Some Pitfalls of Instrument-Based Inference in Structural VARs*

Christian Matthes† Vladimir Menshikov‡ Todd B. Walker§

July 2026

Abstract

Structural VARs are often identified by using instruments derived from the residuals of auxiliary regressions (e.g., Romer and Romer (2004)). We derive analytical characterizations of the instrument's correlation with the true shocks and evaluate the procedure's finite-sample performance using Monte Carlo experiments based on the model of Smets and Wouters (2007). We find that such instruments are meaningfully correlated not only with the monetary policy innovation, but also with other structural shocks, leading to substantial biases and variation in estimated impulse responses. We then examine several proposals from the literature designed to mitigate VAR issues and propose a correction based on instruments for other structural shocks.

*We thank Christiane Baumeister for helpful comments.

†Department of Economics, University of Notre Dame, cmatthes@nd.edu

‡Department of Economics, Indiana University, vlmenshi@iu.edu

§Department of Economics, Indiana University, walkertb@indiana.edu

1 Introduction

Identification of causal effects in macroeconomics is a daunting task. Structural Vector Autoregressions (SVARs) have been used as a primary tool to infer these causal effects since Sims (1980), but the choice of identification assumptions remains an open question. Over the last decade the use of instruments for structural shocks (Mertens and Ravn, 2013; Stock and Watson, 2018; Caldara and Herbst, 2019) has become arguably the most used identification approach, allowing transparent discussions of the underlying identification assumptions. A substantial subset of macroeconomic shock measures and instruments are constructed as residuals of auxiliary regressions or closely related prediction exercises that remove the predictable component of policy actions (Romer and Romer, 2004; Miranda-Agrippino and Ricco, 2021; Aruoba and Drechsel, forthcoming). In this paper, we derive theoretical results on the performance of such a strategy and use the well-known Smets and Wouters (2007) model as a data-generating process to assess the performance of such an identification strategy in practice. In particular, we estimate a monetary policy rule and treat the associated residual as our monetary policy instrument, broadly following the strategy of Romer and Romer (2004); Miranda-Agrippino and Ricco (2021). We then use the internal instrument approach proposed in Plagborg-Møller and Wolf (2021). We compare specifications that differ in the variables included in the VAR (Canova and Ferroni (2022) suggest that the performance of SVARs depends crucially on the size of the VAR), in the number of lags (De Graeve and Westermark (2025) propose to use a large number of lags to improve the performance of SVARs), and in the choice of whether to whiten residuals. We also study the effects of changing the sample size and the role that possible endogeneity in the auxiliary regressions plays. Finally, we suggest a correction that is feasible if reliable proxies for other structural shocks are available.

The same residualization logic appears outside monetary policy. In fiscal applications, Auerbach and Gorodnichenko (2013) construct government-spending innovations by using real-time forecast data to purge policy innovations of predictable components, and Barattieri et al. (2023) estimate sectoral government-spending shocks by residualizing disaggregated procurement changes on industry, aggregate, and expectation controls. In the analysis of trade policy, Barattieri and Cacciatore (2023) purge temporary-trade-barrier measures of past, current, and expected industry conditions and use the residuals as trade-policy shocks. These examples make clear that our analysis speaks to a broader empirical practice: an auxiliary equation is used to isolate an unanticipated policy component, and the estimated residual is then treated as a structural shock, proxy, or instrument in downstream impulse-response analysis.

The closest methodological comparison is Lloyd and Manuel (2024), who study the effects of the common two-step approach where the instrument is first estimated in a separate regression versus a one-step approach. Their contribution is a comparison of estimators: characterizing equivalence, omitted-variable bias, and inference. We ask a different, complementary question. In contrast to our work, they take the residual from the first-stage approach as their shock of interest, whereas we are instead interested in how such a residual is related to the true structural shock of interest. In our Monte Carlo design, the monetary policy shock is known, so we can assess whether an auxiliary policy-rule residual is a valid proxy for that shock, for example when used in an internal-instrument VAR. We show that a bias arises naturally in finite samples, as the residual will generally be a function of all structural shocks present in the data-generating process. This bias is exacerbated by endogeneity issues that are likely present in most macroeconomic applications, whereas in the absence of endogeneity we find no bias, echoing the literature on generated regressors (Pagan, 1984). The key insight is that residualization makes the instrument orthogonal to included observables in sample, not necessarily to structural shocks. This issue is not specific to the internal-instrument VAR used in our application. A local-projection IV (LP-IV) specification that uses the same generated residual as an instrument inherits the same properties: if the residual loads on non-target shocks, the estimated responses trace a composite innovation rather than the intended shock alone. We quantify this contamination and its consequences for impulse-response bias and dispersion.

Our paper is also related to the literature on the relationship between Dynamic Stochastic General Equilibrium (DSGE) models and VARs. Giacomini (2013) provides an overview of the conceptual links between the two approaches. A central lesson from this literature is that VAR-based impulse responses need not recover the structural dynamics of a DSGE: fundamentalness and invertibility conditions matter (Fernández-Villaverde et al., 2007), finite-order VARs may be imperfect approximations to DSGE reduced forms (Ravenna, 2007), and finite samples can generate substantial distortions in DSGE-based Monte Carlo experiments (Erceg et al., 2005; Chari et al., 2008; Christiano et al., 2007). We view our analysis as complementary to this literature. Rather than asking whether a finite-order VAR approximates the DSGE reduced form well, we ask whether a residual-based proxy constructed from an auxiliary policy rule represents the intended structural shock. This distinction matters because instrument contamination and VAR approximation error are separate channels through which DSGE-based impulse responses and VAR-based estimates can differ.

2 OUR SETTING

We study impulse responses estimated via vector autogressions of the following form

$$Y_t = \begin{bmatrix} \widehat{r}_t \\ Z_t \end{bmatrix} = \sum_{\ell=1}^{L} A_\ell Y_{t-\ell} + u_t, \qquad u_t = \begin{bmatrix} u_t^{(r)} \\ u_t^{(Z)} \end{bmatrix}, \qquad E[u_t u_t^\intercal] = \Sigma_u \tag{1}$$

where the forecast error u_t has mean zero and covariance matrix Σ_u. For simplicity, we use demeaned data in our Monte Carlo simulations and thus abstract from an intercept in our VARs. We assume the first element of the $(N+1)$-dimensional vector Y_t, which we denote by $\widehat{r}_t$, is an instrument for the structural shocks of interest, whereas the other elements Z_t are macroeconomic variables. Following Plagborg-Møller and Wolf (2021), we identify impulse responses of interest via a recursive identification scheme. Specifically, in a recursive (triangular) SVAR with the ordering $\left[\widehat{r}_t,\ Z_t^\intercal\right]^\intercal$, the impact (contemporaneous) matrix is $\Omega = \texttt{chol}(\Sigma_u)$, the unique lower-triangular Cholesky decomposition satisfying $\Sigma_u = \Omega\Omega^\intercal$. This restriction implies that the first reduced-form innovation $u_t^{(r)}$ is driven only by the first orthonormal structural shock ε_t^1, whereas $u_t^{(Z)}$ may load on all structural shocks. Let $\{C_h\}_{h\geq 0}$ denote the reduced-form moving-average (MA) coefficients of the VAR, defined by $C_0 = I$ and $C_h = \sum_{\ell=1}^{L} A_\ell C_{h-\ell}$ for $h \geq 1$, then the impulse response of Y at horizon h to a one-unit innovation in ε_t^1 is $\Psi(h) = C_h \Omega e_1$, where e_1 is the first basis vector. Equivalently,

$$\frac{\operatorname{Cov}\left(Y_{t+h}, u_t^{(r)}\right)}{\operatorname{Var}\left(u_t^{(r)}\right)} = \frac{C_h \Sigma_u e_1}{e_1^\top \Sigma_u e_1} = \frac{1}{\omega_{11}} C_h \Omega e_1 \tag{2}$$

where $\omega_{11} := [\Omega]_{11}$. Hence, the impulse response to the shock of interest is identified up to a constant scalar $(1/\omega_{11})$. This scalar factor of proportionality is due to the potential presence of independent measurement error in the instrument in general settings. We discuss later exactly what normalization we use for the impulse responses in our experiments.

As emphasized by Plagborg-Møller and Wolf (2021), this identification relies on contemporaneous covariance restrictions and reduced-form dynamics; it does not require the structural shock of interest to be invertible (i.e., recoverable from present and past reduced-form innovations) – in fact, the presence of any measurement error rules out invertibility. This means that contemporaneous impulse responses are given by $\Omega = \texttt{chol}(\Sigma_u)$ up to the aforementioned normalization and given estimates of Σ_u and $A_1, ..., A_L$, we can then estimate the impulse responses of interest.

We assume the true data-generating process is a dynamic stochastic, general equilibrium (DSGE) model. We then ask whether the VAR methodology can recover impulse responses from data simulated from this DSGE. The instrument is generated via

$$i_t = B^\top Z_t + \varepsilon_t^m \tag{3}$$

where i_t is a variable partially determined by the true shock of interest ε_t^m. If $B = 0$, we observe the shock directly. In our application, this equation will stand in for a monetary policy rule and i_t will be the nominal interest rate. Because, in the DSGE, Z_t generally depends contemporaneously on ε_t^m (see Section 3), estimating (3) by OLS faces an endogeneity problem. However, Carvalho et al. (2021) show that, if the monetary policy shock is not a very important contributor to fluctuations in Z_t, OLS-based estimate of monetary policy rule coefficients outperform standard instrumental variable-based approaches. They verify this with Monte Carlo experiments using the *same* data generating process that we use below–the Smets and Wouters (2007) model.[1] Importantly (for our later discussion), they find that impulse responses using residuals of policy rules based on either GMM or OLS estimates are almost identical. This finding is relevant for our work because we study scenarios where we do not only need to estimate the parameters of the VAR in Equation (1), but also the parameters in Equation (3).

As a first step, it will be useful to define the observed/estimated instrument as follows. Given any estimate $\widehat{B} \in \mathbb{R}^N$, define the *residual-instrument* as

$$\widehat{r}_t(\widehat{B}) := i_t - \widehat{B}^\top Z_t = (B - \widehat{B})^\top Z_t + \varepsilon_t^m \tag{4}$$

[1] Carvalho et al. (2021) state that "In sum, using the Smets and Wouters (2007) model as a laboratory, we find the OLS estimation bias to be small. More importantly, OLS estimates imply model dynamics that are remarkably close to the true ones, with higher precision than dynamics implied by GMM estimates."

This decomposition makes clear that any deviation $\widehat{B} \neq B$ produces contamination: $\widehat{r}_t(\widehat{B}) \neq \varepsilon_t^m$ and, in general, $\widehat{r}_t(\widehat{B})$ is correlated with *other* structural shocks as well. There are two distinct sources: (i) *Endogeneity Bias:* if the estimator targets $B+\delta \neq B$ in population (as OLS does when Z_t is contemporaneously correlated with ε_t^m), then even as $T \to \infty$ the pseudo-instrument converges to $r_t^\star = \varepsilon_t^m - \delta' Z_t$, a mixture of structural shocks. (ii) *Sampling Error:* even for an estimator consistent for B (e.g., a valid IV), finite-sample noise in $\widehat{B}$ leaves a $(B-\widehat{B})^\top Z_t$ term that induces contamination at $O_p(T^{-1/2})$ rates. In either case, the residual-instrument is a noisy proxy for the target shock. We quantify its population correlation with ε_t^m and its induced correlations with other shocks in Section 3.

3 THEORETICAL RESULTS

This section develops a set of theoretical results that clarify how closely the residual-based instrument, $\widehat{r}_t(\widehat{B})$, tracks the true structural shock, ε_t^m. We first study the general case for an arbitrary estimate $(\widehat{B})$ then specialize to the OLS case to quantify both asymptotic and finite-sample correlations. Our goal is to isolate sources of instrument contamination (e.g., sampling noise, endogeneity) without assuming a particular data-generating mechanism beyond linearity and (weak) stationarity.

