Dynamic Identification Using System Projections and Instrumental Variables

Daniel J. Lewis and Karel Mertens

Working Paper 2204 **March 2022**

Research Department

https://doi.org/10.24149/wp2204

Dynamic Identification Using System Projections and Instrumental Variables*

Daniel J. Lewis† and Karel Mertens‡

March 22, 2022

Abstract

We propose System Projections with Instrumental Variables (SP-IV) to estimate dynamic structural relationships. SP-IV replaces lag sequences of instruments in traditional IV with lead sequences of endogenous variables. SP-IV allows the inclusion of controls to weaken exogeneity requirements, can be more efficient than IV with lags, and allows identification over many time horizons without creating many-weak-instruments problems. SP-IV also enables the estimation of structural relationships across impulse responses obtained from local projections or vector autoregressions. We provide a bias-based test for instrument strength, and inference procedures under strong and weak identification. SP-IV outperforms competing estimators of the Phillips Curve parameters in simulations. We estimate the Phillips Curve implied by the main business cycle shock of Angeletos et al. (2020), and find evidence for forward-looking behavior. The data is consistent with weak but also relatively strong cyclical connections between inflation and unemployment.

JEL classification: E3, C32, C36.

Keywords: Structural Equations, Instrumental Variables, Impulse Responses, Robust Inferences, Phillips Curve, Inflation Dynamics.

---

* The views in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Banks of Dallas or New York or the Federal Reserve System. We are grateful for helpful discussions with Kirill Borusyak, Xu Cheng, Andrew Chesher, Richard Crump, Marco Del Negro, Dennis Kristensen, Geert Mesters, José Luis Montiel Olea, Andrei Zeleneev, and seminar participants at UPenn, Penn State, University of Houston, UCL, Boston College, and the Federal Reserve Banks of Kansas City and New York.

† Daniel J. Lewis, Federal Reserve Bank of New York and University of Pennsylvania.

‡ Contact: Karel Mertens, Federal Reserve Bank of Dallas, mertens.karel@gmail.com, tel: +(214) 922-6000.

1 Introduction

This paper is concerned with estimating β in structural time series equations of the form

$$y_t = \beta' Y_t + u_t \ , \tag{1}$$

where y_t is the scalar observation of an outcome variable in period t, Y_t is a $K \times 1$ vector of explanatory variables, u_t is an error term, and the $K \times 1$ vector β contains the structural parameters of interest. The explanatory variables Y_t may contain contemporaneous variables, but also lagged variables or economic agents' expectations of future variables that are not necessarily measured well by the econometrician. We are interested in applications in which $E[Y_t u_t] \neq 0$, such that standard regression techniques yield biased and inconsistent estimates of β because of endogeneity problems.

Equation (1) nests a very wide range of dynamic relationships of interest in macroeconomics. To illustrate the range of difficulties that can arise in the estimation of β, consider the specific example of the Hybrid New Keynesian Phillips Curve (HNKPC),

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f \pi^e_{t+1} + \lambda gap_t + u_t \ , \tag{2}$$

where π_t denotes inflation, π^e_{t+1} is the price setters' period t expectation of inflation in $t+1$, and gap_t is an output gap measure (the difference between the actual level of economic activity and the natural level that would prevail in the absence of price rigidities). Equation (2) maps into the more general estimation problem in (1) after defining $y_t = \pi_t$, $Y_t = [\pi_{t-1}, \pi^e_{t+1}, gap_t]'$ and $\beta = [\gamma_b, \gamma_f, \lambda]'$.

The estimation of $\beta = [\gamma_b, \gamma_f, \lambda]'$ in the HNKPC is complicated by a number of well-known problems that result in $E[Y_t u_t] \neq 0$, see for instance Mavroeidis et al. (2014) or McLeay and Tenreyro (2019) for discussions. A first general source of endogeneity problems is *measurement error*. In practice, both the output gap and price setters' inflation expectations are not observed directly by the econometrician and must be replaced with proxy measures. A second general source of endogeneity problems is *simultaneity*, since the error term generally includes structural disturbances that also influence the endogenous variables in $Y_t = [\pi_{t-1}, \pi^e_{t+1}, gap_t]$. Many theoretical dynamic relationships include expectations terms and other endogenous explanatory variables, and therefore face very similar problems.

A common approach in the literature to address these endogeneity problems is to rely on dynamics for identification, and use lagged economic variables as instrumental variables. In the case of the HNKPC, for example, it is typical to use $gap_{t-1}, gap_{t-2}, \ldots$ and $\pi_{t-2}, \pi_{t-3}, \ldots$, or lags of other readily available macroeconomic variables.[1] Instrument exogeneity in this case requires the assumption that the error term u_t is independent of any of the influences on the instrumenting lagged macroeconomic variables. However, there is generally little theoretical justification for this assumption. Lags of output gaps or inflation, for example, are not valid instruments for (2) in fully specified medium-scale macroeconomic models such as the Smets and Wouters (2007) model. For this reason, Barnichon and Mesters (2020) recently proposed to use the current and lagged values of direct measures of structural shocks from the literature as instrumental variables.[2] Instrument exogeneity in this case requires measures of economic shocks that are independent of both the contemporaneous and lagged macroeconomic influences on the error term u_t. In practice, however, the literature is rarely comfortable with imposing the strong assumption of lag exogeneity on available empirical measures of structural shocks, and typically avoids doing so by including a rich set of lagged macroeconomic controls. Unfortunately, including such controls in an IV regression with lagged shocks as instruments shrinks the explanatory power of the instrument set towards that of only the contemporaneous value of the instrumenting shocks, resulting in weaker or even under-identification.

Even if one is willing to assume that lag sequences of empirical shock measures or macroeconomic variables satisfy the required exogeneity assumptions, in many macroeconomic applications it is challenging to find strong instruments.[3] In practice researchers will therefore often be tempted to include a relatively large number of lags in the instrument set to improve the predictive power of the instruments, which in macroeconomic applications is often distributed over an extended time horizon. However, this can lead to many-weak-instruments problems that can exacerbate small sample bias and render weak-instrument-robust inference methods unreliable.

In this paper, we propose a novel approach to estimating β in the face of the endogeneity problems that are typical in the estimation of dynamic structural equations. Specifically, we replace the single equation (1) by an H-dimensional system in forecast errors for various leads of y_t and Y_t, where H is the number of leads. The forecast errors can be derived from a variety of forecasting models, and in this paper we consider vector autoregressive models (VARs) and local projections (LPs). We then use only the contemporaneous value of the N_z instrumental variables to yield HN_z moment conditions, which can be solved in closed form for β, yielding a restricted two-stage least squares (2SLS) estimator within the system

[1] Galí and Gertler (1999), for example, use four lags of inflation, the labor income share, the output gap, the long-short interest rate spread, wage inflation, and commodity price inflation.

[2] Galí and Gambetti (2020) follow an approach that is closely related to Barnichon and Mesters (2020) to estimate the wage Phillips curve.

[3] For discussions of the weak identification problems that arise in the context of estimating the HNKPC, see Nason and Smith (2008), Kleibergen et al. (2009), Mavroeidis et al. (2014) among others.

of reduced form forecast errors.

The SP-IV approach has several conceptual advantages. First, like Barnichon and Mesters (2020), it can leverage existing identified shocks from the literature, but it requires only the weaker assumption of contemporaneous exogeneity. Our approach is therefore better aligned with the identification assumptions that the literature is typically willing to make about empirical measures of economic shocks. Second, our approach allows exploiting the dynamic relationship between the shocks and the endogenous variables across a large number of horizons H without creating many-weak-instruments problems, since the number of instruments in our setup does not depend on H. Both these advantages arise because our approach is based on lead sequences of the endogenous variables y_t and Y_t, and not on lag sequences of the instrumental variables as in existing approaches.

Barnichon and Mesters (2020) point out that single-equation 2SLS with lag sequences of shocks as instruments is equivalent to a regression of the impulse response function (IRF) of y_t on the IRFs of Y_t, where the IRFs are estimated by regressions of the endogenous variables on a distributed lag of the shocks. We show that SP-IV is similarly equivalent to a regression of the IRF of y_t on the IRFs of Y_t. However, an additional advantage of SP-IV is that these IRFs can be obtained using any valid impulse response estimator and identification scheme. SP-IV therefore allows the estimation of structural relationships across IRFs as they are estimated in practice, which is not typically by distributed lag specifications. The SP-IV inference methods in this paper enable the formal testing of hypotheses about structural relationships across these IRFs, whereas in empirical practice claims about such relationships have typically relied on more informal arguments (e.g., Angeletos et al. (2020)).

