SOME PITFALLS OF INSTRUMENT-BASED INFERENCE IN STRUCTURAL VARS

Christian Matthes* Vladimir Menshikov† Todd B. Walker‡

November 2025

Abstract

Structural VARs are often identified by using instruments derived from the residuals of auxiliary regressions (e.g., Romer and Romer (2004)). We evaluate the finite sample performance of this procedure in a series of Monte Carlo experiments using the Smets and Wouters (2007) model as our laboratory. 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 these issues. In our experimental setting, however, we find that none of these suggested solutions provides a meaningful improvement.

*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 the instruments in macroeconomics are constructed as residuals of auxiliary regressions (Romer and Romer, 2004; Miranda-Agrippino and Ricco, 2021). In this paper, we use the well-known Smets and Wouters (2007) model as a data-generating process to assess the performance of such an identification strategy. In particular, we estimate a correctly specified monetary policy rule and treat a whitened version of that residual as our monetary policy instrument, broadly following the strategy of 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.

In terms of related literature, the closest paper to our 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. In particular, they are interested in possible omitted variable bias when different controls are used in the first and second steps and possible distortions in inference caused by not taking into account estimation uncertainty from the first stage. 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 quantifying the bias induced in estimated responses. Such a bias arises naturally in finite samples, as the residual will almost surely be a function of all structural shocks present in the data-generating process, as we show in this paper.

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 $[\hat{r}_t,\ Z_t^\intercal]^\intercal$, 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,

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

where $\omega_{11} := [\Omega]_{11}$. Hence, the entire path $C_h\Omega e_1$ is identified up to a constant scalar $(1/\omega_{11})$. This scalar factor of proportionality is due to the 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, presence of 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

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

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

[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."

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

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

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-generating process, Z_t, is generated by a (possibly) infinite order moving average*

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

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

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

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

where $h_{0,m}$ is the mth column of H_0.

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

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

We can then compute second moments of our residual instrument $\widehat{r}_t(\widehat{B})$:

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

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

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.

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 true shock 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,

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

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

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

Proof. By ergodicity, $T^{-1} Z^\top Z \to \Sigma_{ZZ}$ and $T^{-1} Z^\top \varepsilon^m \to \Sigma_{Z\varepsilon^m}$; 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 $\widehat{r}_t(\widehat{B}_{\mathrm{OLS}})$ 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

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

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 $\Sigma_{Z\varepsilon^m} = \sigma_m^2 h_{0,m}$ and hence δ and R_m^2; lag coefficients $\{H_\ell : \ell \ge 1\}$ 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 $\widehat{\rho}_T$

We now study the finite-sample correlation between the true shock ε_t^m and the residual-based instrument $\widehat{r}_t$. 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).

Define, as before,

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

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

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

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

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 law and moments. □

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$((T-N)/2, N/2)$ law *does not hold* in general.

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

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

The inequality is strict whenever $0 < N < T$*. Equality holds only in the degenerate case* $N = 0$*, where* $\widehat{\rho}_T \equiv 1$.[3]

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 law, 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 requires no lag feedback)**.** *Suppose ε_t is Gaussian and no lags enter Z_t, i.e. $H_\ell = 0$ for all $\ell \geq 1$ (so $Z_t = H_0\varepsilon_t$). Then, stacking t, $\varepsilon^m = Z\delta + \eta$ with $\eta \sim \mathcal{N}(0, \sigma_e^2 I_T)$ independent of Z, $\sigma_e^2 = \sigma_m^2(1-R_m^2)$. Hence, with $\lambda = \|Z\delta\|^2/\sigma_e^2$,*

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

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 (standard MVN condition-

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

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

ing), hence Lemma 4 applies, giving the noncentral Beta law; cf. Muirhead (2009, §1.2, §1.4), Johnson et al. (1995, Ch. 34). □

We can derive moments of this non-central Beta distribution by exploiting its representation as a mixture of Beta random variables (Johnson et al., 1995) with respect to Poisson random variables. To do so, we let $K \sim \text{Poisson}(\lambda/2)$. This implies that

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

Hence,

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

A complete proof can be found in the appendix. The lower bound follows from Jensen's inequality since $k \mapsto 1/(c+k)$ is convex and decreasing; the upper bound from $K \ge 0$. Thus endogeneity (*larger* λ) depresses the expected finite-sample correlation relative to the central case.

