Federal Reserve Bank of Dallas

Online Appendix to Dynamic Identification Using System Projections on Instrumental Variables

Daniel J. Lewis and Karel Mertens

Working Paper 2204 Appendix   October 2023 (Revised July 2024)

Research Department

https://doi.org/10.24149/wp2204appr1

Dynamic Identification Using System Projections on Instrumental Variables

Daniel Lewis Karel Mertens

ONLINE APPENDIX

Contents

I Testing the Null Hypothesis of Weak Instruments

This section describes the weak instruments test in the SP-IV model discussed in Section 2.2 of the main text. The test nests the popular bias-based test of Stock and Yogo (2005) when $H = 1$. The development of the test is analogous to that of the weak instruments test in Lewis and

Mertens (2022), which extends the Stock and Yogo (2005) test to be robust to autocorrelation and heteroskedasticity. Mathematically, the extension of the Stock and Yogo (2005) test in Lewis and Mertens (2022) closely resembles the extension required for SP-IV to allow $H > 1$.

We first establish some specific notation: $||U||_2$ is the spectral norm of U (the positive square root of the maximum eigenvalue of UU', also the ℓ_2-norm if U is a vector), $\mathbb{P}^n$ is the set of positive definite $n \times n$ matrices, $\mathbb{O}^{n\times m}$ is the set of $n \times m$ orthogonal real matrices U such that $UU' = I_n$, $\mathcal{K}_{n,m}$ denotes the $n \times m$ commutation matrix such that $\mathcal{K}_{n,m} \operatorname{vec}(U) = \operatorname{vec}(U')$ where $U \in \mathbb{R}^{n\times m}$. We also define the special matrix $R_{n,m} = I_n \otimes \operatorname{vec}(I_m)$. The dimension of $R_{n,m}$ is $nm^2 \times n$. For $U \in \mathbb{R}^{nm\times nm}$, the (i, j)-th element of $V = R'_{n,m}(U \otimes I_m)R_{n,m} \in \mathbb{R}^{n\times n}$ is $\operatorname{Tr}(U_{ij})$ where $U_{ij} \in \mathbb{R}^{m\times m}$ is (i, j)-th block of U and $\operatorname{Tr}(\cdot)$ is the trace. For $U \in \mathbb{R}^{nm\times m}$, the i-th element of $V = R'_{n,m} \operatorname{vec}(U') \in \mathbb{R}^n$ is equal to $\operatorname{Tr}(U_i)$ where $U_i \in \mathbb{R}^{m\times m}$ is the i-th row block of U. Note that $R'_{n,m}R_{n,m} = mI_N$.

I.1 Weak IV Representation of the SP-IV Estimator

Using the more general notation for the restriction matrix R defined above, the SP-IV estimator is

$$\text{(I.1)}\quad \hat{\beta} = \left(R'_{K,H}(Y^{\perp}_{H} P_{Z^\perp} Y^{\perp\prime}_{H} \otimes I_H)R_{K,H}\right)^{-1} R'_{K,H} \operatorname{vec}(y^{\perp}_{H} P_{Z^\perp} Y^{\perp\prime}_{H}) ,$$

where $P_{Z^\perp} = Z^{\perp\prime}(Z^\perp Z^{\perp\prime})^{-1}Z^\perp$. As is standard in the literature – see, e.g., Staiger and Stock (1997) – we assume identification but first-stage parameters that are local-to-zero.

Assumption 4. *$\Theta_Y = C/\sqrt{T}$ where $C \in \mathbb{R}^{HK\times N_z}$ is a fixed matrix and $R_{K,H}(CC' \otimes I_H)R_{K,H}$ is of full rank.*

This assumption implies that the instruments are weak under the null hypothesis. The following replace Assumptions 2 and 3 to allow the characterization of the weak instrument asymptotic distribution of $\hat{\beta}$.

Assumption 5. *The following limits hold as $T \to \infty$:*

(5.a) $$u_H^{\perp} u_H^{\perp\prime}/T \xrightarrow{p} \Sigma_{u_H^{\perp}} \in \mathbb{P}^{H},$$
$$u_H^{\perp} v_H^{\perp\prime}/T \xrightarrow{p} \Sigma_{u_H^{\perp} v_H^{\perp}} \in \mathbb{R}^{H\times HK},$$
$$v_H^{\perp} v_H^{\perp\prime}/T \xrightarrow{p} \Sigma_{v_H^{\perp}} \in \mathbb{P}^{HK},$$

(5.b) $$T^{-\frac{1}{2}} \begin{bmatrix} \text{vec}((Z^{\perp}Z^{\perp\prime})^{-\frac{1}{2}} Z^{\perp} w_H^{\perp\prime}) \\ \text{vec}((Z^{\perp}Z^{\perp\prime})^{-\frac{1}{2}} Z^{\perp} v_H^{\perp\prime}) \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, \mathbf{W} \otimes I_{N_z}) \ ,$$

(5.c) *and* $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$

where $\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_{12} \\ \mathbf{W}_{12}' & \mathbf{W}_2 \end{bmatrix} \in \mathbb{P}^{(K+1)H}$.

$w_H^{\perp} = y_H^{\perp} - (\beta' \otimes I_H)\Theta_Y Q^{-\frac{1}{2}} Z^{\perp}$ are the reduced-form errors with covariance matrix $\mathbf{W}_1$, $v_H^{\perp} = Y_H^{\perp} - \Theta_Y Q^{-\frac{1}{2}} Z^{\perp}$ are first-stage error terms with covariance matrix $\mathbf{W}_2$, and $\mathbf{W}$ is the joint covariance of the reduced-form and first-stage errors.