To be as agnostic as possible, we assume the following:

Assumption 1. *The data, Z_t, is generated by a (possibly) infinite order moving average*

$$i_t \;=\; B^\top Z_t \;+\; \varepsilon_t^m, \qquad Z_t \;=\; \sum_{\ell=0}^{\infty} H_\ell\, \varepsilon_{t-\ell}, \qquad t = 1,\dots,T \tag{5}$$

with $Z_t \in \mathbb{R}^N$, $\varepsilon_t = (\varepsilon_t^1,\dots,\varepsilon_t^N)^\top \in \mathbb{R}^N$, and fixed $m \in \{1,\dots,N\}$. Innovations ε_t are i.i.d. with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t\varepsilon_t^\top] = \Sigma_{\varepsilon\varepsilon} = \operatorname{diag}(\sigma_1^2,\dots,\sigma_N^2) \succ 0$. MA coefficients satisfy $\sum_{\ell=0}^{\infty} \|H_\ell\| < \infty$.

Under these conditions, Z_t is strictly stationary and ergodic, square integrable with

$$\Sigma_{ZZ} := \operatorname{Var}(Z_t) = \sum_{\ell \ge 0} H_\ell \Sigma_{\varepsilon\varepsilon} H_\ell^\top \succ 0, \qquad \Sigma_{Z\varepsilon^m} := \operatorname{Cov}(Z_t, \varepsilon_t^m) = \sigma_m^2\, h_{0,m},$$

where $h_{0,m}$ is the mth column of H_0.
We can compute second moments of our residual instrument $\widehat{r}_t(\widehat{B})$ for any given conformable matrix $\widehat{B}$:

$$\operatorname{Cov}\big(\widehat{r}_t(\widehat{B}), Z_t\big) = \Sigma_{ZZ}(B-\widehat{B}) + \Sigma_{Z\varepsilon^m} = \Sigma_{ZZ}(B-\widehat{B}) + \sigma_m^2 h_{0,m} \tag{6}$$

$$\operatorname{Var}\big(\widehat{r}_t(\widehat{B})\big) = (B-\widehat{B})^\top \Sigma_{ZZ}(B-\widehat{B}) + \sigma_m^2 + 2(B-\widehat{B})^\top \Sigma_{Z\varepsilon^m} \tag{7}$$

Equation (6) makes clear that unless $\widehat{B}$ solves the equation $\Sigma_{ZZ}(B-\widehat{B}) + \sigma_m^2 h_{0,m} = 0$, the residual remains correlated with Z_t. Note that the first term inducing a correlation, $\Sigma_{ZZ}(B-\widehat{B})$, will be present even if the covariance between Z_t and ε_t^m is 0; that is, even if the structural shock is exogenous to Z_t in population (so that $\Sigma_{Z\varepsilon^m} = 0$), any estimation error in B generates a correlation term $\Sigma_{ZZ}(B-\widehat{B})$. When $\Sigma_{Z\varepsilon^m} \neq 0$, this bias is compounded by the intrinsic endogeneity of the policy rule itself, producing residuals that conflate the target shock with contemporaneous responses of macroeconomic variables.

To concisely state our results, we stack the sample $\{Z_t\}_{t=1}^T$ into the matrix $Z = [Z_1, \dots, Z_T]^\intercal \in \mathbb{R}^{T\times N}$, and likewise define the vectors $i = (i_1, \dots, i_T)^\intercal$ and $\varepsilon^m = (\varepsilon_1^m, \dots, \varepsilon_T^m)^\intercal$. With this notation, the OLS estimate of the coefficient vector B and the usual projection and annihilator matrices are

$$\widehat{B}_{\text{OLS}} = (Z^\intercal Z)^{-1} Z^\intercal i, \qquad P_Z = Z(Z^\intercal Z)^{-1} Z^\intercal, \qquad M_Z = I_T - P_Z$$

Proposition 1 (Finite-sample identities)**.** *For any realized sample, we have*

$$\widehat{B}_{\text{OLS}} = B + (Z^\intercal Z)^{-1} Z^\intercal \varepsilon^m \tag{8}$$

$$\widehat{r}(\widehat{B}_{\text{OLS}}) = i - Z\widehat{B}_{\text{OLS}} = M_Z\, \varepsilon^m \tag{9}$$

Consequently $Z^\intercal \widehat{r}(\widehat{B}_{\text{OLS}}) = 0$ *exactly.*

Proof. Immediate from $\widehat{B}_{\text{OLS}} = (Z^\intercal Z)^{-1} Z^\intercal (ZB + \varepsilon^m)$ and $I_T - P_Z = M_Z$. □

These results make clear that the residual instrument will always be uncorrelated with the variables Z_t, even if there is endogeneity, i.e. if Z_t is correlated with the true shock ε_t^m. The identities (8)–(9) imply that, with $\widehat{B}_{\text{OLS}}$, the residual instrument is orthogonal (in sample) to the variables, i.e. $Z^\intercal \widehat{r}(\widehat{B}_{\text{OLS}}) = 0$ holds exactly for the realized sample. This is a property of OLS residuals and does not assert zero *population* covariance between $\widehat{r}_t$ and Z_t; in particular, endogeneity can persist in population even though sample orthogonality holds by construction.

3.1 ASYMPTOTIC RESULTS

We next characterize the large-sample behavior of the residual-shock relationship with N fixed and as $T \to \infty$. Absent independent measurement error in the instrument, any population correlation between the instrument and the other (non-targeted) shocks arises from *endogeneity* (i.e., from $\Sigma_{Z\varepsilon^m} \neq 0$), in contrast to the finite-sample distortions.

Proposition 2 (Population projection)**.** *Under Assumption 1 and assuming the ergodic LLN for linear processes,*

$$\widehat{B}_{\text{OLS}} \xrightarrow{p} B + \delta, \qquad \delta := \Sigma_{ZZ}^{-1} \Sigma_{Z\varepsilon^m} = \sigma_m^2\, \Sigma_{ZZ}^{-1} h_{0,m}, \tag{10}$$

$$\widehat{r}_t(\widehat{B}_{\text{OLS}}) \xrightarrow{p} r_t^\star := \varepsilon_t^m - \delta^\intercal Z_t. \tag{11}$$

Moreover,

$$\text{Var}(r_t^\star) = \sigma_m^2 - \Sigma_{Z\varepsilon^m}^\intercal \Sigma_{ZZ}^{-1} \Sigma_{Z\varepsilon^m} = \sigma_m^2 (1 - R_m^2), \tag{12}$$

$$R_m^2 := \frac{\Sigma_{Z\varepsilon^m}^\intercal \Sigma_{ZZ}^{-1} \Sigma_{Z\varepsilon^m}}{\sigma_m^2} = \sigma_m^2\, h_{0,m}^\intercal \Sigma_{ZZ}^{-1} h_{0,m} \in [0,1], \tag{13}$$

$$\text{Cov}(r_t^\star, \varepsilon_t^m) = \text{Var}(r_t^\star), \qquad \text{Corr}(r_t^\star, \varepsilon_t^m) = \sqrt{1 - R_m^2}. \tag{14}$$

Proof. By ergodicity, $T^{-1} Z^\intercal Z \to \Sigma_{ZZ}$ and $T^{-1} Z^\intercal \varepsilon^m \to \Sigma_{Z\varepsilon^m}$; see, e.g., Brockwell and Davis (1991, Ch. 7) or Hamilton (1994, Ch. 3). Then (10) follows by continuous mapping. Next, (11) follows from the definition of $\hat{r}_t(\hat{B}_{\text{OLS}})$ and (10), again by continuous mapping. The variance expression (12) is the Schur complement in the covariance matrix of $(Z_t^\top, \varepsilon_t^m)^\top$, as we show in the appendix. Identities (13)–(14) are immediate from (12). □

The variable R_m^2 is the R^2 of the linear projection of the shock ε_t^m on the macro variables Z_t

$$R_m^2 = \frac{\text{Var}(\delta^\top Z_t)}{\text{Var}(\varepsilon_t^m)}, \qquad \delta = \Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m}$$

It measures the contemporaneous predictability of the target shock from Z_t: $R_m^2 = 0$ means ε_t^m is orthogonal to Z_t, whereas $R_m^2 = 1$ means ε_t^m lies in the span of Z_t almost surely. Only the contemporaneous coefficient $h_{0,m}$ enters $\Sigma_{Z\varepsilon^m} = \sigma_m^2 h_{0,m}$, which partially determines δ and R_m^2; lag coefficients $\{H_\ell : \ell \geq 1\}$ affect Σ_{ZZ} but not $\Sigma_{Z\varepsilon^m}$. We state this explicitly as a corollary to Proposition 2:

Corollary 1 (Degeneracy and exact recovery)**.** *We have $r_t^\star \equiv 0$ iff $R_m^2 = 1$, i.e., if and only if ε_t^m lies almost surely in the span of Z_t. Conversely, $r_t^\star = \varepsilon_t^m$ iff $R_m^2 = 0$, which here is equivalent to $h_{0,m} = 0$.*

3.2 FINITE-SAMPLE DISTRIBUTION OF $\hat{\rho}_T$

We now study the finite-sample correlation between the true shock ε_t^m and the residual-based instrument $\hat{r}_t$. Two cases permit clean results: First, when there is *no endogeneity* ($R_m^2 = 0$), (Z, ε^m) are independent, and OLS estimates yield a tractable distribution. As we discuss below, this is not the most interesting scenario, but provides a tight bound. Second, when endogeneity is present ($R_m^2 > 0$), we retain closed-form characterizations under stronger assumptions (Gaussianity and i.i.d. data). The results we present in this section highlight forces that, when such instruments are used in VARs, interact with the well-known finite sample bias in VARs (Lütkepohl, 2005). Our Monte Carlo results in Section 4 show the combined impact of both of these issues.

Define, as before,

$$\hat{r} = M_Z \varepsilon^m, \qquad \hat{\rho}_T := \text{Corr}(\hat{r}, \varepsilon^m) = \frac{\hat{r}^\top \varepsilon^m}{\|\hat{r}\|\,\|\varepsilon^m\|} = \sqrt{\frac{(\varepsilon^m)^\top M_Z \varepsilon^m}{(\varepsilon^m)^\top \varepsilon^m}}$$

CASE I: $R_m^2 = 0$ (NO POPULATION PREDICTABILITY)

Proposition 3 (Distribution of the finite sample correlation)**.** *If the Z_t process excludes the m-th shock at all lags, i.e. $H_\ell e_m \equiv 0$ for all $\ell \geq 0$,*[2] *then (Z, ε^m) are independent (by independence across components and time), and with ε^m following a spherical distribution (Gaussian, for example)*[3]

$$\hat{\rho}_T^2 \sim \text{Beta}\left(\frac{T-N}{2}, \frac{N}{2}\right), \qquad \mathbb{E}[\hat{\rho}_T^2] = \frac{T-N}{T} \tag{15}$$

Proof. Under $H_\ell e_m \equiv 0$, the stacked ε^m is independent of Z; M_Z is fixed conditionally on Z and independent of ε^m. Apply Lemmas 2 and 3 to obtain the Beta distribution and moments. □

[2] e_m is a selection vector consisting of zeros except for the mth entry, which is equal to 1.

[3] A random variable X follows a spherical distribution if X and HX follow the same distribution for any orthogonal matrix H (Muirhead, 2009).

Note that even though the assumptions for this proposition are not very relevant for macroeconomics (i.e., the shock of interest does not influence macro variables), the proposition is useful in that it shows that even in the absence of endogeneity, the finite sample correlation of the instrument and the shock do not have to be close to 1, especially in small samples and with relatively many controls. In our Monte Carlo experiments, we revisit these correlations.

When $H_\ell e_m \not\equiv 0$ for some $\ell \geq 1$, then Z depends on the *lagged* target shock, so M_Z is a function of (a shift of) ε^m, and the independence needed in Lemma 4 is lost. The central Beta$((T-N)/2, N/2)$ distribution *does not hold* in general.