If some $H_\ell \neq 0$ for $\ell \ge 1$, then Z contains lags of ε^m; in the stacked regression, the projector M_Z depends on the same shock vector being projected. Thus $U = (\varepsilon^m)^\top M_Z \varepsilon^m$ and $V = (\varepsilon^m)^\top P_Z \varepsilon^m$ are quadratic forms in a Gaussian vector with *random* idempotents that are functions of that vector; U and V are not independent central/noncentral χ^2 in general, and the noncentral Beta law does not hold. Lemma 4 requires the projectors be fixed (or independent of Y). Again, while these assumptions might seem restrictive, the results are clean and useful bounds for the simulated results below.

3.3 LARGE-DEVIATION TAIL BOUNDS FOR $\widehat{\rho}_T$

Large-deviation (LD) theory describes how quickly sampling probabilities for rare or "unfavorable" outcomes vanish as the sample size T grows. LD rate functions provide simple sample-size benchmarks and are therefore useful in simulation analysis. For example, given a target correlation threshold c and a probability α, LD theory can find the minimum T required to make $\Pr(\widehat{\rho}_T \le c) \le \alpha$. In our context, these bounds help quantify the two effects: (i) how the "dimensionality ratio" (N/T) alone pushes $\widehat{\rho}_T$ below one even in the absence of endogeneity; and (ii) how endogeneity further depresses the typical correlation level.

Define $\kappa_T := N/T$ and assume $\kappa_T \to \kappa \in [0,1)$ as $T \to \infty$. An LD statement of the form

$$\Pr(\widehat{\rho}_T \le c) \approx \exp\{-T\,I(c)\} \tag{18}$$

means that tail probabilities shrink *exponentially* fast with sample size, at a rate governed by a convex "rate function" $I(\cdot)$: larger $I(c)$ implies faster decay and, therefore, greater reliability for a given T. Formally, such results follow from standard LD tools such as the *Laplace principle* or the *Gärtner–Ellis theorem*; see Dembo and Zeitouni (1998, Ch. 2–3) or Boucheron et al. (2013, Ch. 2).

Under the conditions of Proposition 3 (independence of Z and ε^m and spherical ε^m), $\widehat{\rho}_T^2$ satisfies an

LD principle on $(0,1)$ with speed T and rate function

$$I_\kappa(x) \;=\; \tfrac{1}{2}\left[(1-\kappa)\log\frac{1-\kappa}{x} + \kappa\log\frac{\kappa}{1-x}\right], \qquad x \in (0,1), \tag{19}$$

so that, for any $\varepsilon > 0$,

$$\Pr\Big(|\widehat{\rho}_T - \sqrt{1-\kappa}| > \varepsilon\Big) \lesssim \exp\Big\{-T \inf_{|r-\sqrt{1-\kappa}|>\varepsilon} I_\kappa(r^2)\Big\} \tag{20}$$

This is a standard result for Beta-type statistics (e.g. Muirhead, 2009; Dembo and Zeitouni, 1998). The dimensionality factor ($\kappa = N/T$) pulls the typical correlation down to $\sqrt{1-\kappa}$ and controls the exponential decay of tail probabilities. These LD bounds formalize and extend the finite-sample findings above. The earlier Beta-distribution results provide exact finite-T expectations such as $\mathbb{E}[\widehat{\rho}_T^2] = (T-N)/T$, showing that $\widehat{\rho}_T$ falls below one *in expectation.* The LD results show that this behavior *persists beyond expectation* in that deviations from the typical value $\sqrt{1-\kappa}$ become exponentially rare as T grows.