The SP-IV estimator can be rewritten as

(I.2) $$\hat{\beta} = \left(R_{K,H}'(s_{ZY}s_{ZY}' \otimes I_H)R_{K,H}\right)^{-1} R_{K,H}' \,\text{vec}(s_{Zy}s_{ZY}').$$

where $s_{Zy} = y_H^{\perp} Z^{\perp\prime}(Z^{\perp}Z^{\perp\prime})^{-\frac{1}{2}}$ and $s_{ZY} = Y_H^{\perp} Z^{\perp\prime}(Z^{\perp}Z^{\perp\prime})^{-\frac{1}{2}}$. This alternative expression reformulates $\hat{\beta}$ in terms of random vectors with asymptotic distributions given in Assumption 5. Define the random variables η_1 and η_2 ($H \times N_z$ and $HK \times N_z$ respectively) as

(I.3) $$\begin{bmatrix} \text{vec}(\eta_1) \\ \text{vec}(\eta_2) \end{bmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0}_{HN_z} \\ \text{vec}(C) \end{pmatrix}, \mathbf{S} \otimes I_{N_z} \right)$$

where $\mathbf{S} \in \mathbb{P}^{(K+1)H}$, partitioned as $\mathbf{W}$ with

(I.4)
$$\mathbf{S}_1 = \mathbf{W}_1 + (\beta' \otimes I_H)\mathbf{W}_2(\beta \otimes I_H) - (\beta' \otimes I_H)\mathbf{W}_{12}' - \mathbf{W}_{12}(\beta \otimes I_H),$$
$$\mathbf{S}_{12} = \mathbf{W}_{12} - (\beta' \otimes I_H)\mathbf{W}_2 \ , \ \mathbf{S}_2 = \mathbf{W}_2,$$

such that $\mathbf{S} \otimes I_{N_z}$ is the asymptotic covariance of $T^{-\frac{1}{2}} \left[\ \text{vec}(u_H^{\perp} Z^{\perp\prime}(Z^{\perp}Z^{\perp\prime})^{-\frac{1}{2}})' \quad \text{vec}(Y_H^{\perp} Z^{\perp\prime}(Z^{\perp}Z^{\perp\prime})^{-\frac{1}{2}})' \ \right)'$. Proposition 6 then characterizes the distribution of the random variable $\beta^* = \hat{\beta} - \beta$.

Proposition 6. *Under Assumptions 4 and 5,* $s_{ZY} \xrightarrow{d} \eta_2$ *and* $s_{Zy} \xrightarrow{d} (\beta' \otimes I_H)\eta_2 + \eta_1$*, and thus*

$$\hat{\beta} - \beta \xrightarrow{d} \beta^* = \left(R'_{K,H}(\eta_2\eta_2' \otimes I_H)R_{K,H}\right)^{-1} R'_{K,H}\,\mathrm{vec}(\eta_1\eta_2').$$

Proof. The results follow directly from the stated assumptions, the expression for $\hat{\beta}$ in (I.2), and the continuous mapping theorem. □

Since β^* converges to a quotient of quadratic forms in normal random variables, $\hat{\beta}$ is not a consistent estimator of β. The asymptotic bias of the SP-IV estimator is the expected value $E[\beta^*]$. Before introducing the weak instruments set, we define the concentration matrix for the model.

Definition 1. *The concentration matrix is* $\Lambda = \frac{1}{N_z}\Phi^{-\frac{1}{2}}R_{K,H}(CC' \otimes I_H)R_{K,H}\Phi^{-\frac{1}{2}}$ *where* $\Phi = R'_{K,H}(\mathbf{S}_2 \otimes I_H)R_{K,H}$*.*

I.2 Definition of Weak Instruments

We consider instruments weak when a weighted ℓ_2-norm of the asymptotic bias $E[\beta^*]$ is large relative to a worst-case benchmark.

Definition 2. *The bias criterion is* $B = \mathrm{Tr}(\mathbf{S}_1)^{-\frac{1}{2}}||E\left[\beta^*\right]' \Phi^{\frac{1}{2}}||_2$*.*

Following Stock and Yogo (2005), the ℓ_2-norm in the bias criterion aggregates the K elements of the bias through a quadratic loss function, such that B is weakly positive and penalizes larger biases more heavily. The criterion applies a weighting matrix, Φ, to put the elements of $E\left[\beta^*\right]$ on a comparable scale. The weighting matrix Φ effectively standardizes the regressors in the second stage so that they have unit standard deviations and are orthogonal. The bias criterion also scales by $\mathrm{Tr}(\mathbf{S}_1)$, which is the probability limit of $T^{-1}u_H^{\perp}P_{Z^{\perp}}u_H^{\perp\prime}$. This scaling expresses B as a ratio, relative to the same worst-case bias as in Montiel-Olea and Pflueger (2013), and Lewis and Mertens (2022). The intuition for the worst-case bias is given by the ad-hoc approximation of $E[\beta^*]$ in terms of a ratio of expectations as in Staiger and Stock (1997):

$$E[\beta^*] \approx \frac{\text{vec}(\mathbf{S}_{12})' R_{K,H} \Phi^{-\frac{1}{2}}}{\text{Tr}(\mathbf{S}_1)^{\frac{1}{2}}} (I_K + \Lambda)^{-1} \Phi^{-\frac{1}{2}} \,\text{Tr}(\mathbf{S}_1)^{\frac{1}{2}} \tag{I.5}$$

Using this approximation, the bias criterion in (2) reaches a maximum of unity when the errors $u_H^\perp$ are perfect linear combinations of the second-stage regressors, $v_H^\perp P_Z^\perp$, such that the first term in (I.5) is a $K \times 1$ unit vector, and when the instruments are completely uninformative so the concentration matrix, Λ, is zero.

Definition 3. *The weak instrument set is*

$$\mathbb{B}_\tau(\mathbf{W}) = \{C \in \mathbb{R}^{N\times K}, \beta \in \mathbb{R}^N : B \geq \tau\}. \tag{I.6}$$

The weak instrument set is the set of values for β and the first-stage parameters C such that bias B exceeds a tolerance level τ. This set depends on $\mathbf{W}$, which can be consistently estimated, but also on C, and the K unknown parameters in β.