The inference methods presented in this paper should make SP-IV of practical use in a wide range of settings. We describe standard inference under strong identification, and develop a first-stage test for instrument strength by extending the popular bias-based test in Stock and Yogo (2005) to the SP-IV setting. For practitioners, such a pre-test provides a convenient way to assess whether standard inference will be reliable, or to compare the strength of identifying information contained in different specifications. As instrumental variables are often likely to be weak in practical applications, we develop two weak-instrument-robust inference procedures, one based on Anderson and Rubin's (1949) AR statistic, and another based on an extension of Kleibergen's (2002) KLM statistic.

To demonstrate the potential improvements offered by SP-IV beyond the weaker exogeneity requirements, we present simulation evidence for the estimation of the parameters of the HNKPC in (2) using data generated from the Smets and Wouters (2007) model and instrumental variables that are weak. OLS is strongly biased, and the SP-IV estimators with controls exhibit smaller biases than single-equation 2SLS estimators, including the Almon shrinkage estimator proposed by Barnichon and Mesters (2020). In particular, a VAR implementation of SP-IV has the lowest bias of all estimators we consider, while estimates from SP-IV based on LP implementations have lower variances. Standard Wald inference is badly over-sized for all IV estimators due to weak instruments. We find that both of our proposed robust inference procedures remain relatively well-sized in realistic sample sizes, exhibiting small size distortions only when H is large relative to T. In contrast, single-equation 2SLS with lag sequences as instruments suffers from many-weak-instruments problems that quickly lead to prohibitive over-rejection. The AR test for 2SLS with Almon shrinkage proposed by Barnichon and Mesters (2020) appears to be incorrectly sized asymptotically.

As an empirical application, we estimate the Phillips curve parameters in US data using the Main Business Cycle (MBC) shock of Angeletos et al. (2020) as the instrument. Based on the IRFs to an MBC shock identified in a monthly VAR, our SP-IV inference points towards a greater weight on future inflation than on lagged inflation. At the same time, the SP-IV robust confidence sets for the coefficient on unemployment are consistent with both very weak and fairly strong cyclical responses of inflation. We conclude that the responses to the MBC shock do not necessarily support the conclusion in Angeletos et al. (2020) that inflation dynamics are disconnected from the business cycle. We show in a simple dynamic rational expectations model that the inflation dynamics implied by a Phillips curve parametrized by the SP-IV point estimates closely resemble those estimated in the data. This is not the case for the OLS point estimates or the 2SLS-Almon estimator of Barnichon and Mesters (2020).

For the remainder of the paper, I_N denotes the identity matrix of dimension N, $\otimes$ the Kronecker product, $\mathrm{Tr}(\cdot)$ the trace operator, $\mathrm{vec}(\cdot)$ the vectorization operator, $\mathrm{mineval}\{\cdot\}$ / $\mathrm{maxeval}\{\cdot\}$ the minimum/maximum eigenvalue, $E[X \mid Y]$ the conditional expectation of X given Y, $\xrightarrow{p}$ is convergence in probability, and $\xrightarrow{d}$ is convergence in distribution.

2 System Projections with Instrumental Variables

We begin by reformulating the dynamic relationship of interest in (1) in terms of forecast errors. Taking leads at horizon h of (1) and subtracting the expectation conditional on an information set $\mathcal{I}_{t-1}$ yields

$$y^{\perp}_{t+h} \;=\; \beta' Y^{\perp}_{t+h} + u^{\perp}_{t+h} \;, \tag{3}$$

where $y^{\perp}_{t+h} = y_{t+h} - E[y_{t+h} \mid \mathcal{I}_{t-1}]$, $Y^{\perp}_{t+h} = Y_{t+h} - E[Y_{t+h} \mid \mathcal{I}_{t-1}]$, and $u^{\perp}_{t+h} = u_{t+h} - E[u_{t+h} \mid \mathcal{I}_{t-1}]$. Let z_t denote a $N_z \times 1$ vector of instrumental variables, and define $z^{\perp}_t = z_t - E[z_t \mid$

$\mathcal{I}_{t-1}]$. As explained in the introduction, we focus on applications that rely on dynamics for identification. Instead of the typical approach of imposing orthogonality between z_{t-h} and u_t for various $h \geq 0$, we impose

$$E[u_{t+h}^{\perp} z_t^{\perp}] = 0 \ ; \ h = 0, \ldots, H-1 \ . \tag{4}$$

The orthogonality conditions in (4) naturally furnish a set of HN_z moment conditions that can be used to identify the K elements of β. Let $y_{H,t}^{\perp}$ and $u_{H,t}^{\perp}$ denote the $H \times 1$ vectors in which the $h+1$-th element is $y_{t+h}^{\perp}$ and $u_{t+h}^{\perp}$ respectively. Let $Y_{H,t}$ denote the $HK \times 1$ vector stacking the $H \times 1$ vectors $Y_{H,t}^k$, each of which has $Y_{t+h}^k - E[Y_{t+h}^k \mid \mathcal{I}_{t-1}]$ in the $h+1$-th row, where Y_t^k is the k-th variable in Y_t. Using this notation, the HN_z moment conditions are

$$E[u_{H,t}^{\perp}(\beta) \otimes z_t^{\perp}] = 0 \ , \tag{5}$$

where $u_{H,t}^{\perp}(b) = y_{H,t}^{\perp} - (b' \otimes I_H) Y_{H,t}$, and the true value of b is β.

The moment conditions in (5) must be augmented to include the estimation of the forecast errors. We assume that the information set $\mathcal{I}_{t-1}$ consists of predetermined control variables X_{t-1}, and we consider the class of forecasting models that are linear in X_{t-1}, but possibly nonlinear in a set of parameters collected in the vector d. This class of forecasting models includes local projections (LP) and vector autoregressions (VARs), both of which are widely used in applied macroeconomics. The moment conditions associated with the forecasting step are

$$E\left[\left[y_{H,t}^{\perp\prime}(\zeta), Y_{H,t}^{\perp\prime}(\zeta), z_t^{\perp\prime}(\zeta)\right]' \otimes X_{t-1}\right] = 0, \tag{6}$$

where $y_{H,t}^{\perp}(d)$, $Y_{H,t}^{\perp}(d)$, $z_t^{\perp}(d)$ are functions of parameters d that depend on the forecasting model chosen, and the true value of d is ζ.

The moments in (5) and (6) can be stacked in a moment function $f(b, d, y_{H,t}, Y_{H,t}, z_t, X_{t-1})$ with $E[f(\beta, \zeta, y_{H,t}, Y_{H,t}, z_t, X_{t-1})] = 0$. The associated GMM objective function is given by

(7)

$$F_T(b,d) = \frac{1}{T}\left(\sum_{t=1}^{T} f(b, d, y_{H,t}, Y_{H,t}, z_t, X_{t-1})\right)' \Phi(b,d) \left(\sum_{t=1}^{T} f(b, d, y_{H,t}, Y_{H,t}, z_t, X_{t-1})\right) \ ,$$

where $\Phi(b,d)$ is a positive definite weighting matrix. The forecasting step and the structural estimation stage are separable for estimation purposes. In particular, b does not enter (6). Moreover, the Jacobian of (5) with respect to d is zero in expectation at $d = \zeta$. Therefore, for the remainder of this section, we take the forecast estimates as given, and focus exclusively on the structural estimation step after making the following assumption,

Assumption 1. *There exists a unique solution, ζ, to the first-stage moments (6), and the associated GMM estimator satisfies $\hat{\zeta} \xrightarrow{p} \zeta$ and $\sqrt{T}(\hat{\zeta} - \zeta) \xrightarrow{d} \mathcal{N}(0, V_{fs})$ for some feasible weighting matrix.*

Assumption 1 ensures that estimation error in the forecast errors is asymptotically negligible. Henceforth, we let $\Phi_s(b, d)$ denote the block in the weighting matrix $\Phi(b, d)$ corresponding to the moments of the structural estimation step in (5).