The same logic carries over to the endogenous case. To isolate this effect, consider the standard fixed-N regime (so $\kappa \to 0$). The LD rate now centers at the population limit $\sqrt{1-R_m^2}$ and quantifies how endogeneity sharpens the exponential concentration of $\widehat{\rho}_T$. Under the conditions of Proposition 4 (Gaussianity, $Z_t = H_0\varepsilon_t$), $\widehat{\rho}_T^2$ obeys an LDP with speed T and a convex rate $I_{\kappa,\ell}(\cdot)$ described by the (non-central) mgfs of the underlying quadratic forms and a contraction step (Johnson et al., 1995, Ch. 34); see also Dembo and Zeitouni (1998, Chs. 2–3). Thus, the results for the endogenous case are qualitatively identical to those for the independent case. The only change is that the typical or "most probable" value of $\widehat{\rho}_T$ moves from 1 (the value in the independent, $\kappa = 0$ case) to $\sqrt{1-R_m^2}$, reflecting the fact that endogeneity acts as a distinct source of attenuation. The rate functions $I_\kappa(\cdot)$ and $I_{\kappa,\ell}(\cdot)$ share the same exponential form and depend on T only through the scaling factor with both delivering probabilities of the shape $\Pr(\widehat{\rho}_T \le c) \approx e^{-T\times I(c)}$. Consequently, deviations from $\sqrt{1-R_m^2}$ become exponentially rare at the same speed as in the independence case, but with a steeper rate that increases in the noncentrality parameter $\ell = R_m^2/(1-R_m^2)$.

Under Assumption 1, where Z_t may depend on past shocks, the same large-deviation logic applies. The empirical quadratic forms $(T^{-1}Z^\top Z,\ T^{-1}Z^\top\varepsilon^m,\ T^{-1}(\varepsilon^m)^\top\varepsilon^m)$ satisfy the standard LDP by the Gärtner–Ellis theorem, and the contraction principle implies that $\widehat{\rho}_T$ inherits a rate function $J(\cdot)$ whose unique minimum is again at $\sqrt{1-R_m^2}$. Although J has no simple closed form, its role is the same in that it governs the speed at which the sampling distribution of $\widehat{\rho}_T$ collapses around its population value. The existence of such a rate function justifies the use of $-T^{-1}\log\Pr(\widehat{\rho}_T \le c)$ as an empirical diagnostic in Monte Carlo simulations, since this quantity converges to $J(c)$ as T grows.

Across all three settings–independence, static endogeneity, and dynamic feedback–finite-sample randomness and model dependence affect the *location* of $\widehat{\rho}_T$ but not the *speed* at which it concentrates. Large-deviation principles quantify this concentration explicitly, showing that the apparent fragility of instrument correlations in small samples decays exponentially in T in all three cases. Large-deviation reasoning complements traditional finite-sample Monte Carlo analysis by adding a quantitative measure of *how quickly* small sample irregularities vanish as the available data length increases. In all of our cases studied above, we have exponential decay.

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. 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 0.23 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.[4]

[4] 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.

We know from Plagborg-Møller and Wolf (2021) that with an internal instrument approach we can only identify relative impulse responses due to the presence of measurement error in the instrument. Our exercises abstract from measurement error that is independent of the structural shocks. We therefore present impulse responses to a positive one-standard deviation innovation in the instrument and compare those IRFs to the true impulse responses to a one-standard deviation monetary shock in the Smets and Wouters (2007) model.

	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 monetary shock 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.

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, dotted) are constructed as percentiles of OLS point estimates from the various samples. The left panel 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 right panel of Figure 1 shows results for a larger, seven-variable VAR. This change 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 ($R_m^2 > 0$), the OLS instrument $\hat{r}_t$ is a contaminated proxy ($r_t^\star$) 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 confidence bands, which is also consistent with our theory (Proposition 3): increasing the number of variables N for a fixed sample T increases the dimensionality ratio $\kappa = N/T$, adding to estimation uncertainty.