I.3 Characterizing the Boundary of the Weak Instrument Set

Under Assumptions 4 & 5, the bias criterion in Definition 2 can be decomposed as $B = ||\mathbf{h}\rho||_2$, where

$$\mathbf{h} = HE\left[\left(R'_{K,H}(\mathcal{S}(l+\psi)(l+\psi)'\mathcal{S}' \otimes I_H)R_{K,H}\right)^{-1} R'_{K,H}\left(\mathcal{S}(l+\psi)\psi'\mathcal{S}^{-1} \otimes I_H\right)\right],$$
$$\rho = \left(\Phi^{-\frac{1}{2}} \otimes I_{H^2}\right) \text{vec}\left(\mathbf{S}_{12}\right) / \sqrt{\text{Tr}(\mathbf{S}_1)}\ ,$$

$l = \mathbf{S}_2^{-\frac{1}{2}} C$, $\psi = \mathbf{S}_2^{-\frac{1}{2}}(\eta_2 - C)$, $\text{vec}(\psi) \sim \mathcal{N}(0, I_{KHN_z})$, and $\mathcal{S} = ((\Phi/H)^{-\frac{1}{2}} \otimes I_H)\mathbf{S}_2^{\frac{1}{2}}$. This decomposition is analogous to that of Lemma 1 in Lewis and Mertens (2022). The matrix $\mathbf{h}$ is the expected value of a random matrix that is a function of ψ, a matrix with i.i.d standard normal variables as elements. This expected value – when it exists – also depends on location parameters C and on $\mathbf{W}_2$. The vector ρ depends on $\mathbf{W}$ and β. In general, there is no tractable analytical expression for the integral underlying the expectation in $\mathbf{h}$, which is required to evaluate the bias. Following

Montiel-Olea and Pflueger (2013) and Lewis and Mertens (2022), we adopt a Nagar (1959) approximation to $\mathbf{h}$ around $\psi = 0$, which we denote by $\mathbf{h}_n$. The Nagar bias is defined as $B_n = ||\mathbf{h}_n \rho||_2$. Using the eigenvalue decomposition $\Lambda = Q_\Lambda \mathcal{D}_\Lambda Q'_\Lambda$, the Nagar approximation of $\mathbf{h}$ around $\psi = 0$ is given by

(I.7)

$$\mathbf{h}_n = N_z^{-1} Q_\Lambda \mathcal{D}_\Lambda^{-\frac{1}{2}} M_1 (\mathcal{D}_\Lambda^{-\frac{1}{2}} Q_\Lambda \otimes L_0 \otimes I_K)(I_{KH} \otimes (I_{N_z} \otimes L_0)\mathcal{K}_{N_z,HN_z} R_{H,N_z}) M_2$$

with $L_0 = HN_z^{-\frac{1}{2}} Q'_\Lambda \Lambda^{-\frac{1}{2}} R'_{K,HN_z}(\mathcal{S}\,\mathrm{vec}(l) \otimes I_{HN_z}) \in \mathbb{O}^{K\times HN_z}$, $M_1 = R'_{K,K}\big(I_{K^3} + (\mathcal{K}_{K,K} \otimes I_K)\big)$ and $M_2 = R_{K,H} R'_{K,H}/(K+1) - I_{KH^2}$.

Analogous to Lewis and Mertens (2022), we base our test on the bound

$$B_n \leq \lambda_{\min}^{-1} \mathcal{B}(\mathbf{W}) \; , \tag{I.8}$$

where $\lambda_{\min} = \mathrm{mineval}\{\Lambda\}$ and

(I.9)

$$\mathcal{B}(\mathbf{W}) = (N_z\sqrt{H})^{-1} \sup_{L_0} \{||M_1(I_K \otimes L_0 \otimes I_K)(I_{KH} \otimes (I_{N_z} \otimes L_0)\mathcal{K}_{N_z,HN_z} R_{H,N_z}) M_2 \Psi||_2\} \; ,$$

(I.10)

$$\Psi = \big(\mathcal{S}\mathbf{W}_2^{-\frac{1}{2}} [\mathbf{W}_{12} : \mathbf{W}_2]' \otimes I_H) R_{K+1,H} (R'_{K+1,H}(\mathbf{W} \otimes I_H) R_{K+1,H})^{-\frac{1}{2}} \; .$$

I.4 Null Hypothesis

Given a bias tolerance level τ, the test of the null hypothesis of weak instruments is based on a test of whether the minimum eigenvalue of Λ is less than or equal to a threshold value $\lambda^*_{\min}(\tau)$. More formally, the null and alternative hypotheses for the test are

$$\begin{aligned} &H_0 : \lambda_{\min} \in \mathcal{H}(\mathbf{W}) \;\; \text{vs.} \;\; H_1 : \lambda_{\min} \notin \mathcal{H}(\mathbf{W}), \\ &\text{where} \;\; \mathcal{H}(\mathbf{W}) = \{\lambda_{\min} \in \mathbb{R}_+ : \lambda_{\min} \leq \lambda^*_{\min}(\tau)\}, \end{aligned} \tag{I.11}$$

where $\lambda^*_{\min}(\tau) = \mathcal{B}(\mathbf{W})/\tau$. The null hypothesis is that the minimum eigenvalue of the concentration matrix is in the set of values for which the worst-case Nagar bias is greater than the tolerance level τ. Under the alternative, the minimum eigenvalue is not in that set of values.

I.5 Test Statistic and Critical Values

The following proposition presents our statistic to test the null hypothesis.

Proposition 7. *Define the test statistic*

$$g = N_z^{-1}\,\text{mineval}\{\hat{\Phi}^{-\frac{1}{2}}(Y_H^{\perp}P_{Z^{\perp}}Y_H^{\perp\prime})\hat{\Phi}^{-\frac{1}{2}}\},$$

where $\hat{\Phi} = R'_{K,H}(\hat{\mathbf{W}}_2 \otimes I_H)R_{K,H}$*. Then, under Assumptions 4 and 5,*

$$g \xrightarrow{d} \text{mineval}\{R'_{K,H}(\zeta \otimes I_K)R_{K,H}/(HN_z)\},$$

where the $KH \times KH$ *random matrix* $\zeta = \mathcal{S}(l+\psi)(l+\psi)'\mathcal{S}'$ *has a noncentral Wishart distribution,* $\zeta \sim \mathcal{W}(N_z, \Sigma, \Omega)$*, with* N_z *degrees of freedom, covariance matrix* $\Sigma = \mathcal{S}\mathcal{S}' \in \mathbb{P}^{KH}$*, and a matrix of noncentrality parameters* $\Omega = \Sigma^{-1}\mathcal{S}ll'\mathcal{S}'$*.*[18]

Proof. The proposition follows from Slutsky's theorem, the continuous mapping theorem, and $Y_H^{\perp}P_{Z^{\perp}}Y_H^{\perp\prime} \xrightarrow{d} R'_{K,H}\left(\mathbf{S}_2^{\frac{1}{2}}(l+\psi)(l+\psi)'\mathbf{S}_2^{\frac{1}{2}\prime} \otimes I_K\right)R_{K,H}$, which implies the stated distribution of ζ. □

While ζ has a noncentral Wishart distribution, critical values for the test statistic g require the distribution of $\text{mineval}\{R'_{K,H}(\zeta \otimes I_H)R_{K,H}\}$, which is the minimum eigenvalue of the $K \times K$ matrix consisting of the traces of the $H \times H$ partitions of ζ. To the best of our knowledge, the distribution of this function of ζ is unknown. Moreover, the limiting distribution of g depends in general on all parameters in Σ and Ω, not just on the threshold for $\lambda_{\min}$.