2.1 The SP-IV Estimator

Our baseline estimator is based on the weighting matrix $\Phi_s(b, d) = I_H \otimes Q^{-1}$, where $Q = E[z_t^\perp z_t^{\perp\prime}]$, which standardizes the instruments, $z_t^\perp$. With this choice, the solution for β to the population analog of the GMM problem in (7) has a closed form expression given by

$$(8) \qquad \beta = \left(R'(E[Y_{H,t}^\perp z_t^{\perp\prime}]Q^{-1}E[Y_{H,t}^\perp z_t^{\perp\prime}]' \otimes I_H)R\right)^{-1} R' \operatorname{vec}(E[y_{H,t}^\perp z_t^{\perp\prime}]Q^{-1}E[Y_{H,t}^\perp z_t^{\perp\prime}]') \ ,$$

where $R = I_K \otimes \operatorname{vec}(I_H)$. Let the $H \times T$ matrix $y_H^\perp$, the $HK \times T$ matrix $Y_H^\perp$, and the $N_z \times T$ matrix $Z^\perp$ collect the sample of observations of $y_{H,t}^\perp$, $Y_{H,t}^\perp$, and $z_t^\perp$ respectively. Define the projection matrix $P_{Z^\perp} = Z^{\perp\prime}(Z^\perp Z^{\perp\prime})^{-1}Z^\perp$ and the residualizing matrix $M_{Z^\perp} = I_T - P_{Z^\perp}$. Using this notation, the most general sample analog of (8) is

$$(9) \qquad \hat{\beta} \;=\; \left(R'(Y_H^\perp P_{Z^\perp} Y_H^{\perp\prime} \otimes I_H)R\right)^{-1} R' \operatorname{vec}(y_H^\perp P_{Z^\perp} Y_H^{\perp\prime}) \ ,$$

which minimizes the GMM objective (7) with respect to b, using the sample analog of the weighting matrix, $I_H \otimes (ZZ'/T)^{-1}$. That minimization problem is equivalent to minimizing $\operatorname{Tr}(u_H^\perp P_{Z^\perp} u_H^{\perp\prime})$, or the sum of squared residuals in the system of equations

$$(10) \qquad y_H^\perp \;=\; (\beta' \otimes I_H)Y_H^\perp + u_H^\perp \ ,$$

after projection on the instruments $Z^\perp$. Thus, the estimator in (9) is also the restricted 2SLS estimator of β in the system of equations in (10). For this reason, we refer to $\hat{\beta}$ as the **S**ystem of equations after **P**rojection on the **I**nstrumental **V**ariables (SP-IV) estimator.[4] Note that the parameter restrictions in (10) do not impose any additional restrictions beyond those that are already implied by the original structural equation.

To derive the limiting distribution of $\hat{\beta}$, we make high-level assumptions on the covariances between instruments, $z_t^\perp$, the innovations, $u_{H,t}^\perp$, and the endogenous variables $Y_{H,t}^\perp$.

[4] An equivalent expression for (9) is $\hat{\beta} = \left(\sum_{h=0}^{H-1} Y_h^\perp P_{Z^\perp} Y_h^{\perp\prime}\right)^{-1} \sum_{h=0}^{H-1} Y_h^\perp P_{Z^\perp} y_h^{\perp\prime}$ where $Y_h^\perp$ and $y_h^\perp$ are the rows corresponding to horizon h in $Y_H^\perp$ and $y_H^\perp$, respectively. This alternative expression clarifies $\hat{\beta}$ as a ratio of sums over the terms from the 2SLS estimator for individual horizons.

Assumption 2. *The following limits hold*

(2.a) $$Z^{\perp}Z^{\perp\prime}/T \xrightarrow{p} E[z_t^{\perp}z_t^{\perp\prime}] = Q, \quad \textit{where } Q \textit{ is positive definite},$$

(2.b) $$Y_H^{\perp}Z^{\perp\prime}/T \xrightarrow{p} E[Y_{H,t}^{\perp}z_t^{\perp\prime}] = \Theta_Y Q^{\frac{1}{2}}, \ \textit{a real } HK \times N_z \textit{ matrix},$$

(2.c) $$Z^{\perp}u_H^{\perp\prime}/T \xrightarrow{p} E[z_t^{\perp}u_{H,t}^{\perp\prime}] = 0,$$

and the following rank condition holds

(2.d) $$R'(\Theta_Y\Theta_Y' \otimes I_H)R \textit{ is a fixed matrix with full rank.}$$

The convergence in probability in 2.a-2.c holds under standard primitive conditions and laws of large numbers. Condition 2.a ensures linear independence of the instruments and consistency of the weighting matrix. Condition 2.b states that the covariance between $Y_H^{\perp}$ and $Z^{\perp}$ is consistently estimated. As we will discuss below, the population covariance $\Theta_Y Q^{\frac{1}{2}}$ is a rotation of impulse response coefficients of $Y_t^{\perp}$ to the instruments $z_t^{\perp}$, after standardization, denoted by Θ_Y. Condition 2.c is the exogeneity condition. Finally, the rank condition in 2.d is a sufficient condition for the existence of a unique solution to the moment conditions in (5), and ensures that the denominator of the closed form solution (8) is full rank; together with the definition of Θ_Y, it implies that the instruments are relevant. Jointly, 2.b and 2.d imply that the instruments are strong, an assumption we later relax in Section 3.

The conditions in Assumption 2 closely resemble the usual assumptions for IV under strong identification, see for instance Stock and Yogo (2005). Note, however, that unlike in the traditional IV setting, condition 2.d does not require that there are at least as many instruments as endogenous explanatory variables, $N_z \geq K$. Since $\operatorname{rank}(R'(\Theta_Y\Theta_Y' \otimes I_H)R) = \min\{K, H \operatorname{rank}(\Theta_Y\Theta_Y')\}$, the necessary condition is instead that $HN_z \geq K$, the order condition for (5). Adding leads can therefore make up for $N_z < K$ just as adding more lags can do so in the traditional single-equation 2SLS with lag sequences as instruments.

Assumption 2 and the continuous mapping theorem imply that

(11) $$\begin{aligned} R'(Y_H^{\perp}P_{Z^{\perp}}Y_H^{\perp\prime} \otimes I_H)R &\xrightarrow{p} R'(\Theta_Y\Theta_Y' \otimes I_H)R \ , \\ R' \operatorname{vec}(y_H^{\perp}P_{Z^{\perp}}Y_H^{\perp\prime}) &\xrightarrow{p} R'(\Theta_Y\Theta_Y' \otimes I_H)R\beta \ , \end{aligned}$$

such that the SP-IV estimator in (9) is consistent, $\hat{\beta} \xrightarrow{p} \beta$.

2.2 Exogeneity and Shocks as Instruments

Following the Slutzky-Frisch paradigm of expressing macroeconomic variables as arising from current and past shocks (including measurement error), y_t and Y_t can be expressed in terms of current and past values of a random shock vector ϵ_t, where $E[\epsilon_t] = 0$, $E[\epsilon_t \epsilon_t'] = I_{dim(\epsilon)}$ and $E[\epsilon_t \epsilon_s'] = 0$ for $s \neq t$. Assuming linearity of y_t and Y_t in ϵ_t, equation (1) implies that

$$\tilde{u}_t = u_t - E[u_t] \quad = \quad \mu_0' \epsilon_t + \mu_1' \epsilon_{t-1} + \mu_2' \epsilon_{t-2} + \ldots \tag{12}$$

This representation of the error term helps to clarify that the strictness of the exogeneity assumption in 2.c depends on the information set $\mathcal{I}_{t-1}$ used to form the forecast errors. On one end of the range of information sets is the empy set, $\mathcal{I}_{t-1} = \emptyset$, such that $\hat{\beta}$ is obtained using the current values and leads of $\tilde{y}_t = y_t - E[y_t]$ and $\tilde{Y}_t = Y_t - E[Y_t]$, and the current values of $\tilde{z}_t = z_t - E[z_t]$ as the instruments $z_t^{\perp}$. When $\mathcal{I}_{t-1} = \emptyset$, the forecasts are not based on any additional information, and $\tilde{y}_t$ and $\tilde{Y}_t$ are simply the raw (demeaned) observations. In that case, the exogeneity condition 2.c is satisfied by the (demeaned) instrument $\tilde{z}_t$ when

$$\mu_l' E[\epsilon_{t+h-l} \tilde{z}_t'] = 0 \quad ; \quad l = 0, \ldots, \infty \quad ; \quad h = 0, \ldots, H-1 \; . \tag{13}$$

This condition states that whenever an element of $\tilde{z}_t$ has positive covariance with a past shock, contemporaneous shock, or a shock up to $H-1$ periods in the future, there must be an associated exclusion restriction requiring the corresponding element in μ_l to be zero. Following the terminology in Stock and Watson (2018), we refer to the subset of the conditions in (13) with $l > h$ as *lag exogeneity*, with $l = h$ as *contemporeneous exogeneity*, and with $l < h$ as *lead exogeneity*.

