kappa_3 \approx 1.01$ and excess kurtosis $\kappa_4 \approx 2.70$.
- Shock 2 is drawn from a standardized $\chi^2(7)$ distribution, with $\kappa_3 = \sqrt{8/7} \approx 1.07$ and $\kappa_4 = 12/7 \approx 1.71$.
- Shock 3 is drawn from a standardized $\chi^2(3)$ distribution, with $\kappa_3 = \sqrt{8/3} \approx 1.63$ and $\kappa_4 = 4.00$.

The instruments are $m_{it} = \psi_i \varepsilon_{it} + \sigma_v v_{it}$, where $v_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. The relevance matrix is $\Psi = \text{diag}(0.70, 0.60, 0.80)$ and $\sigma_v = 0.50$ for all the instruments. Therefore, the dimension of the rotational angle parameters is $p_\phi = 3$, with moment conditions $q_{NG} = 9$, $q_P = 6$, giving us the total of $q = 15$ conditions. After estimating the rotation angles $\hat{\phi}$, see Section 2.5, we construct the test statistics J_{joint}, J_{prx}, and J_{NG} as described in Section 4. To obtain bootstrap critical values, we implement a residual-based moving block bootstrap procedure, detailed in Appendix E. We report the Monte Carlo results over sample sizes $T \in \{200, 300, 500, 1000\}$, with $M = 500$ replications and $N_B = 999$ bootstrap replications.

[11] The choice of the underlying distributions is not consequential; the simulation results still hold with other skewed, heavy-tailed distributions including generalized Student-t and other hyperbolic distributions.

Remark 5.1 (Computational Implementation and Asymptotic Invariance)**.** Proposition 2.1 establishes global point identification over the unconstrained rotation space $SO(n)$. In finite samples, however, optimization over higher-order moments is sensitive to initial conditions near Givens gimbal lock boundaries ($\phi = \pi/2$). We initialize the numerical solver at the unweighted non-Gaussian identified preliminary estimate $\hat{\phi}_{\text{prelim}} = \arg\ \min_{\phi} \bar{g}_{NG}(\phi)'\bar{g}_{NG}(\phi)$ and restrict the search to the principal geometric branch $\phi \in \Phi_0 \equiv (-\pi/2,\ \pi/2)^{p_\phi} \subset \Phi$, which eliminates topologically isomorphic aliases without any loss of generality. The criterion is then minimized by a constrained solver on Φ_0, started both at $\hat{\phi}_{\text{prelim}}$ and at draws from the interior of the branch, so that the reported estimate is the best of several local solutions.

Remark 5.2 (Spectral Regularization)**.** The matrices $\mathcal{Q}_P$ and $\mathcal{Q}_{NG}$ for the test statistics (4.6) are constructed from higher-order moment conditions whose finite-sample covariance is usually ill-conditioned. Consequently, inversion without regularization may amplify the smallest eigenvalues, inflating the test statistics and distorting size. We regularize via spectral truncation: eigenvalues of $\mathcal{Q}$ below the threshold $\tau = 10^{-4}$ are set to zero prior to inversion, retaining only the subspace along which the matrix is well-conditioned (Antoine and Renault, 2007). This truncation controls the effective rank of $\mathcal{Q}^+$ without distorting the limiting distribution along the retained eigenvalue directions.

5.2. **Baseline results.**

5.2.1. *Estimates under H_0 and H_1.* Table 1 reports the average estimates and standard errors for ϕ_0, over the simulations, under the baseline DGP. The estimator is effectively unbiased at all sample sizes and standard errors decline monotonically. Table 2 reports estimates under proxy invalidity, where instrument m_2 violates the exogeneity condition: $m_{2t} = 0.4\,\varepsilon_{1t} + \sigma_v v_{2t}$, i.e. it loads on shock 1 rather than the target shock. Table 2 shows severe bias in $\hat{\phi}_2$ and $\hat{\phi}_3$. Both are consistent with the standard result that GMM estimators converge to a pseudo-true value rather than ϕ_0 under moment mis-specification, as shown in White (1982); Hall and Inoue (2003).

	True Value	$T = 200$	$T = 300$	$T = 500$	$T = 1000$
Estimates and their standard errors					
$\hat{\phi}_1$	0.80	0.8046	0.8005	0.8007	0.8006
		(0.0300)	(0.0250)	(0.0201)	(0.0145)
$\hat{\phi}_2$	-0.40	-0.3958	-0.3955	-0.3996	-0.3985
		(0.0315)	(0.0266)	(0.0212)	(0.0153)
$\hat{\phi}_3$	1.20	1.1991	1.2017	1.2025	1.2002
		(0.0310)	(0.0259)	(0.0207)	(0.0150)

Notes: Average estimates and their standard errors are based on $M = 500$ Monte Carlo replications with $B = 999$ bootstrap replications. The DGP is a stationary VAR(2) with $n = 3$ variables and true rotation parameters $\phi_0 = (0.80, -0.40, 1.20)'$. Structural shocks are drawn from NIG, $\chi^2(7)$, and $\chi^2(3)$ distributions. Proxies are valid with relevance $\Psi = (0.70, 0.60, 0.80)'$ and noise $\sigma_v = 0.50$.

TABLE 1. Estimates of ϕ_0 under H_0: Non-Gaussian Shocks with Valid Proxies

	True Value	$T = 200$	$T = 300$	$T = 500$	$T = 1000$
Estimates and their standard errors					
$\hat{\phi}_1$	0.80	0.8277	0.8372	0.8228	0.8034
		(0.0384)	(0.0296)	(0.0219)	(0.0152)
$\hat{\phi}_2$	-0.40	-0.4014	-0.3671	-0.3539	-0.3427
		(0.0475)	(0.0402)	(0.0328)	(0.0234)
$\hat{\phi}_3$	1.20	0.6723	0.8381	0.9966	1.1371
		(0.0494)	(0.0401)	(0.0309)	(0.0223)

Notes: Average estimates and their standard errors are based on $M = 500$ Monte Carlo replications with $B = 999$ bootstrap replications. The DGP is identical to the H_0 specification except that proxy m_2 is mis-specified: $m_{2t} = 0.4\,\varepsilon_{1t} + \sigma_v \nu_{2t}$, violating exogeneity by responding to shock 1 instead of shock 2. The substantial bias in $\hat{\phi}_3$ reflects the contamination of the proxy moment conditions.

TABLE 2. Estimates of ϕ_0 under H_1: Invalid Proxy

5.2.2. *Size under H_0.* Table 3, Panel A reports empirical rejection rates at the 5% nominal level. All three tests are slightly conservative at modest sample sizes. The large mean bootstrap J-statistics in small samples reflect the inefficiency of the estimates when the underlying shocks are heavy-tailed; they converge toward their sample counterparts as T grows. The mean J statistics, both estimated and bootstrap, track the means of their limiting distributions closely: J_{prx} against $\chi^2(q_P)$ and J_{NG} against $\chi^2(q_{NG})$, as stated in Proposition 4.1.

	$T = 200$	$T = 300$	$T = 500$	$T = 1000$
Panel A: Empirical Rejection Rates (nominal size = 0.05)				
Joint specification test	0.034	0.048	0.044	0.056
Proxy validity test	0.044	0.028	0.036	0.036
Non-Gaussianity test	0.024	0.044	0.028	0.044
Panel B: Mean J-statistics				
J_{joint} (bootstrap)	24.471	21.352	17.236	12.714
J_{joint} (estimate)	20.361	18.001	14.077	12.218
J_{prx} (bootstrap)	18.379	13.680	10.315	6.555
J_{prx} (estimate)	14.101	12.455	8.421	6.069
J_{NG} (bootstrap)	32.052	24.031	17.549	11.941
J_{NG} (estimate)	20.170	17.535	13.823	10.989

Notes: Panel A reports the fraction of $M = 500$ Monte Carlo replications in which the null hypothesis is rejected at the 5% nominal level. Panel B reports average J-statistics across replications. "Bootstrap" denotes the average bootstrap J-statistic ($N_B = 999$); "estimate" denotes the average of the J-statistic estimates. The joint test evaluates all overidentifying restrictions; the NG test evaluates non-Gaussianity moment conditions conditional on proxy validity; the proxy validity test evaluates proxy validity conditional on non-Gaussianity.

TABLE 3. Size of Specification Tests under H_0: Non-Gaussian Shocks with Valid Proxies

	$T = 200$	$T = 300$	$T = 500$	$T = 1000$
Panel A: Empirical Rejection Rates (nominal size = 0.05)				
Joint specification test	0.708	0.894	0.972	1.000
Proxy validity test	0.606	0.810	0.926	0.994
Non-Gaussianity test	0.072	0.288	0.642	0.942
Panel B: Mean J-statistics				
J_{joint} (bootstrap)	28.821	24.334	19.889	14.223
J_{joint} (estimate)	90.325	116.339	165.058	292.212
J_{prx} (bootstrap)	30.403	22.313	17.051	9.107
J_{prx} (estimate)	94.347	117.791	218.212	293.421
J_{NG} (bootstrap)	57.127	37.248	23.962	14.197
J_{NG} (estimate)	46.739	61.246	88.026	124.158

Notes: Panel A reports the fraction of $M = 500$ Monte Carlo replications in which the null hypothesis is rejected at the 5% nominal level. Panel B reports average J-statistics across replications. "Bootstrap" denotes the average bootstrap J-statistic ($N_B = 999$); "estimate" denotes the average of the J-statistic estimates. Under H_1, the proxy validity and joint test statistics diverge sharply from their bootstrap counterparts, reflecting growing power as T increases. The NG bootstrap critical values are heavily inflated at small T due to the contamination of the rotation estimates, suppressing NG rejection rates despite the underlying shocks remaining non-Gaussian.

TABLE 4. Power of Specification Tests under H_1: Invalid Proxy

5.2.3. *Power under H_1.* In Table 4, all three tests accumulate power rapidly as T increases. At $T = 500$, rejection rates are 0.972, 0.926, and 0.642 for the joint, proxy validity, and non-Gaussianity tests respectively.

A remark regarding the rejection rates of the non-Gaussianity test is in order. The test for non-Gaussianity rejects despite the true shocks are non-Gaussian. This is due to the contamination in the joint estimate, and it is expected. The joint GMM objective pools proxy and higher-order moment conditions, so the invalid proxy moments displace $\hat{\phi}$ from ϕ_0. The recovered structural shocks $\hat{\varepsilon}_t = \hat{B}_0^{-1}(\hat{\phi})\, u_t$ do not match the true shocks, and their higher-order cross-moments violate the moment conditions even though the underlying DGP is non-Gaussian. This implies that $g_{NG}(y_t, \hat{\phi})$ has nonzero population expectation under H_1, generating power in the non-Gaussianity test through an indirect spillover channel. This mechanism is consistent with the partitioned J-statistic structure: the Schur-complement decomposition yields orthogonal test statistics only under the joint null. Under a mis-specification of the proxy moment conditions, only parameter contamination propagates through the shared space to all components (Newey, 1985).

The power ordering $J_{\text{prx}} \gg J_{\text{NG}}$ is uniform across all sample sizes and supports the theoretical interpretation. The ordering is even more pronounced for the point-estimates of J-statistics, which diverge sharply from their bootstrap counterparts as T increases. The proxy validity statistic reaches 218.1 at $T = 500$ against a bootstrap mean of 17.1, whereas the non-Gaussianity statistic reaches 88.0 against a bootstrap mean of 23.9. When all three tests reject and the J_{prx} dominates, the pattern is consistent with proxy invalidity as the primary source of moment mis-specification.

5.3. **The case of weak instruments.** One of the central concerns in a proxy-SVAR is the sensitivity of inference to the strength of the external instruments. Weak instruments can distort the sampling distribution of GMM-based estimators and invalidate standard asymptotic approximations, see Staiger and Stock (1997); Stock and Wright (2000). While the relevance of instrument is routinely verified in applied work through first-stage F-statistics, limitations of such pre-tests in terms of low power and distorted subsequent inference are documented in Ramey (2016); Lewis and Mertens (2026).

We show that even under weak proxies, the hybrid framework continues to recover the parameters consistently (Proposition 3.1), because the higher-order moment conditions anchor the identification. Though the specification test exhibit asymmetric size distortions and J_{NG} becomes severely conservative, J_{prx} retains reliable calibration and remains a valid test of proxy orthogonality.

We adopt the local-to-zero framework for proxy relevance i.e., a Pitman drift $c/\sqrt{T}$ with $c = 2.24$ across the sample sizes. At these magnitudes, the proxy Jacobian $G_{P,T} = O(T^{-1/2})$ contributes negligible identifying information. Results[12] are reported in Tables 5 and 7.

	True Value	$T = 200$	$T = 300$	$T = 500$	$T = 1000$
Estimates and their standard errors					
$\hat{\phi}_1$	0.80	0.7851	0.7754	0.8019	0.8007
		(0.0597)	(0.0526)	(0.0439)	(0.0332)
$\hat{\phi}_2$	-0.40	-0.3621	-0.3956	-0.3872	-0.3959
		(0.0639)	(0.0580)	(0.0507)	(0.0387)
$\hat{\phi}_3$	1.20	1.1789	1.1870	1.2013	1.2024
		(0.0563)	(0.0492)	(0.0405)	(0.0307)

Notes: Results based on $M = 500$ Monte Carlo replications with $N_B = 999$ bootstrap replications. Proxy relevance follows a Pitman drift: $\Psi_i = c/\sqrt{T}$ with $c = 2.24$ for all instruments, yielding $\Psi \approx (0.158, 0.129, 0.100, 0.071)$ for $T = (200, 300, 500, 1000)$. All other DGP parameters are identical to the strong-proxy H_0 specification.

TABLE 5. Parameter Estimates under H_0: Weak Proxies (Pitman Drift $c = 2.24$)

5.3.1. *Consistent estimation.* Table 5 shows that the joint GMM estimator recovers the parameters consistently despite severe proxy weakness. Though we do note some finite-sample bias at $T = 200$, it reflects the elevated variance of higher-order sample moments under heavy tails. The estimates converge toward ϕ_0 as T increases. This is the simulation evidence for Proposition 3.1. As the proxy block Jacobian $G_{P,T} \xrightarrow{T\to\infty} 0$, the identification remains anchored by the non-Gaussian block. The co-skewness and symmetric co-kurtosis conditions provide full-rank local curvature in a neighborhood of ϕ_0, ensuring the variance

[12]Similar results with strong proxies and weak non-Gaussian shocks can be found in Appendix F. We also provide results for doubly weak-identification, i.e. weak proxies and weak non-Gaussianity. In this case, the estimates remain unbiased with increasing sample size, albeit with larger standard errors compared to strong identification case.

matrix, $\mathcal{V}^*_{NG} = \left(G'_{NG}\,\Sigma^{-1}_{NG}\,G_{NG}\right)^{-1}$ remains asymptotically non-singular. Simultaneously, the shrinking, but a nonzero proxy relevance imposes a strictly positive penalty on each element of the finite discrete set $\mathcal{B} \setminus \{\phi_0\}$. This resolves the global permutation ambiguity that the higher-order moments alone cannot eliminate.

5.3.2. *Robust Anderson-Rubin confidence intervals.* In the presence of weak instruments, standard practice is to construct Anderson-Rubin (AR) confidence sets, see Montiel Olea et al. (2021); Jentsch and Lunsford (2022), by inverting the AR statistic over the parameter space. They maintain correct asymptotic coverage uniformly over proxy strength, including the local-to-zero regime.

To assess the inferential contribution of the non-Gaussian moment conditions under weak proxy identification, we compare the AR confidence sets for the rotation parameters ϕ under two criteria, *Proxy-only* and *Proxy+NG*. The former involves only the proxy moment conditions in the GMM criterion for the AR statistic whereas the latter includes all the moment conditions (non-Gaussian and proxy). This allows us to compare the effect of additional identifying strength of the higher-order moments conditions, in terms of the size of identification-robust confidence sets, under local-to-zero proxy relevance.

For each simulated dataset, the AR statistic is evaluated over a fine p_ϕ-dimension grid of candidate rotation angle vectors $\phi \in \Phi_0$. At each grid point, the AR statistic takes the form

$$AR_T(\phi_{(i)}) = T \cdot \bar{g}(\phi_{(i)})' \hat{S}(\phi_{(i)})^{-1} \bar{g}(\phi_{(i)}), \tag{5.1}$$

where $\bar{g}(\phi_{(i)})$ is the sample moment vector at i^{th} grid point and $\hat{S}_T(\phi_{(i)}) = T^{-1}\sum_{t=1}^{T}(g_t(\phi_{(i)}) - \bar{g}(\phi_{(i)}))(g_t(\phi_{(i)}) - \bar{g}(\phi_{(i)}))'$ is the consistent variance estimator[13]. Under the null that $\phi_{(i)}$ is the true rotation, $AR_T(\phi_{(i)})$ is asymptotically distributed as $\chi^2(q)$, where q is the number of moment conditions. The AR confidence set is the collection of parameter values not rejected by the test at the nominal $1-\alpha$ level :

$$\mathcal{C}_{1-\alpha} = \left\{\phi_{(i)} \in \Phi_0 : AR_T(\phi_{(i)}) \leq \chi^2_{q,\,1-\alpha}\right\}. \tag{5.2}$$

In practice, $\mathcal{C}$ is approximated by evaluating $AR_T(\phi_{(i)})$ over a fine grid of $1,000,000$ (100^{p_ϕ}) candidate vectors covering Φ_0, and retaining all grid points at which the criterion does not exceed the critical value. Given that such grid-based inferential methods are computationally prohibitive, we include the simulation results for a sample size $T = 1000$, for 100 Monte Carlo replications.

The results demonstrate a breakdown of proxy-only identification under weak instruments, and its recovery through non-Gaussianity. For ϕ_3, the *Proxy-only* AR set spans the entire feasible rotation parameter space, indicating that the proxy criterion fails to reject any value in the considered parameter space, and is entirely uninformative. As shown in Table 6, all three *Proxy+NG* sets are centered near the true parameter values (0.80, −0.40, 1.20) and are comparable in length to the HGMM Wald intervals, indicating that the NG conditions

[13]In the empirical application, $\widehat{S}(\phi_{(i)})$ is replaced by the residual based moving block bootstrap estimator $\widehat{S}_{MBB}(\phi_{(i)})$ to account for serial dependence.

Parameter	Mean Estimate	Proxy-only	Proxy+NG	HGMM standard CI
0.80	0.8007	$[0.2221, 1.3328]$	$[0.7140, 0.9044]$	$[0.7447, 0.8567]$
		1.1107	0.1904	0.1120
-0.40	-0.3959	$[-0.8409, 0.0793]$	$[-0.5077, -0.2697]$	$[-0.4603, -0.3305]$
		0.9202	0.2380	0.1298
1.20	1.2024	$[-1.5708, 1.5708]$	$[1.1265, 1.2852]$	$[1.1509, 1.2539]$
		3.1416	0.1587	0.1030

Notes. The 90% AR confidence sets are median sets, across the monte carlo simulations. (with their lengths below). The main comparison is between the *Proxy-only* and *Proxy+NG* criteria. The former involves only the proxy orthogonality moment conditions in the AR statistic whereas the latter includes all the moment conditions (non-Gaussian and proxy). The HGMM standard CI is computed as: $[\hat{\phi} - 1.645 \times \hat{\sigma}_\phi, \hat{\phi} + 1.645 \times \hat{\sigma}_\phi]$, where $\hat{\phi}$ and $\hat{\sigma}_\phi$ are the average of the estimates (across the monte carlo replications) and their standard errors, respectively. Results based on $M = 100$ Monte Carlo replications with $N_B = 999$ bootstrap replications. Proxy relevance follows a Pitman drift: $\Psi_i = c/\sqrt{T}$ and $c = 2.24$, with sample size, $T = 1000$. All other DGP parameters remain unchanged from the H_0 specification.

TABLE 6. Anderson-Rubin 90% Confidence Sets with HGMM standard CIs

restore the information regarding the true parameter. The *Proxy+NG* AR criterion remains informative and well-centered because the higher-order moments conditions identify the rotation parameters independently of proxy strength, as established in Proposition 3.1. The joint hybrid criterion eliminates the identification failure without sacrificing the validity of the identification-robust coverage.

5.3.3. *Asymmetric size distortions in specification tests.* While point identification is preserved, Table 7 reveals a pronounced asymmetry in the size of the specification tests. The asymmetry originates in the conditional structure of each test, i.e., which block of the joint moment system anchors the parameter estimation.

Panel A of Table 7 shows that J_{NG} almost never rejects at all sample sizes, against a nominal significance level of 5%. Panel B identifies the mechanism where the mean sample statistic $\hat{J}_{\text{NG}}$ is stable between 12.5 and 9.9, while their bootstrap analogs are massively inflated at small samples. The bootstrap mean of J_{NG} falls from 147.0 at $T = 200$ to 49.6 at $T = 1000$. This failure to reject is a direct consequence of anchor failure in the conditional testing procedure. The J_{NG} statistic evaluates the higher-order moment conditions with a maintained hypothesis that the identifying strength is provided by the proxy block[14]. Under strong proxy relevance, G_P is full rank and absorbs the full parameter estimation penalty, leaving the higher-order moments as over-identifying conditions. Under weak proxy relevance, $G_{P,T} \approx 0$, the joint GMM criterion obtains identifying curvature from the non-Gaussian block alone. The sample statistic $\hat{J}_{NG}$ is evaluated over a subspace fitted to the higher-order moments, compressing the realized statistic toward $q_{NG} - (p_\phi - rank(G_P))$ (Proposition 4.1).

Simultaneously, since the proxy block contributes near-zero information to the joint long-run covariance matrix, the restricted variance matrix $\mathcal{Q}^*_{NG}$ inherits extreme instability from

[14]This is a specific case where there are sufficient instruments to identify the entire system. As noted in Remark 4.1, if (instruments) $k < n - 1$ (structural shocks), the proxy block cannot provide the entire identification and the distribution of the J_{NG} statistic faces compression in its (asymptotic) degrees of freedom.

	$T = 200$	$T = 300$	$T = 500$	$T = 1000$
Panel A: Empirical Rejection Rates (nominal size = 0.05)				
Joint specification test	0.000	0.008	0.006	0.040
Proxy validity test	0.018	0.032	0.030	0.040
Non-Gaussianity test	0.000	0.000	0.000	0.006
Panel B: Mean J-statistics				
J_{joint} (bootstrap)	25.313	21.309	17.749	14.020
J_{joint} (estimate)	16.524	16.017	13.967	12.958
J_{prx} (bootstrap)	11.527	9.218	7.757	6.814
J_{prx} (estimate)	9.384	8.522	7.344	6.794
J_{NG} (bootstrap)	147.073	113.351	79.632	49.601
J_{NG} (estimate)	12.571	11.910	10.714	9.939

Notes: Panel A reports the fraction of $M = 500$ Monte Carlo replications in which the null hypothesis is rejected at the 5% nominal level. Panel B reports average J-statistics across replications. "Bootstrap" denotes the average bootstrap J-statistic ($N_B = 999$); "estimate" denotes the average of the J-statistic estimates. Under weak proxy asymptotics ($\Delta_i = c/\sqrt{T}$, $c = 2.24$), the proxy moments become asymptotically uninformative. The NG bootstrap critical values are massively inflated due to the near-singular contribution of the proxy block to the long-run covariance matrix.

TABLE 7. Size of Specification Tests under H_0: Weak Proxies (Pitman Drift $c = 2.24$)

these higher-order moments, generating bootstrap values that are orders of magnitude larger than the compressed sample statistic.

In contrast, Panel A of Table 7 shows that J_{prx} maintains better size, improving with T. This is because, analogous to the J_{NG} test, the J_{prx} test relies on the non-Gaussian block as its identifying anchor. The true DGP is non-Gaussian, so G_{NG} maintains full column rank independently of the proxy validity, identifying $\hat{\phi}$ without any contribution from G_P. Hence, J_{prx} evaluates proxy orthogonality against a correctly specified parameter space.