To address both these challenges, we follow Stock and Yogo (2005) and Lewis and Mertens (2022) and obtain critical values from a bounding limiting distribution of g. Specifically, we consider the distribution of $\gamma' R'_{K,H}(\zeta \otimes I_H)R_{K,H}\gamma \geq \text{mineval}\{R'_{K,H}(\zeta \otimes I_H)R_{K,H}\}$ as a bounding distribution, where γ is the eigenvector associated with the minimum eigenvalue

[18] We adopt the notational convention of Muirhead (1982) for the noncentral Wishart distribution.

of Λ and $\gamma'\gamma = 1$. The following theorem is a straightforward extension of Theorem 2 in Lewis and Mertens (2022).

Theorem 1. *For* $\zeta \sim \mathcal{W}(N_z, \Sigma, \Omega)$*,*

(i) The n-th cumulant of $\gamma' R'_{K,H}(\zeta \otimes I_H) R_{K,H}\gamma$ *is*

$$\kappa_n = \; 2^{n-1}(n-1)!\Big(N_z \operatorname{Tr}\left(((\gamma\gamma' \otimes I_H)\Sigma)^n\right) + n \operatorname{Tr}\left(((\gamma\gamma' \otimes I_H)\Sigma)^n \Omega\right)\Big).$$

(ii) The n-th cumulant κ_n *with* $n > 1$ *is bounded by*

$$\begin{aligned}\kappa_n \leq & 2^{n-1}(n-1)!\Big(N_z \operatorname{maxeval}\{R'_{K,H}(\Sigma^n \otimes I_H)R_{K,H}\} \\ & \qquad + nHN_z\lambda_{\min} \operatorname{maxeval}\{\Sigma\}^{n-1}\Big).\end{aligned}$$

Proof. See Lewis and Mertens (2022). □

As in Lewis and Mertens (2022), we consider the class of approximating distributions proposed by Imhof (1961), which match the first three cumulants of an unknown target distribution. We select the Imhof distribution with the largest critical value at significance level α subject to the constraints that the first cumulant, $\kappa_1 = HN_z(1 + \lambda_{\min})$, matches that of the target distribution, and that the second and third cumulants respect the analytical upper bounds on the cumulants of the limiting distribution of g. The resulting critical value is guaranteed to be conservative relative to the unknown critical value from the true limiting distribution of the test statistic, g.

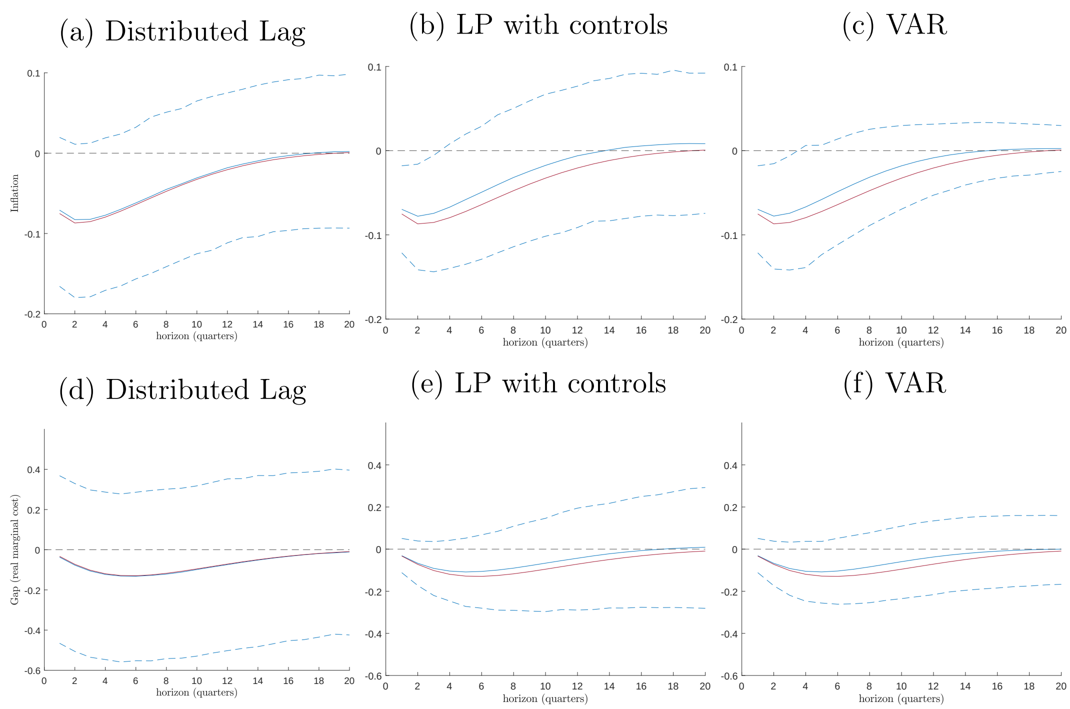
II Additional Simulation Results

II.1 IRF Estimates in the Simulations

Figures II.1 and II.2 show the true model impulse responses to a one s.t.d. contractionary monetary policy shock, together with the mean IRF estimates and 2.5% and 97.5% percentiles, across 5000 simulations from the Smets and Wouters (2007) model discussed in Section 3. The columns show IRFs estimated using a distributed lag specification, local projections with the set of control variables X_{t-1} described in the main text, and a

VAR in X_t with four lags. The top rows in each Figure show the IRFs of inflation, whereas the bottom rows show the IRFs of the output gap (real marginal cost). For brevity, we only show the IRFs associated with the monetary policy shock for $H = 20$ quarters. Results for the other specifications are available on request.