At the other end of the range of information sets is $\mathcal{I}_{t-1} = \mathcal{I}_{t-1}^{full}$, defined as an information set that eliminates the influence of all past values of ϵ_t on y_t and Y_t, such that

$$\bar{u}_t = u_t - E[u_t \mid \mathcal{I}_{t-1}^{full}] = \mu_0' \epsilon_t \; , \tag{14}$$

and conditioning on $\mathcal{I}_{t-1}^{full}$ also eliminates the influence of all past values of ϵ_t on the error term. The exogeneity condition 2.c is satisfied by the fully conditioned instrument $\bar{z}_t$ when

$$\mu_0' E[\epsilon_{t+h} \bar{z}_t'] = 0 \quad ; \quad h = 0, \ldots, H-1 \; , \tag{15}$$

which is analogous to (13) with $l = 0$ only. Equation (15) requires that any contemporanous or future shock that has a nonzero covariance with $\bar{z}_t$ is excluded from $\bar{u}_t$, but compared with $\tilde{u}_t$ it eliminates the same requirement for the entire history of shocks up to period t. With a full information set, the exogeneity conditions are therefore considerably weaker, requiring contemporaneous and lead exogeneity, but not lag exogeneity.

In between the empty and full information sets lies a range of other possible information sets. Each of these sets may purge different combinations of shocks in ϵ_{t-l} for $l > 0$ from u_t, and by doing so may eliminate the corresponding condition of the form $\mu_l' E[\epsilon_{t+h-l} z_t^{\perp\prime}] = 0$, as discussed in Forni and Gambetti (2014), Stock and Watson (2018) or Miranda-Agrippino and Ricco (2019). A large literature focuses on the identification of economic shocks within ϵ_t that can be potential instruments for SP-IV, see Ramey (2016) or Kilian and Lütkepohl (2017) for overviews.

2.3 When to Use SP-IV instead of Single-Equation 2SLS

Our approach based on a system of H leads of y_t and Y_t and the contemporaneous z_t as instruments mirrors the more standard IV approach with a single equation in y_t and Y_t, and H lags of z_t as instruments. In this section, we discuss the circumstances under which SP-IV is preferable over single-equation 2SLS with lag sequences as instruments.

Consider first a single-equation specification with only the (demeaned) raw data, that is the 2SLS regression of $\tilde{y}_t$ on $\tilde{Y}_t$ using $\tilde{z}_{t-h}$, $h = 0, \ldots, H-1$ as instruments. Exogeneity in this case requires that

$$E[\tilde{u}_t \tilde{z}_{t-h}] = 0 \quad ; \quad h = 0, \ldots, H-1 \ , \tag{16}$$

which, given the representation of $\tilde{u}_t$ in terms of shocks in (12), requires that

$$\mu_l' E[\epsilon_{t-l} \tilde{z}_{t-h}'] = 0 \quad ; \quad l = 0, \ldots, \infty \quad ; \quad h = 0, \ldots, H-1 \ . \tag{17}$$

Under stationarity, $E[\epsilon_{t-l} \tilde{z}_{t-h}'] = E[\epsilon_{t+h-l} \tilde{z}_t']$, such that (17) is equivalent to the requirement of lead, contemporaneous, and lag exogeneity in (13). In other words, the exogeneity requirements are the same as for the SP-IV estimator based on the unconditional data ($\mathcal{I}_{t-1} = \emptyset$). If the error term $\tilde{u}_t$ is i.i.d., both estimators are asymptotically equally efficient. In general, however, the SP-IV estimator with unconditional data is likely to be asymptotically less efficient than the single-equation 2SLS estimator. In Appendix A, for example, we prove that single-equation 2SLS is asymptotically more efficient if the error term $\tilde{u}_t$ follows a stable AR(1) process with persistence $\rho > 0$. For specifications without any control variables in either stage, and instruments that satisfy lag exogeneity, we therefore recommend to continue using the single-equation 2SLS estimator.

In practice, however, it is almost always beneficial to add control variables to both IV stages to improve efficiency and/or reduce the strictness of the exogeneity requirements. A key advantage of the SP-IV estimator is that it easily allows such conditioning, up to the point where lag exogeneity is no longer required. This is not the case in the single-equation setting, where conditioning on a full information set generally results in the loss of potentially valuable identifying information. To see this, consider the 2SLS regression of $\bar{y}_t$ on

$\bar{Y}_t$ using $\bar{z}_{t-h}$, $h = 0, \ldots, H-1$ as instruments. This is the case where all variables are orthogonalized to all lagged shocks $\epsilon_{t-1}, \epsilon_{t-2}, \ldots$ by first conditioning on a full information set $\mathcal{I}_{t-1}^{full}$. Such conditioning reduces the residual variance in the two stages, and eliminates the lag exogeneity requirement just as in our approach. However, in the single-equation setting it also means that only the contemporaneous instruments $\bar{z}_t$ remain relevant. By construction, all $\bar{z}_{t-h}$ for $h > 0$ are uncorrelated with $\bar{Y}_t$, and must therefore be dropped from the instrument set. As a result, identification can no longer exploit the information from the full dynamic relationship between z_t and Y_t. When $N_z < K$, dropping the lags from the instrument set also leads to under-identification in the single-equation setting. In SP-IV, the conditioning step has all the usual benefits, but the identifying information from the dynamic relationship between z_t and the endogenous variables is fully preserved.

Under some circumstances, it is possible to partially address lag exogeneity concerns in the single-equation setting by considering the 2SLS regression of $\tilde{y}_t$ on $\tilde{Y}_t$ using $\bar{z}_{t-h}$, $h = 0, \ldots, H-1$ as instruments. That is, z_t is first conditioned on $\mathcal{I}_{t-1}^{full}$, but the second-stage regression still uses the unconditional data to preserve the relevance of lags of $\bar{z}_t$ in the instrument set. This approach is valid as long as $\bar{z}_t$ satisfies

$$(18) \qquad \mu_l' E[\epsilon_{t-l}\bar{z}_{t-h}'] = 0 \;\; ; \;\; l = 0, \ldots, \infty \;\; ; \;\; h = 0, \ldots, H-1 \; ,$$

which under stationarity is equivalent to $\mu_l' E[\epsilon_{t+h-l}\bar{z}_t'] = 0$ for $l = 0, \ldots, \infty$; $h = 0, \ldots, H-1$. Condition (18) is weaker than (13), but stronger than (15). It requires lag exogeneity with respect to the shocks in ϵ_t that are spanned by $\bar{z}_t$, but it does not require lag exogeneity with respect to all other shocks in ϵ_t. Even if the weaker lag exogeneity in (18) is plausibly satisfied, the single-equation setup with 'orthogonalized' instruments can still be inefficient relative to the SP-IV estimator. In Appendix A, for example, we show that when $\tilde{u}_t$ follows an AR(1) process, the SP-IV estimator with all data – not just the instruments – conditioned on $\mathcal{I}_{t-1}^{full}$ is asymptotically more efficient when the error term is sufficiently persistent and H is not too large.

In applications where the number of horizons H is large and lag exogeneity is plausible, single-equation 2SLS is still unlikely to be the better choice if the instruments are weak. As is well known, weak instruments generally introduce bias in 2SLS estimators, and make conventional inference methods unreliable. In those cases, it is necessary to use alternative weak-instrument-robust inference methods, for instance based on inverting the Anderson and Rubin (1949) statistic. As we discuss later, the same weak instrument problems arise for the SP-IV estimator, and in this paper we provide two robust inference methods to handle cases where instruments are weak. When the number of weak instruments is large, the finite sample bias in 2SLS can be large, and robust inference methods can still perform very poorly (e.g., Bekker (1994), Mikusheva (2021)). The SP-IV estimator exploits identifying information from dynamic correlations across H horizons by using only N_z instruments, whereas the corresponding single-equation 2SLS estimator instead uses HN_z instruments. In both cases, the rank condition for identification requires that $HN_z \geq K$. However, when H is large and N_z is small, the SP-IV estimator does not suffer from many-weak-instruments problems.

2.4 Interpretation of SP-IV as a Regression in Impulse Response Space

As mentioned earlier, the large macroeconomic literature identifying economic shocks is a natural source of instruments for estimating the structural parameters in (1). Barnichon and Mesters (2020) provide an appealing interpretation of single-equation 2SLS with a lag sequence of an economic shock as instruments: assuming lag, contemporaneous and lead exogeneity, 2SLS is equivalent to a 'regression in impulse response space', i.e. to regressing the IRF of y_t to the shock on the IRFs of Y_t to the same shock, where the IRF coefficients are estimated in regressions of y_t and Y_t on a distributed lag of the shock.