6. Empirical Applications

We demonstrate the potential of our HGMM framework with two empirical applications. The first application identifies an *oil-news shock* from the OPEC announcements-window oil futures surprise of Känzig (2021) after adopting the corrections suggested by Kilian (2024) (Section 6.2); the second identifies a Euro-area *monetary policy shock* from the high-frequency interest-rate surprises from Altavilla et al. (2019) (Section 6.3). In each case, we report identification-robust confidence sets for the impulse responses under two criteria, *Proxy-only* (the exclusion conditions alone) and *HGMM* (the exclusion conditions combined with the higher-order moment conditions), and we conduct a formal specification test for the validity of the exclusion restrictions. Each illustration is accompanied by a robustness exercise, reported in the appendix: for the oil shock, we re-run the analysis on the Känzig (2021) baseline (original instrument and specification) in Appendix I, and for the monetary policy shock on the Gertler and Karadi (2015) specification in Appendix J.

6.1. **Identification-robust inference.** The headline comparison in both illustrations is between the identification-robust confidence sets of the two criteria. For each rotation in the target-shock subspace, we construct the Anderson–Rubin statistic and retain the rotation if the statistic does not exceed its per-candidate bootstrap quantile. The confidence set for a given impulse response is the projection of this acceptance region onto the corresponding coordinate, obtained by optimizing the impulse response function subject to the acceptance constraint. Because the target-shock responses depend only on the first column of the impact matrix, the inversion reduces to the $n-1$ rotation angles that build the target column, and the remaining columns are not individually considered. This is analogous to the procedure validated in Section 5.3.2; the full construction is given in Appendix G. The dimension of the search is therefore $n-1$ rather than $p_\phi = n(n-1)/2$. For the six-variable oil system, the inversion runs over 5 angles with 16 moment conditions rather than over all the 15 angles, and for the four-variable monetary system over 3 angles with 10 conditions. Throughout, bootstrap quantiles use a residual-based moving-block bootstrap (MBB) that jointly resamples the recentered residuals and the proxy (Brüggemann et al., 2016; Jentsch and Lunsford, 2022) and re-estimates the VAR within each draw, so that first-stage uncertainty propagates into the structural objects. Confidence sets are reported at the nominal 90% level.

6.2. **Oil-news shock.** We revisit the identification of oil-news shocks in Känzig (2021), who use changes in oil futures prices around pre-scheduled OPEC announcements as an external instrument for the structural shock to oil price expectations. The setting is well suited to our framework because the relevance and construction of the monthly instrument have been questioned, most directly by Kilian (2024) but also by Mori and Peersman (2024); Cavaliere et al. (2025). Hence, it is an ideal case in which the identifying information from the higher-order moment conditions can be consequential.

6.2.1. *Specification.* Our main specification adopts the corrections of Kilian (2024), so that the instrument is constructed consistently with the timing of the reduced-form data. The instrument aggregates daily futures-price surprises around OPEC announcements into a monthly series. However, a naive within-month sum is inconsistent with a VAR built on monthly average prices, because a permanent surprise of size δ on trading day d_k of a month with T_t trading days shifts the current month's average by $\frac{T_t - d_k + 1}{T_t}\,\delta$ and the next month's average by $\frac{d_k - 1}{T_t}\,\delta$. The corrected proxy weighs each surprise by its within-month timing and carries the residual weight into the following month. We adopt this temporally weighted instrument, restrict the sample to 1989:M4 onwards where the trading-day data are reliable, and estimate a log-levels VAR with the Kilian and Murphy (2014) oil-market lag order $p = 24$.

The system has the six baseline variables: real oil price, world oil production, world oil inventories, world industrial production, U.S. industrial production, and U.S. CPI, such that $n = 6$ and $k = 1$. The reduced-form VAR is

$$y_t = c + \sum_{j=1}^{p} A_j\, y_{t-j} + u_t, \qquad u_t \sim (0, \Sigma_u), \tag{6.1}$$

estimated on 1989:M4-2017:M12; the structural innovations satisfy $u_t = B\varepsilon_t$, where ε_t meets Assumption 2.3. Following Section 2.1, $B = H\,Q(\phi)$ with H the Cholesky factor of Σ_u and $Q(\phi) \in SO(n)$ governed by the $p_\phi = n(n-1)/2 = 15$ Givens angles, so $\varepsilon_t(\phi) = Q(\phi)'H^{-1}u_t$.

Indexing the oil-news shock as $j = 1$ and letting m_t denote the instrument (robust $F = 8.9$) the moment vector combines the higher-order moment conditions of Section 2.3 with the proxy exclusion conditions, giving $q_P = k(n-1) = 5$ proxy conditions and $q_{NG} = 75$ non-Gaussian conditions: the co-skewness and symmetric co-kurtosis blocks of Section 2.3, together with the asymmetric co-kurtosis[15] conditions appended as a further source of information, so a total of $q = 80$ moment conditions and $p_\phi = 15$ parameters. We estimate $\hat{\phi}$ by continuous-updating GMM and normalize the impact matrix to a positive diagonal; the impulse responses at horizon i are

$$\hat{\Theta}_i = \big(R\,\hat{\mathcal{A}}^i\,R'\big)\hat{B}, \qquad i = 0, 1, \ldots, h, \tag{6.2}$$

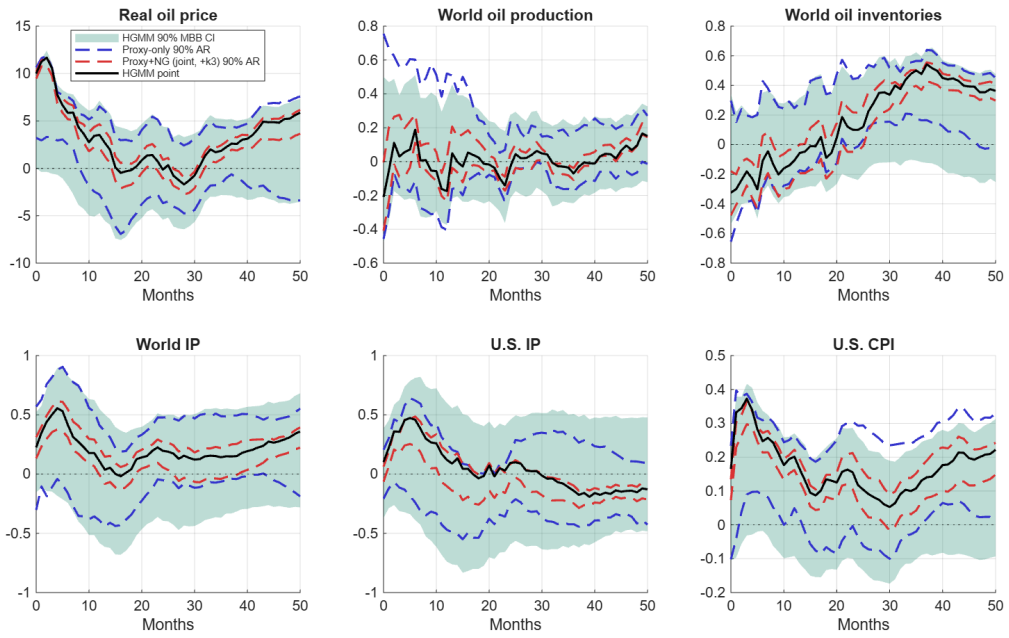
with $R = [I_n, \mathbf{0}_{n\times np}]$ is the selection matrix and $\hat{\mathcal{A}}$ is the companion-form matrix of (6.1). The identification-robust sets compare *Proxy-only* (the five proxy conditions that identify the target shock) with *HGMM* (addition of the target shock's higher-order moment conditions, 16 conditions in total), inverting the AR statistic over the five-dimensional target-shock subspace at 10^5 candidate rotations; the bootstrap uses block length $\ell = 36$ and 9999 replications. Responses are normalized to a 10% increase in the real oil price on impact. Appendix H.1 shows some evidence of non-Gaussianity in the reduced-form innovations, and reports the third and fourth moments of the estimated structural shocks where the joint Gaussianity of the vector is rejected by the pooled Jarque-Bera test ($JB = 35.14$ with 12 $d.f.$, and $p = 0.0004$).

6.2.2. *Results.* Figure 1 plots the 90% identification-robust confidence sets along with their standard MBB confidence bands. An adverse oil *supply news* shock and a positive oil *demand news* shock both raise the real price of oil, so the two are distinguished only by the accompanying responses of oil production, global real activity, and inventories (Kilian and Murphy, 2014). Following a negative supply-news shock, the anticipation of future scarcity raises storage demand on impact, so inventories build immediately, production declines sluggishly but persistently, and global industrial production falls. However, under a demand-news shock, the oil price rises together with global activity, production responds positively to the price incentive, and inventories need not build.

Our estimates match the demand-news pattern, and the inventory response is decisive. World oil inventories are drawn down: the response is broadly negative over the first 12 months, and turns significantly positive only after two years. This is the opposite of the

[15] *Asymmetric co-kurtosis:* the $n(n-1)$ fourth-order conditions $\mathbb{E}[\varepsilon_{i,t}^3\,\varepsilon_{j,t}] = 0$ for all ordered pairs (i,j) with $i \neq j$. They are not necessary for identification and are not part of the moment block of Section 2.3; appending them raises the number of higher-order moment conditions to $2n(n-1) + n(n-1)/2 = 75$. Their sample variance involves the cross moment $\mathbb{E}[\varepsilon_{i,t}^6\varepsilon_{j,t}^2]$, which reduces to the sixth-order marginal requirement $\mathbb{E}|\varepsilon_{i,t}|^{6+3\delta} < \infty$ under the factorization discussed in Section 2.3. Evidence in Appendix H.1 support the inclusion in this specification. Figure H.2 shows the identification-robust sets when the asymmetric co-kurtosis conditions are excluded.

immediate accumulation that defines a supply-news shock. World industrial production rises on impact by 0.23% and is significantly positive; U.S. industrial production is significantly positive over first ten months and U.S. consumer prices increase for more than four years. World oil production shows no persistent decline: it falls on impact, then rises over the following six months, consistent with producers responding to the higher price rather than with an anticipated supply shortfall.



Notes: Each panel superimposes three 90% interval estimates for the impulse response of one variable to the oil news shock. The shaded band is the residual-based moving-block-bootstrap (MBB) percentile confidence interval for the HGMM point estimator. The two dashed lines are the identification-robust AR sets: *Proxy-only* (blue, dashed) inverts the AR statistic using only the five proxy orthogonality conditions that identify the target shock; *HGMM* (red, dashed) additionally imposes the co-skewness, asymmetric co-kurtosis, and co-kurtosis-aggregate conditions of the target shock (16 conditions in total). The solid black line is the HGMM point estimate. Critical values from a residual-based moving-block bootstrap (9999 replications, block length $\ell = 36$). Responses are normalized to a 10% increase in the real oil price on impact. The x-axis is the horizon in months; the y-axis is the response in percent.

FIGURE 1. 90% identification-robust confidence sets with *Proxy-only* and *HGMM* criteria, with standard MBB percentile bands.

Table 8 reports the lengths of the confidence sets under both criteria and Table 9 the corresponding length ratios at horizons 0 to 4. The identifying contribution of the higher-order moment conditions is substantial: the median reduction in set length is 71.6%, and the *HGMM* sets are narrower than the *Proxy-only* sets at every variable-horizon pair, with a median length ratio of 3.5 and a minimum of 1.6. The gain is largest for the real oil price itself, where the impact set contracts from a length of 7.41 under the instrument alone to 0.75 once the higher-order moment conditions are added, a ratio of nearly ten; world industrial production contracts from 0.87 to 0.18 on impact and U.S. CPI from 0.34 to 0.06.

The consequence of augmenting non-Gaussian identification is that the demand-shock characterization above exists only under the hybrid criterion. Table H.2 in Appendix H.2 records the horizons at which each criterion determines the sign of a response. The instrument alone never resolves the sign of the world oil production or U.S. industrial production response; it also never detects the inventory drawdown, since its confidence set for inventories never lies strictly below zero at any horizon and turns positive only after two years. A proxy-only analysis of this specification would therefore be unable to distinguish supply news from demand news at all. Furthermore, the proxy exclusion restriction is not rejected ($\hat{J}_{\text{prx}} = 2.87$, with asymptotic and bootstrap p-values of 0.71 and 0.97, respectively).

	$h=0$		$h=1$		$h=2$		$h=3$		$h=4$	
Variable	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM
Real oil price	7.41	**0.75**	8.65	**0.63**	8.55	**0.88**	7.33	**0.77**	5.11	**0.91**
World oil production	1.21	**0.40**	0.97	**0.43**	0.70	**0.28**	0.71	**0.37**	0.71	**0.29**
World oil inventories	0.95	**0.30**	0.69	**0.21**	0.67	**0.24**	0.59	**0.18**	0.56	**0.24**
World IP	0.87	**0.18**	0.73	**0.17**	0.95	**0.20**	0.93	**0.21**	0.92	**0.24**
U.S. IP	0.41	**0.13**	0.42	**0.12**	0.53	**0.16**	0.55	**0.17**	0.71	**0.22**
U.S. CPI	0.34	**0.06**	0.44	**0.07**	0.34	**0.06**	0.30	**0.07**	0.27	**0.05**

Notes: Each entry is the length (upper minus lower bound) of the 90% identification-robust Anderson–Rubin confidence set for the impulse response of the row variable to the target shock at horizon h (months). *Proxy* inverts the AR statistic using only the five proxy exclusion conditions that identify the target shock; *HGMM* (Proxy+NG) additionally imposes the target shock's higher-order moment conditions, for 16 conditions in total. The smaller length in each (variable, horizon) pair is in **bold**. Responses are normalized to a 10% increase in the real oil price on impact, so entries are in percent. Sets are obtained by inverting the AR statistic over the target-shock rotation subspace with per-candidate moving-block-bootstrap critical values (9999 replications, block length $\ell = 36$); see the construction in Appendix G.

TABLE 8. Length comparison of 90% Anderson–Rubin confidence sets: *Proxy-only* vs *HGMM* (Kilian-corrected specification)

	Ratio at horizon h					Over $h = 0, \ldots, 50$	
Variable	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	Median	Max
Real oil price	9.90	13.82	9.76	9.48	5.61	9.76	13.82
World oil production	3.01	2.25	2.47	1.94	2.46	2.46	3.01
World oil inventories	3.21	3.35	2.81	3.26	2.34	3.21	3.35
World IP	4.90	4.42	4.84	4.38	3.90	4.42	4.90
U.S. IP	3.26	3.42	3.36	3.29	3.28	3.29	3.42
U.S. CPI	5.42	5.97	5.64	4.55	5.56	5.56	5.97

Notes: Each entry is the ratio len(Proxy)/len(HGMM) of the length of the 90% identification-robust Anderson–Rubin confidence set under *Proxy-only* to the corresponding length under *HGMM*, for the impulse response of the row variable to the target shock. Values above one indicate that the *Proxy-only* set is longer. The last two columns summarize the ratio over all 51 horizons $h = 0, \ldots, 50$. Construction as in Table 8.

TABLE 9. Ratio of 90% Anderson–Rubin confidence-set lengths: *Proxy-only* / *HGMM* (Kilian-corrected specification)

Additional results with the specification without the asymmetric co-kurtosis moment block, the MBB percentile bands for the proxy-SVAR estimator, and sensitivity to the bootstrap block length are reported in Appendix H.2. A robustness exercise with application to the Känzig (2021) baseline instrument and specification is reported in Appendix I.

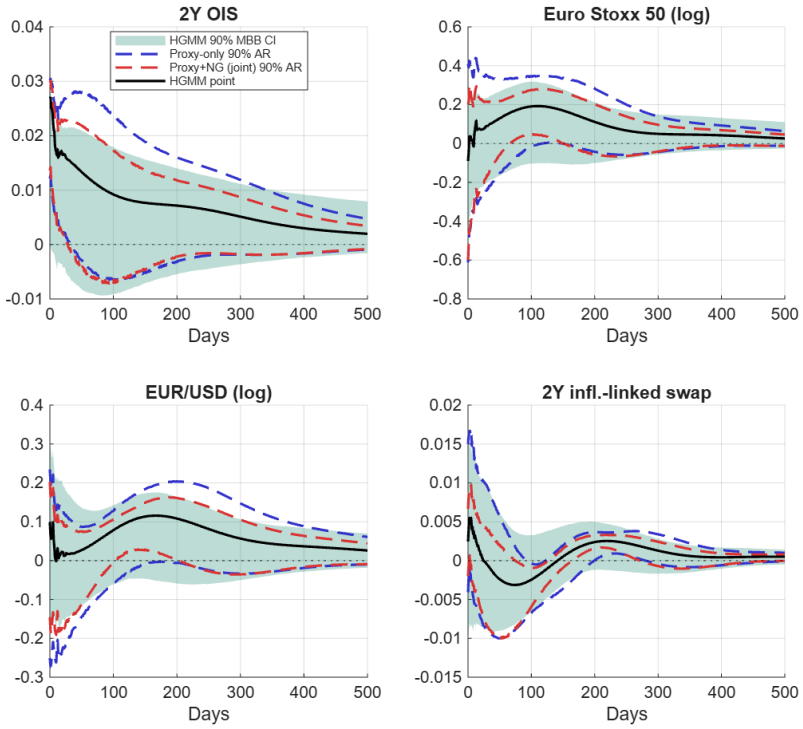
6.3. **Euro-area monetary policy shock.** Our second illustration identifies a Euro-area monetary policy shock using the high-frequency interest-rate surprises of Altavilla et al. (2019) as an external instrument. They measure surprises as changes in overnight-index-swap (OIS) rates within narrow windows around ECB policy communications and rotate the resulting surprises into four orthogonal factors: *target*, *timing*, *forward guidance*, and *quantitative easing*. For the detailed construction of the surprises and the factor rotation, we refer the reader to Altavilla et al. (2019).

6.3.1. *Specification.* We follow the specification of four variables ordered with the policy indicator first: the two-year OIS rate (the target), the log Euro Stoxx 50 equity index, the log EUR/USD exchange rate, and the two-year inflation-linked swap rate, so that $n = 4$ and $k = 1$. The VAR has the form (6.1) with $p = 15$ lags, estimated at daily frequency over 2005:M3-2007:M12. We use the *forward guidance* factor from the press-conference window as the external instrument m_t (robust $F = 9.9$), since every variable in this VAR is a forward-looking asset price, and all four respond to news about the expected path of policy. A surprise that is orthogonal to the contemporaneous policy decision isolates this path component. With $n = 4$, the rotation has $p_\phi = n(n-1)/2 = 6$ Givens angles. We estimate $\hat{\phi}$ by continuous-updating GMM, stacking $q_P = k(n-1) = 3$ proxy exclusion conditions with the $q_{NG} = n(n-1)+n(n-1)/2 = 18$ higher-order moment conditions[16] of Section 2.3. Responses are reported for a one-standard-deviation monetary policy shock. Appendix H.1 reports the corresponding moments of the structural shocks, where all four excess kurtoses are positive, and the pooled Jarque-Bera test rejects their joint Gaussianity.

6.3.2. *Results.* Figure 2 plots the 90% identification-robust confidence sets along with their standard MBB confidence bands. A contractionary forward guidance surprise raises the two-year OIS rate by 2.7 basis points on impact for a one-standard-deviation shock and decays slowly: the response is still 0.7 basis points after 200 business days. Equity prices fall on impact (-0.09 log points), and then rise after two weeks. The euro appreciates from about five months onward, the only horizons at which the exchange rate response is signed. Though the medium-run pattern of a persistent increase in expected rates and a stronger euro is the conventional transmission of a tightening surprise, the delayed increase in equity prices is the information-effect, which the literature associates with hawkish communication that

[16]Unlike the oil-news specification of Section 6.2.1, we do not append the asymmetric co-kurtosis conditions here. Matching a Student-t to the estimated excess kurtosis of the third structural shock gives $\hat{\nu} = 4.7$, which does not clear the $6 + 3\delta$ moment bound required by the asymmetric conditions under the factorization of Section 2.3, whose sample variance would therefore not be consistently estimable. Identification holds on the conditions above in either case; see Appendix H.1.

also conveys a more favorable assessment of the future economic outlook, see Nakamura and Steinsson (2018); Jarociński and Karadi (2020); Miranda-Agrippino and Ricco (2021).



Notes: Each panel superimposes three 90% interval estimates for the impulse response of one variable to the monetary policy shock. The shaded band is the residual-based moving-block-bootstrap (MBB) percentile confidence interval for the HGMM point estimator. The two dashed lines are the identification-robust AR sets: *Proxy-only* (blue, dashed) inverts the AR statistic using only the three proxy-exclusion conditions that identify the target shock; *Proxy+NG* (red, dashed) is the HGMM criterion which additionally imposes the three co-skewness conditions of the target shock and the fourth-order aggregate, for 7 conditions in total. Critical values are from a residual-based moving-block bootstrap (9999 bootstrap replications, block length $\ell = 50$), computed per candidate rotation. The solid black line is the HGMM point estimate. Responses are to a one standard deviation contractionary monetary policy shock; the x-axis is the horizon in business days.

FIGURE 2. 90% identification-robust confidence sets with *Proxy-only* and *HGMM* criteria, with standard MBB percentile bands.

Table 10 reports the lengths of confidence sets at both criteria: *HGMM* and *Proxy-only*, and Table 11 the corresponding length ratios at horizons of 0, 50, 100, 150, and 200 business days. The identifying contribution of the higher-order moment conditions is modest relative to oil-news shock in Section 6.2, but still substantial with a median reduction in length of 26.4%; by variable, the median reduction is 20% for the two-year OIS rate, 30% for the equity index, 29% for the exchange rate, and 39% for the inflation-linked swap, where the sets are more than halved at some horizons. The *HGMM* confidence sets are narrower than the *Proxy-only* sets at every variable-horizon pair. The two sets are not nested as the per-candidate bootstrap critical value rises with the number of conditions imposed. Also, Table H.3 in Appendix H.3 shows these narrower confidence sets allow us to interpret the signs of the point estimates in longer horizons. Furthermore, the proxy exclusion restriction is not

rejected ($\hat{J}_{prx} = 3.18$, with asymptotic and bootstrap p-values of 0.36 and 0.75, respectively). Other results with the MBB percentile bands (including for the proxy-estimator), AR sets with alternate bootstrap block length are reported in Appendix H.3. Appendix J reports another application in which the added higher-order moment conditions are not enough to offset the additional degrees of freedom.

	$h = 0$		$h = 50$		$h = 100$		$h = 150$		$h = 200$	
Variable	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM
2Y OIS	0.0180	**0.0162**	0.0308	**0.0264**	0.0302	**0.0245**	0.0238	**0.0183**	0.0193	**0.0141**
Euro Stoxx 50	1.021	**0.799**	0.484	**0.278**	0.363	**0.230**	0.337	**0.259**	0.328	**0.258**
EUR/USD	0.485	**0.349**	0.231	**0.189**	0.188	**0.111**	0.192	**0.122**	0.211	**0.159**
2Y infl.-linked swap	0.0191	**0.0086**	0.0158	**0.0120**	0.0064	**0.0049**	0.0056	**0.0025**	0.0036	**0.0016**

Notes: Each entry is the length (upper minus lower bound) of the 90% identification-robust Anderson–Rubin confidence set for the impulse response of the row variable to the target shock at horizon h (business days). *Proxy* inverts the AR statistic using only the $q_P = 3$ proxy exclusion conditions that identify the target shock; *HGMM* additionally imposes the higher-order moment conditions, for 7 conditions in total. The smaller length in each (variable, horizon) pair is in **bold**. The interest-rate and swap responses are in percentage points; the equity and exchange-rate responses are in log points. Sets are obtained by inverting the AR statistic over the target-shock rotation subspace with per-candidate moving-block-bootstrap critical values (9999 replications, block length $\ell = 50$); see the construction in Appendix G.