Corollary 2 (Jensen bound for the correlation)**.** *Under the conditions of Proposition 3,* $\widehat{\rho}_T^2 \sim \text{Beta}\big(\frac{T-N}{2}, \frac{N}{2}\big)$ *and*

$$\mathbb{E}[\widehat{\rho}_T^2] = \frac{T-N}{T} \tag{16}$$

Because $x \mapsto \sqrt{x}$ *is concave on* $[0,1]$*, Jensen's inequality yields*

$$\mathbb{E}[\widehat{\rho}_T] = \mathbb{E}\Big[\sqrt{\widehat{\rho}_T^2}\Big] \leq \sqrt{\mathbb{E}[\widehat{\rho}_T^2]} = \sqrt{\frac{T-N}{T}} = \sqrt{1-\frac{N}{T}} \tag{17}$$

The inequality is strict whenever $0 < N < T$*. Equality holds only in the degenerate case* $N = 0$*, where* $\widehat{\rho}_T \equiv 1$.[4]

CASE II: $R_m^2 > 0$ (POSITIVE PREDICTABILITY) We can derive similar results for the case when there is endogeneity. To obtain an exact finite-sample distribution, we assume ε_t are Gaussian and that Z_t has no lag feedback from the shock, i.e., $H_\ell = 0$ for all $\ell \geq 1$ (so $Z_t = H_0 \varepsilon_t$).

Proposition 4 (Noncentral Beta under static Gaussian endogeneity with $K > N$ shocks)**.** *Suppose Assumption 1 holds, except that the shock vector* ε_t *is* K*-dimensional, with* $K > N$*:*

$$\varepsilon_t = (\varepsilon_t^1, \dots, \varepsilon_t^K)^\top \in \mathbb{R}^K, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma_{\varepsilon\varepsilon}), \qquad \Sigma_{\varepsilon\varepsilon} = \text{diag}(\sigma_1^2, \dots, \sigma_K^2) \succ 0,$$

and let the macro variables satisfy

$$Z_t = H_0 \varepsilon_t, \qquad H_0 \in \mathbb{R}^{N \times K},$$

so that

$$\Sigma_{ZZ} := \text{Var}(Z_t) = H_0 \Sigma_{\varepsilon\varepsilon} H_0^\top \succ 0.$$

Fix $m \in \{1, \dots, K\}$ *and write* $\Sigma_{Z\varepsilon^m} := \text{Cov}(Z_t, \varepsilon_t^m) = \sigma_m^2 h_{0,m}$*, where* $h_{0,m}$ *is the* m*th column of* H_0*. Define the population projection coefficient*

$$\delta := \Sigma_{ZZ}^{-1} \Sigma_{Z\varepsilon^m},$$

the associated population R^2

$$R_m^2 := \frac{\Sigma_{Z\varepsilon^m}^\top \Sigma_{ZZ}^{-1} \Sigma_{Z\varepsilon^m}}{\sigma_m^2} \in [0,1],$$

[4] The correlation between the true shock and the instrument is always non-negative, as we prove in the appendix.

and the conditional variance

$$\sigma_e^2 := \mathrm{Var}(\varepsilon_t^m \mid Z_t) = \sigma_m^2(1 - R_m^2).$$

Assume $\sigma_e^2 > 0$ *(equivalently,* $R_m^2 < 1$*).*

Stack $Z = [Z_1, \dots, Z_T]^\top \in \mathbb{R}^{T \times N}$ *and* $\varepsilon^m = (\varepsilon_1^m, \dots, \varepsilon_T^m)^\top \in \mathbb{R}^T$*, and assume* Z *has full column rank* $N < T$*. Then*

$$\varepsilon^m = Z\delta + \eta, \qquad \eta \mid Z \sim \mathcal{N}(0, \sigma_e^2 I_T), \qquad \eta \perp Z.$$

Consequently, with $P_Z := Z(Z^\top Z)^{-1} Z^\top$*,* $M_Z := I_T - P_Z$ *and*

$$\widehat{\rho}_T^2 := \frac{(\varepsilon^m)^\top M_Z \varepsilon^m}{(\varepsilon^m)^\top \varepsilon^m}, \qquad \Lambda_T := \frac{\|Z\delta\|^2}{\sigma_e^2},$$

we have the conditional distribution

$$\widehat{\rho}_T^2 \mid Z \;\sim\; \mathrm{Beta}_{\mathrm{nc}}\left(\frac{T-N}{2}, \frac{N}{2}; \Lambda_T\right).$$

(If instead $\sigma_e^2 = 0$ *(i.e.* $R_m^2 = 1$*), then* $\widehat{\rho}_T^2 \equiv 0$ *almost surely.)*

The restriction $K > N$ is imposed to avoid a "full-revelation" knife-edge, which would hold by construction under the maintained assumptions: if $K = N$ and $Z_t = H_0 \varepsilon_t$ with $\Sigma_{ZZ} = H_0 \Sigma_{\varepsilon\varepsilon} H_0^\top \succ 0$, then H_0 must be invertible and the shocks are perfectly recoverable from Z_t, forcing $R_m^2 = 1$ and $\sigma_e^2 = 0$; avoiding that in the square case would require relaxing the maintained assumption $\Sigma_{ZZ} \succ 0$.

Proof. When $Z_t = H_0 \varepsilon_t$ and ε_t are i.i.d. Gaussian, (Z, ε^m) are jointly normal with block-diagonal time structure; the conditional residual η is i.i.d. Gaussian and independent of Z, hence Lemma 4 applies, giving the noncentral Beta distribution; cf. Muirhead (2009, §1.2, §1.4), Johnson et al. (1995, Ch. 34). □

The preceding result gives an exact finite-sample benchmark for the auxiliary-regression step. Conditional on the realized macro variables, endogeneity changes the central Beta distribution only through the noncentrality parameter Λ_T, which measures how much of shock m is predicted by contemporaneous macro variables. The noncentral Beta distribution has a useful representation as a Poisson mixture of central Beta distributions (Johnson et al., 1995). Let $J \sim \mathrm{Poisson}(\Lambda_T/2)$. Then

$$\widehat{\rho}_T^2 \mid Z \overset{d}{=} Y_J, \qquad Y_j \sim \mathrm{Beta}\left(\frac{T-N}{2}, \frac{N}{2} + j\right),$$

and therefore

$$\mathbb{E}\big[\widehat{\rho}_T^2 \mid Z\big] = \frac{T-N}{2}\,\mathbb{E}\left[\frac{1}{\frac{T}{2} + J}\right], \quad J \sim \mathrm{Poisson}(\Lambda_T/2).$$

Jensen's inequality and the fact that $J \geq 0$ yield sharp bounds:

$$\frac{T-N}{T+\Lambda_T} \leq \mathbb{E}\big[\widehat{\rho}_T^2 \mid Z\big] \leq \frac{T-N}{T}, \qquad \mathbb{E}\big[\widehat{\rho}_T \mid Z\big] \leq \sqrt{\frac{T-N}{T}}.$$

The upper bound is the central benchmark obtained when the macro variables do not predict shock m. The lower bound shows how the same finite sample can generate weaker residual correlations when contemporaneous predictability is stronger.

The scale of Λ_T can be written in terms of the population predictability parameter R_m^2. Suppose the pairs (Z_t, ε_t^m) are mean zero, serially independent over time, and jointly Gaussian, with $\text{Var}(Z_t) = \Sigma_{ZZ} \succ 0$, $\text{Var}(\varepsilon_t^m) = \sigma_m^2$, and $\text{Cov}(Z_t, \varepsilon_t^m) = \Sigma_{Z\varepsilon^m}$. Let

$$\delta = \Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m}, \qquad R_m^2 = \frac{\Sigma_{Z\varepsilon^m}^\top \Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m}}{\sigma_m^2}, \qquad \sigma_e^2 = \sigma_m^2(1-R_m^2),$$

and assume $R_m^2 < 1$. Then

$$\Lambda_T = \frac{\|Z\delta\|^2}{\sigma_e^2} \overset{d}{=} \frac{R_m^2}{1-R_m^2}\chi_T^2, \qquad \frac{\Lambda_T}{T} \xrightarrow{p} \frac{R_m^2}{1-R_m^2}.$$

Thus R_m^2 determines the order of the noncentrality parameter Λ_T. When $R_m^2 = 0$, the central Beta benchmark applies. As R_m^2 rises, the expected squared residual correlation moves down through the lower bound above.

These formulas are not an exact sampling theory for the dynamic DSGE simulations. In the Smets-Wouters model, Z_t generally depends on current and lagged structural shocks. The stacked projector M_Z is then a function of the same shock vector being projected, so the underlying orthogonal components that form $\widehat{\rho}_T^2$ are no longer independent central and noncentral quadratic forms. The benchmark remains useful because it isolates the auxiliary-regression force: finite T and the number of included macro variables mechanically pull residual correlations below one, and contemporaneous predictability of the target shock pulls them down further. The Monte Carlo results below show how much of that benchmark remains visible once the full dynamic data-generating process, and the finite-order VAR approximation are added.

4 MONTE CARLO EXPERIMENTS

In order to assess the performance of the identification approach outlined in the previous section, we simulate 1,000 datasets of length $1{,}000 + M$ from the Smets and Wouters (2007) model, where we discard the initial 1,000 periods as burn-in and vary M to study the effect of the sample size on our results. The parameter values are given by the fixed parameters and the posterior mode estimates reported in Smets and Wouters (2007). We present all the parameter values in Table 2 in the Appendix. In particular, the monetary policy rule in that model is given by

$$r_t = \rho r_{t-1} + (1-\rho)\left[r_\pi \pi_t + r_y\left(y_t - y_t^p\right)\right] + r_{\Delta y}\left[\left(y_t - y_t^p\right) - \left(y_{t-1} - y_{t-1}^p\right)\right] + \varepsilon_t^r,$$

where y_t is output in the economy, and y_t^p is (potential) output under flexible prices. The monetary policy shocks (ε_t^r) follow an autoregressive process with an i.i.d. innovation

$$\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \eta_t^r.$$

We estimate this rule via OLS, following Carvalho et al. (2021). We then either directly use this residual (possibly autocorrelated) as our instrument or, following Miranda-Agrippino and Ricco (2021), estimate an AR(1) process on the policy rule residual and take the residual of that AR(1) regression as our instrument (our benchmark). For each MC sample, we estimate the VAR via OLS and obtain the associated point estimate of impulse responses to a monetary instrument. For each response variable and each horizon, we then compute the median response and the 5%, 95% percentiles across all MC repetitions. As a benchmark we use a three-variable VAR with 4 lags with the instrument ordered first and also including inflation and output. The default sample size is $M = 200$.

4.1 RESULTS

We first describe the correlation between the true monetary shock in the Smets and Wouters (2007) model and our estimated instrument (Panel A of Table 1). Our estimated monetary policy rule has the correct functional form, so any differences come from either estimation error or endogeneity issues, as discussed above. The first column shows that, at face value, our estimation strategy is successful–the estimated instrument is highly correlated with the shock of interest (0.9). However, the remaining volatility in the instrument is due to correlation with the other structural shocks; specifically, productivity shocks and risk premium shocks. Other shocks can have a sizable correlation with the instrument in specific samples as well, with the maximum absolute correlation being 0.48 across MC samples, shocks, and sample sizes. Importantly, doubling the sample size does not mitigate these effects. Although the large correlation with the true monetary shock speaks well of the OLS-based strategy (Carvalho et al., 2021) we use here, the high correlations with some of the other structural shocks necessitates careful inspection of estimated impulse responses.[5]

We know from Plagborg-Møller and Wolf (2021) that with an internal instrument approach, we can only identify relative impulse responses if measurement error is present in the instrument. Our exercises abstract from measurement error that is independent of the structural shocks. We therefore scale the VAR impulse responses so that the impact response of the estimated instrument equals the true impact response of the monetary policy shock process to a one-standard deviation monetary policy innovation in the Smets and Wouters (2007) model.[6]

Benchmark. We focus on the impulse responses of output and inflation throughout. We plot the "true" median impulse response from the SW model (solid, black); the median estimated response of the simulations (dashed, blue), calculated as the median response across simulations; and error bands (blue,

[5] It might be useful to point out here that Carvalho et al. (2021) present an application where they build residual-based instruments from policy rules estimated with either OLS or IV and find basically indistinguishable results.