SP-IV similarly has the interpretation of a regression in impulse response space when the instruments are measures of economic shocks. Suppose that the instruments are an identified rotation of the first N_z shocks that drive the endogenous variables y_t and Y_t, that is $z_t^\perp = Q^{\frac{1}{2}}\epsilon_t^{1:N_z}$. Consider

$$(19) \qquad \hat{\Theta}_Y = \frac{Y_H^\perp Z^{\perp\prime}}{T}\left(\frac{Z^\perp Z^{\perp\prime}}{T}\right)^{-\frac{1}{2}} \quad ; \quad \hat{\Theta}_y = \frac{y_H^\perp Z^{\perp\prime}}{T}\left(\frac{Z^\perp Z^{\perp\prime}}{T}\right)^{-\frac{1}{2}} ,$$

which are the OLS coefficients in the regression of $Y_H^\perp$ and $y_H^\perp$ on the standardized instruments $\left(Z^\perp Z^{\perp\prime}/T\right)^{-\frac{1}{2}} Z^\perp$. Under conditions 2.a and 2.b, $\hat{\Theta}_Y \xrightarrow{p} \Theta_Y$, and since $z_t^\perp = Q^{\frac{1}{2}}\epsilon_t^{1:N_z}$ each of the columns Θ_Y collects the HK IRF coefficients of the K variables in Y_t to the corresponding structural shock in $\epsilon_t^{1:N_z}$. Similarly, assuming that $y_H^\perp Z^{\perp\prime}/T \xrightarrow{p} \Theta_y$, we have that $\hat{\Theta}_y \xrightarrow{p} \Theta_y$ where Θ_y contains the IRF coefficients of y_t to the structural shocks in $\epsilon_t^{1:N_z}$.

Using the elements of $\hat{\Theta}_y$, construct the $HN_z \times 1$ vector $\hat{\mathbf{\Theta}}_y$ by stacking the N_z vectors with the H IRF coefficients of y_t. Similarly, construct the $HN_z \times K$ matrix $\hat{\mathbf{\Theta}}_Y$ by stacking the vectors of IRF coefficients for Y_t. Formally,

$$(20) \hat{\mathbf{\Theta}}_Y = ((Z^\perp Z^{\perp\prime}/T)^{-\frac{1}{2}} Z^\perp \otimes I_H/T)\mathbf{Y}_H^\perp \quad ; \quad \hat{\mathbf{\Theta}}_y = ((Z^\perp Z^{\perp\prime}/T)^{-\frac{1}{2}} Z^\perp \otimes I_H/T)\mathbf{y}_H^\perp ,$$

where $\mathbf{y}_H^\perp = \text{vec}\left(y_H^\perp\right)$ is $TH \times 1$ and $\mathbf{Y}_H^\perp = [\text{vec}(Y_{H,1}^\perp), \ldots, \text{vec}(Y_{H,K}^\perp)]$ is $TH \times K$. With these definitions, $\hat{\beta}$ in (9) can equivalently be expressed as

$$(21) \qquad \hat{\beta} = (\mathbf{Y}_H^{\perp\prime}(P_{Z^\perp} \otimes I_H)\mathbf{Y}_H^\perp)^{-1}\mathbf{Y}_H^{\perp\prime}(P_{Z^\perp} \otimes I_H)\mathbf{y}_H^\perp = (\hat{\mathbf{\Theta}}_Y'\hat{\mathbf{\Theta}}_Y)^{-1}\hat{\mathbf{\Theta}}_Y'\hat{\mathbf{\Theta}}_y ,$$

which clarifies the interpretation of SP-IV as a projection of estimated IRFs, i.e. of $\hat{\mathbf{\Theta}}_y$ on $\hat{\mathbf{\Theta}}_Y$. Whereas single-equation 2SLS relies on IRFs estimated by a distributed lag specification, SP-IV can be based on the IRFs from more commonly used approaches, such as VARs or LPs. Moreover, these IRFs can be identified in a variety of ways. They can be based on external instruments, but also any other VAR or LP-based identification scheme, see Kilian and Lütkepohl (2017) and Plagborg-Møller and Wolf (2021) for overviews. Finally, note that SP-IV does not necessarily require using the identifying information over all horizons $h = 0, \ldots, H-1$, but can also be based on any subset of horizons.

2.5 Inference for SP-IV under Strong Identification

When identification is strong, which is the case under the conditions in Assumption 2, inference can proceed using methods analogous to those for single-horizon 2SLS estimators. To do so, we make the following additional high-level assumption,

Assumption 3. $T^{-1/2}\,\mathrm{vec}(Z^{\perp}u_H^{\perp\prime}) \xrightarrow{d} N(0, (\Sigma_{u_H^{\perp}} \otimes Q))$, *where* $\Sigma_{u_H^{\perp}}$ *is full rank.*

Rearranging the expression for $\hat{\beta}$ in (9) and using the fact that $\mathrm{vec}(\beta' \otimes I_H) = R\beta$,

$$\hat{\beta} - \beta = \left(R'(Y_H^{\perp} P_{Z^{\perp}} Y_H^{\perp\prime} \otimes I_H)R\right)^{-1} R'\,\mathrm{vec}(u_H^{\perp} P_{Z^{\perp}} Y_H^{\perp\prime}) \; . \tag{22}$$

Under Assumptions 2 and 3,

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_\beta) \; , \tag{23}$$

where

$$V_\beta = (R'(\Theta_Y \Theta_Y' \otimes I_H)R)^{-1} R' \left(\Theta_Y \Theta_Y' \otimes \Sigma_{u_H^{\perp}}\right) R\,(R'(\Theta_Y \Theta_Y' \otimes I_H)R)^{-1} \, . \tag{24}$$

A consistent estimator of V_β can be obtained by replacing $\Sigma_{u_H^{\perp}}$ with a consistent estimator, $\hat{\Sigma}_{u_H^{\perp}}$, and $\Theta_Y \Theta_Y'$ with $Y_H^{\perp} P_{Z^{\perp}} Y_H^{\perp\prime}$, and inference can be based on the standard Wald statistic.[5]

2.6 Generalized SP-IV

The efficient GMM estimator of β arises from using the weighting matrix $\Phi_s(\beta, \zeta) = (\Sigma_{u_H^{\perp}}^{-1} \otimes Q^{-1})$. This estimator is also the 'Generalized Least Squares' version of SP-IV that minimizes

[5] Note that, given the forecasting model in (6) and Assumption 1, estimation error in the forecast errors does not impact the asymptotic variance of $\hat{\beta}$. As in the standard IV setting, this is an application of the Frisch-Waugh-Lovell theorem. More formally, the expected Jacobian of the moments with respect to the model parameters is block-diagonal, since the entries corresponding to derivatives of second-stage moments with respect to first-stage parameters all feature products of control variables X_{t-1} and forecast errors $y_{H,t}^{\perp}, Y_{H,t}^{\perp}, z_t^{\perp}$, which are orthogonal by construction.

$\mathrm{Tr}\left((u_H^{\perp} P_{Z^{\perp}} u_H^{\perp\prime})\Sigma_{u_H^{\perp}}^{-1}\right)$. Given $\Sigma_{u_H^{\perp}}$, the estimator is available in closed form as

$$(25) \qquad \hat{\beta}_G = \left(R'\left(Y_H^{\perp} P_{Z^{\perp}} Y_H^{\perp\prime} \otimes \Sigma_{u_H^{\perp}}^{-1}\right) R\right)^{-1} R'\left(Y_H^{\perp} P_{Z^{\perp}} \otimes \Sigma_{u_H^{\perp}}^{-1}\right) \mathrm{vec}(y_H^{\perp} P_{Z^{\perp}}) \; .$$

We replace Assumption 2.d by

Assumption 2.d′. *$R'(\Theta_Y \Theta_Y' \otimes \Sigma_{u_H^{\perp}}^{-1})R$ is a fixed matrix with full rank.*

Under Assumptions 2.a-2.c, Assumption 2.d′ and Assumption 3,

$$(26) \qquad \sqrt{T}(\hat{\beta}_G - \beta) \xrightarrow{d} N(0, V_{\beta_G}) \quad , \quad V_{\beta_G} = \left(R'\left(\Theta_Y \Theta_Y' \otimes \Sigma_{u_H^{\perp}}^{-1}\right) R\right)^{-1} .$$

The Generalized SP-IV estimator is feasible only after replacing $\Sigma_{u_H^{\perp}}$ with a consistent estimator, either in a two-step or iterated procedure.

3 SP-IV with Weak Instruments

So far, we have assumed that identification is strong. In this section, we consider settings where identification is weak, which is often more realistic in macroeconomic applications. We first derive a bias-based test of instrument strength based on the first stage that is similar to the popular Stock and Yogo (2005) test for single-equation IV. Next, we describe inference procedures for the structural parameters that are robust to weak identification based on Anderson and Rubin (1949) and Kleibergen (2002).

3.1 A First-Stage Test for Weak Instruments

To model weak identification, we replace Assumptions 2 and 3 by Assumption 4.