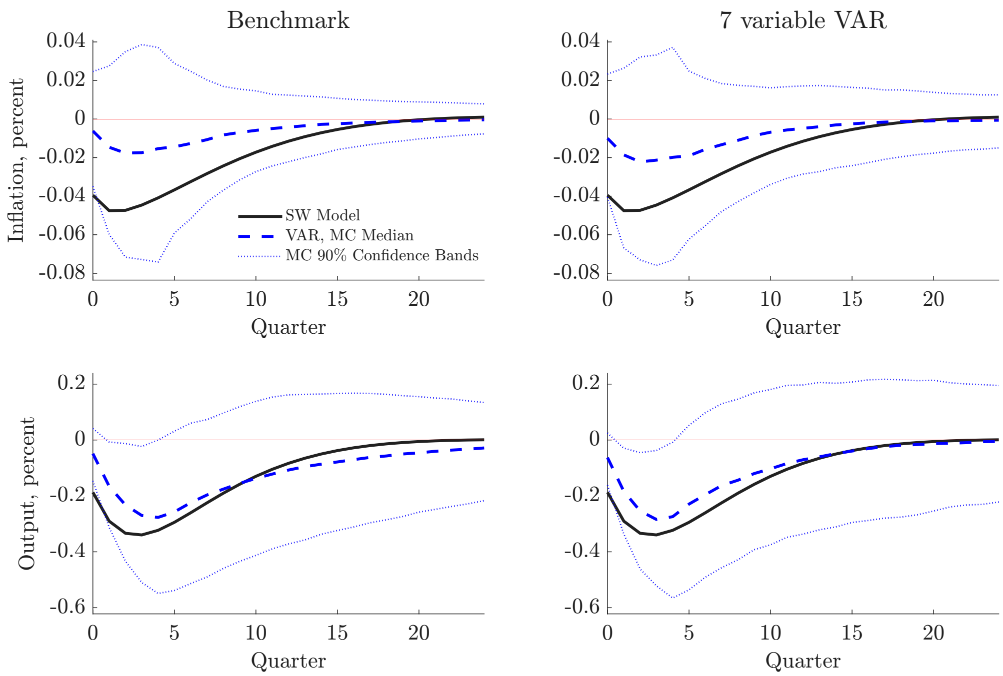

Figure 1: IRFs to one standard deviation monetary policy shock identified via estimated instrument. Benchmark three variable VAR vs seven variable VAR.

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.[5] 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 precisely the "dimensionality curse" principle explored in Proposition 3 and the LDP bounds. Just as a high $\kappa = N/T$ ratio

[5] 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.

degrades the instrument's quality, a high parameter-to-observation ratio in the VAR estimation causes the dynamic estimates (A_ℓ) to become 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.