FIGURE II.1: True and Estimated IRFs in Simulations, Small Sample ($T = 250$)

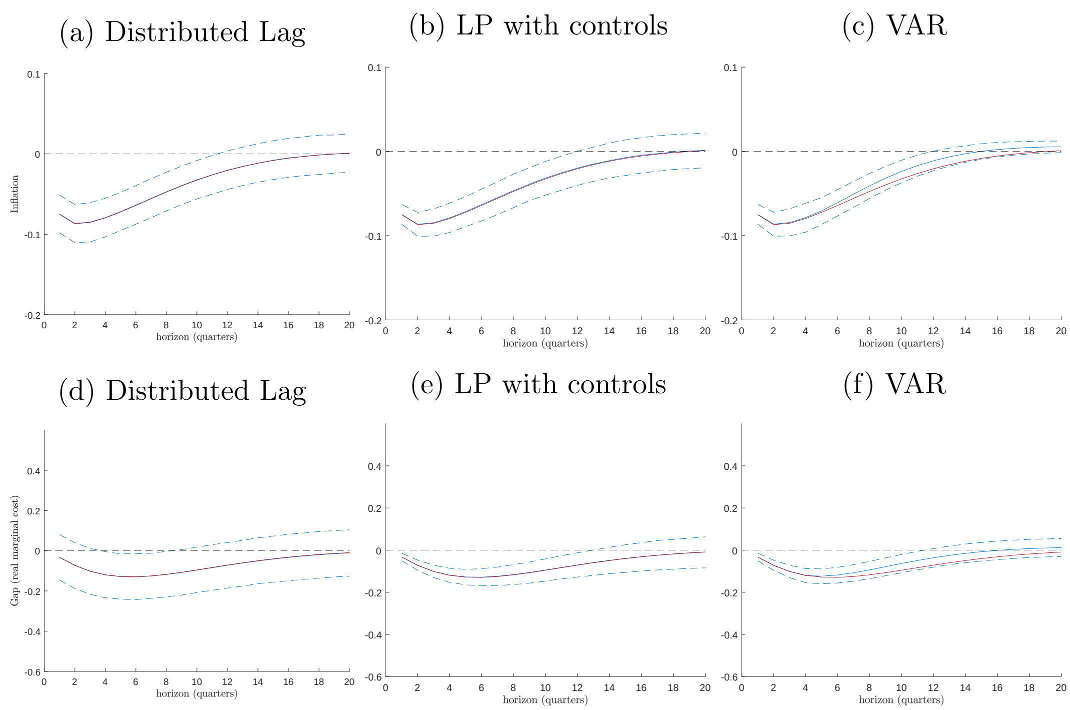


Notes: Figures show IRFs to a one s.t.d. contractionary monetary policy shock in data generated by the Smets and Wouters (2007) model. Red lines show the true IRFs. Blue lines show the mean and 2.5% and 97.5% percentiles of the estimated IRFs across 5000 samples.

Figure II.1 shows the IRF estimates in a small sample with $T = 250$. The DL estimates display smaller small-sample bias than the LP and VAR estimates but have a wider 95% range at shorter horizons. Consistent with Li et al. (2021), the VAR estimates have a narrower range than LP with controls, particularly at longer horizons.

Figure II.2 shows the IRF estimates in a larger sample with $T = 5000$. The DL and LP estimates show essentially no bias for $T = 5000$. Consistent with Montiel Olea and Plagborg-Møller (2021), the VAR estimates

FIGURE II.2: True and Estimated IRFs in Simulations, Large Sample ($T = 5000$)



Notes: See Figure II.1

show no bias for horizons up to the lag length of the VAR (four). Given that the Stock and Watson (2012) model does not have a finite-order VAR representation in X_t, the restrictions implied by the finite-order VAR model result in bias in the IRF estimates at horizons beyond the lag length of the VAR.

II.2 Simulation Results Using Three Instruments ($N_z = 3$)

This section presents the simulation results for specifications using three instruments. Besides the monetary policy shock, the additional instruments are the government spending shock and the risk premium shock from the Smets and Wouters (2007) model. These additional shocks also satisfy the exogeneity requirements for estimating the parameters of the Phillips curve in the data-generating process, both for 2SLS with DL instruments and the SP-IV estimators.

Panel a. of Table II.1 reports the mean estimates across 5000 Monte Carlo samples, Panel b. shows the standard deviations. The results are qualitatively similar to those reported in Tables 2, 3 in the main text, which show results for simulations with only the monetary policy shock as an instrument.

More specifically, the relative performance of the various estimators in terms of bias and variance remains the same with three instruments. In general, the bias improvements from using the IV estimators relative to OLS are smaller with three instruments. However, the comparison of panel b. in Table II.1 and Table 3 in the main text shows that using additional instruments lowers the variance of all the estimators. Therefore, the choice of the number of instruments involves a bias-variance trade-off, at least in data generated from the SW model.

Table II.2 shows the empirical rejection rates for the specifications that use three instruments. The Table repeats the rejection rates for $H = 20$, $N_z = 3$ that are also in Table 4 in the main text and are discussed there. The first three columns additionally report the results for $H = 8$ and $N_z = 3$ for completeness. The general conclusions remain qualitatively the same, although the size distortions related to the many-moments problem are naturally quantitatively less pronounced. The robust SP-IV tests also appear somewhat less affected by many moment distortions than the robust tests for the single equation specification with DL instruments (referred to in the Table as AR SE-IV and KLM SE-IV).

II.3 Simulation Results for Generalized SP-IV estimators

This section presents simulation results for the (feasible) generalized SP-IV estimators based on a 2-step procedure. First, we estimate the baseline SP-IV estimators and estimate the covariance matrix $\hat{\Sigma}_u^{\perp}$ using (19). Then, we use the latter to obtain the generalized SP-IV estimators as in (B.1). The generalized SP-IV estimators are also the feasible 2-step efficient GMM estimators.