[6] Forni et al. (2024) show how to extend the standard proxy-VAR approach to allow for non-invertibility.

	Monetary Policy	Productivity	Risk Premium	Spending	Investment-specific	Price Mark-up	Wage Mark-up
Panel A: Benchmark Monetary Policy Rule							
100	0.92 (0.85, 0.97); 0.99	-0.13 (-0.28, 0.02); -0.46	0.27 (0.15, 0.37); 0.48	0.06 (-0.11, 0.24); 0.44	0.06 (-0.11, 0.22); 0.40	0.03 (-0.10, 0.16); 0.29	0.03 (-0.13, 0.18); 0.37
150	0.93 (0.88, 0.96); 0.99	-0.13 (-0.25, 0.00); -0.37	0.26 (0.17, 0.36); 0.43	0.07 (-0.07, 0.20); 0.32	0.06 (-0.08, 0.20); 0.33	0.03 (-0.07, 0.15); 0.25	0.03 (-0.10, 0.16); 0.32
200	0.93 (0.89, 0.96); 0.98	-0.13 (-0.24, -0.03); -0.38	0.26 (0.18, 0.34); 0.41	0.06 (-0.05, 0.18); 0.26	0.06 (-0.06, 0.19); 0.26	0.04 (-0.05, 0.13); 0.23	0.03 (-0.08, 0.14); 0.23
Panel B: Backward-Looking Monetary Policy Rule							
100	0.98 (0.95, 1.00); 1.00	-0.00 (-0.16, 0.16); -0.37	-0.00 (-0.17, 0.17); 0.33	-0.00 (-0.17, 0.16); -0.31	0.01 (-0.16, 0.17); 0.35	-0.00 (-0.17, 0.17); 0.38	0.00 (-0.16, 0.16); 0.37
150	0.99 (0.96, 1.00); 1.00	0.00 (-0.13, 0.14); -0.27	0.00 (-0.14, 0.14); -0.25	0.00 (-0.13, 0.14); -0.27	0.00 (-0.14, 0.14); -0.32	0.00 (-0.13, 0.13); -0.28	-0.00 (-0.13, 0.14); 0.29
200	0.99 (0.97, 1.00); 1.00	-0.00 (-0.12, 0.11); -0.21	0.00 (-0.12, 0.12); -0.26	0.00 (-0.12, 0.13); 0.27	-0.00 (-0.12, 0.11); -0.21	-0.00 (-0.12, 0.12); 0.24	0.00 (-0.12, 0.12); 0.24

Table 1: Correlation between true shocks and instrument, for three sample sizes across 1,000 MC datasets. Entries show the median, as well as 5th and 95th percentiles (in parentheses) and the largest correlation in absolute value (next to the parentheses) across the MC repetitions.

dotted) are constructed as percentiles of OLS point estimates from the various samples.[7] The first column of Figure 1 plots the impulse response from our benchmark specification, i.e., three variables (instrument, inflation, output), with 4 lags and 200 observations. Two points are noteworthy. First, there is substantial variation across Monte Carlo samples, including significant probability that a “price puzzle" emerges, i.e. an increase in inflation after a contractionary monetary policy shock. This is somewhat reminiscent of the issues that have plagued structural VARs identified via sign restrictions (Wolf, 2020). Second, the magnitude of the effect on inflation is underestimated in most samples. This is most severe on impact, where the median impulse response is more than four times smaller than the true effect in absolute value. Canova and Ferroni (2022) highlight that small VARs can often lead to distorted estimates of impulse responses, even though their main focus is on identification schemes other than those that use instruments. This leads to our first robustness exercise.

Adding Variables. The second column of Figure 1 shows results for a larger, seven-variable VAR. The larger VAR makes minimal difference to the median impulse response, a result that may seem surprising given findings like Canova and Ferroni (2022) on the importance of VAR specification. However, the primary source of bias here is not omitted-variable misspecification in the VAR, but rather the instrument contamination identified in Proposition 2. Because the policy rule is endogenous ($R^2_m > 0$), the OLS instrument $\hat{r}_t$ is a contaminated proxy ($r^\star_t$) that mixes the true monetary shock with other structural shocks (as confirmed in Table 1).

Since both the 3-variable and 7-variable VARs are identified using the same contaminated instrument, both trace out the impulse response to the same incorrect proxy. The only notable change is a slight widening of the error bands: increasing the number of variables for a fixed sample T increases the number of parameters, adding to estimation uncertainty. The third column reports an oracle version of the benchmark three-variable VAR that uses the true monetary policy innovation as the instrument.

[7] From the literature on generated regressors, we know that the presence of such generated regressors widens error bands (Pagan, 1984), an argument that would also apply to our setting if we reported error bands or coverage for a given sample. We

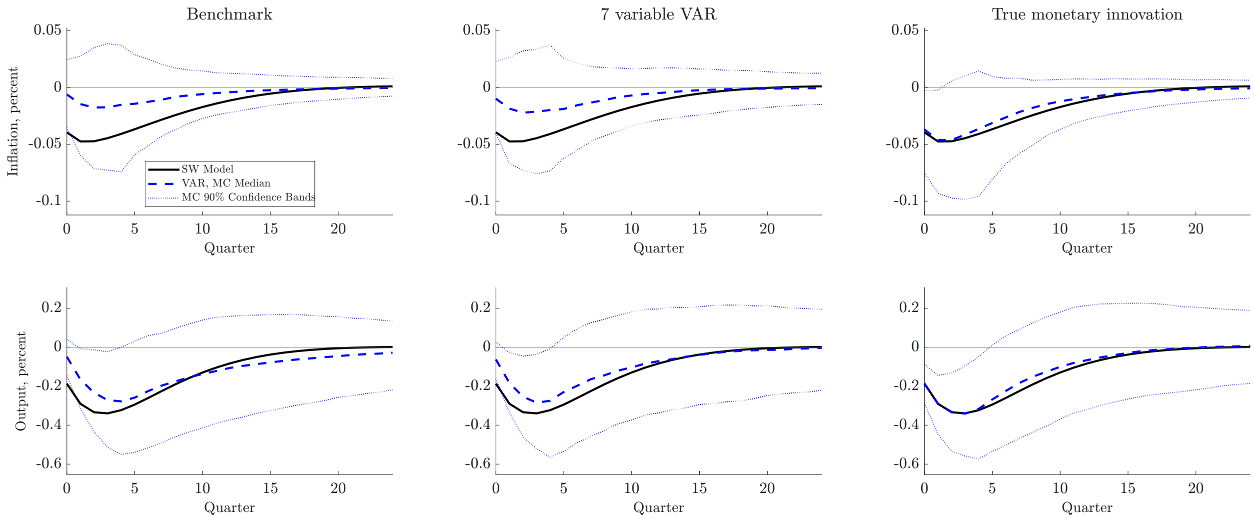

Figure 1: IRFs scaled so that the impact response of the estimated instrument equals the true impact response of the monetary policy shock process to a one-standard deviation monetary policy innovation. Benchmark three variable VAR, seven variable VAR, and oracle three variable VAR using the true monetary policy innovation as the instrument.

Increase Lag Length. Next, we ask what happens if we increase the lag length. De Graeve and Westermark (2025) have shown that this can substantially reduce both bias and variance of estimated impulse responses, as the variance reduction from correcting misspecification can be large.[8] We check this in Figure 2, where we increase the number of lags in our VAR from 4 to 16.

This result, however, provides a clear example of the finite-sample trade-offs discussed above. While adding lags may reduce misspecification bias, it dramatically increases the number of parameters to be estimated (from $k \approx 3 \times 4 = 12$ parameters per equation to $k \approx 3 \times 16 = 48$). This is the same finite-sample logic as in Proposition 3: for a fixed T, adding more estimated objects lowers precision. In the auxiliary regression, a higher $\kappa = N/T$ mechanically weakens the residual correlation; in the VAR, a high parameter-to-observation ratio makes the dynamic estimates (A_ℓ) imprecise.

With our fixed sample of $T = 200$, the (potential) gain from reduced bias is swamped by an increase in estimation variance. This manifests in Figure 2 as substantially wider error bands and a more pronounced price puzzle, as the over-parameterized VAR has poor finite-sample properties.

Sample Size. Our benchmark sample size $T = 200$ is quite large for macro applications (50 years of quarterly data). While not reported here, we did analyze the impact of reducing the sample size to $T = 100$. As the correlations in Table 1 suggests, it had negligible impact on our impulse response functions.

Removing Autocorrelation. Finally, we ask if removing the autocorrelation via an estimated AR model as in Miranda-Agrippino and Ricco (2021) has any effect. Figure 3 shows what happens if we do not

instead focus here on showing the dispersion of point estimates across Monte Carlo samples.

[8] One of the applications in De Graeve and Westermark (2025) uses more lags in the Miranda-Agrippino and Ricco (2021) application that uses a residual-based instrument for a monetary policy shock.

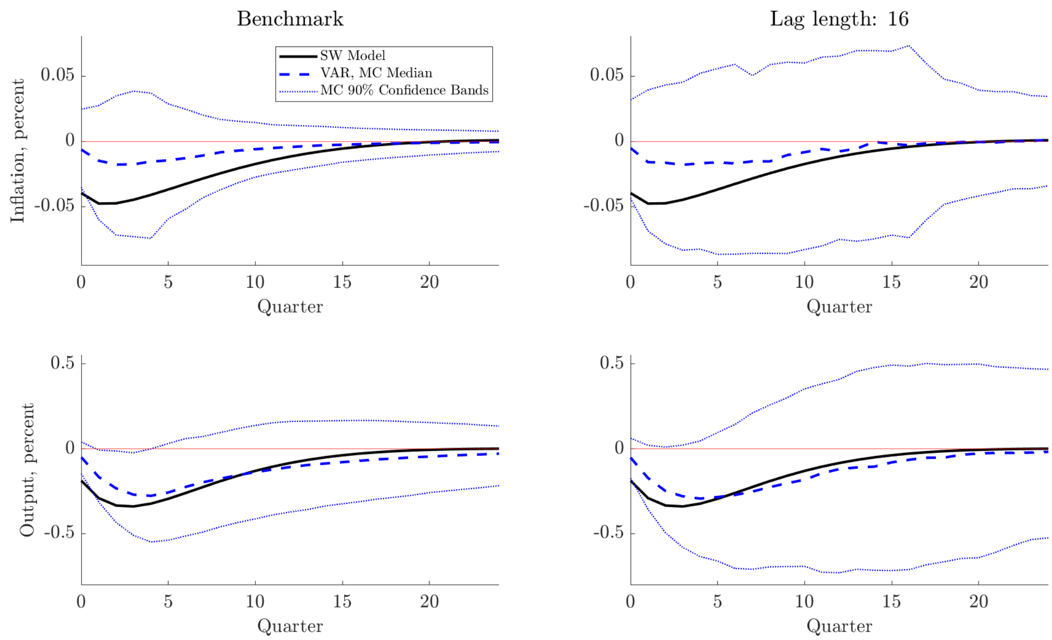

Figure 2: IRFs scaled so that the impact response of the estimated instrument equals the true impact response of the monetary policy shock process to a one-standard deviation monetary policy innovation. Benchmark three variable VAR with four lags vs three variable VAR with 16 lags.

pre-whiten the instrument - we use this alternative instrument in both the 3 and 7 variable VARs.

We find no meaningful difference. This is to be expected, since our internal instrument VAR approach automatically accounts for lags of the instrument. The VAR estimation process is a "whitening" filter by construction. Manually pre-whitening $\hat{r}_t$ with a simple AR(1) is therefore redundant. The VAR's lag structure already accounts for the instrument's autocorrelation (and all other linear-dynamic relationships), isolating the same innovation $u_t^{(r)}$. We suspect that such a whitening procedure is helpful in external VAR settings such as Caldara and Herbst (2019), where the instrument's own dynamics are not explicitly modeled as part of the VAR system.