TABLE 10. Length comparison of 90% Anderson–Rubin confidence sets: *Proxy-only* vs *HGMM* (Altavilla monetary-policy specification)

	Ratio at horizon h					Over $h = 0, \ldots, 200$	
Variable	$h = 0$	$h = 50$	$h = 100$	$h = 150$	$h = 200$	Median	Max
2Y OIS	1.11	1.17	1.23	1.30	1.37	1.25	1.38
Euro Stoxx 50	1.28	1.74	1.58	1.30	1.27	1.43	1.86
EUR/USD	1.39	1.23	1.69	1.56	1.32	1.41	1.87
2Y infl.-linked swap	2.22	1.32	1.31	2.24	2.25	1.64	2.33

Notes: Each entry is the ratio len(Proxy)/len(HGMM) of the length of the 90% identification-robust Anderson–Rubin confidence set under *Proxy-only* to the corresponding length under *HGMM*, for the impulse response of the row variable to the target shock. Values above one indicate that the *Proxy-only* set is longer. The last two columns summarize the ratio over all 201 horizons $h = 0, \ldots, 200$. Construction as in Table 10.

TABLE 11. Ratio of 90% Anderson–Rubin confidence-set lengths: *Proxy-only* / *HGMM* (Altavilla monetary-policy specification)

7. CONCLUSION

This paper combines identification through external instruments with identification through non-Gaussianity, stacking higher-order moment conditions and proxy exclusion restrictions into a single over-identified GMM criterion. With non-Gaussian structural shocks, the higher-order moment conditions eliminate the continuous rotational indeterminacy, and reduce the identified set to the finite discrete group $\mathcal{B}$; a sufficient number of valid proxies then selects a unique element from $\mathcal{B}$. The proxy-targeted columns of B_0 are globally point-identified and the remaining columns are identified up to signed permutation. Neither channel achieves

this alone: this HGMM estimator is robust to weak instruments, and it does not require an arbitrary labeling convention to give the estimated shocks an economic interpretation.

Under local-to-zero proxy relevance the estimator remains consistent, and the identification-robust Anderson–Rubin confidence sets are substantially narrower than those based on the proxy exclusion conditions alone. The hybrid estimator is also more efficient than the standalone non-Gaussian GMM estimator, which is less obvious: even when the proxies lose their identifying power, their long-run covariance with the non-Gaussian moment conditions continues to sharpen the effective weighting of the non-Gaussian block.

We propose two mutually orthogonal specification tests, J_{prx} and J_{NG}, which assess the identification assumptions of each channel separately. Under the joint null both test statistics follow χ^2 limits. The effective degrees of freedom compress below the nominal count by $p_\phi - \text{rank}(G_{\text{complement}})$ whenever the complementary Jacobian block is rank-deficient. A residual-based moving block bootstrap with Hall–Horowitz recentering supplies critical values that adapt both to this compression and to the fourth-order cumulants in the weighting matrix.

Some extensions can be pursued. The first is to explore the estimator when either of the identification source is weak: both channels do not enter the criterion at the same informative level, so this problem does not reduce to a familiar weak-identification analysis with two sources, and requires further work. The second is partial identification through non-Gaussianity, where only some shocks satisfy the higher-order moment conditions; this would involve estimation of the target column directly. The reduction used for the identification-robust sets shows that the target-shock responses are a function of the unit vector q_1 alone. This implies that a criterion built from the moment conditions involving the target shock could be minimized over the sphere $\mathcal{S}^{n-1}$ without ever forming the full rotation. This would make the dimension of the estimation problem grow linearly rather than quadratically in n, at the cost of giving up the joint identification of the remaining columns. It is a natural route to apply this framework to larger applications than those considered here.

References

Altavilla, C., Brugnolini, L., Gürkaynak, R. S., Motto, R., and Ragusa, G. (2019). Measuring euro area monetary policy. *Journal of Monetary Economics*, 108:162–179. 4, 28, 33

Andrews, D. W. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica*, 70(1):119–162. 21, 57

Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, 59(3):817–858. 11

Andrews, D. W. K. (1992). Generic Uniform Convergence. *Econometric Theory*, 8(2):241–257. 50

Andrews, D. W. K. (1994). Empirical process methods in econometrics. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, volume 4, chapter 37, pages 2247–2294. Elsevier. 50

Andrews, D. W. K. and Cheng, X. (2012). Estimation and Inference With Weak, Semi-Strong, and Strong Identification. *Econometrica*, 80(5):2153–2211. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA9456. 50

Andrews, D. W. K., Cheng, X., and Guggenberger, P. (2020). Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics*, 218(2):496–531. 50

Angelini, G., Cavaliere, G., and Fanelli, L. (2024a). An identification and testing strategy for proxy-SVARs with weak proxies. *Journal of Econometrics*. Accepted: 2024-01-08T13:16:27Z. 5, 60

Angelini, G. and Fanelli, L. (2019). Exogenous uncertainty and the identification of structural vector autoregressions with external instruments. *Journal of Applied Econometrics*, 34(6):951–971. 10

Angelini, G., Fanelli, L., and Neri, L. (2024b). INVALID PROXIES AND VOLATILITY CHANGES. 5

Antoine, B. and Renault, E. (2007). Efficient GMM with multiple rates of convergence and applications to nearly-weak identification. 22

Arias, J. E., Rubio-Ramírez, J. F., and Waggoner, D. F. (2021). Inference in Bayesian Proxy-SVARs. *Journal of Econometrics*, 225(1):88–106. 5

Barth III, M. J. and Ramey, V. A. (2000). The Cost Channel of Monetary Transmission. 74

Bauer, M. D. and Swanson, E. T. (2022). A Reassessment of Monetary Policy Surprises and High-Frequency Identification. 17

Baumeister, C. and Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, 83(5):1963–1999. 5

Bekaert, G., Engstrom, E., and Ermolov, A. (2021). Macro risks and the term structure of interest rates. *Journal of Financial Economics*, 141(2):479–504. 74

Bruns, M., Lütkepohl, H., and McNeil, J. (2025). Avoiding Unintentionally Correlated Shocks in Proxy Vector Autoregressive Analysis. *Journal of Business & Economic Statistics*, 43(4):1119–1131. _eprint: https://doi.org/10.1080/07350015.2025.2476699. 5

Brüggemann, R., Jentsch, C., and Trenkler, C. (2016). Inference in VARs with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 191(1):69–85. 3, 21, 29, 57

Carriero, A., Marcellino, M., and Tornese, T. (2024). Blended identification in structural VARs. *Journal of Monetary Economics*, 146:103581. 5

Cavaliere, G., Fanelli, L., and Georgiev, I. (2025). Bootstrap Diagnostic Tests. arXiv:2509.01351 [econ]. 29

Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284. 61

Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing*, 36(3):287–314. 2, 6, 7, 47, 61

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press, Oxford. 50

Davidson, R. and MacKinnon, J. G. (2008). Bootstrap inference in a linear equation estimated by instrumental variables. *Econometrics Journal*, 11(3):443–477. 60

Drautzburg, T. and Wright, J. H. (2023). Refining Set-Identification in VARs through Independence. 59, 61

Dufour, J.-M. (1990). Exact tests and confidence sets in linear regressions with autocorrelated errors. *Econometrica*, 58(2):475–494. 61

Dufour, J.-M. and Taamouti, M. (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica*, 73(4):1351–1365. 61

Fagiolo, G., Napoletano, M., and Roventini, A. (2008). Are output growth-rate distributions fat-tailed? some evidence from oecd countries. *Journal of Applied Econometrics*, 23(5):639–669. 74

Fisher, J. D. and Peters, R. (2010). Using Stock Returns to Identify Government Spending Shocks. *The Economic Journal*, 120(544):414–436. 74

Gertler, M. and Karadi, P. (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics*, 7(1):44–76. 28, 71, 72, 74

Giacomini, R., Kitagawa, T., and Read, M. (2022). Narrative Restrictions and Proxies. 5

Golub, G., Klema, V., and Stewart, G. W. (1977). Rosetak document 4: Rank degeneracies and least square problems. Working Paper 165, National Bureau of Economic Research. 54

Gonçalves, S. and White, H. (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics*, 119(1):199–219. 57

Gonçalves, S. and Kilian, L. (2005). Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form. 60

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017). Statistical inference for independent component analysis: Application to structural VAR models. *Journal of Econometrics*, 196(1):111–126. 2, 74

Granziera, E., Moon, H. R., and Schorfheide, F. (2018). Inference for VARs identified with sign restrictions. *Quantitative Economics*, 9(3):1087–1121. 61

Gregory, A. W., McNeil, J., and Smith, G. W. (2024). US fiscal policy shocks: Proxy-SVAR overidentification via GMM. *Journal of Applied Econometrics*, 39(4):607–619. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.3038. 5

Guay, A. (2021). Identification of structural vector autoregressions through higher unconditional moments. *Journal of Econometrics*, 225(1):27–46. 7

Guvenen, F., Karahan, F., Ozkan, S., and Song, J. (2021). What do data on millions of u.s. workers reveal about lifecycle earnings dynamics? *Econometrica*, 89(5):2303–2339. 74

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press. Google-Books-ID: 8YkSDAAAQBAJ. 12

Hall, A. R. and Inoue, A. (2003). The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics*, 114(2):361–394. 22

Hall, P. and Horowitz, J. L. (1996). Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. *Econometrica*, 64(4):891–916. 4, 21, 60

Hansen, B. E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics and Statistics*, 81(4):594–607. 60

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, pages 1029–1054. 10, 18

Hoesch, L., Lee, A., and Mesters, G. (2024). Locally robust inference for non-Gaussian SVAR models. *Quantitative Economics*, 15(2):523–570. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE2274. 17, 59

Hyvärinen, A. (2013). Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110534. 7

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability and Its Applications*, 7(4):349–382. 49

Jarociński, M. and Karadi, P. (2020). Deconstructing monetary policy surprises: The role of information shocks. *American Economic Journal: Macroeconomics*, 12(2):1–43. 34

Jentsch, C. and Lunsford, K. G. (2022). Asymptotically Valid Bootstrap Inference for Proxy SVARs. *Journal of Business & Economic Statistics*, 40(4):1876–1891. _eprint: https://doi.org/10.1080/07350015.2021.1990770. 21, 26, 29, 57, 60

Kaido, H., Molinari, F., and Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*, 87(4):1397–1432. 61

Keweloh, S. A. (2021). A Generalized Method of Moments Estimator for Structural Vector Autoregressions Based on Higher Moments. *Journal of Business & Economic Statistics*, 39(3):772–782. _eprint: https://doi.org/10.1080/07350015.2020.1730858. 2, 6, 9

Keweloh, S. A., Klein, M., and Prüser, J. (2025). Estimating fiscal multipliers by combining statistical identification with potentially endogenous proxies. *The Econometrics Journal*, page utaf027. 5

Kilian, L. (2024). How to construct monthly VAR proxies based on daily surprises in futures markets. *Journal of Economic Dynamics and Control*, 168:104966. 4, 28, 29

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis:*. Cambridge University Press, 1 edition. 7

Kilian, L. and Murphy, D. P. (2014). The Role of Inventories and Speculative Trading in the Global Market for Crude Oil. *Journal of Applied Econometrics*, 29(3):454–478. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.2322. 29, 30

Kleffe, J. (1989). Magnus, jan r., h. neudecker: Matrix differential calculus with applications in statistics and econometrics. j. wiley & sons, chichester – new york – brisbane – toronto – singapore 1988, xvii, 393 pp., £24.50. *Biometrical Journal*, 31(7):885–886. 14

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70(5):1781–1803. 60

Komunjer, I. (2012). GLOBAL IDENTIFICATION IN NONLINEAR MODELS WITH MOMENTRESTRICTIONS. *Econometric Theory*, 28(4):719–729. 2, 10

Känzig, D. R. (2021). The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements. *American Economic Review*, 111(4):1092–1125. 4, 28, 29, 33, 68, 70, 74

Lanne, M., Liu, K., and Luoto, J. (2023). Identifying Structural Vector Autoregression via Leptokurtic Economic Shocks. *Journal of Business & Economic Statistics*, 41(4):1341–1351. _eprint: https://doi.org/10.1080/07350015.2022.2134872. 6, 8, 47

Lanne, M. and Luoto, J. (2021). GMM Estimation of Non-Gaussian Structural Vector Autoregression. *Journal of Business & Economic Statistics*, 39(1):69–81. _eprint: https://doi.org/10.1080/07350015.2019.1629940. 2, 6, 7, 9

Lanne, M., Meitz, M., and Saikkonen, P. (2017). Identification and estimation of non-Gaussian structural vector autoregressions. *Journal of Econometrics*, 196(2):288–304. 74

Lewis, D. (2024). Identification based on higher moments. Technical report, Cemmap. 6, 7

Lewis, D. J. and Mertens, K. (2026). Weak Instrument Bias in Impulse Response Estimators. *Federal Reserve Bank of Dallas, Working Papers*, 2026(2601). 2, 25

Liu, K. (2025). Non-Gaussian Structural Vector Autoregression with Unknown Break Points. 9

Ludvigson, S. C., Ma, S., and Ng, S. (2021). Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response? *American Economic Journal: Macroeconomics*, 13(4):369–410. 2

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer, Berlin [u.a.]. 6, 14, 56

Matteson, D. S. and Tsay, R. S. (2011). Dynamic Orthogonal Components for Multivariate Time Series. *Journal of the American Statistical Association*. 6, 7, 14

Matteson, D. S. and Tsay, R. S. (2013). Independent Component Analysis via Distance Covariance. arXiv:1306.4911 [stat]. 7

Mertens, K. and 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. 2

Mesters, G. and Zwiernik, P. (2024). Non-Independent Components Analysis. arXiv:2206.13668 [math]. 7, 9

Mikusheva, A. (2007). Uniform Inference in Autoregressive Models. *Econometrica*, 75(5):1411–1452. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0262.2007.00798.x. 60

Miranda-Agrippino, S. and Ricco, G. (2021). The Transmission of Monetary Policy Shocks. *American Economic Journal: Macroeconomics*, 13(3):74–107. 17, 34

Montiel Olea, J. L., Stock, J. H., and Watson, M. W. (2021). Inference in Structural Vector Autoregressions identified with an external instrument. *Journal of Econometrics*, 225(1):74–87. 2, 5, 14, 17, 26, 59, 60, 74

Moreira, M. J., Porter, J. R., and Suarez, G. A. (2009). Bootstrap validity for the score test when instruments may be weak. *Journal of Econometrics*, 149(1):52–64. 60

Mori, L. and Peersman, G. (2024). Estimating the Macroeconomic Effects of Oil Supply News. 29

Murphy, K. M. and Topel, R. H. (1985). Estimation and Inference in Two-Step Econometric Models. *Journal of Business & Economic Statistics*, 3(4):370–379. 6, 13

Nakamura, E. and Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: The information effect. *The Quarterly Journal of Economics*, 133(3):1283–1330. 34

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics*, 29(3):229–256. 3, 12, 18, 24

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier. 13, 49, 50

Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703–708. 11

Newey, W. K. and Windmeijer, F. (2009). Generalized Method of Moments with Many Weak Moment Conditions. *Econometrica*, 77(3):687–719. 14

Ramey, V. A. (2016). Chapter 2 - Macroeconomic Shocks and Their Propagation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 71–162. Elsevier. 2, 25, 74

Romer, C. D. and Romer, D. H. (2010). The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *American Economic Review*, 100(3):763–801. 74

Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica*, 39(3):577–591. 10

Salgado, S., Guvenen, F., and Bloom, N. (2019). Skewed Business Cycles. 74

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586. 2, 25

Stock, J. H. and Watson, M. W. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *The Economic Journal*, 128(610):917–948. 2, 7, 74

Stock, J. H. and Wright, J. H. (2000). GMM with Weak Identification. *Econometrica*, 68(5):1055–1096. 25

Stock, J. H. and Yogo, M. (2002). Testing for Weak Instruments in Linear IV Regression. 74

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25. 22

Appendix A. Preliminary notations and key matrices

We collect preliminary notations and asymptotic conventions used throughout the appendix. We also derive key matrices used for the proofs, and the moment conditions are stated again for the reader's convenience. We adopt the basic SVAR parameterization of Section 2.1.

A.1. **Preliminaries.** Let X_T be a sequence of random vectors and X a random vector on a common probability space with measure $\mathbb{P}$. We write $X_T \xrightarrow{p} X$ for convergence in probability and $X_T \xrightarrow{d} X$ for convergence in distribution. For a sequence of random variables Y_T, $Y_T = o_p(1)$ denotes convergence to zero in probability and $Y_T = O_p(1)$ denotes stochastic boundedness: for every $\epsilon > 0$ there are a finite $M_\epsilon > 0$ and an integer T_0 with $\mathbb{P}(\|Y_T\| > M_\epsilon) < \epsilon$ for all $T > T_0$. For a deterministic sequence a_T, $O(a_T)$ denotes a deterministic sequence bounded by a constant multiple of a_T. We denote $Z \sim \mathcal{N}(\mu, \Sigma)$ for the multivariate Gaussian with mean μ and covariance Σ, and $\chi^2(r)$ for the chi-squared distribution with r degrees of freedom.

Let v be a vector and A a real matrix. We define $\|v\| \coloneqq \sqrt{v^\top v}$ for the Euclidean norm and $\|A\| \coloneqq \sup_{\|x\|=1} \|Ax\|$ for the induced spectral norm, $\mathrm{vec}(A)$ for the column-wise vectorization of A, and $\otimes$ for the Kronecker product. The column space and null space of A are $\mathrm{col}(A)$ and $\mathrm{null}(A)$, and A^+ is the Moore–Penrose pseudo-inverse. For symmetric A, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the smallest and largest eigenvalues.

Definiteness is stated through the quadratic form $x^\top Ax$ for symmetric A: $A \succ 0$ means $x^\top Ax > 0$ for all $x \neq 0$, and $A \succeq 0$ means $x^\top Ax \geq 0$ for all x. For a stationary process X_t, the long-run variance $\sum_{h=-\infty}^{\infty} \mathrm{Cov}(X_t, X_{t-h})$ is the asymptotic variance of $T^{-1/2} \sum_t X_t$; it reduces to $\mathrm{Var}(X_t)$ when X_t is serially uncorrelated.

A.2. **Moment conditions.** Let $g_t(\phi) = (g_{NG,t}(\phi)', g_{P,t}(\phi)')'$ collect the non-Gaussian (NG) and proxy (P) moment contributions, with dimensions q_{NG} and $q_P = k(n-1)$, respectively, and total $q = q_{NG} + q_P$. Define:

$$\begin{aligned}
\bar{g}_T(\phi) &= T^{-1} \sum_{t=1}^{T} g_t(\phi) = \big(\bar{g}_{NG,T}(\phi)',\ \bar{g}_{P,T}(\phi)'\big)', \qquad \bar{g}_T : \Phi \to \mathbb{R}^q, \\
g(\phi) &= \mathbb{E}\big[\bar{g}_T(\phi)\big], \\
G(\phi) &= \frac{\partial g(\phi)}{\partial \phi'} \in \mathbb{R}^{q \times p_\phi}, \qquad G \equiv G(\phi_0) = \big(G_{NG}', G_P'\big)',
\end{aligned}$$

where $G_P = \partial g_P(\phi_0)/\partial \phi'$ and $G_{NG} = \partial g_{NG}(\phi_0)/\partial \phi'$. The consistent sample Jacobian is $\hat{G}_T(\phi)$.

The long-run variance of the moments is $S = \mathrm{Avar}(T^{1/2} \bar{g}_T(\phi_0)) > 0$, estimated consistently by $\hat{S}_T$. Conformably with the q_{NG}/q_P partition, we denote

$$S = \begin{pmatrix} S_{NN} & S_{NP} \\ S_{PN} & S_{PP} \end{pmatrix}, \qquad S_{PN} \equiv S_{P,NG}, \quad S_{NP} = S_{PN}',$$

where $S_{PP} \in \mathbb{R}^{q_P \times q_P}$, $S_{NN} \in \mathbb{R}^{q_{NG} \times q_{NG}}$, $S_{PN} \in \mathbb{R}^{q_P \times q_{NG}}$. The efficient weighting matrix is $W = S^{-1}$, with $\hat{W}_T = \hat{S}_T^{-1}$.

A.3. **Key derived matrices.**

- *Schur complements* of S:

$$\Sigma_P \equiv S_{PP} - S_{PN} S_{NN}^{-1} S_{NP} > 0, \qquad \Sigma_{NG} \equiv S_{NN} - S_{NP} S_{PP}^{-1} S_{PN} > 0. \tag{A.1}$$

  Positive definiteness follows from $S > 0$.

- *Annihilator matrix*:

$$\Omega \equiv S - G(G' S^{-1} G)^{-1} G' \succeq 0, \quad \operatorname{rank}(\Omega) = q - p_\phi, \tag{A.2}$$

  and the sample-level annihilator operator $M \equiv I_q - G(G' S^{-1} G)^{-1} G' S^{-1}$, satisfying $T^{1/2} \bar{g}_T(\hat{\phi}_T) = M\, T^{1/2} \bar{g}_T(\phi_0) + o_p(1)$.

- *Orthogonalizing (Frisch-Waugh) operators*:

$$L \equiv [I_{q_P},\ -S_{PN} S_{NN}^{-1}] \in \mathbb{R}^{q_P \times q}, \qquad K \equiv [-S_{NP} S_{PP}^{-1},\ I_{q_{NG}}] \in \mathbb{R}^{q_{NG} \times q}. \tag{A.3}$$

- *Partialled (projected) Jacobians*:

$$\tilde{G}_P \equiv LG\ ;\ \tilde{G}_{NG} \equiv KG \tag{A.4}$$

- *Orthogonalized sample moments*:

$$\tilde{g}_{P,T} \equiv L \bar{g}_T(\hat{\phi}_T), \qquad \tilde{g}_{NG,T} \equiv K \bar{g}_T(\hat{\phi}_T). \tag{A.5}$$

- *Asymptotic variance matrices of orthogonalized moments*:

$$\mathcal{Q}_P \equiv L \Omega L', \qquad \mathcal{Q}_{NG} \equiv K \Omega K'. \tag{A.6}$$

  These are the limiting covariance matrices of $T^{1/2} \tilde{g}_{P,T}(\hat{\phi}_T)$ and $T^{1/2} \tilde{g}_{NG,T}(\hat{\phi}_T)$, respectively (see Appendix B).

Appendix B. Proofs

This section contains the proofs for all the lemmas, propositions and corollaries stated in the paper.

B.1. **Proof of Proposition 2.1.** For the main proof, we use the following Lemma B.1:

Lemma B.1 (Non-Gaussian identification)**.** *Let $B_0 = HQ_0$ with $Q_0 \in SO(n)$, and let $Q(\phi) \in SO(n)$ be the Givens parametrization with $p_\phi = n(n-1)/2$ angles. Let ϕ_0 satisfy $Q(\phi_0) = Q_0$ and the differential of $\phi \mapsto Q(\phi)$ is injective at ϕ_0. Let the non-Gaussian moment block stack the co-skewness and co-kurtosis conditions $f_{ij}^{(1)}$ and $f_{ij}^{(2)}$ in (B.1)–(B.2) for all ordered pairs $i \neq j$. Let $\mathcal{S}_n = \{P\Lambda : P \in \mathcal{P}_n,\ \Lambda \in \mathcal{D}_n^{\pm}\}$ denote the $n!\,2^n$ signed permutation matrices, and let $\mathcal{B} = \{B_0 S : S \in \mathcal{S}_n\}$ be the signed-permutation orbit of B_0. Under Assumptions 2.3 and 2.4:*

(i) the population higher-order moment conditions hold at ϕ if and only if $B(\phi) \in \mathcal{B}$. Hence, the non-Gaussian identified set within the $SO(n)$ parametrization is $\{B_0 S : S \in \mathcal{S}_n,\ \det S = 1\}$, which has $n!\,2^{n-1}$ elements; without the determinant restriction it is the full orbit $\mathcal{B}$, with $n!\,2^n$ elements.