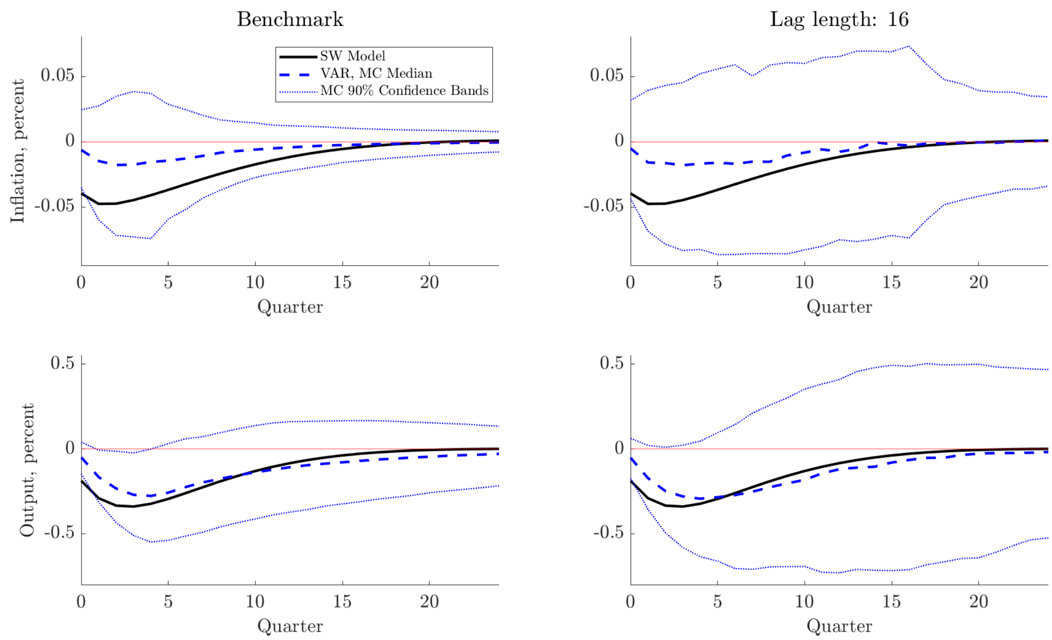

Figure 2: IRFs to one standard deviation monetary policy shock identified via estimated instrument. Benchmark three variable VAR with four lags vs three variable VAR with 16 lags.

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. (Results available upon request.)

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 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 exact 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.

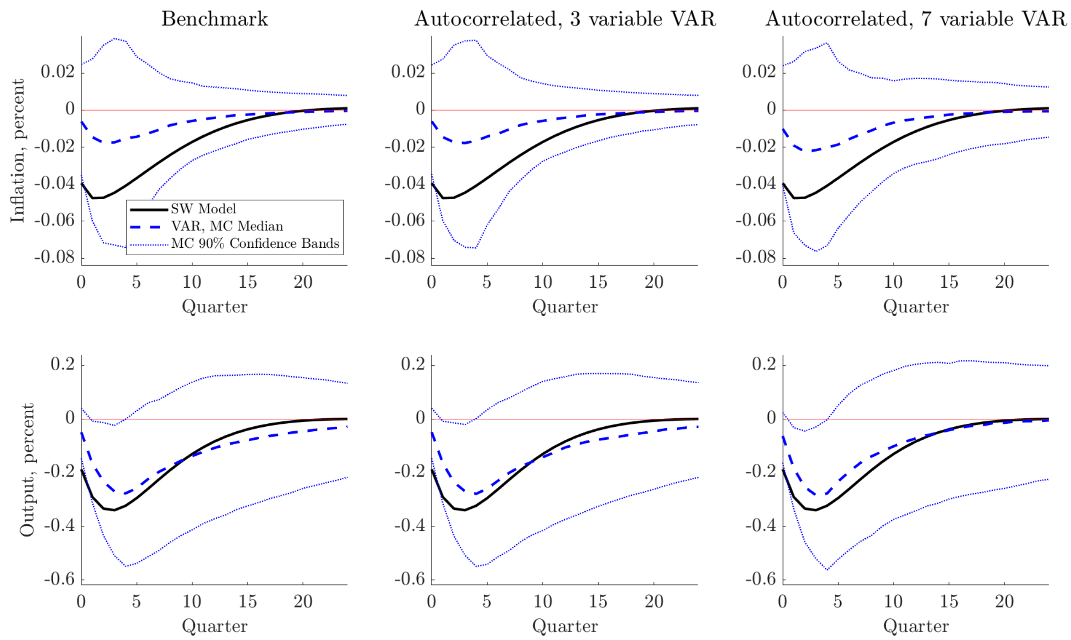

Figure 3: IRFs to one standard deviation monetary policy shock identified via estimated instrument. Benchmark three variable VAR with whitened instrument vs VARs with autocorrelated instruments.