4.1.1 THE ROLE OF ENDOGENEITY

A natural question is how our results are driven by the two sources of contamination identified in Section 3: (i) endogeneity bias, and (ii) finite-sample estimation error. One possibility would be to re-estimate the equation that gives us the instrument using instrumental variables instead of OLS. However, Carvalho et al. (2021) show that OLS provides estimates of the policy rule coefficients that are for all practical purposes at least as good as IV-based estimates with common instruments for the policy rule estimation. Hence, instead we take a different approach and modify the data-generating process so that the policy rule is purely backward looking, removing endogeneity and thus any doubt that OLS-based estimation provides sensible estimates. We adopt the following backward looking version of the monetary policy rule:

$$r_t = \rho r_{t-1} + (1-\rho)\left[r_\pi \pi_{t-1} + r_y\left(y_{t-1} - y^p_{t-1}\right)\right] + r_{\Delta y}\left[\left(y_{t-1} - y^p_{t-1}\right) - \left(y_{t-2} - y^p_{t-2}\right)\right] + \varepsilon^r_t.$$

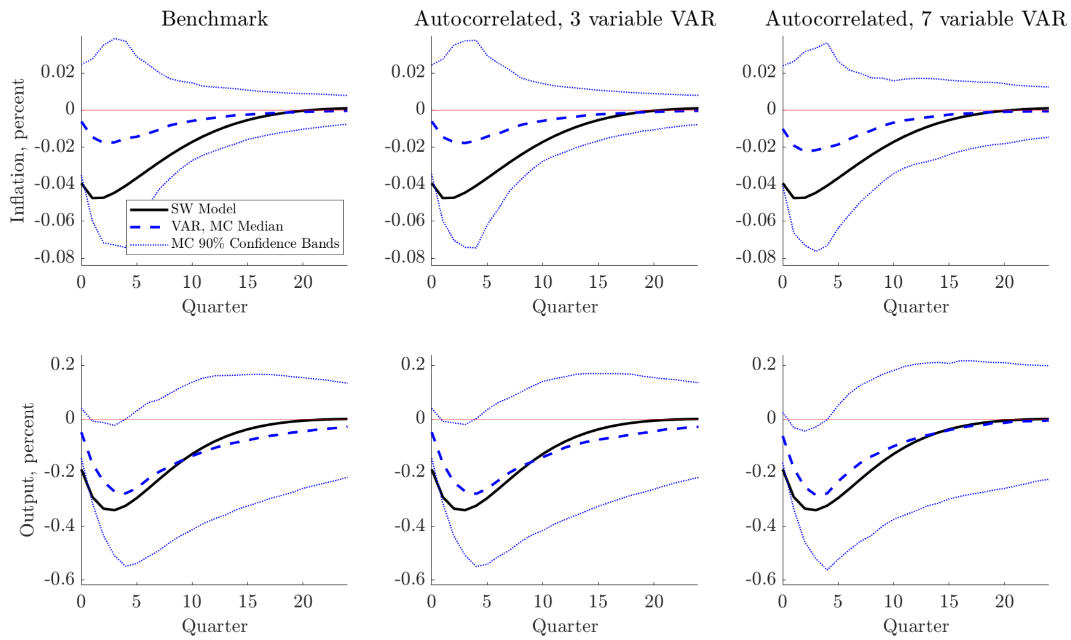

Figure 3: IRFs scaled so that the impact response of the estimated instrument equals the true impact response of the monetary policy shock process to a one-standard deviation monetary policy innovation. Benchmark three variable VAR with whitened instrument vs VARs with autocorrelated instruments.

By construction, the monetary shock is now contemporaneously orthogonal to the variables in the rule, forcing $R_m^2 = 0$. We then estimate a correctly specified backward-looking rule via OLS to generate the instrument.

Panel B of Table 1 shows the result of this change. The median correlation between the instrument and the true monetary shock (column 1) is now ≈ 0.99. Critically, the median correlation with all other structural shocks (columns 2-7) is now exactly zero. This confirms that by removing endogeneity, we have eliminated the population-level OLS bias (δ) that was contaminating the instrument.

Figure 4 shows the associated impulse responses. Two results stand out. First, the median IRF (the population-level result) now correctly lines up with the true SW model. This is the direct result of removing endogeneity as shown in Proposition 2. By forcing $R_m^2 = 0$, the OLS instrument $\hat{r}_t$ now converges to the true shock ε_t^m, not the contaminated proxy $r_t^\star$. Second, the dispersion in the estimates remains quite large. Many point estimates still show a price puzzle, as evidenced by the wide percentile bands. This result is consistent with Proposition 3 and the finite-sample benchmark above. Even with endogeneity removed, $\kappa = N/T > 0$ leaves residual sampling error in any single sample, leading to wide dispersion.

4.1.2 A possible (partial) correction

The results above suggest a possible correction. If the generated monetary policy instrument is contaminated by other structural shocks, one can try to remove the component of the instrument that is predictable from proxies for those nuisance shocks. Empirically, this can be relevant in situations where a researcher has access to credible instruments for other shocks that are not bases on first stage regressions such as, say, those based on high frequency changes in asset

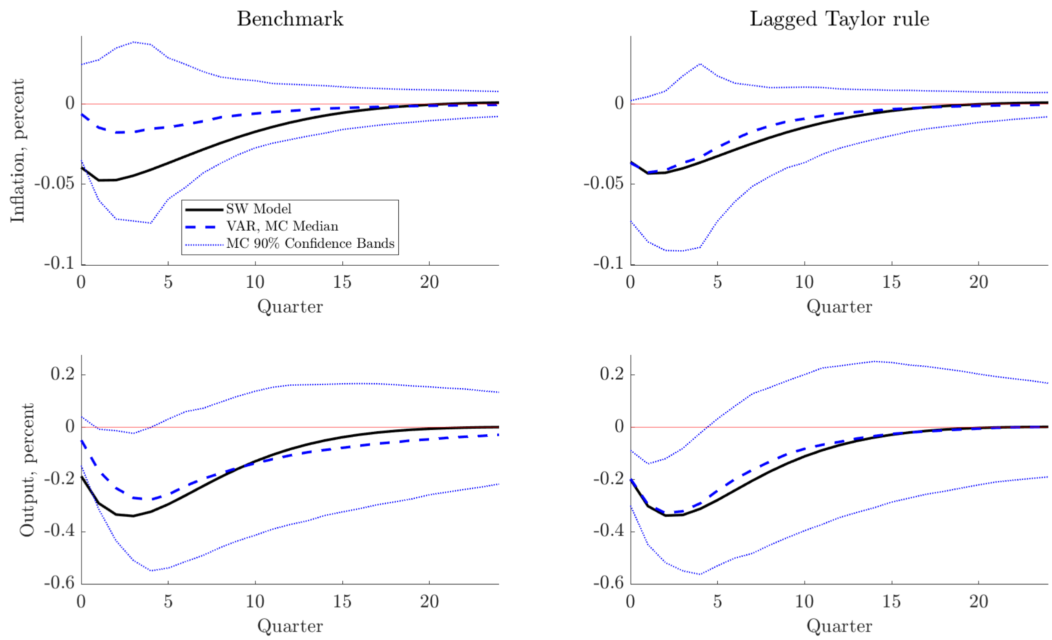

Figure 4: IRFs scaled so that the impact response of the estimated instrument equals the true impact response of the monetary policy shock process to a one-standard deviation monetary policy innovation. Benchmark three variable VAR vs 3 variable VAR for data-generating process with backward-looking monetary policy rule.

prices if those can be credibly and uniquely linked to another structural shock. This correction targets the consequence of the endogeneity problem in our setting: the OLS residual is not only a noisy measure of the monetary policy shock, but also loads on other structural shocks.

We implement this idea in the same Monte Carlo design used above. For each simulated sample from the Smets and Wouters (2007) model, we estimate the benchmark monetary policy rule by OLS and construct the original generated monetary instrument, denoted $\hat{r}_t^m$. Let $\mathcal{J}$ represent the set of the six non-monetary structural shocks reported in Table 1 and write ε_t^j for the corresponding jth standardized structural shock. In our Monte Carlo, we assume for simplicity that the other structural shocks are directly observable (an "oracle" exercise). More generally, the econometrician may observe a noisy but uncontaminated proxy

$$q_{j,t}(\lambda_j) = \varepsilon_t^j + \sqrt{\lambda_j}\xi_{j,t}, \qquad \xi_{j,t} \sim N(0,1), \qquad \xi_{j,t} \perp \{\varepsilon_\tau^\ell\}_{\ell,\tau}, \tag{18}$$

where λ_j is the noise-to-signal ratio and any noise shocks are uncorrelated across proxies. When $\lambda_j = 0$, the proxy is the standardized true structural shock. For any set of proxies for other structural shocks $S \subseteq \mathcal{J}$, we residualize the generated monetary instrument by estimating

$$\hat{r}_t^m = \alpha_S + \Pi_S^\top q_{S,t} + \tilde{r}_{S,t}^m, \tag{19}$$

where $q_{S,t}$ collects the proxies $q_{j,t}(\lambda_j)$ for $j \in S$. We then use $\tilde{r}_{S,t}^m$ as the internal instrument in the same

VAR specification used in the benchmark results.

Suppose the original generated instrument has the linear decomposition

$$\widehat{r}_t^m = c_m \varepsilon_t^m + \sum_{j \in \mathscr{J}} c_j \varepsilon_t^j + e_t, \tag{20}$$

where the structural shocks are mutually orthogonal and e_t is a measurement error that is orthogonal to the structural shocks and to the proxy errors.[9] For an included structural shock $j \in S$, $\text{Var}(q_{j,t}) = 1 + \lambda_j$ and $\text{Cov}(q_{j,t}, \widehat{r}_t^m) = c_j$. The population projection coefficient is therefore

$$\pi_j^* = \frac{\text{Cov}(q_{j,t}, \widehat{r}_t^m)}{\text{Var}(q_{j,t})} = \frac{c_j}{1+\lambda_j}, \qquad j \in S. \tag{21}$$

Substituting these coefficients into the population residual gives

$$\widetilde{r}_{S,t}^{m,*} = c_m \varepsilon_t^m + \sum_{\ell \in \mathscr{J} \setminus S} c_\ell \varepsilon_t^\ell + \sum_{j \in S} c_j \frac{\lambda_j}{1+\lambda_j} \varepsilon_t^j + e_t - \sum_{j \in S} \frac{c_j \sqrt{\lambda_j}}{1+\lambda_j} \xi_{j,t}. \tag{22}$$

Thus residualizing on a proxy for structural shock j removes the fraction $1/(1+\lambda_j)$ of that shock's loading and leaves the fraction $\lambda_j/(1+\lambda_j)$. The oracle case, $\lambda_j = 0$, removes the included nuisance shock completely. With noisy proxies, residualization attenuates rather than eliminates the loading on that structural shock, and it also introduces proxy noise into the residualized instrument. Structural shocks not included in S keep their original loadings.

The key distinction is what residualization makes orthogonal. In population,

$$\text{Cov}(\widetilde{r}_{S,t}^{m,*}, q_{j,t}) = 0, \qquad j \in S, \tag{23}$$

but

$$\text{Cov}(\widetilde{r}_{S,t}^{m,*}, \varepsilon_t^j) = c_j \frac{\lambda_j}{1+\lambda_j}, \qquad j \in S. \tag{24}$$

Residualization therefore makes the instrument orthogonal to the included proxy, not necessarily to the true structural shock. Only the oracle case $\lambda_j = 0$ makes the residualized instrument orthogonal to the true included structural shock. This distinction matters for empirical applications: residualizing on a noisy nuisance instrument removes only the part of the contamination predictable from that proxy.