Assumption 4. *The following limits hold*

$$(4.a) \qquad \left(u_H^{\perp} u_H^{\perp\prime}/T, u_H^{\perp} v_H^{\perp\prime}/T, v_H^{\perp} v_H^{\perp\prime}/T\right) \xrightarrow{p} \left(\Sigma_{u_H^{\perp}}, \Sigma_{u_H^{\perp} v_H^{\perp}}, \Sigma_{v_H^{\perp}}\right) \; ,$$

$$(4.b) \qquad T^{-1/2}\left[\mathrm{vec}(Z^{\perp} u_H^{\perp\prime})', \mathrm{vec}(Z^{\perp} v_H^{\perp\prime})'\right]' \xrightarrow{d} \left[\mathrm{vec}(\Psi_{Zu})', \mathrm{vec}(\Psi_{Zv})'\right]' \sim \mathcal{N}(0, Q \otimes \Sigma^{\perp}) \; ,$$

$$\Sigma^{\perp} = \begin{bmatrix} \Sigma_{u_H^{\perp}} & \Sigma_{u_H^{\perp} v_H^{\perp}} \\ \Sigma_{u_H^{\perp} v_H^{\perp}}' & \Sigma_{v_H^{\perp}} \end{bmatrix} \; \textit{non-singular} \; ,$$

and $v_H^{\perp} = Y_H^{\perp} - \Theta_Y Q^{-\frac{1}{2}} Z^{\perp}$ denote the first-stage error terms. Assumption 4.a requires that sample averages of the errors are consistent for their variances. Note that, in contrast to Stock and Yogo (2005), we allow for autocorrelation in both the first stage and structural equation errors by allowing correlation across horizons in v_H and u_H ($\Sigma_{u_H^{\perp}}$ and $\Sigma_{v_H^{\perp}}$ need not be diagonal). Assumption 4.b requires that a central limit theorem applies to suitably scaled sums, with a variance structured to reflect homoskedasticity of the errors conditional on $Z^{\perp}$.

We use the conventional local-to-zero asymptotic embedding for weak identification,

Assumption 5. $\Theta_Y = C/\sqrt{T}$ *where* C *is a fixed* $HK \times N_z$ *matrix.*

Assumption 5 implies that all coefficients in Θ_Y are on the same order.

To provide a benchmark for the SP-IV estimator, we define the 'system' version of the OLS estimator (SP-OLS) that minimizes $\mathrm{Tr}(u_H^{\perp} u_H^{\perp\prime})$, given by

$$\hat{\beta}_{\text{SP-OLS}} = \left(R'(Y_H^{\perp} Y_H^{\perp\prime} \otimes I_H)R\right)^{-1} R' \operatorname{vec}(y_H^{\perp} Y_H^{\perp\prime}). \tag{27}$$

Under Assumptions 4 and 5, $Y_H^{\perp} Y_H^{\perp\prime}/T \xrightarrow{p} \Sigma_{v_H^{\perp}}$ and $u_H^{\perp} Y_H^{\perp\prime}/T \xrightarrow{p} \Sigma_{u_H^{\perp} v_H^{\perp}}$, and the asymptotic bias of the SP-OLS estimator is

$$\hat{\beta}_{\text{SP-OLS}} - \beta \xrightarrow{p} \left(R'(\Sigma_{v_H^{\perp}} \otimes I_H)R\right)^{-1} R' \operatorname{vec}(\Sigma_{u_H^{\perp} v_H^{\perp}}). \tag{28}$$

Under Assumptions 4 and 5, $Y_H^{\perp} P_{Z^{\perp}} Y_H^{\perp\prime} \xrightarrow{d} \Sigma_{v_H^{\perp}}^{\frac{1}{2}}(L+Z_v)(L+Z_v)'\Sigma_{v_H^{\perp}}^{\frac{1}{2}}$ and $u_H^{\perp} P_{Z^{\perp}} Y_H^{\perp\prime} \xrightarrow{d} \Sigma_{u_H^{\perp}}^{\frac{1}{2}} Z_u (L+Z_v)'\Sigma_{v_H^{\perp}}^{\frac{1}{2}}$, where $L = \Sigma_{v_H^{\perp}}^{-\frac{1}{2}} C'$, $Z_u = \Sigma_{u_H^{\perp}}^{-\frac{1}{2}} \Psi_{Zu}' Q^{-\frac{1}{2}}$ and $Z_v = \Sigma_{v_H^{\perp}}^{-\frac{1}{2}} \Psi_{Zv}' Q^{-\frac{1}{2}}$, $[\operatorname{vec}(Z_u), \operatorname{vec}(Z_v)] \sim N(0, I_{N_Z} \otimes \bar{\Sigma})$, and $\bar{\Sigma} = \begin{bmatrix} I_H & \Sigma_{u_H^{\perp}}^{-\frac{1}{2}} \Sigma_{u_H^{\perp} v_H^{\perp}} \Sigma_{v_H^{\perp}}^{-\frac{1}{2}} \\ \Sigma_{v_H^{\perp}}^{-\frac{1}{2}} \Sigma_{u_H^{\perp} v_H^{\perp}}' \Sigma_{u_H^{\perp}}^{-\frac{1}{2}} & I_{HK} \end{bmatrix}$.

The asymptotic behavior of the SP-IV estimator is therefore

$$\hat{\beta} - \beta \xrightarrow{d} \left(R'\left(\Sigma_{v_H^{\perp}}^{\frac{1}{2}}(L+Z_v)(L+Z_v)'\Sigma_{v_H^{\perp}}^{\frac{1}{2}} \otimes I_H\right)R\right)^{-1} R' \operatorname{vec}\left(\Sigma_{u_H^{\perp}}^{\frac{1}{2}} Z_u (L+Z_v)'\Sigma_{v_H^{\perp}}^{\frac{1}{2}}\right). \tag{29}$$

Equation (29) states that $\hat{\beta}$ converges to a quotient of quadratic forms in normal random variables, and is therefore not consistent.

Deriving the Bias

To account for more than one endogenous regressor ($K \geq 1$) we consider the inner product of the asymptotic bias, weighted by $\Omega = R'(\Sigma_{v_H^{\perp}} \otimes I_H)R$. This weighting scheme mirrors that of Stock and Yogo (2005). Using (28), the weighted squared SP-OLS bias is

$$(E\hat{\beta}_{\text{SP-OLS}} - \beta)'\Omega(E\hat{\beta}_{\text{SP-OLS}} - \beta) = \rho' R R' \rho \leq H\rho'\rho\ , \tag{30}$$

where $\rho = ((\Omega^{-\frac{1}{2}} \otimes I_H) \otimes I_H) \operatorname{vec}(\Sigma_{u_H^{\perp} v_H^{\perp}})$ and the inequality provides the worst-case bias for SP-OLS for a given value of ρ. The weighted squared SP-IV bias is

$$(E\hat{\beta} - \beta)'\Omega(E\hat{\beta} - \beta) = \rho' \mathbf{h}' \mathbf{h} \rho\ . \tag{31}$$

where $\mathbf{h} = E\left[\left(R'\left(S(L+Z_v)(L+Z_v)'S'\otimes I_H\right)R\right)^{-1}R'\left(S(L+Z_v)Z_v'S^{-1}\otimes I_H\right)\right]$ and $S = (\Omega^{-\frac{1}{2}}\otimes I_H)\Sigma_{v_H^\perp}^{\frac{1}{2}}$. The squared SP-IV bias relative to the worst-case squared SP-OLS bias is therefore

$$B^2 = \frac{1}{H}\frac{\rho'\mathbf{h}'\mathbf{h}\rho}{\rho'\rho} \leq \frac{\text{maxeval}\left\{\mathbf{h}'\mathbf{h}\right\}}{H}, \tag{32}$$

where the bound is independent of ρ and depends only on $\Sigma_{v_H^\perp}$ and L.

In the single-equation case, Stock and Yogo (2005) show that when the number of instruments is large, the maximum 2SLS asymptotic bias is a decreasing function of the minimum eigenvalue of a concentration matrix. In our setting, the same approximation yields

$$\text{maxeval}\left\{\mathbf{h}'\mathbf{h}\right\}/H = \text{maxeval}\left\{\mathbf{h}\mathbf{h}'\right\}/H \approx \left(1+\text{mineval}\{\Lambda\}\right)^{-2}, \tag{33}$$

where $\Lambda = R'(SLL'S'\otimes I_H)R/N_z$ is the concentration matrix, and the approximate maximum bias depends only on its minimum eigenvalue. As in Stock and Yogo (2005), this dependence on the minimum eigenvalue of Λ motivates our bias-based test of instrument strength. We define the weak instrument set as $\mathcal{B} = \{L, \Sigma_{v_H^\perp} : |B| > \xi\}$, i.e. as the set of instruments associated with asymptotic relative bias greater than a tolerance level, ξ.