Table II.3 reports the standard deviations of the estimates in the simulations. The generalized SP-IV, or "GSP-IV", estimators are, in theory, asymptotically more efficient than our baseline estimators. However, the

Table II.1: Mean and Variance of parameter estimates, $N_z = 3$

a. Mean Parameter Estimates

	$T = 250$			$T = 500$			$T = 5000$		
Estimator	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
True Value	0.15	0.85	0.05	0.15	0.85	0.05	0.15	0.85	0.05
OLS	0.47	0.47	0.00	0.48	0.48	0.00	0.48	0.48	0.00
$H = 8$									
2SLS	0.40	0.56	0.00	0.36	0.63	0.00	0.23	0.82	0.02
SP-IV LP	0.39	0.56	0.00	0.36	0.63	0.00	0.22	0.82	0.02
SP-IV LP-C	0.40	0.55	0.02	0.36	0.63	0.03	0.20	0.81	0.04
SP-IV VAR	0.34	0.69	0.01	0.29	0.75	0.02	0.20	0.83	0.04
$H = 20$									
2SLS	0.45	0.51	0.00	0.43	0.55	0.00	0.28	0.76	0.01
SP-IV LP	0.44	0.51	0.00	0.42	0.56	0.00	0.28	0.76	0.01
SP-IV LP-C	0.44	0.51	0.01	0.43	0.56	0.01	0.27	0.76	0.02
SP-IV VAR	0.35	0.69	0.01	0.31	0.74	0.01	0.23	0.82	0.02

b. Standard Deviation of Parameter Estimates

	$T = 250$			$T = 500$			$T = 5000$		
Estimator	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
$H = 8$									
2SLS	0.09	0.10	0.03	0.08	0.09	0.03	0.06	0.05	0.02
SP-IV LP	0.09	0.10	0.04	0.09	0.09	0.03	0.06	0.05	0.02
SP-IV LP-C	0.09	0.10	0.06	0.09	0.09	0.05	0.06	0.05	0.03
SP-IV VAR	0.11	0.13	0.05	0.11	0.11	0.05	0.06	0.05	0.03
$H = 20$									
2SLS	0.04	0.04	0.01	0.04	0.04	0.01	0.04	0.04	0.01
SP-IV LP	0.05	0.05	0.02	0.04	0.04	0.01	0.04	0.04	0.01
SP-IV LP-C	0.04	0.05	0.02	0.04	0.04	0.02	0.04	0.04	0.01
SP-IV VAR	0.09	0.11	0.02	0.09	0.10	0.02	0.05	0.04	0.02

Notes: The first row in Panel a. contains the true parameter values $\beta = [\gamma_b, \gamma_f, \lambda]'$ of (2) in the Smets and Wouters (2007) model. The other rows show the mean (Panel a.) and standard deviation (Panel b.) of estimates across 5000 Monte Carlo samples of size T and with $h = 0, \ldots, H-1$. All IV estimators use the true monetary policy shock, government spending shock, and risk premium shock in the model as instruments. SP-IV LP and LP-C denote implementations based on LPs without and with X_{t-1} as controls, respectively. SP-IV VAR denotes implementation with a VAR for X_t with four lags.

Table II.2: Empirical size of nominal 5% tests, $N_z = 3$

	$H=8$			$H=20$		
	$T=250$	$T=500$	$T=5000$	$T=250$	$T=500$	$T=5000$
Wald 2SLS	83.1	79.2	58.9	100.0	99.9	94.3
Wald SP-IV LP	84.3	80.4	60.4	100.0	99.9	93.8
Wald SP-IV LP-C	75.8	62.4	22.7	100.0	99.8	83.0
Wald SP-IV VAR	39.2	28.3	13.3	86.7	76.7	54.1
AR SE-IV	16.9	11.4	4.3	60.0	36.3	6.4
AR SP-IV LP	7.0	5.7	4.7	14.3	8.0	5.0
AR SP-IV LP-C	7.0	5.6	4.5	16.9	9.2	5.1
AR SP-IV VAR	3.9	5.1	4.8	6.5	5.2	4.6
KLM SE-IV	2.7	4.3	4.3	0.0	7.2	5.0
KLM SP-IV LP	5.7	5.2	5.3	7.6	6.5	5.3
KLM SP-IV LP-C	7.3	5.5	5.6	11.4	7.6	6.1
KLM SP-IV VAR	6.9	6.6	4.9	11.7	8.5	5.5

Notes: The table shows empirical rejection rates of nominal 5% tests of the true values of $\beta = [\gamma_b, \gamma_f, \lambda]'$ in 5000 Monte Carlo samples from the Smets and Wouters (2007) model. All IV estimators are based on $h = 0, ..., H-1$ and use the true monetary policy shock, government spending shock, and risk premium shock in the model as instruments. SP-IV LP and LP-C denote implementations based on local projections without and with X_{t-1} (described in the text) as controls, respectively. SP-IV VAR denotes implementation with a vector autoregression for X_t with four lags. Robust tests for 2SLS/SE-IV use a HAR Newey-West variance matrix with Sun (2014) fixed-b critical values; inference procedures for SP-IV are described in Section 2 and Appendix A.

feasible versions do not generally improve performance in practice, at least not in realistic sample sizes and for our data-generating process. For $N_z = 1$, most GSP-IV variances slightly exceed those of their SP-IV counterparts in Table 3 in the main text. With more instruments ($N_z = 3$), there is some sporadic evidence of small efficiency gains of GSP-IV relative to their SP-IV counterparts in Panel b. in Table 3. The fact that GSP-IV does not consistently provide efficiency gains (and frequently fares slightly worse) in small samples likely results from estimation error in the $H \times H$ weighting matrix, which itself depends on the estimate $\hat{\beta}$, which is only weakly identified.