In the Monte Carlo exercise, we contrast our benchmark with the case of full residualization uses $S = \mathscr{J}$. Figure 5 reports the oracle case with $\lambda_j = 0$ for all $j \in S$. The residualized instrument is much cleaner: the median correlation with the true monetary policy shock rises from about 0.93 to 0.97, while the median nuisance-shock correlations are eliminated. The associated impulse responses move closer to the true Smets and Wouters (2007) responses for both inflation and output, especially on impact. The percentile bands remain wide, however, so residualization addresses the contamination problem without removing the finite-sample dispersion documented above.

[9] In our Monte Carlo experiments, we have $e_t = 0 \forall t$.

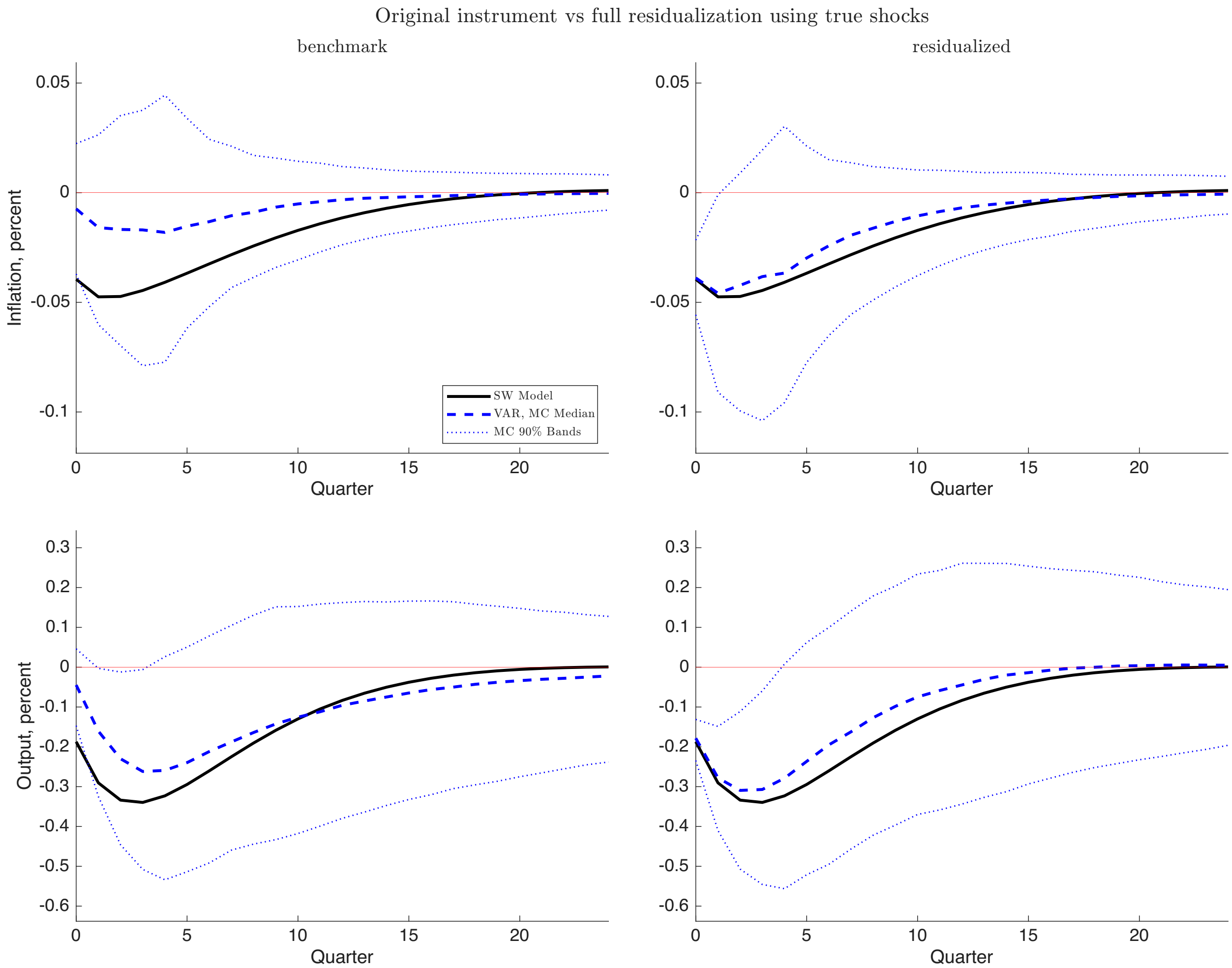

Figure 5: IRFs scaled so that the impact response of the estimated instrument equals the true impact response of the monetary policy shock process to a one-standard deviation monetary policy shock. Benchmark generated instrument versus full residualization using the true non-monetary structural shocks.

The oracle nature of this exercise matters. Equations (23)–(24) show that a noisy proxy can deliver exact orthogonality with respect to the proxy while leaving contamination from the true nuisance shock. Thus the result should be read as an upper bound on what residualization can achieve with informative nuisance-shock instruments, not as evidence that the endogeneity problem can be solved mechanically in applications, even if relevant proxies are available.

5 CONCLUSION

Endogeneity is pervasive in macroeconomics. Our benchmark results and those extensions that do not remove endogeneity are useful characterizations of the issues that researchers developing an instrumental variable approach must confront. Many instruments rely on estimating auxiliary regressions where similar issues will arise, even if they do not fit exactly in the residual-based framework studied here (Bu et al., 2021). The instrument, while highly correlated with the "true" shock, may still be compromised through contamination by other shocks. We show that various suggestions in the literature to improve the performance of structural VARs, such as increasing the lag length or using larger VARs, do not substantially fix these issues. An analogous issue arises for generated instruments based on fitted values or principal components. Such procedures can reduce idiosyncratic measurement error in noisy observables, but they do not by themselves establish proxy validity for a particular structural shock. If the underlying observables load on several structural innovations, the generated instrument inherits those loadings. The resulting proxy is then relevant for the target shock but not exogenous to non-target shocks, and impulse responses identify a response to a composite innovation. This logic is closely related to the high-frequency monetary-policy literature: for example, Nakamura and Steinsson (2018) construct a policy-news measure from the first principal component of interest-rate surprises and explicitly model the fact that FOMC announcements may reveal information about non-monetary fundamentals as well as monetary policy.

REFERENCES

Aruoba, S. Borağan and Thomas Drechsel, "Identifying Monetary Policy Shocks: A Natural Language Approach," *American Economic Journal: Macroeconomics*, forthcoming.

Auerbach, Alan J. and Yuriy Gorodnichenko, "Output Spillovers from Fiscal Policy," *American Economic Review*, 2013, *103* (3), 141–146.

Barattieri, Alessandro and Matteo Cacciatore, "Self-Harming Trade Policy? Protectionism and Production Networks," *American Economic Journal: Macroeconomics*, 2023, *15* (2), 97–128.

_ , _ , and Nora Traum, "Estimating the Effects of Government Spending Through the Production Network," NBER Working Paper 31680, National Bureau of Economic Research September 2023.

Boyd, Stephen and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

Brockwell, Peter J. and Richard A. Davis, *Time Series: Theory and Methods*, 2 ed., New York: Springer, 1991.

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

Caldara, Dario and Edward Herbst, "Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs," *American Economic Journal: Macroeconomics*, January 2019, *11* (1), 157–192.

Canova, Fabio and Filippo Ferroni, "Mind the Gap! Stylized Dynamic Facts and Structural Models," *American Economic Journal: Macroeconomics*, October 2022, *14* (4), 104–135.

Carvalho, Carlos, Fernanda Nechio, and Tiago Tristão, "Taylor rule estimation by OLS," *Journal of Monetary Economics*, November 2021, *124*, 140–154.

Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan, "Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theory?," *Journal of Monetary Economics*, 2008, *55* (8), 1337–1352.

Christiano, Lawrence J., Martin Eichenbaum, and Robert Vigfusson, "Assessing Structural VARs," in Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, eds., *NBER Macroeconomics Annual 2006*, Vol. 21, Cambridge, MA: MIT Press, 2007, pp. 1–106.

Craig, Allen T., "Note on the Independence of Certain Quadratic Forms," *Annals of Mathematical Statistics*, 1943, *14* (2), 195–197.

De Graeve, Ferre and Andreas Westermark, "Long-lag VARs," *Journal of Monetary Economics*, 2025. Forthcoming.

Erceg, Christopher J., Luca Guerrieri, and Christopher Gust, "Can Long-Run Restrictions Identify Technology Shocks?," *International Journal of Central Banking*, 2005, *1* (2), 123–157.

Fang, Kai-Tai, Samuel Kotz, and Kai W. Ng, *Symmetric Multivariate and Related Distributions*, London: Chapman and Hall, 1990.

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent, and Mark W. Watson, "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 2007, *97*, 1021–1024.

Forni, Mario, Luca Gambetti, and Giovanni Ricco, "Generalised External-Instrument SVAR Analysis," September 2024.

Giacomini, Raffaella, "The Relationship Between DSGE and VAR Models," in "VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims" Advances in Econometrics, Emerald Group Publishing Limited, 2013, pp. 1–25.

Hamilton, James D., *Time Series Analysis*, Princeton, NJ: Princeton University Press, 1994.

Horn, Roger A. and Charles R. Johnson, *Matrix Analysis*, 2 ed., Cambridge: Cambridge University Press, 2013.

Johnson, Norman L., Samuel Kotz, and Narayanaswamy Balakrishnan, *Continuous Univariate Distributions, Volume 2*, 2 ed., New York: Wiley, 1995.

_ , _ , and _ , *Continuous Multivariate Distributions, Volume 1: Models and Applications*, 2 ed., New York: Wiley-Interscience, 2000.

Laha, R. G., "On the Stochastic Independence of Two Second-Degree Polynomial Statistics in Normally Distributed Variates," *Annals of Mathematical Statistics*, 1956, *27* (3), 790–796.

Li, Chi-Kwong, "A Simple Proof of the Craig–Sakamoto Theorem," *Linear Algebra and its Applications*, 2000, *321*, 281–283.

Lloyd, Simon and Ed Manuel, "Controls, not shocks: estimating dynamic causal effects in macroeconomics," Staff Working Paper 1079, Bank of England, London July 2024.

Lütkepohl, Helmut, *New Introduction to Multiple Time Series Analysis*, Berlin, Heidelberg: Springer, 2005.

Mertens, Karel and Morten O. Ravn, "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *American Economic Review*, 2013, *103*, 1212–1247.

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

Muirhead, Robb J., *Aspects of Multivariate Statistical Theory*, 2 ed., Hoboken, NJ: Wiley, 2009.

Nakamura, Emi and Jón Steinsson, "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics*, 2018, *133*, 1283–1330.

Ogawa, J. and I. Olkin, "A Tale of Two Countries: The Craig–Sakamoto–Matusita Theorem," *Journal of Statistical Planning and Inference*, 2008, *138* (7), 1906–1915.

Pagan, Adrian, "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 1984, *25* (1), 221–247.

Plagborg-Møller, Mikkel and Christian K. Wolf, "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, March 2021, *89* (2), 955–980.

Ravenna, Federico, "Vector Autoregressions and Reduced Form Representations of DSGE Models," *Journal of Monetary Economics*, 2007, *54* (7), 2048–2064.

Romer, C. and D. Romer, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 2004, *94* (4), 1055–84.

Seber, George A. F. and Alan J. Lee, *Linear Regression Analysis*, 2 ed., Hoboken, NJ: Wiley-Interscience, 2003.

Sims, Christopher A., "Macroeconomics and Reality," *Econometrica*, 1980, *48*, 1–48.

Smets, Frank and Rafael Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, June 2007, *97* (3), 586–606.

Stock, James H. and Mark W. Watson, "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *Economic Journal*, May 2018, *128* (610), 917–948.

Wolf, Christian K., "SVAR (Mis)identification and the Real Effects of Monetary Policy Shocks," *American Economic Journal: Macroeconomics*, October 2020, *12* (4), 1–32.

Zhang, Fuzhen, *The Schur Complement and Its Applications*, Springer, 2005.