Test Statistic

Our test statistic is based on an extension of the Cragg and Donald (1993) statistic. Under weak instrument asymptotics

$$\Gamma = \Omega^{-\frac{1}{2}}R'(Y_H^\perp P_{Z^\perp}Y_H^{\perp\prime}\otimes I_H)R\Omega^{-\frac{1}{2}} \xrightarrow{d} R'(W\otimes I_H)R\,. \tag{34}$$

The random matrix $W = S(L+Z_v)(L+Z_v)'S'$ has a noncentral Wishart distribution, $W \sim \mathcal{W}(N_z, SS', D)$, with N_z degrees of freedom, covariance matrix SS' and noncentrality parameter $D = (SS')^{-1}SLL'S'$, using the notation in Muirhead (1982). Motivated by (33), we characterize the weak instrument set in terms of g_{min}, the minimum eigenvalue of Γ:

$$g_{min} = N_z^{-1}\,\text{mineval}\{\Gamma\} \;\xrightarrow{d}\; N_z^{-1}\,\text{mineval}\{R'(W\otimes I_H)R\}\,. \tag{35}$$

We compute $\ell_{min}(\xi)$, the threshold for the minimum eigenvalue of Λ associated with a level of bias ξ, numerically. Specifically, we draw a large number of $HK\times N_z$ matrices L_0 such that $SL_0L_0'S'$ has rank $\min\{HK, N_z\}$ and $R'(SL_0L_0'S'\otimes I_H)R/N_z$ has a minimum eigenvalue of 1. For each L_0, we set $L = \sqrt{x}L_0$ and compute $\mathbf{h}$ as a function of x by simulation. Next, we solve for the value of x such that $|B| = \xi$. The value of $\ell_{min}(\xi)$ is the supremum of the solutions for x over all draws of L_0.

While W has a noncentral Wishart distribution, the critical values for the test statistic g_{min} for a given bias tolerance ξ requires the distribution of $N_z^{-1}\,\text{mineval}\{R'(W \otimes I_H)R\}$. The form of this distribution is in general unknown, and in addition depends on all parameters of the concentration matrix (not just its minimum eigenvalue). In Appendix B, we derive upper bounds for the first three cumulants of $g_{\min}$ that only depend on $\ell_{min}(\xi)$ and S, and we show that a distribution proposed by Imhof (1961) matching these upper bounds is a conservative limiting distribution for g_{min} in the right tail. When $H = 1$, our test statistic corresponds to the usual Cragg and Donald (1993) statistic $N_z^{-1}\,\text{mineval}\{\Sigma_{v_H^\perp}^{-1/2} Y_H^\perp P_{Z^\perp} Y_H^{\perp\prime} \Sigma_{v_H^\perp}^{-1/2}\}$, as in Stock and Yogo (2005). When $H = 1$ or when SS' is the identity matrix (i.e. no serial correlation in the first-stage errors), the bounding cumulants correspond to those of a noncentral χ^2 with noncentrality $N_z \ell_{min}(\xi)$ and N_z degrees of freedom, which in those cases can be used for the critical values instead of the Imhof (1961) approximation.

3.2 Weak-Instrument Robust Inference for SP-IV

Under weak identification, the Wald statistic in Section 2.5 is not valid, and leads to empirical rejection rates that generally exceed the desired nominal levels, see for instance the simulations in Section 5. We describe two test statistics (with appropriate limiting distributions) that are robust to weak identification and are asymptotically correctly sized regardless of the strength of identification.

AR Statistic

The first test for SP-IV that is robust to weak identification is based on the Anderson and Rubin (1949) statistic for the setting with multiple outcome variables y_{t+h} with $h = 0, \ldots, H-1$. In particular, the AR test statistic associated with the SP-IV estimator in (9) and its limiting distribution under the null hypothesis are given by

$$AR(b) = (T - d_{AR})\,\text{Tr}\left(u_H^\perp(b) P_{Z^\perp} u_H^\perp(b)' \left(u_H^\perp(b) M_{Z^\perp} u_H^\perp(b)'\right)^{-1}\right)\;;\; AR(\beta) \xrightarrow{d} \chi^2_{HN_z}\,, \tag{36}$$

where $u_H^\perp(b) = y_H^\perp - (b' \otimes I_H) Y_H^\perp$, $d_{AR} = N_z + N_x$ is a degrees of freedom correction, and N_x is the number of predetermined control variables. Just as the conventional AR statistic for single-equation 2SLS, the AR statistic in (36) exploits the testable restriction that the instruments are uncorrelated with the residuals $u_H^\perp$ under the null hypothesis. In Appendix C.1, we show that $AR(\beta)$ has a χ^2 limiting distribution with HN_z degrees of freedom. For the Generalized SP-IV estimator, the AR statistic is asymptotically equivalent to the S statistic of Stock and Wright (2000), which can be used directly for inference by evaluating the GMM objective in (7) using the efficient weighting matrix.

KLM Statistic

While the AR statistic is robust to instrument strength, it can have poor power in practice when over-identifying restrictions are present. Kleibergen (2002) proposes the KLM (or K) statistic as an alternative to the AR statistic with better power properties in various settings, see also the overview in Andrews et al. (2019).

The degrees of freedom of the χ^2 limiting distribution of the AR statistic increase with the number of instruments – and, in our setting – also with the number of horizons. When the number of instruments (times the number of horizons in our case) is much larger than the number of structural parameters, the AR statistic has low power. The KLM statistic remedies this loss of power without introducing nuisance parameters. Instead of the covariance of $u_H^{\perp}$ and $(Z^{\perp}Z^{\perp\prime})^{-1/2}Z^{\perp}$, the numerator of the KLM statistic features the covariance of $u_H^{\perp}$ and the projection of a transformation of $Y_H^{\perp}$ on $(Z^{\perp}Z^{\perp\prime})^{-1/2}Z^{\perp}$. The transformation is necessary since, considering $u_H^{\perp}Z^{\perp\prime}(Z^{\perp}Z^{\perp\prime})^{-1/2}\hat{\Theta}_Y'$, $\hat{\Theta}_Y$ is in general correlated with $u_H^{\perp}Z^{\perp\prime}$ even asymptotically; the transformation ensures that the (consistent) estimate of Θ_Y is asymptotically independent of $u_H^{\perp}Z^{\perp\prime}$. This independence means that the limiting distribution does not depend on Θ_Y, and thus also not on the strength of the instruments.

The KLM statistic is a score statistic based on the derivative of the AR statistic in (36) with respect to b, accounting for the dependence of the variance estimator in the denominator on b. This step gives rise to the particular projection of $Y_H^{\perp}$ on $Z^{\perp}$ that the statistic exploits. The test statistic is a quadratic form of this score, normalized by the variance of the score. Based on this logic, the KLM statistic for our setting is:

$$
\begin{aligned}
K(b) &= (T-d_K)R'(\Xi^{-1}u_H^{\perp}(b)\check{Y}_H'\otimes I_H)R \\
&\quad\times\left(R'(\check{Y}_H\check{Y}_H'\otimes\Xi^{-1}u_H^{\perp}(b)u_H^{\perp\prime}(b)\Xi^{-1})R\right)^{-1} \\
&\quad\times R'(\check{Y}_H u_H^{\perp\prime}(b)\Xi^{-1}\otimes I_H)R\ , \\
K(\beta) &\xrightarrow{d}\chi_K^2\ ,
\end{aligned}
\tag{37}
$$

where $\check{Y}_H = Y_H^{\perp}P_{Z^{\perp}} - \check{v}_H^{\perp}\check{u}_H^{\perp\prime}(b)\left(\check{u}_H^{\perp}(b)\check{u}_H^{\perp\prime}(b)\right)^{-1}u_H^{\perp}(b)P_{Z^{\perp}}$ is the projection of $Y^{\perp}$ on $Z^{\perp}$, $\Xi = u_H^{\perp}(b)M_{Z^{\perp}}u_H^{\perp\prime}(b)$, $\check{v}_H^{\perp} = v_H^{\perp}M_{Z^{\perp}}$, and $\check{u}_H^{\perp}(b) = u_H^{\perp}(b)M_{Z^{\perp}}$. $d_K = N_z + N_x$ is a degrees of freedom correction and N_x is the number of predetermined control variables. Appendix C.2 derives $K(b)$ and shows that the stated limiting distribution holds under strong, weak, and non-identification. The Generalized SP-IV estimator coincides with the efficient GMM estimator, so in that case inference can instead proceed directly with the KLM statistic for GMM described in Kleibergen (2005).

4 A Guide For Implementing SP-IV

This section describes how to implement SP-IV based on two of the most widely used forecasting/IRF methods, Jordà (2005) local projections and vector autoregressive models (VARs).[6] SP-IV is highly flexible and, if concerned about misspecification, other models can also be used to estimate the conditional expectations entering the forecast errors.[7] We also discuss the implementation of the proposed inference procedures.