(ii) $G_{NG}(\phi_0)$ has full column rank p_ϕ;

The restriction $\det Q_0 = 1$ is without loss of generality: if $\det Q_0 = -1$, we can replace B_0 by the observationally equivalent $B_0\Lambda_1$, where Λ_1 flips the sign of the first shock.

Proof. Since $u_t = B_0\varepsilon_t = HQ_0\varepsilon_t$, the candidate shocks are $\varepsilon_t(\phi) = R(\phi)\varepsilon_t$, where $R(\phi) \equiv Q(\phi)'Q_0 \in SO(n)$ and $R(\phi_0) = I$. Let us denote r_i' as the i-th row of R, so that $\varepsilon_{i,t}(\phi) = r_i'\varepsilon_t$, and for $i \neq j$ set $X = \varepsilon_{i,t}(\phi)$ and $Y = \varepsilon_{j,t}(\phi)$. By Assumption 2.3, the components of ε_t have mean zero, unit variance and are mutually uncorrelated, with finite fourth-order moments. Hence, $\mathbb{E}[X^2] = \|r_i\|^2 = 1$, $\mathbb{E}[Y^2] = \|r_j\|^2 = 1$ and $\mathbb{E}[XY] = r_i'r_j = 0$ for every ϕ. For mean-zero variables, the moment–cumulant identities are:

$$\mathbb{E}[X^2Y] = \mathrm{cum}(X,X,Y), \qquad \mathbb{E}[X^2Y^2] = \mathrm{cum}(X,X,Y,Y) + \mathbb{E}[X^2]\,\mathbb{E}[Y^2] + 2\,\mathbb{E}[XY]^2.$$

By Assumption 2.3, the third- and fourth-order cumulant tensors of ε_t are diagonal, with diagonal entries $\kappa_{3,k} = \mathbb{E}[\varepsilon_{k,t}^3]$ (the skewness, by unit variance) and $\kappa_{4,k} = \mathbb{E}[\varepsilon_{k,t}^4] - 3$ (the excess kurtosis). Since X and Y are linear combinations of the components of ε_t, multilinearity of cumulants gives:

$$\mathrm{cum}(X,X,Y) = \sum_k \kappa_{3,k}\, r_{ik}^2 r_{jk}, \qquad \mathrm{cum}(X,X,Y,Y) = \sum_k \kappa_{4,k}\, r_{ik}^2 r_{jk}^2,$$

because every cross term involves an off-diagonal cumulant, which is zero. Substituting, and using $\mathbb{E}[X^2]\mathbb{E}[Y^2] + 2\mathbb{E}[XY]^2 = 1$, we obtain for $i \neq j$

$$f_{ij}^{(1)}(R) \equiv \mathbb{E}[\varepsilon_{i,t}^2(\phi)\varepsilon_{j,t}(\phi)] = \sum_k \kappa_{3,k}\, r_{ik}^2 r_{jk}, \tag{B.1}$$

$$f_{ij}^{(2)}(R) \equiv \mathbb{E}[\varepsilon_{i,t}^2(\phi)\varepsilon_{j,t}^2(\phi)] - 1 = \sum_k \kappa_{4,k}\, r_{ik}^2 r_{jk}^2. \tag{B.2}$$

Sufficiency. Let $R \in \mathcal{S}_n$, so that $r_i' = \pm e_{\pi(i)}'$ for a permutation π of $\{1,\dots,n\}$. For $i \neq j$ we have $\pi(i) \neq \pi(j)$, so for every k at most one of r_{ik} and r_{jk} is nonzero and $r_{ik}r_{jk} = 0$. Hence, $r_{ik}^2 r_{jk} = r_{ik}(r_{ik}r_{jk}) = 0$ and $r_{ik}^2 r_{jk}^2 = (r_{ik}r_{jk})^2 = 0$ for every k, so every term in (B.1) and (B.2) vanishes, for any values of $\kappa_{3,k}$ and $\kappa_{4,k}$.

Necessity. Let $R \in SO(n)$ satisfy $f_{ij}^{(2)}(R) = 0$ for all $i \neq j$, and write $s_{ik} = r_{ik}^2 \geq 0$. By Assumption 2.4(b), at most one shock is mesokurtic and the remaining $n-1$ excess kurtoses share a common sign. Without loss of generality this sign is positive (otherwise multiply (B.2) by -1), so $\kappa_{4,k} \geq 0$ for all k. For any pair $i \neq j$,

$$0 = f_{ij}^{(2)}(R) = \sum_k \kappa_{4,k}\, s_{ik}s_{jk} \implies s_{ik}s_{jk} = 0 \quad \text{for every non-mesokurtic } k,$$

since a sum of non-negative terms is zero only if every term is zero. Hence, each of the $n-1$ non-mesokurtic columns of R has at most one nonzero entry. The columns of an orthogonal matrix have unit norm, so $\sum_i s_{ik} = 1$ and that entry equals ± 1. These $n-1$ columns are mutually orthogonal signed unit vectors and are therefore supported on distinct rows. The remaining column, of unit norm and orthogonal to them, is the signed unit vector on the remaining row. Thus, $R \in \mathcal{S}_n$. This proves (i).

Local coordinates. Since $R(\phi_0) = I$, we can write $R(\phi) = \exp(\Phi(\phi))$, where $\Phi(\phi)$ is skew-symmetric, smooth in ϕ, and $\Phi(\phi_0) = 0$. Let $\vartheta(\phi) = (\Phi_{kl}(\phi))_{k<l} \in \mathbb{R}^{p_\phi}$ and $J = \partial\vartheta/\partial\phi' \,|_{\phi_0}$. This implies $\partial R/\partial\phi_m|_{\phi_0} = \partial\Phi/\partial\phi_m|_{\phi_0}$. Moreover, $\partial R/\partial\phi_m = (\partial Q/\partial\phi_m)'Q_0$ with Q_0 invertible. Hence, J is nonsingular if and only if the differential of $\phi \mapsto Q(\phi)$ is injective at ϕ_0. Writing the moment vector as $g(\vartheta(\phi))$, the chain rule gives $G_{NG}(\phi_0) = G_\vartheta J$ with $G_\vartheta = \partial g/\partial\vartheta' \,|_{\vartheta=0}$. It therefore suffices to show that G_ϑ has full column rank.

Expansion. For $k < l$, set $\vartheta_{kl} = \Phi_{kl}$, so that $\Phi_{lk} = -\vartheta_{kl}$, and substitute $r_{ik} = \delta_{ik} + \Phi_{ik} + O(\|\Phi\|^2)$. For the co-kurtosis conditions,

$$f_{ij}^{(2)} = \sum_k \kappa_{4,k}(\delta_{ik} + \Phi_{ik})^2(\delta_{jk} + \Phi_{jk})^2 + O(\|\Phi\|^2) = O(\|\Phi\|^2), \qquad i \neq j,$$

because every first-order term carries the factor $\delta_{ik}\delta_{jk} = 0$. Hence, the co-kurtosis rows of G_ϑ are zero. For the co-skewness conditions,

$$f_{ij}^{(1)} = \sum_k \kappa_{3,k}(\delta_{ik} + \Phi_{ik})^2(\delta_{jk} + \Phi_{jk}) + O(\|\Phi\|^2) = \kappa_{3,i}\Phi_{ji} + O(\|\Phi\|^2).$$

Thus, each coordinate ϑ_{kl}, $k < l$, enters in the following two conditions,

$$f_{kl}^{(1)} = -\kappa_{3,k}\,\vartheta_{kl} + O(\|\Phi\|^2), \qquad f_{lk}^{(1)} = \kappa_{3,l}\,\vartheta_{kl} + O(\|\Phi\|^2),$$

so the column of G_ϑ for ϑ_{kl} equals $(-\kappa_{3,k},\, \kappa_{3,l})'$ on rows $f_{kl}^{(1)}, f_{lk}^{(1)}$ and zero elsewhere. Each ordered pair (i,j) belongs to exactly one unordered pair $\{k,l\}$, so distinct columns have disjoint supports. The columns are therefore linearly independent if and only if each is nonzero, i.e. $(\kappa_{3,k}, \kappa_{3,l}) \neq (0,0)$ for every $k < l$. By Assumption 2.4(a), at most one skewness is zero, so this condition holds. Hence G_ϑ has full column rank p_ϕ, and so does $G_{NG}(\phi_0) = G_\vartheta J$. This proves (ii). □

Remark B.1 (Local versus global identification)**.** The two parts of Lemma B.1 rest on different moments. The global characterization in (i) uses only the co-kurtosis conditions; the common-sign restriction is that of Lanne et al. (2023), who assume leptokurtic shocks, and the signed-permutation indeterminacy is the classical result of Comon (1994). Local identification in (ii) comes entirely from the co-skewness conditions, because the co-kurtosis block is flat at ϕ_0. Consequently, when two or more shocks are close to symmetric, $G_{NG}(\phi_0)$ is close to rank deficient even though (ii) continues to hold: ϕ_0 remains globally identified up to the orbit, but first-order local identification is weak. Applying the same argument at any other element of the identified set, with the skewness coefficients permuted and sign-changed, shows that (i) holds there as well, provided the parametrization is regular at that point.

Now we provide proof of the main proposition:

Proof. By Lemma B.1, any $\tilde{B}_0 \in \mathcal{B}$ takes the form $\tilde{B}_0 = B_0 P \Lambda$ with corresponding alternative shocks

$$\tilde{\varepsilon}_t = \tilde{B}_0^{-1} u_t = \Lambda P' \varepsilon_t, \tag{B.3}$$

where $\Lambda^{-1} = \Lambda$ and $P^{-1} = P'$. Since $\mathcal{B}$ is the non-Gaussian identified set, it suffices to show that the proxy exclusion conditions of Assumption 2.2 and the sign normalization of Assumption 2.8 jointly select a unique element from $\mathcal{B}$ for the target columns, and characterize the indeterminacy for the remaining columns.

For $\tilde{B}_0$ to satisfy Assumption 2.2, the cross-moment $\mathbb{E}[m_t \tilde{\varepsilon}_t']$ must preserve the required block-zero structure. Substituting (B.3) and using $\mathbb{E}[m_t \varepsilon_t'] = [\Psi \;\; 0]$ with Ψ, a $k \times k$ nonsingular diagonal matrix,

$$\mathbb{E}[m_t \tilde{\varepsilon}_t'] = [\,\Psi \;\; 0\,]\, P\Lambda = \left[\, \Psi P_{11}\Lambda_1 \quad \Psi P_{12}\Lambda_2 \,\right], \tag{B.4}$$

where P and Λ are partitioned conformably with block sizes $(k, n-k)$. The $k \times (n-k)$ zero-block requirement in Assumption 2.2 forces $\Psi P_{12}\Lambda_2 = 0$. Since Ψ and Λ_2 are both nonsingular, this gives $P_{12} = 0_{k\times(n-k)}$, and the permutation structure of P then implies $P_{21} = 0_{(n-k)\times k}$. Hence, P is block-diagonal:

$$P = \begin{pmatrix} P_{11} & 0 \\ 0 & P_{22} \end{pmatrix}, \qquad P_{11} \in \mathcal{P}_k, \quad P_{22} \in \mathcal{P}_{n-k}. \tag{B.5}$$

The relevance block of (B.4) reduces to $\Psi P_{11}\Lambda_1 = \tilde{\Psi}$, where $\tilde{\Psi}$ is $k \times k$ diagonal matrix. The product ΨP_{11} is a scaled permutation matrix with non-zero entry ψ_i at position $(i, \sigma(i))$, where σ is the permutation induced by P_{11}. For $\Psi P_{11}\Lambda_1$ to be diagonal, each non-zero entry must lie on the main diagonal, requiring $\sigma(i) = i$ for all $i = 1, \ldots, k$. Hence, $P_{11} = I_k$. Substituting, $\Psi\Lambda_1 = \tilde{\Psi}$ implies $\psi_i \lambda_i = \tilde{\psi}_i > 0$; since $B_{ii} > 0$ (sign-normalized), we obtain $\lambda_i = +1$ for all i, so $\Lambda_1 = I_k$. The allowable transformation therefore reduces to

$$P\Lambda = \begin{bmatrix} I_k & 0 \\ 0 & P_{22}\Lambda_2 \end{bmatrix}, \qquad P_{22} \in \mathcal{P}_{n-k}, \quad \Lambda_2 \in \mathcal{D}_{n-k}^{\pm}, \tag{B.6}$$

and any admissible $\tilde{B}_0 \in \mathcal{B}$ takes the form

$$\tilde{B}_0 = B_0 P \Lambda = \begin{bmatrix} B_{0,1} & B_{0,2} P_{22} \Lambda_2 \end{bmatrix}. \tag{B.7}$$

Since every admissible $\tilde{B}_0$ shares the first k columns $B_{0,1}$ regardless of the choice of P_{22} and Λ_2, the target sub-matrix $B_{0,1}$ is globally point-identified for all $1 \leq k \leq n$. This establishes result (i).

The remaining identification result (ii) depends on the cardinality of the set of admissible (P_{22}, Λ_2) pairs, which is governed by k relative to n.

Case $k = n$. Both P_{22} and Λ_2 are vacuous (the block is empty), so (B.6) gives $P\Lambda = I_n$ directly. Hence, $\tilde{B}_0 = B_0$ and B_0 is globally point-identified.

Case $k = n - 1$. $P_{22} \in \mathcal{P}_1$ forces $P_{22} = (1)$, so $P = I_n$. The remaining sign $\lambda_n \in \{-1, +1\}$ enters the (n, n) entry of $\tilde{B}_0$ as $\lambda_n [B_0]_{nn}$. Assumption 2.8 requires this entry to be positive, which holds if and only if $\lambda_n = +1$. Hence, $\Lambda_2 = (1)$, $P\Lambda = I_n$, and B_0 is globally point-identified.

Case $1 \leq k < n - 1$. The residual permutation P_{22} ranges over $\mathcal{P}_{n-k}$, which has cardinality $(n-k)!$, and Λ_2 ranges over $\mathcal{D}^{\pm}_{n-k}$, which has cardinality 2^{n-k}. The proxy and sign normalization conditions place no further restrictions on (P_{22}, Λ_2), so the equivalence class $\mathcal{E}_{n-k}$ has cardinality at most $(n-k)! \cdot 2^{n-k}$. The residual sub-matrix $B_{0,2}$ is not point-identified; it is identified only up to $\mathcal{E}_{n-k}$. This establishes result (ii) in all cases. □

B.2. **Reversion to set identification under proxy irrelevance.**

Corollary B.2. *Suppose the conditions of Proposition 2.1 hold, but Assumption 2.2 is replaced by strict irrelevance:* $\mathbb{E}[m_t \varepsilon_t'] = 0_{k \times n}$. *Then* $\mathbb{E}[g_{P,t}(\phi)] = 0$ *for every* $\phi \in \Phi$, *the proxy moment block imposes no identifying restrictions, and the identified set under the joint moment system coincides exactly with* $\mathcal{B}$. *The parameter* ϕ_0 *is globally set-identified with identified set of cardinality* $|\mathcal{B}| \leq n! \cdot 2^n$.

Proof. Under strict irrelevance, $\mathbb{E}[m_t u_t'] = \mathbb{E}[m_t \varepsilon_t'] B_0' = 0$. For any $\phi \in \Phi$, using the notation of Section A.3:

$$\mathbb{E}[m_t \varepsilon_t(\phi)'] = \mathbb{E}[m_t u_t'](H^{-1})' Q(\phi) = 0_{k \times n}.$$

Accordingly, $\mathbb{E}[g_{P,t}(\phi)] = 0$ for every ϕ. The proxy conditions are trivially satisfied throughout Φ and contribute no restrictions beyond those already imposed by Assumptions 2.3 and 2.4. The identified set therefore equals $\mathcal{B}$, from which no element can be excluded. □

B.3. **Proof of Proposition 2.2.**

Proof. Substituting $G_{NG} = \tilde{G}_{NG} + S_{NP} S_{PP}^{-1} G_P$ and applying the block-matrix inversion identity $[S^{-1}]_{NN} = \Sigma_{NG}^{-1}$ (see Section A) yields

$$I_{joint} \;=\; G' S^{-1} G \;=\; \underbrace{G_P' S_{PP}^{-1} G_P}_{I_P} \;+\; \underbrace{\tilde{G}_{NG}' \Sigma_{NG}^{-1} \tilde{G}_{NG}}_{\Delta I_{NG}}.$$

$\Sigma_{NG} = S_{NN} - S_{NP}S_{PP}^{-1}S_{PN}$ is the Schur complement of S_{PP} in S. Since $S \succ 0$ by the regularity conditions in Section 2.6.1, $S_{PP} \succ 0$ and $\Sigma_{NG} \succ 0$; hence $\Delta I_{NG} = \tilde{G}'_{NG}\Sigma_{NG}^{-1}\tilde{G}_{NG} \succeq 0$, and $I_{joint} = I_P + \Delta I_{NG} \succeq I_P \succ 0$. Matrix inversion is order-reversing on the positive definite cone: $A \succeq B \succ 0$ implies $B^{-1} \succeq A^{-1}$. Therefore, $\mathcal{V}_P = I_P^{-1} \succeq I_{joint}^{-1} = \mathcal{V}_{joint}$. □

B.4. **Proof of Proposition 3.1.**

Proof. **Part (i)** Under the Pitman sequence only the first-moment relevance $\mathbb{E}[m_{t,T}\varepsilon_{\text{target},t}]$ drifts to zero so the population proxy Jacobian $G_{P,T} = \partial\,\mathbb{E}[g_P(\phi_0)]/\partial\phi'$ satisfies $G_{P,T} = O(T^{-1/2})$. But the proxy's own second moment $\mathbb{E}[m_{t,T}^2]$ is bounded away from zero, so the proxy-block covariance $S_{PP,T}$ stays positive definite with $S_{PP,T}^{-1}$ bounded (and $S_{NN} \succ 0$), and hence $\mathcal{I}_{P,T} = G'_{P,T}S_{PP}^{-1}G_{P,T} = O(T^{-1}) \to 0$ in operator norm. Using the additive decomposition (2.18),

$$\mathcal{I}_{\text{joint},T} = \underbrace{G'_{P,T}\,S_{PP}^{-1}\,G_{P,T}}_{\to 0} + \tilde{G}'_{NG,T}\,\Sigma_{NG}^{-1}\,\tilde{G}_{NG,T}, \qquad \tilde{G}_{NG,T} = G_{NG} - S_{NP}S_{PP}^{-1}G_{P,T}.$$

Since $G_{P,T} \to 0$, we have $\tilde{G}_{NG,T} \to G_{NG}$, and therefore $\mathcal{I}_{\text{joint},T} \to \mathcal{I}^* = G'_{NG}\Sigma_{NG}^{-1}G_{NG}$ in operator norm. By Lemma B.1, G_{NG} has full column rank at ϕ_0 and $\Sigma_{NG} \succ 0$, so $\mathcal{I}^* \succ 0$.

Part (ii) Note that under local-to-zero relevance the exclusion conditions remain correctly specified, i.e., $\mathbb{E}[g_P(\phi_0)] = 0$ holds. Under the regularity conditions of Section 2.6.1 and Assumptions 2.2–2.4, the criterion $\bar{g}_T(\phi)'\hat{S}^{-1}\bar{g}_T(\phi)$ obeys a uniform law of large numbers with a limit uniquely minimized at ϕ_0. The consistency theorem of Newey and McFadden (1994, Thm. 2.6) then gives $\hat{\phi}_T \xrightarrow{p} \phi_0$. Combining the central limit theorem $\sqrt{T}\,\bar{g}_T(\phi_0) \xrightarrow{d} N(0, S)$, which holds under condition 2 of Section 2.6.1 by the CLT for α-mixing sequences (Ibragimov, 1962), gives us

$$\sqrt{T}\,(\hat{\phi}_T - \phi_0) \;\xrightarrow{d}\; \mathcal{N}\big(0,\,(\mathcal{I}^*)^{-1}\big), \qquad \mathcal{V}_{NG}^* = (\mathcal{I}^*)^{-1}.$$

Part (iii) The Schur complement satisfies $\Sigma_{NG} = S_{NN} - S_{NP}S_{PP}^{-1}S_{PN} \preceq S_{NN}$, hence $\Sigma_{NG}^{-1} \succeq S_{NN}^{-1}$ and $\mathcal{I}^* = G'_{NG}\Sigma_{NG}^{-1}G_{NG} \succeq G'_{NG}S_{NN}^{-1}G_{NG}$. The order-reversing property of inversion on the positive-definite cone gives $\mathcal{V}_{NG}^* = (\mathcal{I}^*)^{-1} \preceq (G'_{NG}S_{NN}^{-1}G_{NG})^{-1} = \mathcal{V}_{NG}$, with equality if $S_{NP} = 0$. □

Remark B.2. Part (iii) gives $\mathcal{V}_{NG}^* \preceq \mathcal{V}_{NG}$, with equality if and only if $S_{NP} = 0$. The gain is governed by S_{NP}, not by the relevance matrix Ψ. If the proxy depends on the target shock only linearly, $m_t = \Psi_T\varepsilon_{1,t} + \eta_t$, then $S_{NP,T} \to 0$ as $\Psi_T \to 0$ and the gain vanishes in the local-to-zero limit. If the proxy also loads on higher powers of the target shock, S_{NP} can stay bounded away from zero while $\Psi_T \to 0$, and the gain survives in the limit.

B.5. **Proof of Proposition 3.2.** First we prove the Lemma B.3, then we proceed with the proof of the proposition.

Lemma B.3 (Uniform information bound)**.** *Under Assumption 3.1,*

$$\inf_{\gamma\in\mathcal{P}} \lambda_{\min}\big(G'S^{-1}G\big) \;\geq\; \underline{\kappa} \;\equiv\; \underline{g}^{\,2}/\bar{s} \;>\; 0.$$

Proof. By the information decomposition (2.18), written with the non-Gaussian block marginal,

$$G'S^{-1}G = G'_{NG}S_{NN}^{-1}G_{NG} + \tilde{G}'_P\Sigma_P^{-1}\tilde{G}_P \;\succeq\; G'_{NG}S_{NN}^{-1}G_{NG},$$

since $\tilde{G}'_P\Sigma_P^{-1}\tilde{G}_P \succeq 0$. The dropped term is the only component which is dependent on the proxy strength, so the bound is independent of Ψ. Then $G'_{NG}S_{NN}^{-1}G_{NG} \succeq \lambda_{\max}(S_{NN})^{-1}\, G'_{NG}G_{NG} \succeq (\sigma_{\min}(G_{NG})^2/\bar{s})\, I_{p_\phi} \succeq (\underline{g}^2/\bar{s})I_{p_\phi}$, using $\lambda_{\max}(S_{NN}) \leq \bar{s}$ and Assumption 3.1. □

Now we show the main proof of the proposition:

Sketch of the proof. Lemma B.3 places every $\gamma \in \mathcal{P}$ in the strongly identified regime of Andrews and Cheng (2012), with the parameter of interest ϕ strongly identified by the non-Gaussian block irrespective of Ψ. The four standard steps are then uniform over $\mathcal{P}$. *(i) Uniform consistency.* The criterion $\bar{g}_T(\phi)'\hat{S}^{-1}\bar{g}_T(\phi)$ converges uniformly to a limit with a well-separated minimiser at ϕ_0 (uniform LLN, Andrews, 1992; separation from the curvature bound of Lemma B.3), so $\sup_{\gamma\in\mathcal{P}}\|\hat{\phi}_T - \phi_0\| = o_p(1)$. *(ii) Uniform linearization.* A mean-value expansion of the first-order conditions, with $G'S^{-1}G$ uniformly nonsingular (Lemma B.3) and uniform stochastic equicontinuity (Andrews, 1994), gives $\sqrt{T}(\hat{\phi}_T - \phi_0) = -(G'S^{-1}G)^{-1}G'S^{-1}T^{-1/2}\sum_t(g_t + \Gamma_\theta\zeta_t) + r_T$, $\sup_{\gamma\in\mathcal{P}}\|r_T\| = o_p(1)$, where ζ_t is the first-stage influence function of Section 2.6.3. *(iii) Uniform CLT.* The central limit theorem for triangular arrays of near-epoch dependent functions of α-mixing processes (Davidson, 1994), applied with the mixing coefficients and $(2+\delta)$-moment bounds that are uniform over $\mathcal{P}$ by Assumption 3.1, yields $T^{-1/2}\sum_t(g_t + \Gamma_\theta\zeta_t) \xrightarrow{d} \mathcal{N}(0, \Omega_\psi)$ uniformly, hence $\sqrt{T}(\hat{\phi}_T - \phi_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_\phi)$ uniformly, with $\mathcal{V}_\phi$ as in (2.21). *(iv) Uniform Studentization.* $\hat{\mathcal{V}}_\phi \xrightarrow{p} \mathcal{V}_\phi$ uniformly (with the inverse uniformly continuous by Lemma B.3) over $\mathcal{P}$ by condition 4 of the regularity conditions 2.6.1 and Assumption 3.1, and $R\mathcal{V}_\phi R'$ is uniformly nonsingular, so by the generic uniformity results of Andrews et al. (2020), $W_T \xrightarrow{d} \chi^2_d$ uniformly over $\mathcal{P}$. The coverage statement follows by test inversion. □

B.6. **Proof of Proposition 3.3.**

Proof. Throughout, the estimator uses the efficient weight $W = S^{-1}$, and $\mathbb{E}[g(\phi_0)] = (\,0',\, \mu'_T\,)' = (\,0',\, c'/\sqrt{T}\,)'$ with $\mathbb{E}[g_{NG}(\phi_0)] = 0$ and $\mathbb{E}[g_P(\phi_0)] = \mu_T = c/\sqrt{T}$.