A SOME USEFUL THEOREMS

Lemma 1 (Idempotents and orthogonal projectors)**.** *A square matrix P is idempotent if $P^2 = P$. If Z has full column rank, then $P_Z := Z(Z^\top Z)^{-1}Z^\top$ and $M_Z := I_T - P_Z$ are symmetric idempotents; moreover, P_Z is the orthogonal projector onto $\mathrm{col}(Z)$ and M_Z onto $\mathrm{col}(Z)^\perp$. See Seber and Lee (2003, Appendix B).*

Lemma 2 (Spectral theorem representation)**.** *If S is real symmetric then $S = Q\Lambda Q^\top$ with orthogonal Q; if additionally $S^2 = S$, then Λ has only 0's and 1's. Writing $Q = [Q_1\ Q_2]$ for eigenvectors of eigenvalues 1 and 0 respectively, $S = Q_1Q_1^\top$ and $I_T - S = Q_2Q_2^\top$. Apply to $S = P_Z$ to obtain $P_Z = Q_1Q_1^\top$ and $M_Z = Q_2Q_2^\top$. See Horn and Johnson (2013, Thm. 4.1.5).*

Lemma 3 (Spherical radius–direction factorization)**.** *If Y is spherically symmetric in $\mathbb{R}^T$, then $Y \overset{d}{=} RU$ with $R \geq 0$, $U \sim \mathrm{Unif}(\mathbb{S}^{T-1})$, where $\mathbb{S}^{T-1}$ is the unit sphere in $\mathbb{R}^T$, $R \perp\!\!\!\perp U$. Moreover $(U_1^2, \ldots, U_T^2) \sim \mathrm{Dirichlet}(\frac{1}{2}, \ldots, \frac{1}{2})$, so the sum of any k coordinates is $\mathrm{Beta}(k/2, (T-k)/2)$. See Muirhead (2009, §1.5), Fang et al. (1990, Ch. 2), Johnson et al. (2000, Ch. 49).*

Lemma 4 (Quadratic forms in Gaussian vectors)**.** *Let $Y \sim \mathcal{N}_T(\mu, \Sigma)$ with $\Sigma \succ 0$.*

1. *If A is symmetric and idempotent with $\mathrm{rank}(A) = r$, then in the spherical case $\Sigma = \sigma^2 I_T$,*

$$\frac{Y^\top AY}{\sigma^2} \sim \chi_r^2\left(\frac{\mu^\top A\mu}{\sigma^2}\right).$$

   More generally, $Y^\top BY$ is noncentral χ^2 iff $B\Sigma$ is idempotent (Muirhead, 2009, Thm. 1.4.2).

2. *(**Independence**). If A, B are symmetric (constant) matrices, then*

$$Y^\top AY \perp Y^\top BY \iff A\Sigma B = 0.$$

   In the case $\Sigma = \sigma^2 I_T$, this reduces to $AB = 0$ (Craig, 1943; Laha, 1956; Li, 2000; Ogawa and Olkin, 2008). In particular, if A and B are orthogonal projectors onto orthogonal subspaces, then $Y^\top AY$ and $Y^\top BY$ are independent.

When A, B are random but independent of Y, the statement in (2) holds conditionally on (A, B).

Proposition 5 (Nonnegativity of the OLS residual–shock correlation)**.** *Let $W \in \mathbb{R}^{T\times N}$ have full column rank and define the projection $P_W := W(W^\top W)^{-1}W^\top$ and the residual-maker $M_W := I_T - P_W$. For any nonzero vector $y \in \mathbb{R}^T$, set $r := M_W y$ and*

$$\widehat{\rho}_T := \mathrm{Corr}(r, y) = \frac{r^\top y}{\|r\|\,\|y\|}.$$

Then

$$\widehat{\rho}_T = \frac{\|r\|}{\|y\|} \in [0, 1].$$

Consequently $\widehat{\rho}_T$ *equals the positive square root*

$$\widehat{\rho}_T = \sqrt{\frac{y^\top M_W y}{y^\top y}}.$$

Moreover, $\widehat{\rho}_T = 0$ *iff* $r = 0$ *(equivalently* $y \in \mathrm{col}(W)$*), and* $\widehat{\rho}_T = 1$ *iff* $P_W y = 0$ *(equivalently* $y \perp \mathrm{col}(W)$*).*

Proof. Since M_W is a symmetric idempotent matrix (orthogonal projector onto $\mathrm{col}(W)^\perp$), we have $M_W^\top = M_W$ and $M_W^2 = M_W$ (see, e.g., Seber and Lee, 2003, Appendix B). With $r = M_W y$,

$$r^\top y = y^\top M_W y = y^\top M_W^2 y = (M_W y)^\top (M_W y) = r^\top r = \|r\|^2.$$

Hence

$$\widehat{\rho}_T = \frac{r^\top y}{\|r\|\,\|y\|} = \frac{\|r\|^2}{\|r\|\,\|y\|} = \frac{\|r\|}{\|y\|} \geq 0.$$

Because r is an orthogonal projection residual, the Pythagorean decomposition $\|y\|^2 = \|P_W y\|^2 + \|M_W y\|^2$ gives $\|r\| \leq \|y\|$, so $\widehat{\rho}_T \leq 1$. The characterization of the equality cases is immediate from $r = 0$ and $P_W y = 0$. □

B COVARIANCES AND SCHUR COMPLEMENTS

Definition 1 (Schur complement)**.** *Let* $\Sigma = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ *be a block matrix with* $A \in \mathbb{R}^{N\times N}$ *invertible. The Schur complement of* A *in* Σ *is*

$$S_{D\cdot A} := D - C A^{-1} B.$$

When Σ *is symmetric (so* $C = B^\top$*), this reduces to* $S_{D\cdot A} = D - B^\top A^{-1} B$*.*

Remark 1 (Basic facts)**.** *If* $\Sigma \succeq 0$ *and* $A \succ 0$*, then its Schur complement* $S_{D\cdot A} \succeq 0$*. Moreover,*

$$\det(\Sigma) = \det(A)\,\det(S_{D\cdot A}), \qquad \Sigma^{-1} = \begin{bmatrix} A^{-1} + A^{-1} B S_{D\cdot A}^{-1} B^\top A^{-1} & -A^{-1} B S_{D\cdot A}^{-1} \\ -S_{D\cdot A}^{-1} B^\top A^{-1} & S_{D\cdot A}^{-1} \end{bmatrix},$$

whenever the inverses exist. See, e.g., Horn and Johnson (2013), Boyd and Vandenberghe (2004), Zhang (2005).

Proposition 6 (Variance as a Schur complement)**.** *Let the zero-mean random vector* $(Z_t^\top, y_t)^\top \in \mathbb{R}^{N+1}$ *have block covariance*

$$\mathrm{Var}\begin{pmatrix} Z_t \\ y_t \end{pmatrix} = \begin{bmatrix} \Sigma_{ZZ} & \Sigma_{Zy} \\ \Sigma_{yZ} & \sigma_{yy} \end{bmatrix}, \quad \Sigma_{ZZ} \succ 0.$$

Let $\beta := \Sigma_{ZZ}^{-1}\Sigma_{Zy}$ be the population linear projection of y_t on Z_t, and define the innovation (projection residual) $r_t^\star := y_t - \beta^\top Z_t$. Then

$$\mathrm{Var}(r_t^\star) \;=\; \sigma_{yy} - \Sigma_{yZ}\,\Sigma_{ZZ}^{-1}\,\Sigma_{Zy},$$

which is exactly the Schur complement of Σ_{ZZ} in the block covariance above.

Proof. Using $\beta = \Sigma_{ZZ}^{-1}\Sigma_{Zy}$,

$$\begin{aligned}\mathrm{Var}(y_t - \beta^\top Z_t) &= \mathrm{Var}(y_t) - 2\,\beta^\top \mathrm{Cov}(Z_t, y_t) + \beta^\top \mathrm{Var}(Z_t)\beta \\ &= \sigma_{yy} - 2\,\Sigma_{Zy}^\top \Sigma_{ZZ}^{-1}\Sigma_{Zy} + \Sigma_{Zy}^\top \Sigma_{ZZ}^{-1}\Sigma_{ZZ}\Sigma_{ZZ}^{-1}\Sigma_{Zy} \\ &= \sigma_{yy} - \Sigma_{yZ}\Sigma_{ZZ}^{-1}\Sigma_{Zy}.\end{aligned}$$

By Definition 1 (with $A = \Sigma_{ZZ}$, $D = \sigma_{yy}$, $B = \Sigma_{Zy}$), this equals the Schur complement $S_{D\cdot A}$. □

Remark 2 (Connection to conditional variance)**.** *If (Z_t, y_t) is jointly Gaussian, then $\mathrm{Var}(y_t \mid Z_t) = \sigma_{yy} - \Sigma_{yZ}\Sigma_{ZZ}^{-1}\Sigma_{Zy}$, so the Schur complement equals the conditional variance. The positivity $S_{D\cdot A} \ge 0$ implies $0 \le R^2 \le 1$, with equality cases corresponding to perfect (in)predictability.*

Our application. Set $y_t = \varepsilon_t^m$ and Z_t as in the dynamic setup. Then

$$\begin{bmatrix} \Sigma_{ZZ} & \Sigma_{Z\varepsilon^m} \\ \Sigma_{\varepsilon^m Z} & \sigma_m^2 \end{bmatrix} \quad\Rightarrow\quad \mathrm{Var}(r_t^\star) = \sigma_m^2 - \Sigma_{Z\varepsilon^m}^\top\,\Sigma_{ZZ}^{-1}\,\Sigma_{Z\varepsilon^m},$$

which is exactly equation (12).

Lemma 5 (Deriving (13)–(14) from (12))**.** *Recall (12):*

$$\mathrm{Var}(r_t^\star) = \sigma_m^2 - \Sigma_{Z\varepsilon^m}^\top\,\Sigma_{ZZ}^{-1}\,\Sigma_{Z\varepsilon^m}. \tag{12}$$

Then:

1. *Expression for R_m^2 ((13)). Define the population R^2 of the linear projection of ε_t^m on Z_t as*

$$R_m^2 \;:=\; \frac{\mathrm{Var}\big(\delta^\top Z_t\big)}{\mathrm{Var}(\varepsilon_t^m)}, \qquad \delta \;:=\; \Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m}.$$

   Because $\mathrm{Var}(\delta^\top Z_t) = \delta^\top\Sigma_{ZZ}\delta$,

$$R_m^2 \;=\; \frac{\delta^\top\Sigma_{ZZ}\delta}{\sigma_m^2} \;=\; \frac{\Sigma_{Z\varepsilon^m}^\top\,\Sigma_{ZZ}^{-1}\,\Sigma_{Z\varepsilon^m}}{\sigma_m^2}.$$

   Under the MA representation $Z_t = \sum_{\ell\ge0} H_\ell\varepsilon_{t-\ell}$ with independent components and serial indepen-

dence of (ε_t),

$$\Sigma_{Z\varepsilon^m} = \operatorname{Cov}\Big(\sum_{\ell\ge 0} H_\ell \varepsilon_{t-\ell},\, \varepsilon_t^m\Big) = \sum_{\ell\ge 0} H_\ell \operatorname{Cov}(\varepsilon_{t-\ell}, \varepsilon_t^m) = H_0 \operatorname{Cov}(\varepsilon_t, \varepsilon_t^m) = H_0\, \sigma_m^2 e_m = \sigma_m^2 h_{0,m}.$$

Substituting gives the stated form

$$R_m^2 = \sigma_m^2\, h_{0,m}^\top \Sigma_{ZZ}^{-1} h_{0,m} \in [0,1]. \tag{13}$$

2. *Covariance with the target shock ((14), first part). Using* $r_t^\star = \varepsilon_t^m - \delta^\top Z_t$ *and* $\delta = \Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m}$,

$$\begin{aligned}\operatorname{Cov}(r_t^\star, \varepsilon_t^m) &= \operatorname{Var}(\varepsilon_t^m) - \delta^\top \operatorname{Cov}(Z_t, \varepsilon_t^m)\\ &= \sigma_m^2 - \delta^\top \Sigma_{Z\varepsilon^m}\\ &= \sigma_m^2 - \Sigma_{Z\varepsilon^m}^\top \Sigma_{ZZ}^{-1} \Sigma_{Z\varepsilon^m}\\ &= \operatorname{Var}(r_t^\star) \qquad \text{by (12)}.\end{aligned}$$