4.1.1 REMOVING 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:

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

By construction, the monetary shock is now contemporaneously orthogonal to the variables in the rule, forcing $R^2_m = 0$. 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 our LDP bounds. Even with endogeneity solved, the "dimensionality curse" ($\kappa = N/T > 0$) remains. This finite-sample error ensures that the instrument in any single sample is still noisy, leading to wide dispersion.

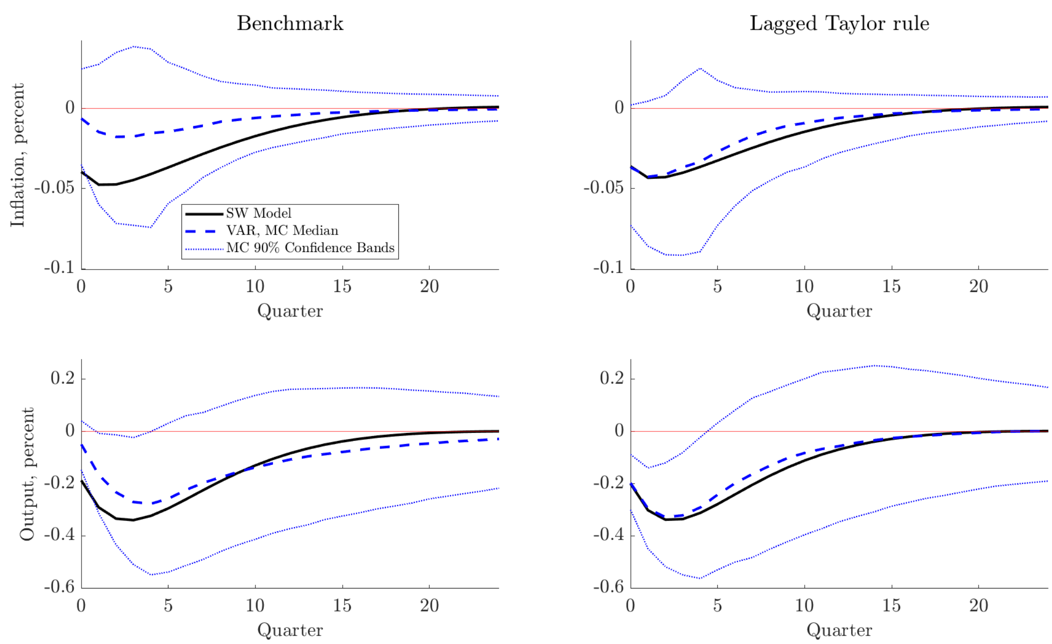

Figure 4: IRFs to one standard deviation monetary policy shock identified via estimated instrument. Benchmark three variable VAR vs 3 variable VAR for data-generating process with backward-looking monetary policy rule.

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). Instrument, while highly correlated with "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.

REFERENCES

Boucheron, Stéphane, Gábor Lugosi, and Pascal Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, 2013.

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.

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.

Dembo, Amir and Ofer Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer, 1998.

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

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.

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.

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.

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

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 $\operatorname{col}(Z)$ and M_Z onto $\operatorname{col}(Z)^\perp$. 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 $\operatorname{rank}(A) = r$, 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*

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

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*

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

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,*

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

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 $S_{D\cdot A} \geq 0$ implies* $0 \leq R^2 \leq 1$*, 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 $Z_t = \sum_{\ell\geq 0} H_\ell \varepsilon_{t-\ell}$ with independent components and serial indepen-

dence of (ε_t)*,*

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

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

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

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

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

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

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

hence

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

Lemma 6 (Noncentral-Beta via Poisson mixture)**.** *Let* $Z \in \mathbb{R}^{T\times N}$ *have full column rank* $N < T$*. Let* $\varepsilon = (\varepsilon_1^m, \ldots, \varepsilon_T^m)^\top$ *be jointly Gaussian with* Z *and write*