Under the regularity conditions of Newey and McFadden (1994), the pseudo-true parameter $\phi_{*,T}$ is an interior minimizer of the population criterion $\mathcal{Q}_T(\phi) = \mathbb{E}[g(\phi)]'S^{-1}\mathbb{E}[g(\phi)]$ and therefore satisfies the population first-order conditions

$$G(\phi_{*,T})'\, S^{-1}\, \mathbb{E}[g(\phi_{*,T})] = 0, \tag{B.8}$$

where $G(\phi) \equiv \partial\mathbb{E}[g(\phi)]/\partial\phi'$ is partitioned conformably as $G(\phi) = (G_{NG}(\phi)', G_P(\phi)')'$. Since $\mathbb{E}[g(\phi_0)] = O(T^{-1/2})$, the criterion is minimised within an $O(T^{-1/2})$ neighborhood of ϕ_0, so $\phi_{*,T} - \phi_0 = O(T^{-1/2})$.

Let $\delta_T = \phi_{*,T} - \phi_0 = O(T^{-1/2})$. A first-order mean-value expansion of $\mathbb{E}[g]$ around ϕ_0 gives

$$\mathbb{E}[g(\phi_{*,T})] = \mathbb{E}[g(\phi_0)] + G(\phi_0)\,\delta_T + O(\|\delta_T\|^2) = \mathbb{E}[g(\phi_0)] + G(\phi_0)\,\delta_T + O(T^{-1}), \tag{B.9}$$

and the corresponding expansion of the Jacobian gives

$$G(\phi_{*,T}) = G(\phi_0) + O(\|\delta_T\|) = G(\phi_0) + O(T^{-1/2}). \tag{B.10}$$

Substituting (B.9)–(B.10) into (B.8) and expanding,

$$G(\phi_0)'S^{-1}\mathbb{E}[g(\phi_0)] + G(\phi_0)'S^{-1}G(\phi_0)\,\delta_T + O(T^{-1/2})\cdot O(T^{-1/2}) + O(T^{-1}) = 0, \tag{B.11}$$

where the third term is the cross-product of the $O(T^{-1/2})$ Jacobian remainder in (B.10) with $\mathbb{E}[g(\phi_0)] = O(T^{-1/2})$. All remainder terms are therefore $O(T^{-1})$.

Since $W = S^{-1}$ is not block-diagonal, the contamination is transmitted through the off-diagonal blocks of S^{-1}. Writing the inverse via the Schur complement of S_{NN}, $\Sigma_P \equiv S_{PP} - S_{PN}S_{NN}^{-1}S_{NP}$,

$$S^{-1} = \begin{pmatrix} S_{NN}^{-1} + S_{NN}^{-1}S_{NP}\Sigma_P^{-1}S_{PN}S_{NN}^{-1} & -S_{NN}^{-1}S_{NP}\Sigma_P^{-1} \\ -\Sigma_P^{-1}S_{PN}S_{NN}^{-1} & \Sigma_P^{-1} \end{pmatrix}, \tag{B.12}$$

so that, using $\mathbb{E}[g(\phi_0)] = (0', \mu_T')'$,

$$S^{-1}\mathbb{E}[g(\phi_0)] = \begin{pmatrix} -S_{NN}^{-1}S_{NP}\Sigma_P^{-1}\mu_T \\ \Sigma_P^{-1}\mu_T \end{pmatrix}. \tag{B.13}$$

Pre-multiplying by $G(\phi_0)' = (G_{NG}', G_P')$ and collecting terms,

$$G(\phi_0)'S^{-1}\mathbb{E}[g(\phi_0)] = \big(G_P - S_{PN}S_{NN}^{-1}G_{NG}\big)'\Sigma_P^{-1}\mu_T = \tilde{G}_P'\,\Sigma_P^{-1}\,\mu_T, \tag{B.14}$$

where $\tilde{G}_P = G_P - S_{PN}S_{NN}^{-1}G_{NG}$ is the partialed proxy Jacobian and we used $(S_{PN}S_{NN}^{-1}G_{NG})' = G_{NG}'S_{NN}^{-1}S_{NP}$ ($S_{NN} = S_{NN}'$, $S_{PN}' = S_{NP}$). The same identity applied to $G(\phi_0)'S^{-1}G(\phi_0)$ yields the (non-Gaussian-marginal) information decomposition

$$G'S^{-1}G = G_{NG}'S_{NN}^{-1}G_{NG} + \tilde{G}_P'\,\Sigma_P^{-1}\,\tilde{G}_P, \tag{B.15}$$

the symmetric counterpart of (2.18) with the roles of the two moment blocks interchanged. By Lemma B.1, G_{NG} has full column rank, so $G_{NG}'S_{NN}^{-1}G_{NG} \succ 0$ and hence $G'S^{-1}G \succ 0$.

Substituting (B.14) into (B.11),

$$\big(G'S^{-1}G\big)\,\delta_T = -\,\tilde{G}_P'\,\Sigma_P^{-1}\,\mu_T + O(T^{-1}) = -\frac{1}{\sqrt{T}}\,\tilde{G}_P'\,\Sigma_P^{-1}\,c + O(T^{-1}). \tag{B.16}$$

Multiplying by $\sqrt{T}$, setting $\mathbb{B}_T = \sqrt{T}\,\delta_T$, and pre-multiplying by $(G'S^{-1}G)^{-1}$,

$$\mathbb{B}_T = -\big(G'S^{-1}G\big)^{-1}\,\tilde{G}_P'\,\Sigma_P^{-1}\,c + O(T^{-1/2}), \tag{B.17}$$

which establishes (3.5). Taking $T \to \infty$ and recalling $\delta_T = \mathbb{B}_T/\sqrt{T} \to 0$ confirms $\phi_{*,T} \to \phi_0$, so $\hat{\phi}_T$ remains consistent for ϕ_0 under local contamination.

Apply the central limit theorem of Section 2.6.1 (condition 2) to the recentered moment vector $\big(g_{NG}(\phi_0)',\, g_P(\phi_0) - \mu_T\big)'$, which has mean zero and satisfies $T^{-1/2}\sum_t\big(g_t - \mathbb{E}[g_t]\big) \xrightarrow{d} \mathcal{N}(0, S)$. For the efficiently weighted estimator the influence-function representation is $\sqrt{T}(\hat{\phi}_T - \phi_0) = -(G'S^{-1}G)^{-1}G'S^{-1}\,T^{1/2}\bar{g}_T(\phi_0) + o_p(1)$, whose stochastic part has

asymptotic variance $(G'S^{-1}G)^{-1}G'S^{-1}\,S\,S^{-1}G(G'S^{-1}G)^{-1} = (G'S^{-1}G)^{-1}$, while the deterministic drift $\mu_T = c/\sqrt{T}$ enters through (B.14) as the non-centrality $-(G'S^{-1}G)^{-1}\tilde{G}_P'\Sigma_P^{-1}c$. Hence

$$\sqrt{T}(\hat{\phi}_T - \phi_0) \xrightarrow{d} \mathcal{N}\Big(-(G'S^{-1}G)^{-1}\tilde{G}_P'\Sigma_P^{-1}c,\ (G'S^{-1}G)^{-1}\Big),$$

When the innovations are estimated, the variance $(G'S^{-1}G)^{-1}$ carries the generated-regressor correction of Section 2.6.3; this leaves the non-centrality unchanged.

Taking Euclidean norms in (B.17) and applying sub-multiplicativity of the spectral norm,

$$\|\mathbb{B}_T\|_2 \le \big\|(G'S^{-1}G)^{-1}\big\|_2 \big\|\tilde{G}_P'\Sigma_P^{-1}c\big\|_2 + O(T^{-1/2}) = \frac{\big\|\tilde{G}_P'\Sigma_P^{-1}c\big\|_2}{\lambda_{\min}(G'S^{-1}G)} + O(T^{-1/2}), \tag{B.18}$$

using $\|A^{-1}\|_2 = 1/\lambda_{\min}(A)$ for symmetric positive-definite A. By the decomposition (B.15) and $\tilde{G}_P'\Sigma_P^{-1}\tilde{G}_P \succeq 0$,

$$\lambda_{\min}\big(G'S^{-1}G\big) \ \ge\ \lambda_{\min}\big(G_{NG}'S_{NN}^{-1}G_{NG}\big) \ >\ 0, \tag{B.19}$$

so replacing the denominator in (B.18) with $\lambda_{\min}(G_{NG}'S_{NN}^{-1}G_{NG})$ yields the valid upper bound (3.6). The minorant $\lambda_{\min}(G_{NG}'S_{NN}^{-1}G_{NG})$ is the marginal non-Gaussian information and coincides with the anchor of Lemma B.3. □

B.7. **Limiting distributions of the test statistics.**

B.7.1. *Setup.* All notation is as defined in Section A. We additionally maintain Assumption B.1 below, which holds under Assumptions 2.3–2.4 by Lemma B.1.

Assumption B.1. $G_{NG} \in \mathbb{R}^{q_{NG}\times p_\phi}$ has full column rank p_ϕ, with $q_{NG} \ge p_\phi$.

Assumption B.1 implies G has full column rank, so $G'S^{-1}G \succ 0$ and the annihilator $\Omega = S - G(G'S^{-1}G)^{-1}G'$ is positive semidefinite of rank $q - p_\phi$.

B.7.2. *Properties of L and K.* As defined in Section A Equation (A.3),

Lemma B.4. *Under $S > 0$:*

$$LS = [\Sigma_P,\ 0], \quad KS = [0,\ \Sigma_{NG}], \tag{B.20}$$

$$SL' = (\Sigma_P', 0)', \quad SK' = (0', \Sigma_{NG}')', \tag{B.21}$$

$$\mathrm{Cov}(Lg(\phi_0),\ Kg(\phi_0)) = LSK' = 0. \tag{B.22}$$

Proof. Direct block multiplication establishes (B.20):

$$LS = [I_{q_P},\ -S_{PN}S_{NN}^{-1}]\begin{bmatrix} S_{PP} & S_{PN} \\ S_{NP} & S_{NN}\end{bmatrix} = [S_{PP} - S_{PN}S_{NN}^{-1}S_{NP},\ 0] = [\Sigma_P,\ 0],$$

$$KS = [-S_{NP}S_{PP}^{-1},\ I_{q_{NG}}]\begin{bmatrix} S_{PP} & S_{PN} \\ S_{NP} & S_{NN}\end{bmatrix} = [0,\ S_{NN} - S_{NP}S_{PP}^{-1}S_{PN}] = [0,\ \Sigma_{NG}].$$

Equations (B.21) follow by transposition. For (B.22): $LSK' = [\Sigma_P,\ 0](0', \Sigma_{NG}')' = 0$. □

B.7.3. *Projected annihilator representations.*

Lemma B.5. *Under assumptions of Proposition 2.1 and B.1, with annihilator* $M = I_q - G(G'S^{-1}G)^{-1}G'S^{-1}$*:*

$$T^{1/2}\tilde{g}_{P,T}(\hat{\phi}_T) = LM\,T^{1/2}\bar{g}_T(\phi_0)+o_p(1), \qquad T^{1/2}\tilde{g}_{NG,T}(\hat{\phi}_T) = KM\,T^{1/2}\bar{g}_T(\phi_0)+o_p(1). \tag{B.23}$$

The asymptotic variances and cross-covariance are

$$\mathcal{Q}_P \equiv L\Omega L' = \Sigma_P - \tilde{G}_P(G'S^{-1}G)^{-1}\tilde{G}_P', \tag{B.24}$$

$$\mathcal{Q}_{NG} \equiv K\Omega K' = \Sigma_{NG} - \tilde{G}_{NG}(G'S^{-1}G)^{-1}\tilde{G}_{NG}', \tag{B.25}$$

$$L\Omega K' = -\tilde{G}_P(G'S^{-1}G)^{-1}\tilde{G}_{NG}'. \tag{B.26}$$

Proof. Pre-multiplying the full annihilator representation $T^{1/2}\bar{g}_T(\hat{\phi}_T) = M\,T^{1/2}\bar{g}_T(\phi_0)+o_p(1)$ by L and K gives (B.23). For the variances, use $\Omega = MSM' = S - G(G'S^{-1}G)^{-1}G'$:

$$L\Omega L' = LSL' - LG(G'S^{-1}G)^{-1}G'L'.$$

From (B.21), $LSL' = [I_{q_P}, -S_{PN}S_{NN}^{-1}](\Sigma_P', 0)' = \Sigma_P$. With $LG = \tilde{G}_P$, equation (B.24) follows. The derivation of (B.25) is identical with K in place of L, using $KSK' = \Sigma_{NG}$ from (B.20). For (B.26): $L\Omega K' = LSK' - LG(G'S^{-1}G)^{-1}G'K' = 0 - \tilde{G}_P(G'S^{-1}G)^{-1}\tilde{G}_{NG}'$, where $LSK' = 0$ by (B.22). □

Remark B.3 (Non-independence)**.** Equation (B.26) shows that J_{prx} and J_{NG} are not asymptotically independent in general: their limiting covariance is zero if and only if $\tilde{G}_P(G'S^{-1}G)^{-1}\tilde{G}_{NG}' = 0$, which is not implied by the maintained assumptions.

B.7.4. *Ranks of* $\mathcal{Q}_P$ *and* $\mathcal{Q}_{NG}$*.* The null space of Ω is $\text{null}(\Omega) = \text{col}(S^{-1}G)$ with dimension p_ϕ, since $\Omega v = 0 \Leftrightarrow v = S^{-1}G\alpha$ for $\alpha = (G'S^{-1}G)^{-1}G'v$.

Lemma B.6 (Full rank of $\mathcal{Q}_P$)**.** *Under Assumption B.1,* $\mathcal{Q}_P \succ 0$ *and* $\text{rank}(\mathcal{Q}_P) = q_P$*.*

Proof. Let $\mathcal{Q}_P v = 0$. Since $\mathcal{Q}_P \succeq 0$, this gives $(L'v)'\Omega(L'v) = 0$, so $L'v \in \text{null}(\Omega) = \text{col}(S^{-1}G)$. Thus $L'v = S^{-1}G\alpha$ for some α. Pre-multiplying by S and applying (B.21):

$$\begin{pmatrix}\Sigma_P v\\ 0\end{pmatrix} = SL'v = G\alpha = \begin{pmatrix}G_P\alpha\\ G_{NG}\alpha\end{pmatrix}.$$

The lower block gives $G_{NG}\alpha = 0$. Since G_{NG} has full column rank (Assumption B.1), $\alpha = 0$. Then $\Sigma_P v = 0$, and $\Sigma_P \succ 0$ gives $v = 0$. □

Lemma B.7 (Rank of $\mathcal{Q}_{NG}$)**.** *Under Assumption B.1,*

$$\text{rank}(\mathcal{Q}_{NG}) = q_{NG} - (p_\phi - \text{rank}(G_P)). \tag{B.27}$$

In particular, $\mathcal{Q}_{NG} \succ 0$ *if and only if* $\text{rank}(G_P) = p_\phi$*.*

Proof. Let $\mathcal{Q}_{NG}v = 0$. The same argument gives $K'v = S^{-1}G\alpha$, and applying (B.21):

$$\begin{pmatrix}0\\ \Sigma_{NG}v\end{pmatrix} = SK'v = G\alpha = \begin{pmatrix}G_P\alpha\\ G_{NG}\alpha\end{pmatrix}.$$

This yields the system

$$G_P\alpha = 0, \qquad G_{NG}\alpha = \Sigma_{NG}v. \tag{B.28}$$

The first equation confines α to $\mathcal{N}_P \equiv \text{null}(G_P)$, which has dimension $p_\phi - \text{rank}(G_P)$.

The map $\varphi : \mathcal{N}_P \to \mathbb{R}^{q_{NG}}$ defined by $\varphi(\alpha) = \Sigma_{NG}^{-1}G_{NG}\alpha$ is a bijection onto $\text{null}(\mathcal{Q}_{NG})$:

- *Injective*: if $\varphi(\alpha) = 0$ then $G_{NG}\alpha = 0$; full column rank of G_{NG} (Assumption B.1) forces $\alpha = 0$.
- *Surjective*: every $v \in \text{null}(\mathcal{Q}_{NG})$ satisfies (B.28) for some $\alpha \in \mathcal{N}_P$, and $v = \Sigma_{NG}^{-1}G_{NG}\alpha = \varphi(\alpha)$.

Therefore $\dim(\text{null}(\mathcal{Q}_{NG})) = \dim(\mathcal{N}_P) = p_\phi - \text{rank}(G_P)$, giving (B.27). □

Remark B.4 (Degrees-of-freedom compression)**.** In a SVAR with n variables and k proxy instruments, $p_\phi = n(n-1)/2$ and $q_P = k(n-1)$, we have $\text{rank}(G_P) < p_\phi$ generically, and $\mathcal{Q}_{NG}$ is rank-deficient. As a concrete example, for $n = 3$ and $k = 1$: $q_P = 2 < p_\phi = 3$, so $\text{rank}(\mathcal{Q}_{NG}) \leq q_{NG} - 1$.

B.7.5. *Asymptotic distributions.*

Lemma B.8 (Quadratic form under singular normal)**.** *Let $Z \sim \mathcal{N}(0, A)$ with $A \succeq 0$ of rank r. Then $Z'A^+Z \sim \chi^2(r)$.*

Proof. Let $A = U\Lambda U'$ with $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$, $\lambda_i > 0$, U orthogonal, and $A^+ = U\Lambda^+U'$. Set $W = U'Z \sim \mathcal{N}(0, \Lambda)$ and partition $W = (W_1', W_2')'$ with $W_1 \in \mathbb{R}^r$. Since $\text{Var}(W_2) = 0$, Z is supported on $\text{col}(A)$, so $W_2 = U_2'Z = 0$ a.s. Therefore:

$$Z'A^+Z = W'\Lambda^+W = \sum_{i=1}^{r} \frac{W_{1,i}^2}{\lambda_i}.$$

Each $W_{1,i}/\sqrt{\lambda_i} \sim \mathcal{N}(0,1)$ independently, so the sum is $\chi^2(r)$. □

Proof of Proposition 4.1

Proof. Part (i). By (B.23) and the CLT from the regularity conditions 2.6.1, $T^{1/2}\tilde{g}_{P,T}(\hat{\phi}_T) \xrightarrow{d} Z_P \sim \mathcal{N}(0, \mathcal{Q}_P)$, since $LMS(LM)' = L\Omega L' = \mathcal{Q}_P$. By Lemma B.6, $\mathcal{Q}_P \succ 0$, so $\hat{\mathcal{Q}}_P^{-1} \xrightarrow{p} \mathcal{Q}_P^{-1}$ and the continuous mapping theorem gives $J_{\text{prx}} \xrightarrow{d} Z_P'\mathcal{Q}_P^{-1}Z_P = \|\mathcal{Q}_P^{-1/2}Z_P\|^2 \sim \chi^2(q_P)$.

Part (ii). Identical to Part (i) with K, $\mathcal{Q}_{NG}$, q_{NG} in place of L, $\mathcal{Q}_P$, q_P, using $\mathcal{Q}_{NG} \succ 0$ from Lemma B.7.

Part (iii). We have $T^{1/2}\tilde{g}_{NG,T}(\hat{\phi}_T) \xrightarrow{d} Z_{NG} \sim \mathcal{N}(0, \mathcal{Q}_{NG})$ with $\text{rank}(\mathcal{Q}_{NG}) = r_{NG}$ by Lemma B.7.

Let $\mathcal{Q}_{NG} = V\Lambda_{NG}V'$ be the spectral decomposition, with $\Lambda_{NG} = \text{diag}(\mu_1, \ldots, \mu_{r_{NG}}, 0, \ldots, 0)$, $\mu_i > 0$, and Π the projector onto $\text{col}(\mathcal{Q}_{NG})$ (spanned by the first r_{NG} columns of V). The thresholded estimator $\hat{\Pi}$ retains eigenvectors of $\hat{\mathcal{Q}}_{NG}$ with eigenvalues $\geq \tau_T$. Since the positive eigenvalues of $\mathcal{Q}_{NG}$ are bounded away from zero and $\hat{\mathcal{Q}}_{NG} \xrightarrow{p} \mathcal{Q}_{NG}$, Weyl's inequality implies $\hat{\mu}_i \xrightarrow{p} \mu_i > 0$ for $i \leq r_{NG}$ and $\hat{\mu}_j \xrightarrow{p} 0$ for $j > r_{NG}$. The rate condition $\epsilon_T \to 0$, $T^{1/2}\epsilon_T \to \infty$ ensures $\hat{\Pi} \xrightarrow{p} \Pi$ and $\hat{\Pi}\hat{\mathcal{Q}}_{NG}^{+}\hat{\Pi} \xrightarrow{p} \mathcal{Q}_{NG}^{+}$ (Golub et al., 1977) Theorem 2.3.

Since Z_{NG} is supported on $\mathrm{col}(\mathcal{Q}_{NG}) = \mathrm{col}(\Pi)$ a.s., we have $\Pi Z_{NG} = Z_{NG}$ a.s. The continuous mapping theorem then gives

$$J_{NG} \xrightarrow{d} Z'_{NG}\mathcal{Q}^{+}_{NG}Z_{NG} \sim \chi^2(r_{NG}),$$

where the last step applies Lemma B.8 with $A = \mathcal{Q}_{NG}$ and $r = r_{NG}$. □

Remark B.5 (Structural interpretation of the asymmetry)**.** J_{prx} uses the jointly identified estimator $\hat{\phi}_T$; the full-rank G_{NG} absorbs all p_ϕ estimation degrees of freedom, leaving the proxy moments free to span all q_P dimensions. For J_{NG}, only $\mathrm{rank}(G_P)$ of the p_ϕ identification directions are supplied by the proxy block; the remaining $p_\phi - \mathrm{rank}(G_P)$ directions compress the effective degrees of freedom below q_{NG}.