For brevity, we do not report the simulation results for the bias and empirical rejection rates, but they are available on request. The results are comparable overall to the regular SP-IV estimators discussed in the main

Table II.3: Standard deviation of parameter estimates, GSP-IV

	$T = 250$			$T = 500$			$T = 5000$		
	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
$H = 8, N_z = 1$									
GSP-IV LP	0.33	0.46	0.24	0.27	0.40	0.24	0.12	0.08	0.09
GSP-IV LP-C	0.36	0.33	0.31	0.31	0.22	0.28	0.12	0.06	0.08
GSP-IV VAR	0.36	0.41	0.34	0.33	0.27	0.32	0.13	0.06	0.09
$H = 20, N_z = 1$									
GSP-IV LP	0.15	0.18	0.07	0.12	0.14	0.06	0.07	0.05	0.03
GSP-IV LP-C	0.10	0.11	0.06	0.09	0.09	0.05	0.08	0.05	0.03
GSP-IV VAR	0.24	0.28	0.14	0.21	0.19	0.13	0.12	0.06	0.06
$H = 8, N_z = 3$									
GSP-IV LP	0.11	0.12	0.04	0.09	0.10	0.03	0.06	0.05	0.02
GSP-IV LP-C	0.10	0.10	0.05	0.09	0.08	0.05	0.06	0.05	0.03
GSP-IV VAR	0.11	0.12	0.05	0.10	0.11	0.05	0.06	0.05	0.03
$H = 20, N_z = 3$									
GSP-IV LP	0.04	0.04	0.01	0.03	0.03	0.01	0.04	0.04	0.01
GSP-IV LP-C	0.03	0.03	0.01	0.03	0.03	0.01	0.04	0.04	0.01
GSP-IV VAR	0.07	0.09	0.02	0.07	0.09	0.02	0.05	0.04	0.02

Notes: Rows show standard deviations across 5000 Monte Carlo samples of size T with $h = 0, \ldots, H-1$. $N_z = 1$ estimators use the monetary policy shock as an instrument; $N_z = 3$ add the government spending and risk premium shocks as instruments. GSP-IV is the (feasible) generalized estimator in (B.1), obtained in a two-step procedure using (19).

text. The GSP-IV estimators consistently show somewhat greater bias than their SP-IV counterparts when additional instruments are included. In sum, at least in our setting, the simulation results offer little motivation to prefer GSP-IV over SP-IV in practice.

II.4 Simulation Results for Alternative 2SLS Specifications

This section presents simulation results for alternative 2SLS estimators that incorporate controls.

As mentioned in the main text, we consider three alternative versions of 2SLS estimation with controls. The first, labeled 2SLS-C, adds X_{t-1} as controls to both stages of 2SLS with DL instruments. The second, labeled 2SLS-CL, adds X_{t-H} as controls to both stages of 2SLS with DL instruments. The third, labeled 2SLS-CZ, does not add any controls to 2SLS, but uses a DL of $z_t^\perp$ – the residual in the regression of z_t on X_{t-1} – as the instruments.

Table II.4 reports mean estimates of the different versions of 2SLS in the simulations for $T = 5000$ with the lag endogenous monetary policy instrument as in Section 3.1. For comparison, the Table also repeats the results for the implementations of SP-IV based on LPs with X_{t-1} as controls and a VAR in X_t with four lags. As the Table shows, 2SLS-C reduces overall bias in the parameter estimates relative to 2SLS, but still produces a severely biased estimate for the slope of the Phillips Curve, λ. As explained in the main text, the problem of including X_{t-1} as controls to address lag endogeneity is that doing so also diminishes the explanatory power of the lags of the instruments in the first stage. As a result, identification weakens to the point where, even in large samples, there remains a strong bias in λ. Adding X_{t-H} instead as controls (2SLS-CL) avoids that problem but also does not fully insulate 2SLS from the bias due to lag endogeneity. Table II.4 shows that 2SLS-CL generates bias improvements relative to 2SLS, but not to the same extent as the SP-IV LP-C and VAR estimators. Finally, the only version of 2SLS that is successful in removing the lag endogeneity bias is the version in which z_t is first orthogonalized to X_{t-1} (2SLS-CZ). Table II.4 shows that, in large samples, 2SLS-CZ generates, on average, the same parameter estimates as SP-IV using LPs with controls.

Table II.5 reports mean estimates of the different versions of 2SLS in the same simulations with the fully exogenous monetary policy instrument as in Section 3.2. As the Table shows, 2SLS-C produces strong bias in λ in all sample sizes. The reason is again that adding X_{t-1} as controls greatly weakens the identifying information from the lags of the instrument. The 2SLS-CL estimates based on adding X_{t-H} as controls lead to some small sample bias improvements relative to 2SLS without any controls, but these improvements are not as large as for SP-IV since estimation is based exclusively on H-step ahead forecast errors. Finally, when the instrument z_t is fully exogenous already, the 2SLS-CZ estimates with a DL of residualized

Table II.4: Mean Parameter estimates, Alternative 2SLS Specifications, Lag Endogenous Instrument, $T = 5000$

Estimator	γ_b	γ_f	λ
True Value	0.15	0.85	0.05
OLS	0.48	0.48	0.00
$H = 8$			
2SLS	0.27	0.58	-0.09
2SLS-C	0.19	0.87	-0.06
2SLS-CL	0.20	0.83	0.01
2SLS-CZ	0.16	0.84	0.05
SP-IV LP-C	0.16	0.84	0.05
SP-IV VAR	0.12	0.83	0.09
$H = 20$			
2SLS	0.24	0.76	-0.02
2SLS-C	0.21	0.84	-0.06
2SLS-CL	0.23	0.81	0.01
2SLS-CZ	0.23	0.81	0.02
SP-IV LP-C	0.23	0.81	0.02
SP-IV VAR	0.17	0.83	0.05

Notes: The first row contains the true parameter values $\beta = [\gamma_b, \gamma_f, \lambda]'$ of (2) in the Smets and Wouters (2007) model. The other rows show the mean estimates across 5000 Monte Carlo samples of size T and with $h = 0, \ldots, H-1$. 2SLS-C adds X_{t-1} as controls to both stages, 2SLS-CL adds X_{t-H} as controls to both stages, 2SLS-CZ is 2SLS a DL of $z_t^{\perp}$ as instruments instead of z_t. SP-IV LP-C and VAR denote implementations based on LP with X_{t-1} and a VAR in X_t with four lags, respectively.

shocks $z_t^{\perp}$ offer no further improvement relative to 2SLS with a DL of z_t in small samples: the mean parameter estimates for 2SLS and 2SLS-CZ are essentially the same across all specifications; unlike SP-IV with controls, there is no improvement in the effective instrument strength.