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

Let $a := (T-N)/2$, $b := N/2$, $\delta := \Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon^m}$, $\sigma_e^2 := \mathrm{Var}(\varepsilon_t^m)(1 - R_m^2)$*, and* $\lambda := \|Z\delta\|^2/\sigma_e^2 \geq 0$*. Then, conditional on* Z*,*

$$\widehat{\rho}_T^2 \mid Z \;\sim\; \mathrm{Beta}_{\mathrm{nc}}(a,\; b;\; \lambda) \quad \textit{and} \quad \widehat{\rho}_T^2 \;\overset{d}{=}\; Y_K,\;\; Y_k \sim \mathrm{Beta}(a, b+k),\;\; K \sim \mathrm{Poisson}\Big(\frac{\lambda}{2}\Big),$$

with K *independent of* $\{Y_k\}_{k\geq 0}$*.*

Proof. Write $U := \varepsilon^\top M_Z \varepsilon$ and $V := \varepsilon^\top P_Z \varepsilon$. Under joint normality and orthogonality of M_Z and P_Z, $U \sim$

$\sigma_e^2\chi^2_{T-N}$, $V \sim \sigma_e^2\chi'^2_N(\lambda)$, and $U \perp V$. Therefore $\hat\rho_T^2 = U/(U+V) \mid Z \sim \mathrm{Beta}_{\mathrm{nc}}(a,b;\lambda)$. 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). $\square$

Corollary 3 (Bounds for the first moment of $\hat\rho_T^2$)**.** *Under the conditions of Lemma 6,*

$$\frac{T-N}{T+\lambda} \le \mathbb{E}\big[\hat\rho_T^2 \,|\, Z\big] \le \frac{T-N}{T} \tag{21}$$

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

$$\mathbb{E}\big[\hat\rho_T^2 \mid Z\big] = \mathbb{E}\Big[\mathbb{E}\big[\hat\rho_T^2 \mid K, Z\big] \,\Big|\, Z\Big] = \mathbb{E}\Big[\frac{a}{a+b+K}\Big], \qquad K \sim \mathrm{Poisson}\Big(\frac{\lambda}{2}\Big).$$

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

$$\frac{1}{a+b+K} \le \frac{1}{a+b} \quad\Longrightarrow\quad \mathbb{E}\big[\hat\rho_T^2 \mid Z\big] \le \frac{a}{a+b} = \frac{T-N}{T}.$$

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

$$\mathbb{E}\Big[\frac{1}{a+b+K}\Big] \ge \frac{1}{a+b+\mathbb{E}[K]} = \frac{1}{a+b+\lambda/2},$$

so

$$\mathbb{E}\big[\hat\rho_T^2 \mid Z\big] \ge \frac{a}{a+b+\lambda/2} = \frac{(T-N)/2}{T/2+\lambda/2} = \frac{T-N}{T+\lambda},$$

which is the claimed lower bound. $\square$

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

$$\mathbb{E}\big[\hat\rho_T \,|\, Z\big] \le \sqrt{\mathbb{E}\big[\hat\rho_T^2 \mid Z\big]} \le \sqrt{\frac{T-N}{T}} \tag{22}$$

Proof. The function $x \mapsto \sqrt{x}$ is concave on $[0,1]$, so by Jensen, $\mathbb{E}[\hat\rho_T \mid Z] \le \sqrt{\mathbb{E}[\hat\rho_T^2 \mid Z]}$. Combining this with the upper bound in (21) yields (22). $\square$

Remark 3 (Tightness and special cases)**.** *(i) The upper bound in* (21) *is tight at* $\lambda = 0$ *(the central case* $R_m^2 = 0$*), where* $\hat\rho_T^2$ *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*:* $\partial\{(T-N)/(T+\lambda)\}/\partial\lambda < 0$*. (iii) The correlation bound* (22) *improves (weakly) with* T *and worsens with* N *but does not depend on* λ*.*

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, risk premium 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.