Remark B.6 (Symmetric contamination)**.** By the same argument, proxy mis-specification ($\mathbb{E}[g_P(\phi_0)] \neq 0$) drives $\hat{\phi}_T$ to ϕ^{**} with $\mathbb{E}[g_{NG}(\phi^{**})] \neq 0$ generically, inflating J_{NG}. This non-separability is why both statistics are constructed from the jointly identified $\hat{\phi}_T$.

B.7.6. *Summary of degrees of freedom.* Table B.1 collects the degrees of freedom for each test statistic.

Statistic	d.f. (general)	d.f. ($\mathrm{rank}(G_P) = p_\phi$)
J_{joint}	$q - p_\phi$	$q - p_\phi$
J_{prx}	q_P	q_P
J_{NG}	$q_{NG} - (p_\phi - \mathrm{rank}(G_P))$	q_{NG}

TABLE B.1. Effective degrees of freedom for each test statistic. The general column applies whenever assumptions for Proposition 2.1 and Assumption B.1 hold.

The sum of the subset degrees of freedom satisfies

$$q_P + \big(q_{NG} - (p_\phi - \mathrm{rank}(G_P))\big) = (q - p_\phi) + \mathrm{rank}(G_P).$$

This exceeds the full-test degrees of freedom $q - p_\phi$ whenever $\mathrm{rank}(G_P) > 0$, reflecting the fact that J_{prx} and J_{NG} are not asymptotically independent (see the Remark following Lemma B.5): the additive decomposition $J_{\mathrm{full}} = J_{\mathrm{prx}} + J_{NG}$ holds in distribution if and only if $\tilde{G}_P(G'S^{-1}G)^{-1}\tilde{G}'_{NG} = 0$, which is not generically satisfied.

APPENDIX C. GENERATED-REGRESSOR CORRECTION

The first-stage estimator admits the expansion $\sqrt{T}(\hat{\theta}_T - \theta_0) = T^{-1/2}\sum_t \zeta_t + o_p(1)$, with $\mathcal{V}_\theta = \mathbb{E}[\zeta_t\zeta_t']$. The long-run variance of the combined score $g_t + \Gamma_\theta\zeta_t$ is

$$\Omega_\psi \equiv \sum_{h=-\infty}^{\infty} \mathbb{E}\big[(g_t + \Gamma_\theta\zeta_t)(g_{t-h} + \Gamma_\theta\zeta_{t-h})'\big] = S + \Gamma_\theta\mathcal{V}_\theta\Gamma'_\theta + \Gamma_\theta\mathcal{C}' + \mathcal{C}\Gamma'_\theta, \tag{C.1}$$

$$\mathcal{C} \equiv \sum_{h\geq 0} \mathbb{E}\big[g_t(\psi_0)\,\zeta'_{t-h}\big],$$

where S is the long-run variance of Section 2.6.1 and $\mathcal{C}$ is the long-run cross-covariance between the structural moments and the first-stage score. The sum defining $\mathcal{C}$ runs over $h \geq 0$ only, because ζ_t is a martingale difference: $\mathbb{E}[g_t \zeta_{t+j}'] = 0$ for $j \geq 1$. The off-diagonal block of (2.20) is

$$\mathcal{V}_{\theta\phi} = -\left(\mathcal{V}_\theta \Gamma_\theta' + \mathcal{C}'\right) S^{-1} G_0 \left(G_0' S^{-1} G_0\right)^{-1}. \tag{C.2}$$

The cross-covariance $\mathcal{C}$ is generically nonzero under non-Gaussian identification. For the co-skewness coordinate $g = \varepsilon_i^2 \varepsilon_j$ $(i \neq j)$ and $u_t = B_0 \varepsilon_t$, $\mathbb{E}[u_{k,t}\, \varepsilon_{i,t}^2 \varepsilon_{j,t}] = B_{0,kj}\, \mathbb{E}[\varepsilon_{j,t}^2 \varepsilon_{i,t}^2] = B_{0,kj} \neq 0$, so both $\mathcal{C}$ and $\mathcal{V}_{\theta\phi}$ must be retained.

The estimator $\hat{\Omega}_\psi$ in (2.22) is

$$\hat{\Omega}_\psi = \hat{S} + \hat{\Gamma}_\theta \hat{\mathcal{V}}_\theta \hat{\Gamma}_\theta' + \hat{\Gamma}_\theta \hat{\mathcal{C}}' + \hat{\mathcal{C}} \hat{\Gamma}_\theta', \tag{C.3}$$

with

$$\begin{aligned} \hat{G} &= T^{-1} \textstyle\sum_t \partial g_t(\hat{\psi})/\partial \phi', & \hat{\Gamma}_\theta &= T^{-1} \textstyle\sum_t \partial g_t(\hat{\psi})/\partial \theta', \\ \hat{\zeta}_t &= \left(T^{-1} \textstyle\sum_s x_s x_s'\right)^{-1} x_t \otimes \hat{u}_t, & \hat{\mathcal{C}} &= \textstyle\sum_{h=0}^{T-1} \omega_{h,T}\, T^{-1} \sum_{t=h+1}^{T} g_t(\hat{\psi})\, \hat{\zeta}_{t-h}', \end{aligned} \tag{C.4}$$

where $\omega_{h,T}$ are the kernel weights of $\hat{S}_T$.

Appendix D. Inference for structural impulse-response functions

The structural VMA representation is $y_t = \mu + \sum_{h=0}^{\infty} \Theta_h \varepsilon_{t-h}$, where $\Theta_h = \Xi_h B_0$ (see Section A for the definitions of Θ_h, Ξ_h, and related objects). The reduced-form VMA coefficients satisfy the recursion

$$\Xi_h = \sum_{j=1}^{\min(h,p)} A_j \Xi_{h-j}, \qquad \Xi_0 = I_n, \tag{D.1}$$

and $[\Theta_h]_{ij}$ measures the response of variable i to a unit impulse in structural shock j at horizon h.

Since $\Theta_h = \Xi_h(\theta) B_0(\theta, \phi)$ is a smooth function of ψ, the composite Jacobian $\mathcal{J}_{\Theta_h} = \partial \operatorname{vec}(\Theta_h)/\partial \psi'$ follows from the product rule:

$$\mathcal{J}_{\Theta_h} = (B_0' \otimes I_n) \frac{\partial \operatorname{vec}(\Xi_h)}{\partial \psi'} + (I_n \otimes \Xi_h)\, \mathcal{J}_{B_0}. \tag{D.2}$$

Since Ξ_h depends only on θ, we have $\partial \operatorname{vec}(\Xi_h)/\partial \phi' = 0$. Differentiating the VMA recursion with respect to θ (Lütkepohl, 2005) gives

$$\frac{\partial \operatorname{vec}(\Xi_h)}{\partial \theta'} = \sum_{j=1}^{\min(h,p)} \left[(I_n \otimes A_j) \frac{\partial \operatorname{vec}(\Xi_{h-j})}{\partial \theta'} + (\Xi_{h-j}' \otimes I_n)\, E_j \right], \tag{D.3}$$

where $(\Xi_{h-j}' \otimes I_n) E_j$ captures the direct effect of the j-th lag matrix on Ξ_h, and $(I_n \otimes A_j)$ propagates the derivative of Ξ_{h-j} forward through the recursion. The delta method applied to the joint asymptotic distribution of $\hat{\psi}$ with Jacobian $\mathcal{J}_{\Theta_h}$ yields

$$\sqrt{T} \operatorname{vec}(\hat{\Theta}_h - \Theta_h) \xrightarrow{d} \mathcal{N}\left(0,\ \mathcal{J}_{\Theta_h} \mathcal{V}_\psi \mathcal{J}_{\Theta_h}'\right). \tag{D.4}$$

Appendix E. Residual-Based moving block bootstrap for HGMM estimator

The bootstrap follows Brüggemann et al. (2016). Their validity results cover statistics that are smooth functions of second moments of (u_t, m_t); for the higher-order moment conditions used here, validity of the moving-block bootstrap for GMM statistics of α-mixing data follows from Gonçalves and White (2004), under the moment and mixing conditions of Section 2.6.1. The procedure operates on the joint residual-proxy pairs

$$z_t = (\hat{u}_t', m_t')',$$

where $\hat{u}_t = y_t - \hat{\nu} - \sum_{j=1}^{p} \hat{A}_j y_{t-j}$ for $t = 1, \dots, T$ are OLS residuals. Construct the $T - \ell + 1$ overlapping blocks

$$\mathcal{L}_i = \big(z_i,\ z_{i+1},\ \dots,\ z_{i+\ell-1}\big), \qquad i = 1, \dots, T - \ell + 1, \tag{E.1}$$

where the block length ℓ satisfies $\ell \to \infty$ and $\ell^3/T \to 0$ as $T \to \infty$. Following Jentsch and Lunsford (2022), we set $\ell = \lfloor 5.03\, T^{1/4} \rfloor$.

For each replication $b = 1, \dots, B$, execute the following steps.

(1) **Block resampling.** Draw $b_T = \lceil T/\ell \rceil$ blocks independently with replacement from $\{\mathcal{L}_1, \dots, \mathcal{L}_{T-\ell+1}\}$. Concatenate and trim to length T, yielding the bootstrap joint sequence $\{z_t^*\}_{t=1}^T = \{(\hat{u}_t^{*\prime}, m_t^{*\prime})'\}_{t=1}^T$. By construction, the bootstrap blocks preserve the empirical joint dependence between $\hat{u}_t$ and m_t within each block.

(2) **Bootstrap standardised innovations.** Recursively generate the bootstrap time series $\hat{y}_t^*$ using the autoregressive coefficients and the resampled residuals; then demean the bootstrap residuals $\hat{u}_t^*$ and compute the bootstrap residual covariance $\hat{\Sigma}^* = T^{-1}(U^{*\prime})U^*$. With the lower-triangular Cholesky factor H^* satisfying $H^*(H^*)' = \hat{\Sigma}^*$, the bootstrap standardised innovations are

$$\tilde{U}^* = (H^*)^{-1}(U^*)'. \tag{E.2}$$

(3) **k-step Newton–Raphson bootstrap estimation.** Initialize the bootstrap parameter at the original estimate, $\phi_0^* = \hat{\phi}_T$, and iterate the Newton–Raphson update using the bootstrap moment matrix $G_m(\phi^*, \tilde{U}^*, M^*)$ and the fixed original-data weighting matrix $W = \hat{S}^{-1}$. The bootstrap gradient and Hessian are

$$\nabla = G^{*\prime} W g^*, \qquad \mathcal{H} = G^{*\prime} W G^*, \tag{E.3}$$

with update $\phi^* \leftarrow \phi^* - \mathcal{H}^{-1}\nabla$ until both the step norm and gradient norm fall below tolerance $\epsilon = 10^{-6}$, or after a maximum of 100 iterations. This k-step procedure follows Andrews (2002) and achieves first-order asymptotic equivalence to the full two-step bootstrap estimator at substantially reduced computational cost.

(4) **Recentering and bootstrap weighting matrix.** Evaluate the recentered bootstrap moment vector

$$g^*_{\text{final}} = \bar{G}_m^* - \bar{g}_T(\hat{\phi}_T), \tag{E.4}$$

and the bootstrap weighting matrix

$$S^* = \frac{1}{T}(G_m^* - \bar{G}_m^*)'(G_m^* - \bar{G}_m^*), \qquad W^* = (S^*)^{-1}. \tag{E.5}$$

(5) **Bootstrap Schur complements and projected Jacobians.** Partition S^* and G^* conformably with block sizes (q_{NG}, q_P) and compute the bootstrap projection operators

$$P_P^* = S_{PN}^*(S_{NN}^*)^{-1}, \qquad P_{NG}^* = S_{NP}^*(S_{PP}^*)^{-1}. \tag{E.6}$$

The bootstrap projected Jacobians and Schur complements are

$$\tilde{G}_P^* = G_P^* - P_P^* G_{NG}^*, \qquad \tilde{G}_{NG}^* = G_{NG}^* - P_{NG}^* G_P^*, \tag{E.7}$$

$$\Sigma_P^* = S_{PP}^* - P_P^* S_{NP}^*, \qquad \Sigma_{NG}^* = S_{NN}^* - P_{NG}^* S_{PN}^*. \tag{E.8}$$

(6) **Parameter-estimation correction.** Apply the estimation-degrees-of-freedom correction to each Schur complement:

$$\mathcal{Q}_P^* = \Sigma_P^* - \tilde{G}_P^*(G^{*\prime}W^*G^*)^{-1}(\tilde{G}_P^*)', \tag{E.9}$$

$$\mathcal{Q}_{NG}^* = \Sigma_{NG}^* - \tilde{G}_{NG}^*(G^{*\prime}W^*G^*)^{-1}(\tilde{G}_{NG}^*)'. \tag{E.10}$$

(7) **Bootstrap test statistics.** The bootstrap orthogonalized moments are

$$g_P^* = g_P^{*,\text{fin}} - P_P^* g_{NG}^{*,\text{fin}}, \qquad g_{NG}^* = g_{NG}^{*,\text{fin}} - P_{NG}^* g_P^{*,\text{fin}}. \tag{E.11}$$

The bootstrap test statistics are

$$J_{\text{joint}}^* = T \cdot (g_{\text{final}}^*)' W^* g_{\text{final}}^*, \tag{E.12}$$

$$J_{\text{prx}}^* = T \cdot (g_P^*)' (\mathcal{Q}_P^*)^+ g_P^*, \tag{E.13}$$

$$J_{NG}^* = T \cdot (g_{NG}^*)' (\mathcal{Q}_{NG}^*)^+ g_{NG}^*. \tag{E.14}$$

Appendix F. Estimates under Weak non-Gaussianity

	True Value	$T = 200$	$T = 300$	$T = 500$	$T = 1000$
Estimates and their standard errors					
$\hat{\phi}_1$	0.80	0.7881	0.7996	0.8009	0.8002
		(0.0356)	(0.0295)	(0.0231)	(0.0165)
$\hat{\phi}_2$	−0.40	−0.3976	−0.3992	−0.3996	−0.3999
		(0.0358)	(0.0299)	(0.0235)	(0.0168)
$\hat{\phi}_3$	1.20	1.2013	1.1993	1.2044	1.1989
		(0.0373)	(0.0312)	(0.0243)	(0.0174)

Notes: Results based on $M = 500$ Monte Carlo replications with $N_B = 999$ bootstrap replications. Structural shocks follow a Pitman mixture: $\varepsilon_{kt} = w \cdot \varepsilon_{kt}^{NG} + (1-w) \cdot z_{kt}$, where $w = c/\sqrt{T}$ with $c = 2.24$, ε^{NG} are the baseline non-Gaussian shocks, and $z_{kt} \sim N(0,1)$. As T increases, the shocks approach Gaussianity ($w \approx 0.158, 0.129, 0.100, 0.071, 0.022$). Proxies are valid with relevance $\Psi = (0.70, 0.60, 0.80)'$ and noise $\sigma_v = 0.50$.

Table F.1. Parameter Estimates under H_0: Weak non-Gaussianity (Pitman Drift $c = 2.24$)

	True Value	$T = 200$	$T = 300$	$T = 500$	$T = 1000$
Estimates and their standard errors					
$\hat{\phi}_1$	0.80	0.7619	0.7942	0.8001	0.7984
		(0.1218)	(0.1256)	(0.1275)	(0.1314)
$\hat{\phi}_2$	−0.40	−0.3776	−0.3931	−0.3685	−0.3962
		(0.1092)	(0.1117)	(0.1163)	(0.1207)
$\hat{\phi}_3$	1.20	1.0839	1.1105	1.1493	1.1229
		(0.1202)	(0.1275)	(0.1273)	(0.1320)

Notes: Results based on $M = 500$ Monte Carlo replications with $N_B = 999$ bootstrap replications. Structural shocks follow a Pitman mixture: $\varepsilon_{kt} = w \cdot \varepsilon_{kt}^{NG} + (1-w) \cdot z_{kt}$, where $w = c/\sqrt{T}$ with $c = 2.24$, ε^{NG} are the baseline non-Gaussian shocks, and $z_{kt} \sim N(0,1)$. As T increases, the shocks approach Gaussianity ($w \approx 0.158, 0.129, 0.100, 0.071$). Proxy relevance follows a Pitman drift: $\Psi_i = c/\sqrt{T}$ with $c = 2.24$ for all instruments, yielding $\Psi \approx (0.158, 0.129, 0.100, 0.071)$ for $T = (200, 300, 500, 1000)$.

TABLE F.2. Parameter Estimates under H_0: Weak Proxies and Weak non-Gaussianity (Pitman Drift $c = 2.24$)

APPENDIX G. CONSTRUCTION OF IDENTIFICATION-ROBUST CONFIDENCE SETS

We construct the confidence sets by inverting an identification-robust test statistic over structural rotation angles of the target shock, as in the weak-identification-robust SVAR literature (Montiel Olea et al., 2021; Drautzburg and Wright, 2023; Hoesch et al., 2024).

The impulse responses of the target shock depend on $B = H\,Q(\phi)$ only through its first column,

$$b_1 \;=\; H\,q_1(\phi_{(1)}), \qquad q_1(\phi_{(1)}) \;=\; Q(\phi)\,e_1. \tag{G.1}$$

Under the Givens rotations parameterization of Section 2.1, q_1 depends only on the first $n-1$ angles $\phi_{(1)}$; the remaining angles rotate the non-target shocks among themselves. Following Montiel Olea et al. (2021), we invert the test over $\phi_{(1)} \in (-\pi/2, \pi/2)^{n-1}$ only, on which q_1 ranges over the hemisphere $q_{1,1} > 0$. Since H is lower triangular with positive diagonal, this fixes a positive impact of the shock on its own target variable, matching the normalization of the point estimate.

For a candidate q_1, the target shock is $e_{1,t}(q_1) = q_1'\tilde{u}_t$ and the remaining shocks are $e_{\perp,t}(q_1) = N(q_1)'\tilde{u}_t$, where $\tilde{u}_t = H^{-1}u_t$ and the columns of $N(q_1)$ are an orthonormal basis of the orthogonal complement of q_1. The *Proxy-only* criterion uses the proxy exclusion conditions; *HGMM* adds the higher-order conditions of Section 2.3 evaluated at the target shock:

$$\bar{g}(q_1) = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} m_t\, e_{\perp,t}(q_1) \\ e_{1,t}^2(q_1)\, e_{\perp,t}(q_1) \\ e_{1,t}^3(q_1)\, e_{\perp,t}(q_1) \\ e_{1,t}^2(q_1)\big(\tilde{u}_t'\tilde{u}_t - e_{1,t}^2(q_1)\big) - (n-1) \end{bmatrix}. \tag{G.2}$$

The blocks are, in order, the proxy exclusion conditions $\mathbb{E}[m_t e_{j,t}] = 0$, the co-skewness conditions $\mathbb{E}[e_{1,t}^2 e_{j,t}] = 0$, the co-kurtosis conditions $\mathbb{E}[e_{1,t}^3 e_{j,t}] = 0$, and the aggregated symmetric

co-kurtosis condition $\mathbb{E}\big[e_{1,t}^2\|e_{\perp,t}\|^2\big] = n-1$, which uses $\|e_{\perp,t}\|^2 = \tilde{u}_t'\tilde{u}_t - e_{1,t}^2$. All hold at the true rotation under Assumption 2.3, unit-variance shocks and proxy exogeneity. This gives $n-1$ conditions under *Proxy-only* and $3(n-1)+1$ under *HGMM*: 5 and 16 in the oil system, 3 and 10 in the monetary system.[17]

The reduction to $\phi_{(1)}$ is exact. Replacing $N(q_1)$ by $N(q_1)R$, with R orthogonal, pre-multiplies the first three blocks by R' and leaves the last unchanged, so $\bar{g}(q_1)$ and its bootstrap counterparts are transformed by a common nonsingular matrix P, and $\widehat{S}(q_1)$ by $P(\cdot)P'$. The statistic (G.3) and its bootstrap distribution are invariant to such transformations.

The Anderson–Rubin statistic is

$$\mathrm{AR}(q_1) \;=\; T\,\bar{g}(q_1)'\,\widehat{S}(q_1)^{-1}\,\bar{g}(q_1), \tag{G.3}$$

with $\widehat{S}(q_1)$ the long-run variance of the moments evaluated at the hypothesized rotation q_1. It involves no estimated Jacobian, so under the null $\sqrt{T}\,\bar{g}(q_1)$ is asymptotically normal regardless of identification strength and (G.3) is asymptotically pivotal.

We depart from the exact implementation in two respects. First, $\widehat{S}(q_1)$ is estimated by the residual-based moving-block bootstrap of Section 4.4 (Jentsch and Lunsford, 2022), which resamples the recentered residuals and the proxy jointly in blocks of length ℓ and re-estimates the VAR in each draw. The estimate therefore accounts for estimation of $(\hat{\mathcal{A}}, H)$, which Montiel Olea et al. (2021) handle by the delta method. With $Z_b^*(q_1) = \sqrt{T}\big(\bar{g}_b^*(q_1) - \bar{g}(q_1)\big)$ the bootstrap moments recentered at the sample moments (Hall and Horowitz, 1996),

$$\widehat{S}(q_1) = \frac{1}{B-1}\sum_{b=1}^{B} Z_b^*(q_1)\,Z_b^*(q_1)', \qquad \mathrm{AR}_b^*(q_1) = Z_b^*(q_1)'\,\widehat{S}(q_1)^{-1}\,Z_b^*(q_1), \tag{G.4}$$

where the first expression is a bootstrap variance estimator (Gonçalves and Kilian, 2005). Second, the critical value $q_\alpha(q_1)$ is the $(1-\alpha)$ quantile of $\{\mathrm{AR}_b^*(q_1)\}_{b=1}^B$, computed separately at each candidate, as in the grid bootstrap (Hansen, 1999; Mikusheva, 2007). Bootstrap critical values for Anderson–Rubin-type statistics remain valid under weak instruments (Moreira et al., 2009; Davidson and MacKinnon, 2008), and grid moving-block bootstrap AR sets are used in proxy SVARs by Angelini et al. (2024a). The confidence set is

$$\mathcal{C}_{1-\alpha} = \Big\{\, \phi_{(1)} \in (-\tfrac{\pi}{2}, \tfrac{\pi}{2})^{n-1} \;:\; \mathrm{AR}\big(q_1(\phi_{(1)})\big) \le q_\alpha\big(q_1(\phi_{(1)})\big) \Big\}. \tag{G.5}$$

Because the critical value is computed at each q_1 and is typically larger when more conditions enter, the *HGMM* and *Proxy-only* sets need not be nested: moment conditions that are nearly uninformative in some directions raise the critical value more than the statistic (Kleibergen, 2002).

The response of variable v at horizon h is linear in q_1,

$$\mathrm{IRF}_v(h; q_1) = e_v'\,\big(R\,\hat{\mathcal{A}}^h R'\big)H\,q_1 \;=\; c_{v,h}'\,q_1, \tag{G.6}$$

[17] The individual symmetric terms $\mathbb{E}[e_{1,t}^2 e_{j,t}^2] - 1$ are not used, because a change of basis $N(q_1)$ mixes them with the cross terms $\mathbb{E}[e_{1,t}^2 e_{j,t} e_{k,t}]$; only their sum is basis-free. As a robustness check, we omit the asymmetric co-kurtosis conditions, leaving $2(n-1)+1$ conditions under *HGMM*; see Figure H.2.

and its confidence interval is the projection of $\mathcal{C}_{1-\alpha}$, $\left[\min_{\mathcal{C}_{1-\alpha}} c'_{v,h} q_1,\ \max_{\mathcal{C}_{1-\alpha}} c'_{v,h} q_1\right]$, as in Dufour (1990); Dufour and Taamouti (2005); Granziera et al. (2018). Since $\mathcal{C}_{1-\alpha}$ need not be convex or connected (Drautzburg and Wright, 2023) and a full grid is infeasible, we compute each endpoint by constrained optimization (Kaido et al., 2019). We evaluate (G.5) at 10^5 rotations, half drawn uniformly over the box and half concentrated around the point estimate (Chernozhukov et al., 2007), and then optimize (G.6) subject to $\mathrm{AR}(q_1) \leq q_\alpha(q_1)$ from the most extreme accepted points.