II.5 Measuring Effective Instrument Strength

This Section provides formal measures of the effective instrument strength for the various estimators. We do so by approximating the concentration matrix, Λ, associated with each estimator. For 2SLS estimators, this is the object derived in Lewis and Mertens (2022), which extends Stock and Yogo (2005) to be robust to heteroskedasticity and autocorrelation; for SP-IV, it

Table II.5: Mean Parameter estimates, Alternative 2SLS Specifications, Fully Exogenous Instruments

	$T=250$			$T=500$			$T=5000$		
Estimator	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
True Value	0.15	0.85	0.05	0.15	0.85	0.05	0.15	0.85	0.05
OLS	0.47	0.47	0.00	0.48	0.48	0.00	0.48	0.48	0.00
$H=8$									
2SLS	0.27	0.51	0.01	0.24	0.61	0.00	0.17	0.83	0.04
2SLS-C	0.31	0.69	-0.05	0.27	0.76	-0.04	0.18	0.89	-0.11
2SLS-CL	0.30	0.58	0.02	0.26	0.69	0.03	0.16	0.83	0.05
2SLS-CZ	0.27	0.52	0.01	0.24	0.62	0.01	0.16	0.84	0.05
SP-IV LP-C	0.29	0.64	0.05	0.25	0.74	0.04	0.16	0.84	0.05
SP-IV VAR	0.22	0.80	0.03	0.18	0.84	0.05	0.12	0.83	0.09
$H=20$									
2SLS	0.39	0.53	0.00	0.36	0.61	0.00	0.23	0.80	0.01
2SLS-C	0.37	0.57	-0.07	0.33	0.64	-0.06	0.21	0.85	-0.08
2SLS-CL	0.40	0.53	0.00	0.37	0.61	0.00	0.23	0.80	0.02
2SLS-CZ	0.39	0.53	0.00	0.36	0.61	0.00	0.23	0.81	0.02
SP-IV LP-C	0.41	0.55	0.01	0.37	0.64	0.01	0.23	0.81	0.02
SP-IV VAR	0.27	0.80	0.01	0.23	0.84	0.02	0.17	0.83	0.05

Notes: The first row contains the true parameter values $\beta = [\gamma_b, \gamma_f, \lambda]'$ of (2) in the Smets and Wouters (2007) model. The other rows show the mean estimates across 5000 Monte Carlo samples of size T and with $h = 0, \ldots, H-1$. 2SLS-C adds X_{t-1} as controls to both stages, 2SLS-CL adds X_{t-H} as controls to both stages, 2SLS-CZ is 2SLS a DL of $z_t^{\perp}$ as instruments instead of z_t. SP-IV LP-C and VAR denote implementations based on LP with X_{t-1} and a VAR in X_t with four lags, respectively.

is defined in Definition 1. We approximate it by setting $T = 1,000,000$ in the Smets and Wouters (2007) model to estimate the first-stage coefficients Θ_Y and relevant error covariance matrices. The minimum eigenvalue of the concentration matrix is a sufficient statistic for the worst-case bias of the respective IV estimators, which motivates the proposed test statistic, g_{min}. We report this minimum eigenvalue relative to the $\tau = 0.10$ and $\alpha = 0.05$ critical values (as described in Lewis and Mertens (2022) and Section I, respectively), specific to each estimator and specification, to provide a metric that is on comparable scale across models. Table II.6 reports the results for $T = 500$, and these results scale linearly with T.

The measures of instrument strength reported in Table II.6 are overall

Table II.6: Measures of instrument strength, $T = 500$

	$N_z = 1$	$N_z = 3$
$H = 8$		
2SLS	0.02	0.02
SPIV LP	0.02	0.02
SP-IV LP-C	0.17	0.17
SP-IV VAR	0.12	0.13
$H = 20$		
2SLS	0.02	0.01
SPIV LP	0.02	0.01
SP-IV LP-C	0.08	0.05
SP-IV VAR	0.03	0.06

Notes: The table reports an approximation of the minimum eigenvalue of the concentration matrix, Λ, associated with the respective estimator, relative to the associated critical value for the bias-based first-stage test. For 2SLS, Λ and critical values are defined as in Lewis and Mertens (2022) and for SP-IV as in Definition 1 and Section I. The true value of Θ_Y is approximated using a sample of $T = 1,000,000$ from the Smets and Wouters (2007) model.

relatively small, and it may therefore appear surprising that the performance of the estimators is not considerably worse. However, the tests for instrument strength are based on controlling the *worst case* bias, and there is no reason to believe the model parameters, i.e. eigenstructure of Θ_Y, covariances of errors) are in the neighborhood of this worst case.

The results in Table II.6 are qualitatively consistent with the theoretical predictions: SP-IV LP-C and SP-IV VAR have effectively stronger instruments due to the inclusion of controls, whereas 2SLS and SP-IV LP (without controls) they are essentially identical. SP-IV LP-C appears to make marginally better use of the instruments than SP-IV VAR, although the population measure of instrument strength masks the difficulty in estimating the IRFs efficiently in small samples. $H = 20$ delivers weaker identification than $H = 8$, almost certainly due to the flattening (and noisier estimation) of IRFs at longer horizons. The effect of additional instruments depends on the estimator considered: additional shocks provide more identifying variation, but the critical values to which the measures are benchmarked also increase with N_z.

Online Appendix References

Imhof, J. P. (1961). Computing the Distribution of Quadratic Forms in Normal Variables. *Biometrika*, *48*(3-4), 419–426.

Lewis, D. J., & Mertens, K. (2022). *A Robust Test for Weak Instruments with Multiple Endogenous Regressors* (Staff Reports No. 1020). Federal Reserve Bank of New York.

Li, D., Plagborg-Moller, M., & Wolf, C. K. (2021). *Local Projections vs. VARs: Lessons From Thousands of DGPs* (Papers No. 2104.00655). arXiv.org.

Montiel-Olea, J. L., & Pflueger, C. (2013). A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics*, *31*(3), 358–369.

Montiel Olea, J. L., & Plagborg-Møller, M. (2021). Local Projection Inference is Simpler and More Robust Than You Think. *Econometrica*, *89*(4), 1789–1823.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley; Sons Ltd.

Nagar, A. L. (1959). The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations. *Econometrica*, *27*(4), 575–595.

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

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

Stock, J., & Yogo, M. (2005). Testing for weak instruments in linear iv regression. In D. W. Andrews (Ed.), *Identification and Inference for Econometric Models* (pp. 80–108). Cambridge University Press.

Stock, J. H., & Watson, M. W. (2012). Disentangling the Channels of the 2007-2009 Recession. *Brookings Papers on Economic Activity, Spring 2012*, 81–135.

Sun, Y. (2014). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics*, *178*(P3), 659–677.