3. *Correlation with the target shock ((14), second part). By the previous step,*

$$\operatorname{Corr}(r_t^\star, \varepsilon_t^m) = \frac{\operatorname{Cov}(r_t^\star, \varepsilon_t^m)}{\sqrt{\operatorname{Var}(r_t^\star)}\,\sqrt{\operatorname{Var}(\varepsilon_t^m)}} = \frac{\operatorname{Var}(r_t^\star)}{\sqrt{\operatorname{Var}(r_t^\star)}\,\sqrt{\sigma_m^2}} = \sqrt{\frac{\operatorname{Var}(r_t^\star)}{\sigma_m^2}}.$$

Use (12) *and the definition of* R_m^2 *just derived:*

$$\frac{\operatorname{Var}(r_t^\star)}{\sigma_m^2} = 1 - \frac{\Sigma_{Z\varepsilon^m}^\top \Sigma_{ZZ}^{-1} \Sigma_{Z\varepsilon^m}}{\sigma_m^2} = 1 - R_m^2,$$

hence

$$\operatorname{Corr}(r_t^\star, \varepsilon_t^m) = \sqrt{1 - R_m^2}. \tag{14}$$

Lemma 6 (Noncentral-Beta via Poisson mixture)**.** *Let* $Z \in \mathbb{R}^{T\times N}$ *have full column rank* $N < T$*. Let* $\varepsilon = (\varepsilon_1^m, \dots, \varepsilon_T^m)^\top$*. Suppose that, conditional on* Z,

$$\varepsilon = Z\delta + \eta, \qquad \eta \mid Z \sim \mathcal{N}(0, \sigma_e^2 I_T), \qquad \sigma_e^2 > 0.$$

Write

$$\widehat{\rho}_T^2 \equiv \frac{\varepsilon^\top M_Z \varepsilon}{\varepsilon^\top \varepsilon}.$$

Let $a := (T-N)/2$, $b := N/2$, *and* $\Lambda_T := \|Z\delta\|^2/\sigma_e^2 \ge 0$*. Then, conditional on* Z,

$$\widehat{\rho}_T^2 \mid Z \;\sim\; \operatorname{Beta}_{\text{nc}}(a,\, b;\, \Lambda_T) \quad \textit{and} \quad \widehat{\rho}_T^2 \overset{d}{=} Y_J,\;\; Y_j \sim \operatorname{Beta}(a, b+j),\;\; J \sim \operatorname{Poisson}\Big(\tfrac{\Lambda_T}{2}\Big),$$

with J independent of $\{Y_j\}_{j\geq 0}$.

Proof. Write $U := \varepsilon^\top M_Z \varepsilon$ and $V := \varepsilon^\top P_Z \varepsilon$. Conditional on Z, $M_Z Z\delta = 0$ and $P_Z Z\delta = Z\delta$. Since M_Z and P_Z are orthogonal idempotent matrices with ranks $T-N$ and N, $U \sim \sigma_e^2 \chi^2_{T-N}$, $V \sim \sigma_e^2 \chi'^2_N(\Lambda_T)$, and $U \perp V$. Therefore $\widehat{\rho}^2_T = U/(U+V) \mid Z \sim \text{Beta}_{\text{nc}}(a,b;\Lambda_T)$. The Poisson-mixture representation follows from the decomposition of a noncentral χ^2 as a Poisson mixture of central χ^2 variables and the induced mixture for the ratio; see, e.g., the noncentral Beta chapter in Johnson et al. (1995) and regression R^2 sampling theory in Muirhead (2009). □

Lemma 7 (Distribution of the noncentrality under serial independence and Gaussianity)**.** *Suppose the rows* (Z_t, ε_t^m) *are mean zero, serially independent over time, and jointly Gaussian, with*

$$\text{Var}(Z_t) = \Sigma_{ZZ} \succ 0, \qquad \text{Var}(\varepsilon_t^m) = \sigma_m^2, \qquad \text{Cov}(Z_t, \varepsilon_t^m) = \Sigma_{Z\varepsilon^m}.$$

Let

$$\delta = \Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m}, \qquad R_m^2 = \frac{\Sigma_{Z\varepsilon^m}^\top \Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m}}{\sigma_m^2}, \qquad \sigma_e^2 = \sigma_m^2(1-R_m^2),$$

and assume $R_m^2 < 1$. *Then*

$$\Lambda_T = \frac{\|Z\delta\|^2}{\sigma_e^2} \stackrel{d}{=} \frac{R_m^2}{1-R_m^2}\chi^2_T.$$

Consequently,

$$\frac{\Lambda_T}{T} \xrightarrow{p} \frac{R_m^2}{1-R_m^2}.$$

Proof. Under joint Gaussianity,

$$\varepsilon_t^m = Z_t^\top \delta + \eta_t,$$

where η_t is independent of Z_t and has variance σ_e^2. The predictable component is $Z_t^\top\delta$. Since Z_t is Gaussian with mean zero and covariance matrix Σ_{ZZ},

$$Z_t^\top \delta \sim \mathcal{N}\big(0, \delta^\top \Sigma_{ZZ}\delta\big).$$

Moreover,

$$\begin{aligned}\delta^\top \Sigma_{ZZ}\delta &= (\Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m})^\top \Sigma_{ZZ}(\Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m}) \\ &= \Sigma_{Z\varepsilon^m}^\top \Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m} = \sigma_m^2 R_m^2.\end{aligned}$$

Thus

$$\frac{Z_t^\top\delta}{\sigma_e} \sim \mathcal{N}\left(0, \frac{R_m^2}{1-R_m^2}\right).$$

The rows are serially independent, so the variables $Z_t^\top \delta/\sigma_e$ are independent Gaussian random variables.

Therefore

$$\Lambda_T = \frac{\|Z\delta\|^2}{\sigma_e^2} = \sum_{t=1}^{T}\left(\frac{Z_t^\top \delta}{\sigma_e}\right)^2 \stackrel{d}{=} \frac{R_m^2}{1-R_m^2}\chi_T^2.$$

Finally, $\chi_T^2/T \xrightarrow{p} 1$ by the law of large numbers, since a chi-square random variable with T degrees of freedom is the sum of T independent squared standard normal random variables. Hence

$$\frac{\Lambda_T}{T} \xrightarrow{p} \frac{R_m^2}{1-R_m^2}.$$

□

Corollary 3 (Bounds for the first moment of $\widehat{\rho}_T^2$)**.** *Under the conditions of Lemma 6,*

$$\frac{T-N}{T+\Lambda_T} \le \mathbb{E}\big[\widehat{\rho}_T^2 \mid Z\big] \le \frac{T-N}{T} \tag{25}$$

Proof. By Lemma 6 and the law of iterated expectations,

$$\mathbb{E}\big[\widehat{\rho}_T^2 \mid Z\big] = \mathbb{E}\Big[\mathbb{E}\big[\widehat{\rho}_T^2 \mid J, Z\big] \,\Big|\, Z\Big] = \mathbb{E}\left[\frac{a}{a+b+J}\right], \qquad J \sim \text{Poisson}\left(\tfrac{\Lambda_T}{2}\right).$$

Upper bound. Since $J \ge 0$ almost surely,

$$\frac{1}{a+b+J} \le \frac{1}{a+b} \quad\Longrightarrow\quad \mathbb{E}\big[\widehat{\rho}_T^2 \mid Z\big] \le \frac{a}{a+b} = \frac{T-N}{T}.$$

Lower bound. The map $x \mapsto 1/x$ is convex on $(0,\infty)$. Hence, by Jensen,

$$\mathbb{E}\left[\frac{1}{a+b+J}\right] \ge \frac{1}{a+b+\mathbb{E}[J]} = \frac{1}{a+b+\Lambda_T/2},$$

so

$$\mathbb{E}\big[\widehat{\rho}_T^2 \mid Z\big] \ge \frac{a}{a+b+\Lambda_T/2} = \frac{(T-N)/2}{T/2+\Lambda_T/2} = \frac{T-N}{T+\Lambda_T},$$

which is the claimed lower bound. □

Corollary 4 (A universal bound for the expected correlation)**.** *Under the conditions of Lemma 6,*

$$\mathbb{E}\big[\widehat{\rho}_T \mid Z\big] \le \sqrt{\mathbb{E}\big[\widehat{\rho}_T^2 \mid Z\big]} \le \sqrt{\frac{T-N}{T}} \tag{26}$$

Proof. The function $x \mapsto \sqrt{x}$ is concave on $[0,1]$, so by Jensen, $\mathbb{E}[\widehat{\rho}_T \mid Z] \le \sqrt{\mathbb{E}[\widehat{\rho}_T^2 \mid Z]}$. Combining this with the upper bound in (25) yields (26). □

Remark 3 (Tightness and special cases). *(i) The upper bound in* (25) *is tight at* $\Lambda_T = 0$ *(the central case* $R^2_m = 0$*), where* $\widehat{\rho}^2_T$ *is* $\mathrm{Beta}((T-N)/2, N/2)$ *with mean* $(T-N)/T$. *(ii) For fixed* (T, N), *the lower bound decreases monotonically in* Λ_T: $\partial\{(T-N)/(T+\Lambda_T)\}/\partial\Lambda_T < 0$. *(iii) The correlation bound* (26) *improves (weakly) with* T *and worsens with* N *but does not depend on* Λ_T.

C PARAMETER VALUES IN MODEL SIMULATIONS

Parameter	Notation	Value
Curvature Kimball aggregator wages	ε_w	10.00
Feedback technology on exogenous spending	ρ_{ga}	0.53
Curvature Kimball aggregator prices	ε_p	10.00
Steady state hours	$\bar{l}$	-0.10
Steady state inflation rate	$\bar{\pi}$	0.82
Time preference rate in percent	$100(\beta^{-1}-1)$	0.16
Coefficient on MA term, wage markup	μ_w	0.89
Coefficient on MA term, price markup	μ_p	0.74
Capital share	α	0.19
Capacity utilization cost	ψ	0.55
Investment adjustment cost	φ	5.49
Depreciation rate	δ	0.03
Risk aversion	σ_c	1.40
External habit degree	λ	0.71
Fixed cost share	ϕ_p	1.61
Indexation to past wages	ι_w	0.59
Calvo parameter wages	ξ_w	0.74
Indexation to past prices	ι_p	0.23
Calvo parameter prices	ξ_p	0.66
Frisch elasticity	σ_l	1.92
Gross markup wages	ϕ_w	1.50
Taylor rule inflation feedback	r_π	2.03
Taylor rule output growth feedback	$r_{\Delta y}$	0.22
Taylor rule output level feedback	r_y	0.08
Interest rate persistence	ρ	0.82
Persistence, productivity shock	ρ_a	0.96
Persistence, risk premium shock	ρ_b	0.18
Persistence, spending shock	ρ_g	0.98
Persistence, investment-specific technology shock	ρ_i	0.71
Persistence, monetary policy shock	ρ_r	0.13
Persistence, price markup shock	ρ_p	0.90
Persistence, wage markup shock	ρ_w	0.97
Net growth rate in percent	$\bar{\gamma}$	0.43
Steady state exogenous spending share	$\frac{\bar{g}}{\bar{y}}$	0.18
Standard deviation, productivity shock	σ_a	0.45
Standard deviation, risk premium shock	σ_b	0.24
Standard deviation, spending shock	σ_g	0.52
Standard deviation, investment-specific technology shock	σ_I	0.45
Standard deviation, monetary policy shock	σ_r	0.24
Standard deviation, price mark-up shock	σ_p	0.14
Standard deviation, wage mark-up shock	σ_w	0.24

Table 2: Parameter values used in simulations of the Smets and Wouters model.