The exact projection is conservative by construction. The computed interval is an inner approximation to it, and the endpoints are stable when the number of draws is increased. Because the shock has unit variance, $|c'_{v,h} q_1| \leq \|c_{v,h}\|$, so every interval is bounded. An interval that reaches the boundary of the box, where $q_{1,1} \to 0$, cannot exclude a zero impact on the target variable; we report it as reaching the normalization boundary.

Appendix H. Additional Results

This appendix collects further results for the two main illustrations. For each we report some evidence regarding the non-Gaussianity of the reduced-form innovations and the estimated structural shocks. We also provide impulse responses of the target shock estimated by the hybrid GMM criterion, with moving-block-bootstrap percentile confidence intervals, alongside the single-proxy SVAR benchmark; both estimators are computed on the same bootstrap replications, so the two comparisons differ only in the identification step. The point estimates and their percentile bands lead to the same economic reading as the identification-robust sets in the main text.

H.1. **Evidence for non-Gaussianity in the structural shocks.** Panel A of Table H.1 reports the third and fourth moments of the reduced-form innovations and of the structural shocks recovered at the estimate. Identification requires Assumption 2.4: at most one shock may have zero skewness, and at least $n-1$ shocks must have nonzero excess kurtosis of a common sign. Five of the six excess kurtoses are positive and one, the world IP shock, is negative, so part (b) holds with exactly $n-1$ shocks sharing a sign. The weaker condition of Comon (1994), that at most one shock have both cumulants equal to zero holds comfortably. Taken as a system, joint Gaussianity of the shock vector is rejected decisively (pooled $JB = 35.14$ on 12 degrees of freedom, $p = 0.0004$).

Matching a Student-t to the estimated kurtoses gives us $\hat{\nu} \geq 9.9$ for every shock, so the sixth- and eighth-order moments that govern the asymptotic variance of the fourth-order conditions are finite. This allows the asymmetric co-kurtosis conditions to be appended in Section 6.2.1 for oil-shock illustration; they are an addition to the identifying block, not a necessary requirement for it.

Panel B of Table H.1 reports the same statistics for the daily financial VAR. The normality is rejected for all four structural shocks, and on the raw statistics the euro-area shocks are further from Gaussian than the oil-market shocks of Panel A. The identifying gain reported in Section 6.3.2 is nonetheless the smaller of the two since the additional moment block

	Reduced-form $\hat{u}_t$		Structural $\hat{\varepsilon}_t(\hat{\phi})$	
Variable / shock	Skew	Ex. kurt	Skew	Ex. kurt
Panel A: oil news shock, Kilian-corrected specification				
1 Real oil price / **target**	−0.084	+0.043	−0.042	+0.004
2 World oil production	−0.014	−0.139	+0.150	+0.102
3 World oil inventories	−0.146	+0.003	−0.246	+0.016
4 World IP	+0.059	+0.311	−0.083	−0.289
5 U.S. IP	−0.295	+1.862	−0.308	+1.021
6 U.S. CPI	+0.165	+1.141	+0.035	+0.860
Pooled $JB = 35.14$ on 12 d.f., $p = 0.0004$				
Panel B: monetary policy shock, Altavilla specification				
1 2Y OIS / **target**	−0.179	+1.883	0.105	+1.258
2 Euro Stoxx 50 (log)	0.241	0.657	−0.433	+0.733
3 EUR/USD (log)	0.252	0.533	+0.424	+1.031
4 2Y infl.-linked swap	0.038	12.146	−0.334	+10.307
Pooled $JB = 2572.49$ on 8 d.f., $p < 10^{-16}$				

Notes: Sample skewness and excess kurtosis of the reduced-form residuals $\hat{u}_t$ of (6.1) and of the structural shocks $\hat{\varepsilon}_t(\hat{\phi}) = Q(\hat{\phi})' H^{-1} \hat{u}_t$ recovered at the HGMM estimate. Identification requires Assumption 2.4: at most one shock with zero skewness, and at least $n - 1$ shocks with nonzero excess kurtosis of a common sign. The pooled statistic, the sum of the marginal JB statistics, is distributed χ^2_{2n} under joint Gaussianity of the shock vector and rejects decisively in both applications.

TABLE H.1. Evidence of non-Gaussianity in the reduced-form innovations and estimated structural shocks

(asymmetric co-kurtosis) that the oil application appends is not available here. Under the factorization of Section 2.3, $\mathbb{E}\|g_t\|^{2+\delta} < \infty$ needs $4 + 2\delta$ marginal moments for $\mathbb{E}[\varepsilon_i^2 \varepsilon_j^2] = 1$ but $6 + 3\delta$ for $\mathbb{E}[\varepsilon_i^3 \varepsilon_j] = 0$; without independence both bounds double and the ordering is unchanged. An estimated structural shock, with excess kurtosis of over 10 and hence $\hat{\nu} \approx 4.5$, clears the first threshold and fails the second: the identifying conditions remain admissible, but the asymmetric co-kurtosis conditions cannot be appended as their variance is undefined[18]. The euro-area criterion therefore rests on the identifying conditions of co-skewness and symmetric co-kurtosis alone.

H.2. **Oil news shock (Kilian (2024)-corrected specification).** Table H.2 records the horizons at which each criterion determines the sign of a response. Figure H.1 reports the HGMM and proxy-SVAR responses with MBB percentile bands. Figure H.2 shows the identification-robust sets when the asymmetric co-kurtosis conditions $\mathbb{E}[\varepsilon_i^3 \varepsilon_j]$ are omitted from the non-Gaussian block, for both estimation and inference. The ranking of the two criteria is unaffected, the *HGMM* set remains the shorter one at 277 of the 306 (variable, horizon) pairs, but the magnitude of the gain is smaller. The median reduction in length falls from 71.6% to 14.8%. The asymmetric block therefore carries a large share of the identifying information in

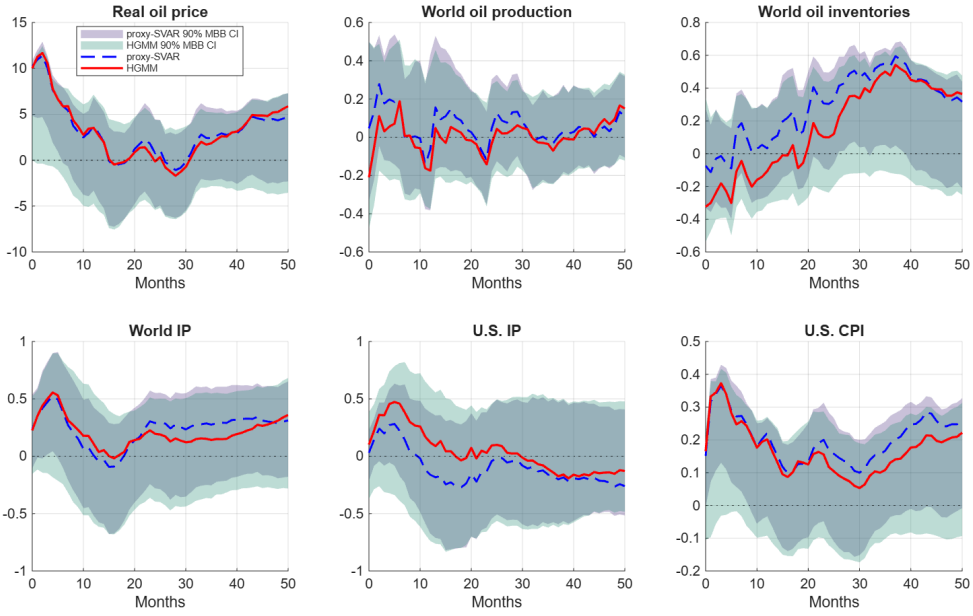
[18]The symmetric co-kurtosis conditions clear $4 + 2\delta$ but not the $8 + 4\delta$ bound that holds without the factorization, so the euro-area results may maintain weak tail dependence across shocks.

this application, which is consistent with the moment-existence argument of Appendix H.1: the oil-market shocks show finite eighth moments and can support fourth-order conditions, whereas the euro-area shocks cannot. Figure H.3 repeats the main-text comparison at the shorter bootstrap block length $\ell = 25$.

	Horizons with $0 \notin$ CS		Sign-resolved horizon ranges	
Variable	Proxy	HGMM	Proxy	HGMM
Real oil price	9	**41**	$[0, 8]$ (+)	$[0, 14]$, $[21, 22]$ (+), $[26, 30]$ (−), $[32, 50]$ (+)
World oil production	0	**23**	–	$[22, 23]$ (−), $[29, 31]$ (+), $[32, 36]$ (−), $[41, 50]$ (+)
World oil inventories	23	**37**	$[25, 46]$ (+)	$[0, 5]$, $[8, 12]$ (−), $[26, 50]$ (+)
World IP	1	**31**	–	$[0, 11]$, $[19, 25]$, $[39, 50]$ (+)
U.S. IP	0	**31**	–	$[1, 10]$ (+), $[31, 50]$ (−)
U.S. CPI	26	**49**	$[2, 13]$, $[36, 50]$ (+)	$[0, 29]$, $[32, 50]$ (+)

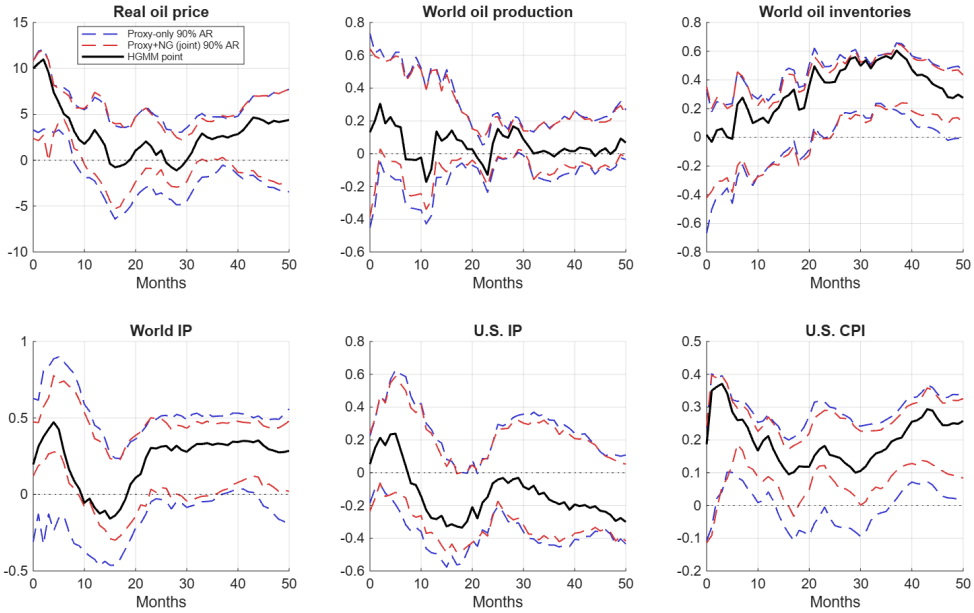
Notes: Columns 2–3 count the horizons $h \in \{0, \ldots, 50\}$ (out of 51) at which the 90% identification-robust confidence set excludes zero, so that the sign of the response is determined. Columns 4–5 report every maximal block of consecutive such horizons, in horizon order, with the sign of the resolved response as a superscript. Out of the 306 (variable, horizon) pairs, *HGMM* resolves a sign that *Proxy-only* leaves open at 154 pairs, while the reverse occurs at a single pair. Under *Proxy-only* the sign of the world oil production and U.S. industrial production responses is undetermined at every horizon, and the world oil inventory response is never resolved as negative. Construction as in Table 8.

TABLE H.2. Horizons at which the sign of the response is resolved: *Proxy-only* vs *HGMM* (Kilian-corrected specification)



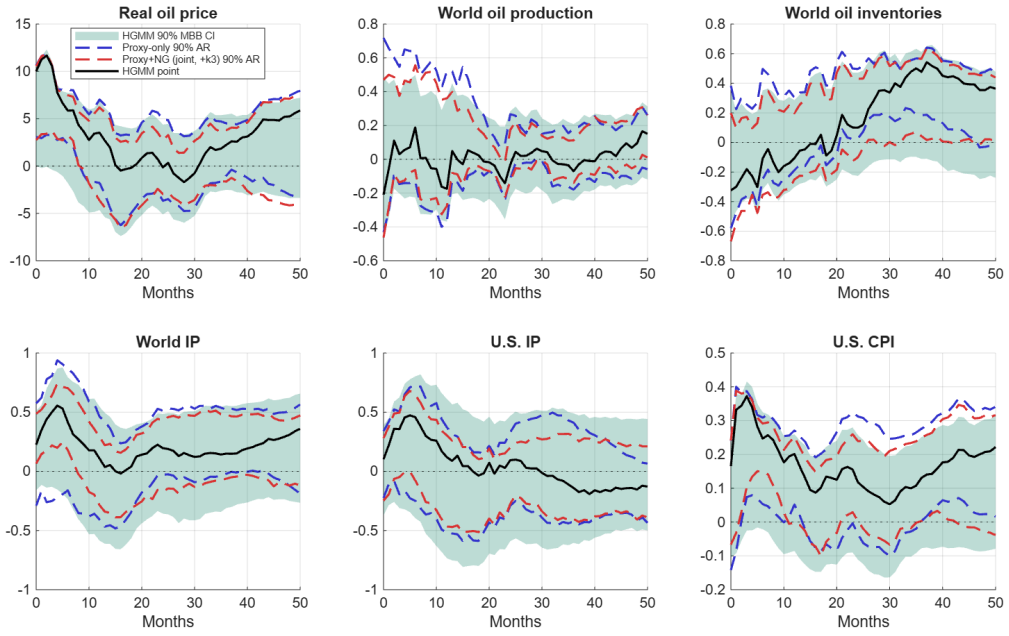
Notes: Point impulse responses of the oil news shock estimated by the hybrid GMM criterion (HGMM, solid red) and by the single-proxy SVAR (proxy-SVAR, dashed blue), each with 90% residual-based moving-block-bootstrap percentile confidence intervals (shaded; HGMM in green, proxy-SVAR in purple). Both estimators are computed on every bootstrap draw from the same resampled residuals, so the two comparisons differ only in the identification step. Bootstrap: 9999 replications, block length $\ell = 36$, VAR re-estimated within each draw. Responses are normalized to a 10% increase in the real oil price on impact; the x-axis is the horizon in months, the y-axis the response in percent.

FIGURE H.1. Impulse responses to the oil news shock: HGMM vs proxy-SVAR, with moving-block-bootstrap percentile confidence intervals (Kilian-corrected specification).



Notes: Each panel plots the 90% identification-robust Anderson–Rubin (AR) confidence set for the impulse response of one variable to the oil news shock, at horizons 0–50 months. *Proxy-only* (blue, dashed) inverts the AR statistic using only the five proxy-orthogonality conditions that identify the target shock; *HGMM* (red, dashed) additionally imposes the co-skewness and (only) symmetric co-kurtosis-aggregate conditions of the target shock (11 conditions in total). The solid black line is the HGMM point estimate. Critical values from a residual-based moving-block bootstrap (9999 replications, block length $\ell = 36$). Responses are normalized to a 10% increase in the real oil price on impact. The x-axis is the horizon in months; the y-axis is the response in percent.

FIGURE H.2. 90% Anderson–Rubin confidence sets for the oil news shock: *Proxy-only* vs *HGMM* (without asymmetric co-kurtosis conditions).



Notes: Each panel superimposes three 90% interval estimates for the impulse response of one variable to the oil news shock. The shaded band is the residual-based moving-block-bootstrap (MBB) percentile confidence interval for the HGMM point estimator. The two dashed lines are the identification-robust AR sets: *Proxy-only* (blue, dashed) inverts the AR statistic using only the five proxy orthogonality conditions that identify the target shock; *HGMM* (red, dashed) additionally imposes the higher-order moment conditions of the target shock (16 conditions in total). The solid black line is the HGMM point estimate. Critical values from a residual-based moving-block bootstrap (9999 replications, block length $\ell = 25$). Responses are normalized to a 10% increase in the real oil price on impact. The x-axis is the horizon in months; the y-axis is the response in percent.

FIGURE H.3. 90% identification-robust confidence sets with *Proxy-only* and *HGMM* criteria, with standard MBB percentile bands (block length $\ell = 25$)

H.3. **Monetary policy shock (Altavilla (2019) specification).** Table H.3 reports the horizons $h \in \{0, \ldots, 200\}$ (out of 201) at which the 90% identification-robust confidence sets excludes zero, so that the sign of the response is determined. Under *Proxy-only*, the exchange rate response is not signed at any horizon, while under *HGMM* it is signed over $[103, 200]$. The delayed equity appreciation is resolved from $h = 67$ under *HGMM* but only from $h = 119$ under *Proxy-only*, a difference of roughly two months, and the inflation-linked swap is signed over $[76, 126]$ against $[94, 121]$. The policy rate is the exception: both criteria sign it only over the first six weeks, and *Proxy-only* does so at four more horizons than *HGMM*.

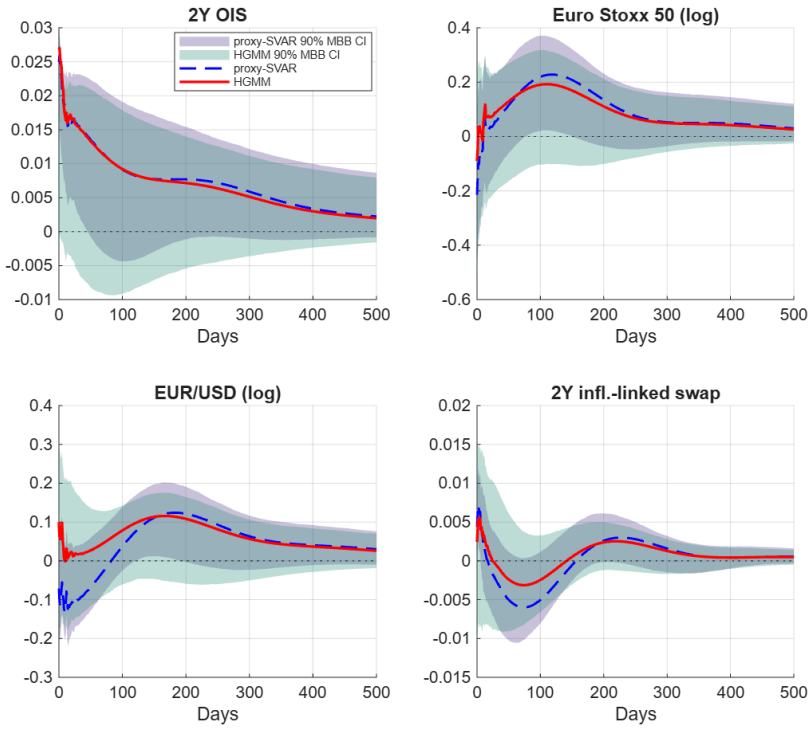
Figure H.4 reports the HGMM and proxy-SVAR responses with percentile bands. Consistent with the identification-robust sets in the main text, the hybrid and proxy-SVAR point estimates are close, with modest sharpening of confidence sets.

Figure H.5 reports the AR sets with standard MBB CIs, for alternate bootstrap block length of 35. The shorter block widens both criteria's sets and compresses the gain, from a median reduction of 26.4% to 10.2%, but the median reduction stays positive for all four variables.

	Horizons with $0 \notin$ CS		Sign-resolved horizon ranges	
Variable	Proxy	HGMM	Proxy	HGMM
2Y OIS	**33**	29	$[0, 32]$ (+)	$[0, 28]$ (+)
Euro Stoxx 50	31	**84**	$[119, 150]$ (+)	$[67, 150]$ (+)
EUR/USD	0	**98**	–	$[103, 200]$ (+)
2Y infl.-linked swap	28	**88**	$[94, 121]$ (−)	$[76, 126]$ (−), $[165, 200]$ (+)

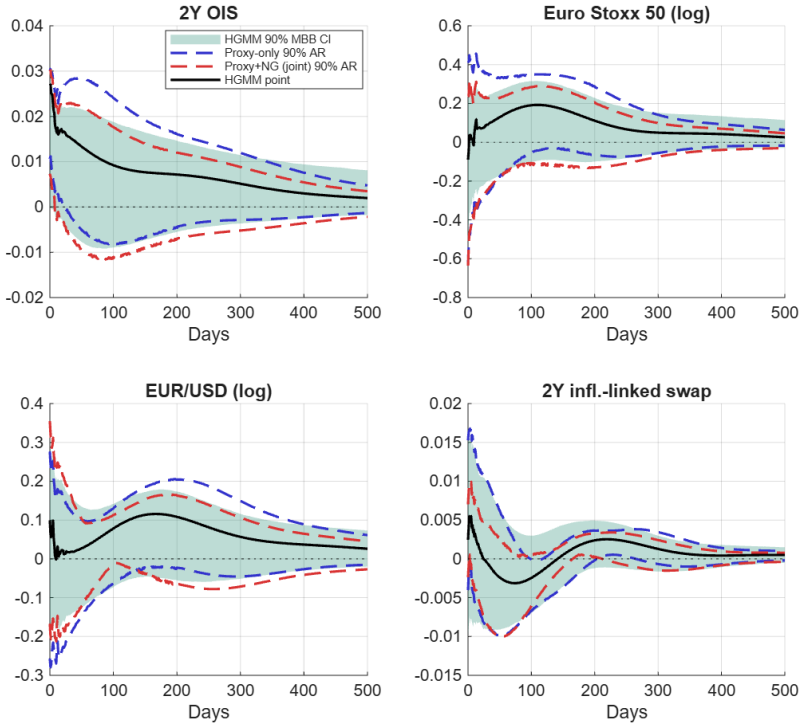
Notes: Columns 2–3 count the horizons $h \in \{0, \ldots, 200\}$ (out of 201) at which the 90% identification-robust confidence set excludes zero, so that the sign of the response is determined. Columns 4–5 report the corresponding horizon ranges, with the sign of the resolved response in parentheses; isolated horizons within a reported block are suppressed. Out of the 804 (variable, horizon) pairs, *HGMM* resolves a sign that *Proxy-only* leaves open at 211 pairs, while the reverse occurs at 4 pairs, all for the two-year OIS rate. Construction as in Table 10.

TABLE H.3. Horizons at which the sign of the response is resolved: *Proxy-only* vs *HGMM* (Altavilla monetary-policy specification)



Notes: Point impulse responses of the monetary policy shock estimated by the hybrid GMM criterion (HGMM, solid red) and by the single-proxy SVAR (proxy-SVAR, dashed blue), each with 90% residual-based moving-block-bootstrap percentile confidence intervals (shaded; HGMM in green, proxy-SVAR in purple). Both estimators are computed on every bootstrap draw from the same resampled residuals, so the two comparisons differ only in the identification step. Bootstrap: 9999 replications, block length $\ell = 50$, VAR re-estimated within each draw. Responses are to a one-standard-deviation monetary policy shock; the x-axis is the horizon in business days.

FIGURE H.4. Impulse responses to the euro-area monetary policy shock: HGMM vs proxy-SVAR, with moving-block-bootstrap percentile confidence intervals.



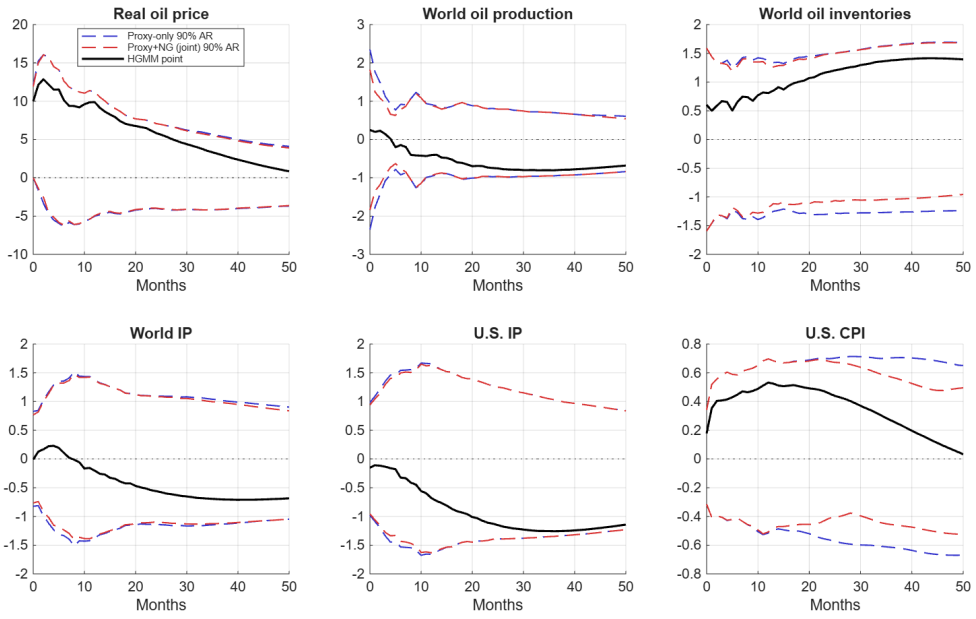
Notes: Each panel superimposes three 90% interval estimates for the impulse response of one variable to the monetary policy shock. The shaded band is the residual-based moving-block-bootstrap (MBB) percentile confidence interval for the HGMM point estimator. The two dashed lines are the identification-robust AR sets: *Proxy-only* (blue, dashed) inverts the AR statistic using only the three proxy-exclusion conditions that identify the target shock; *Proxy+NG* (red, dashed) is the HGMM criterion which additionally imposes the three co-skewness conditions of the target shock and the fourth-order aggregate, for 7 conditions in total. Critical values are from a residual-based moving-block bootstrap (9999 bootstrap replications, block length $\ell = 35$), computed per candidate rotation. The solid black line is the HGMM point estimate. Responses are to a one standard deviation contractionary monetary policy shock; the x-axis is the horizon in business days.

FIGURE H.5. 90% identification-robust confidence sets with *Proxy-only* and *HGMM* criteria, with standard MBB percentile bands (block length $\ell = 35$).

APPENDIX I. ROBUSTNESS: THE KÄNZIG BASELINE (OIL NEWS SHOCK)

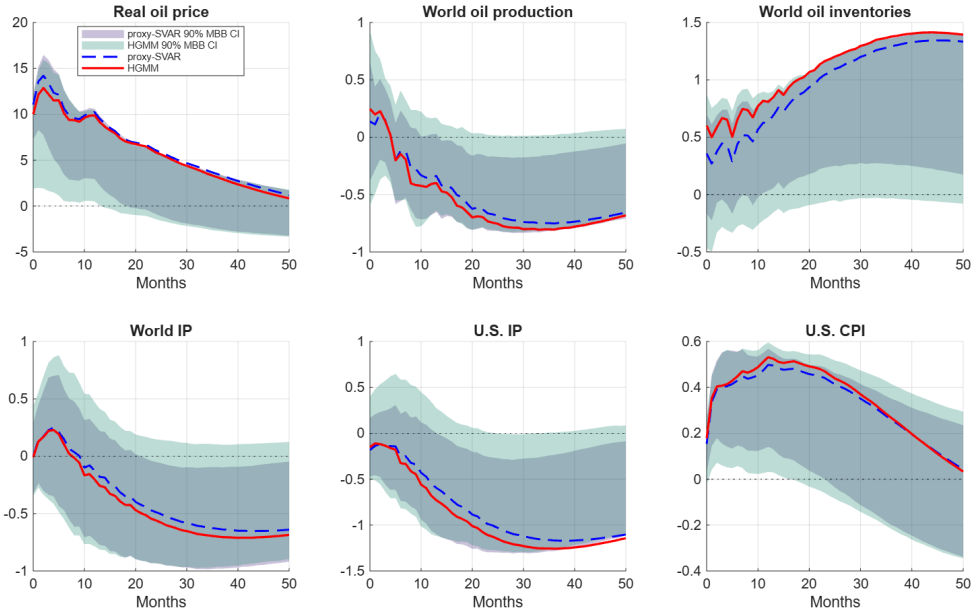
As a robustness check on the oil illustration, we re-run the analysis on the Känzig (2021) baseline: the full 1974:M1–2017:M12 sample, $p = 12$ lags, and the original OPEC-window instrument (without the Kilian corrections). Figure I.1 and Table I.1 show the same direction of effect, more modestly: across the 306 (variable, horizon) pairs the *HGMM* set is strictly shorter at 246, of equal length at 60, and longer at none; the median reduction in length is 2.3% and the largest is 27.0%. The sharpening is concentrated at short horizons for world oil production, and at long horizons for U.S. CPI, where the median reduction across horizons is 15.6%. Neither criterion determines the sign of any response at any horizon in this specification, against 212 sign-resolved horizons under the Kilian-corrected instrument. This is consistent with the theory that the payoff of sharper inference depends on the marginal identifying content of those conditions offsetting the cost of the additional degrees of freedom.

The proxy exclusion restriction is not rejected ($\hat{J}_{\text{prx}} = 1.90$, with asymptotic and bootstrap p-values of 0.86 and 0.99).



Notes: Each panel plots the 90% identification-robust Anderson–Rubin (AR) confidence set for the impulse response of one variable to the oil news shock, at horizons 0–50 months for the Känzig (2021) baseline specification (VAR sample 1974:M1–2017:M12, $p = 12$ lags, monthly OPEC-window surprise summed within the month). *Proxy-only* (blue, dashed) uses the five proxy-exogeneity conditions; *HGMM* (red, dashed) additionally imposes the target shock's co-skewness conditions and the co-kurtosis aggregate (11 conditions in total). The solid black line is the HGMM point estimate. Critical values are from a residual-based moving-block bootstrap (9999 replications, block length $\ell = 36$); responses are normalized to a 10% increase in the real oil price on impact.

FIGURE I.1. 90% Anderson–Rubin confidence sets for the oil news shock: *Proxy-only* vs *HGMM* (Känzig baseline specification).



Notes: As in Figure H.1, for the Känzig (2021) baseline specification. HGMM (solid red) and proxy-SVAR (dashed blue) with 90% moving-block-bootstrap percentile confidence intervals (shaded), computed from the same resampled residuals (9999 replications, block length $\ell = 36$, VAR re-estimated per draw). Responses are normalized to a 10% increase in the real oil price on impact.

FIGURE I.2. Impulse responses to the oil news shock: HGMM vs proxy-SVAR, with moving-block-bootstrap percentile confidence intervals (Känzig baseline specification).

	$h=0$		$h=1$		$h=2$		$h=3$		$h=4$		$h=5$	
Variable	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM
Real oil price	12.11	**11.92**	16.89	**16.39**	19.39	**18.57**	20.40	**20.27**	20.04	**19.65**	20.02	**19.91**
World oil production	4.70	**3.66**	3.53	**2.59**	2.90	**2.25**	2.20	**1.86**	1.85	**1.39**	1.55	**1.26**
World oil inventories	3.18	3.18	2.90	2.90	2.66	**2.64**	2.65	2.65	2.76	**2.67**	2.47	**2.37**
World IP	1.65	**1.53**	1.66	**1.56**	2.02	**1.92**	2.28	**2.15**	2.52	**2.44**	2.66	**2.50**
U.S. IP	1.96	**1.89**	2.19	**2.11**	2.42	**2.30**	2.69	**2.57**	2.91	**2.74**	2.91	**2.75**
U.S. CPI	0.66	0.66	0.93	0.93	0.95	0.95	0.99	0.99	1.03	1.03	1.01	1.01

Notes: Each entry is the length (upper minus lower bound) of the 90% identification-robust Anderson–Rubin confidence set for the impulse response of the row variable to the target shock at horizon h (months), on the Känzig (2021) baseline: the full 1974:M1–2017:M12 sample, $p = 12$ lags, and the original OPEC-window instrument. *Proxy* inverts the AR statistic using only the five proxy exclusion conditions that identify the target shock; *HGMM* (Proxy+NG) additionally imposes the target shock's five co-skewness conditions and the co-kurtosis aggregate, for 11 conditions in total. The smaller length in each (variable, horizon) pair is in **bold**. Responses are normalized to a 10% increase in the real oil price on impact. Sets are obtained by inverting the AR statistic over the target-shock rotation subspace with per-candidate moving-block-bootstrap critical values; see the construction in Appendix G.

TABLE I.1. Length comparison of 90% Anderson–Rubin confidence sets: *Proxy-only* vs *HGMM* (Känzig baseline specification)

	Ratio at horizon h						Over $h = 0, \dots, 50$	
Variable	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	Median	Max
Real oil price	1.02	1.03	1.04	1.01	1.02	1.01	1.01	1.04
World oil production	1.28	1.36	1.29	1.18	1.33	1.23	1.00	1.36
World oil inventories	1.00	1.00	1.01	1.00	1.03	1.04	1.09	1.11
World IP	1.08	1.06	1.05	1.06	1.03	1.06	1.03	1.08
U.S. IP	1.04	1.04	1.05	1.05	1.06	1.06	1.00	1.06
U.S. CPI	1.00	1.00	1.00	1.00	1.00	1.00	1.18	1.37

Notes: Each entry is the ratio len(Proxy)/len(HGMM) for the impulse response of the row variable to the target shock. Values above one indicate that the *Proxy-only* set is longer. The last two columns summarize the ratio over all 51 horizons $h = 0, \dots, 50$. Construction as in Table I.1.

TABLE I.2. Ratio of 90% Anderson–Rubin confidence-set lengths: *Proxy-only* / *HGMM* (Känzig baseline specification)

APPENDIX J. ROBUSTNESS: IDENTIFICATION OF A MONETARY POLICY SHOCK, GERTLER–KARADI (2015)

As a robustness check on the monetary policy illustration, we apply the hybrid GMM framework to the Gertler and Karadi (2015) system. This is a case in which the higher-order moment conditions carry too little signal to contribute for the degrees of freedom they add, and we report it because it shows the criterion does not invent identifying content in such cases.

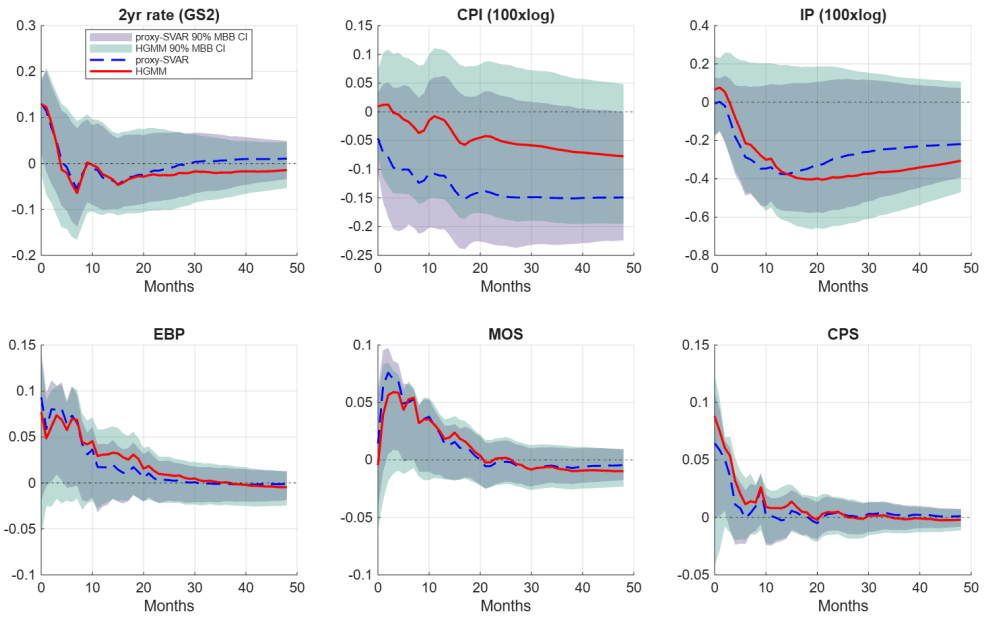
J.1. **Specification.** The specification contains six variables ordered with the policy indicator first: the two-year Treasury rate[19] (the policy indicator), the consumer price index and industrial production (both in $100 \times$ log levels), the excess bond premium, and two credit spreads. The reduced-form VAR is estimated at monthly frequency with $p = 12$ lags on 1979:M7–2012:M6. The instrument is the high-frequency fed-funds futures surprise (FF4), available on 1991:M1–2012:M6; with $n = 6$ and $k = 1$ this gives $p_\phi = 15$ rotation angles, $q_P = 5$ proxy conditions and $q_{NG} = 75$ higher-order moment conditions, for $q = 80$ in total. Inference is at the nominal 90% confidence level using a residual-based MBB with block length $\ell = 15$ and 9999 bootstrap replications (VAR re-estimated within each draw); the identification-robust sets invert the AR statistic over the target-shock subspace at 10^5 candidate rotations. Impulse responses are reported for a one-standard-deviation monetary policy shock.

J.2. **Results.** Figure J.2 and Table J.1 report the identification-robust confidence sets. In contrast to the oil applications, the *Proxy-only* sets are not improved on, and adding the higher-order moment conditions does not narrow them; across the 294 (variable, horizon) pairs the *HGMM* set is shorter at 46, of equal length at 47, and longer at 201; the median length change is +2.2% and the maximum ratio anywhere is 1.04. Neither criterion determines the sign of any response at any horizon. Again, this is consistent as when the non-Gaussianity of the shocks does not translate into robust sample-wide moment restrictions,

[19] In one of the specifications of the original paper the authors use the one-year Treasury rate as the policy indicator; the choice is not consequential for the relative (non-)informativeness of the higher-order moment conditions.

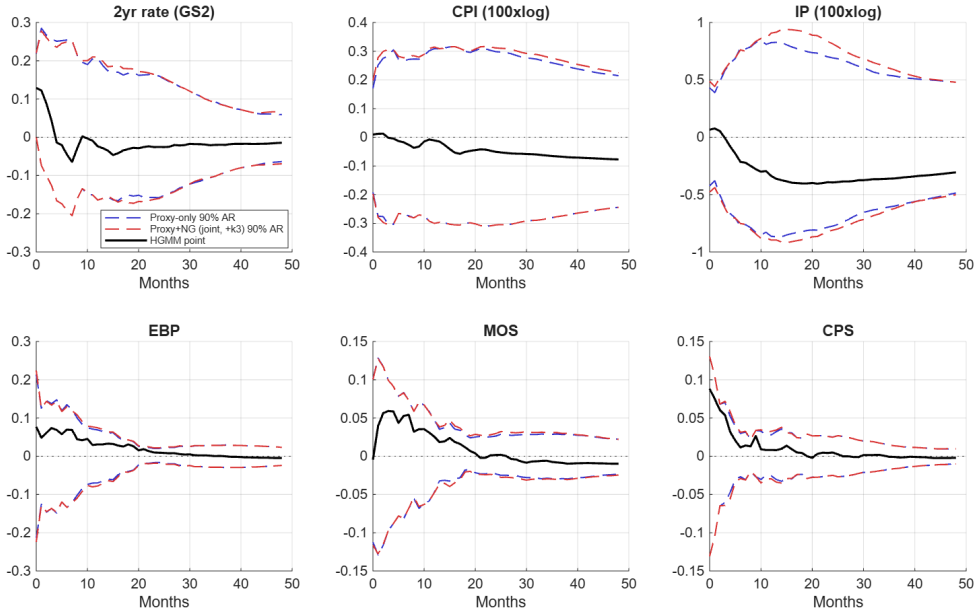
the extra conditions add degrees of freedom to the AR statistic without a corresponding increase in identifying signal, so the joint set cannot improve on the proxy-only set. We regard this as reassuring. The hybrid criterion tightens the confidence sets in the oil and euro-area applications, where the higher-order moment conditions are informative. A criterion that narrowed the sets in every application regardless of the data would be the worrying outcome.

The point responses remain those of a contractionary monetary policy shock and are similar across the hybrid and proxy-SVAR estimators (Figure J.1): the policy rate rises on impact, industrial production and the price level decline, and the excess bond premium and credit spreads widen, consistent with Gertler and Karadi (2015). The specification test does not reject the proxy exclusion restriction ($\hat{J}_{\text{prx}} = 3.09$, with asymptotic and bootstrap p-values of 0.68 and 0.99).



Notes: As in Figure H.1, for the Gertler–Karadi (2015) system. HGMM (solid red) and proxy-SVAR (dashed blue) impulse responses to a one-standard-deviation monetary policy shock, with 90% moving-block-bootstrap percentile confidence intervals (shaded), computed from the same resampled residuals (9999 replications, block length $\ell = 15$, VAR re-estimated per draw). The x-axis is the horizon in months.

FIGURE J.1. Impulse responses to the monetary policy shock: HGMM vs proxy-SVAR, with moving-block-bootstrap percentile confidence intervals.



Notes: 90% identification-robust Anderson–Rubin confidence sets for the impulse responses to the monetary policy shock in the Gertler–Karadi (2015) system (monthly VAR, $p = 12$, 1979:M7–2012:M6; FF4 high-frequency surprise, 1991:M1–2012:M6). *Proxy-only* (blue, dashed) uses the five proxy orthogonality conditions; *HGMM* (red, dashed) adds the co-skewness, asymmetric co-kurtosis, and co-kurtosis aggregate conditions of the target shock. The solid black line is the HGMM point estimate. Critical values are from a residual-based moving-block bootstrap (9999 replications, block length $\ell = 15$). Responses are to a one-standard-deviation monetary policy shock; the x-axis is the horizon in months.

FIGURE J.2. 90% Anderson–Rubin confidence sets for the monetary policy shock: *Proxy-only* vs *HGMM* (Gertler–Karadi application).

	$h=0$		$h=1$		$h=2$		$h=3$		$h=4$		$h=5$	
Variable	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM	Proxy	HGMM
2yr rate (GS2)	0.22	0.22	0.36	**0.35**	0.37	**0.36**	0.38	**0.37**	0.42	**0.40**	0.43	**0.42**
CPI (100xlog)	**0.36**	0.40	**0.53**	0.56	**0.55**	0.60	**0.59**	0.61	0.61	0.61	**0.54**	0.55
IP (100xlog)	**0.85**	0.97	**0.77**	0.88	**0.99**	1.04	**1.19**	1.21	**1.31**	1.35	**1.39**	1.40
EBP	**0.43**	0.45	0.25	0.25	0.29	0.29	0.27	0.27	0.30	**0.29**	0.24	0.24
MOS	**0.21**	0.22	0.26	0.26	0.23	0.23	0.20	0.20	0.18	0.18	0.16	0.16
CPS	0.26	0.26	0.21	0.21	0.13	0.13	**0.13**	0.14	**0.10**	0.11	0.07	0.07

Notes: Each entry is the length (upper minus lower bound) of the 90% identification-robust Anderson–Rubin confidence set for the impulse response of the row variable to the target shock at horizon h (months). *Proxy* inverts the AR statistic using only the five proxy exclusion conditions that identify the target shock; *HGMM* (Proxy+NG) additionally imposes the target shock's five co-skewness conditions, five asymmetric co-kurtosis conditions, and the co-kurtosis aggregate, for 16 conditions in total. The smaller length in each (variable, horizon) pair is in **bold**. Responses are to a one-standard-deviation monetary policy shock. Sets are obtained by inverting the AR statistic over the target-shock rotation subspace with per-candidate moving-block-bootstrap critical values; see the construction in Appendix G.

TABLE J.1. Length comparison of 90% Anderson–Rubin confidence sets: *Proxy-only* vs *HGMM* (Gertler–Karadi application)

	Ratio at horizon h						Over $h = 0, \dots, 48$	
Variable	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	Median	Max
2yr rate (GS2)	1.00	1.02	1.02	1.03	1.04	1.02	0.99	1.04
CPI (100xlog)	0.92	0.95	0.93	0.96	1.00	0.98	0.98	1.00
IP (100xlog)	0.88	0.87	0.95	0.98	0.97	1.00	0.94	1.03
EBP	0.95	0.99	1.01	1.01	1.02	1.01	0.99	1.02
MOS	0.99	1.01	1.00	1.00	1.00	1.00	0.96	1.02
CPS	1.00	1.00	1.00	0.96	0.87	0.93	1.00	1.02

Notes: Each entry is the ratio len(Proxy)/len(HGMM) for the impulse response of the row variable to the target shock. Values *below* one indicate that the *HGMM* set is the longer of the two. The last two columns summarize the ratio over all 49 horizons $h = 0, \dots, 48$. Construction as in Table J.1.

TABLE J.2. Ratio of 90% Anderson–Rubin confidence-set lengths: *Proxy-only* / *HGMM* (Gertler–Karadi application)

APPENDIX K. EVIDENCE OF WEAK INSTRUMENTS AND NON-GAUSSIANITY IN THE MACROECONOMIC LITERATURE

Study	Identified shock (instrument)	F	Robust/eff. F	Source of statistic
Gertler and Karadi (2015)	Monetary policy (HF FF4 surprise)	21.5	17.5	own first stage (one-year rate)
same instrument, LP-IV, no controls	Monetary policy	1.7	—	Stock and Watson (2018), Tab. 1[a]
Montiel Olea et al. (2021)	Oil supply (illustration)	—	9.4	robust stat. < 10; $\xi_1 = 4.4$
Känzig (2021)	Oil supply news (OPEC HF surprise)	22.7	10.6	own first stage
Romer and Romer (2010)	Tax (narrative)	1.6	—	Ramey (2016)[b]
Government-spending proxies[c]	Fiscal (narrative / HF)	< 10	—	Ramey (2016), p. 120

Notes: "F" is the standard (homoskedastic) first-stage F; "Robust/eff. F" is the heteroskedasticity-robust or Montiel Olea et al. (2021)-type effective first-stage statistic. The Stock and Yogo (2002) 10% critical value is ≈ 10.3; the common rule of thumb is $F > 10$. Leading instruments are strong in their original specification, but relevance is fragile under robust statistics or alternative specifications. [a] Rises to $F = 23.7$ once four lags of Y_t and Z_t are included as controls. [b] First stage of tax revenue on the Romer and Romer (2010) narrative shocks. [c] Fisher and Peters (2010), Ben Zeev & Pappa (2017), and Barth III and Ramey (2000); see Ramey (2016), p. 120, Fig. 4 Panel B.

TABLE K.1. Instrument relevance in leading proxy-SVARs

Study	Series / shocks	Non-Gaussianity measure (estimate)
Lanne et al. (2017)	Macro SVAR structural errors	Gaussianity rejected; ML favors heavy-tailed errors
Gouriéroux et al. (2017)	SVAR shocks (ICA)	Components significantly non-Gaussian (≤ 1 Gaussian)
Bekaert et al. (2021)	Aggregate supply & demand shocks	Demand shocks negatively skewed, leptokurtic; higher moments sig.
Salgado et al. (2019)	Output / employment / sales growth	Strongly procyclical skewness; left tail widens in recessions
Fagiolo et al. (2008)	OECD output growth (US GDP, IP)	Laplace tails (Subbotin $b \approx 1$); excess kurtosis ≈ 3
Guvenen et al. (2021)	Labor-income shocks	Left-skewed; kurtosis up to ≈ 18 (vs 3 Gaussian)

*Notes:*The first three rows concern structural SVAR shocks directly; the remainder document non-Gaussianity at the aggregate-output and labor-income level. A Gaussian benchmark has skewness 0 and kurtosis 3.

TABLE K.2. non-Gaussianity in macroeconomic and financial shocks