Let y_H denote the $H \times T$ matrix of leads of the outcome variable, i.e. with y_{t+h} in the $h+1$-th row and t-th column. Let Y_H be the $HK \times T$ matrix vertically stacking the $H \times T$ matrices Y_H^k for $k = 1, \ldots, K$, each of which has Y_{t+h}^k in the $h+1$-th row and t-th column, and Y_t^k is the k-th variable in the vector Y_t. Let X_t be the period t observation of an $N_x \times 1$ collection of predetermined control variables (including a constant). By assumption, X_t is generally a function of all current and all lagged values of the full set of shocks ϵ_t that drive y_t and Y_t, but X_t is independent of all future values of ϵ_t. Note that X_t can include not only current values, but also lags of y_t, Y_t, Z_t, or any other time series.

4.1 Implementation Using Local Projections

Define the $N_x \times T$ matrix X with controls X_{t-1} in the t-th column, and the projection matrix $P_X = X'(XX')^{-1}X$ and residualizing matrix $M_X = I_T - P_X$. Using a direct forecasting approach, the forecast errors after projection on X_{t-1} are given by

$$y_H^{\perp} = y_H M_X \quad , \quad Y_H^{\perp} = Y_H M_X \quad , \quad Z^{\perp} = Z M_X \ , \tag{38}$$

which can be used in (9) to obtain the SP-IV estimator $\hat{\beta}$. By the Frisch-Waugh-Lovell theorem, this direct forecasting approach is equivalent to estimating Jordà (2005) local projections of y_{t+h} and Y_{t+h} on z_t and X_{t-1} for $h = 0, \ldots, H-1$, using the estimated coefficients on z_t to construct the rows of $\hat{\boldsymbol{\Theta}}_y$ and $\hat{\boldsymbol{\Theta}}_Y$, and subsequently constructing the SP-IV estimator using the alternative expression for $\hat{\beta}$ in (21). When z_t are measures of economic shocks conditional on X_{t-1}, the LP estimates are IRF coefficients representing the dynamic causal effects of the shocks. Some studies estimate IRFs by local projections of an endogenous outcome variable at $t+h$ on an endogenous explanatory variable $\mathcal{Y}_t$ and controls X_{t-1} using z_t as instruments, a procedure often referred to as 'LP-IV'. Such IRFs can be used for identification in the SP-IV estimator exactly as described above, i.e. by the associated reduced form projections of the outcome variables directly on z_t and X_{t-1}.

[6] For recent assessments of both methods, see Stock and Watson (2018), Montiel Olea and Plagborg-Møller (2021), Plagborg-Møller and Wolf (2021) or Li et al. (2021).

[7] Depending on the model, however, it may be necessary to adjust inference for estimation error in the first-stage, since the expected Jacobian may no longer have the structure described in footnote 5.

4.2 Implementation Using Vector Autoregressions

Suppose that y_t, and the elements of Y_t and Z_t are – possibly together with other variables – all contained in X_t, and that X_t evolves according to a VAR,

$$X_t = AX_{t-1} + w_t \ . \tag{39}$$

The representation in terms of a VAR of order one is without loss of generality, as any VAR of order p can be rewritten as a VAR of order one. As before, let X denote the $N_x \times T$ matrix with X_{t-1} in the t-th column, and let X^f denote the $N_x \times T$ matrix with X_t in the t-th column. The standard estimator of A is $\hat{A} = X^f X'(XX')^{-1}$, leading to the h-step ahead forecast errors

$$X^{\perp}_{t+h} = \sum_{j=0}^{h} \hat{A}^{h-j}\hat{w}_{t+j} \quad , \quad \hat{w}_t = X_t - \hat{A}X_{t-1} \ . \tag{40}$$

The appropriate selection of elements in $X^{\perp}_{t+h}$ leads to $y^{\perp}_H$, $Y^{\perp}_H$ and $Z^{\perp}$, which can be used to obtain the SP-IV estimator $\hat{\beta}$ in (9). 'Structural' VARs are VARs in which researchers make assumptions to identify columns of B in $w_t = B\epsilon_t$, allowing the estimation of IRFs that are interpretable as dynamic causal effects of the associated economic shocks in ϵ_t. If $\hat{\epsilon}^{1:Nz}_t$ are the N_z identified shocks in the structural VAR, it is possible to use $z^{\perp}_t = \hat{\epsilon}^{1:Nz}_t$ to form $Z^{\perp}$, and use these shock estimates for identification in the SP-IV estimator. This procedure also nests the case of identification with 'external instruments', which can be directly included in the VAR and combined with zero restrictions in B, or used indirectly as instruments to identify columns in B as in the 'proxy SVAR' or 'SVAR-IV' approach, see Mertens and Ravn (2013), Stock (2008), Stock and Watson (2012), and Stock and Watson (2018). Note that (19), or equivalently (20), are consistent estimators of the IRFs associated with $\hat{\epsilon}^{1:Nz}_t$. In finite samples, however, even under (39) these IRF estimates will not be numerically identical to the structural VAR impulse responses obtained from $\hat{\Theta}^{VAR}_{X,h} = \hat{A}^h B^{1:Nz}$, $h = 0, \ldots, H-1$, where $B^{1:Nz}$ denotes the first N_z columns of B. The reason is that the restrictions implied by the VAR dynamics are imposed on the reduced form forecast errors, but (19) or (20) do not impose the same VAR dynamics on the IRFs. An alternative implementation of SP-IV with structural VARs is to select the elements corresponding to y_t and Y_t in $\hat{\Theta}^{VAR}_{X,h}$ to form $\hat{\mathbf{\Theta}}_y$ and $\hat{\mathbf{\Theta}}_Y$, and then obtain the SP-IV estimator from the regression of impulse responses as in (21). This alternative implementation imposes the VAR dynamics on both the reduced form forecast errors as well as on the impulse responses. In general, imposing the VAR dynamics is easily done in all formulas above by replacing $y^{\perp}_H P_{Z^{\perp}} Y^{\perp}_H$ by $\hat{\Theta}^{VAR}_y \hat{\Theta}^{VAR\prime}_Y$ and $Y^{\perp}_H P_{Z^{\perp}} Y^{\perp}_H$ by $\hat{\Theta}^{VAR}_Y \hat{\Theta}^{VAR\prime}_Y$, where $\hat{\Theta}^{VAR}_Y$ is the $HK \times N_z$ matrix stacking the K blocks of the VAR IRF coefficients of Y_t, and $\hat{\Theta}^{VAR}_y$ contains the $H \times N_z$ VAR IRF coefficients of y_t.[8] When comfortable imposing VAR dynamics, it makes sense to impose these restrictions consistently, and we therefore recommend this second implementation in practical applications of SP-IV with VAR-based IRFs.

4.3 Inference in Practice

4.3.1 Estimating the Error Variances

Given $y_H^{\perp}$, $Y_H^{\perp}$, $\hat{\Theta}_Y$ and $\hat{\beta}$ as obtained from either the LP or VAR implementation, the error term of the structural equation can be obtained by $\hat{u}_H^{\perp} = y_H^{\perp} - (\hat{\beta}' \otimes I_H)Y_H^{\perp}$, whereas the first-stage error terms can be obtained by $\hat{v}_H^{\perp} = Y_H^{\perp} - \hat{\Theta}_Y(ZM_XZ'/T)^{-\frac{1}{2}}ZM_X$. The error variance can be consistently estimated by

$$\hat{\Sigma}_{u_H^{\perp}} = \frac{\hat{u}_H^{\perp}\hat{u}_H^{\perp\prime}}{T - N_x - K} \quad ; \quad \hat{\Sigma}_{v_H^{\perp}} = \frac{\hat{v}_H^{\perp}\hat{v}_H^{\perp\prime}}{T - N_x - N_z} \ , \tag{41}$$

where N_x is the dimension of X_t (including a constant) in the LP or VAR. Note that (41) embeds any serial correlation up to H horizons. The variance estimators in (41) could be replaced with a heteroskedasticity and autocorrelation robust (HAR) estimator such as Newey and West (1987) or Lazarus et al. (2018). However, pre-multiplying the serially correlated errors in our estimators by $\Sigma_{u_H^{\perp}}^{-1/2}$ or $\hat{\Sigma}_{v_H^{\perp}}^{-1/2}$ (as, in effect, the denominators of our test statistics achieve) already has a pre-whitening effect, and the resulting errors become white noise when the order of serial correlation is less than H. Moreover, when K and H are relatively large, it is difficult to implement orthogonal-series HAR estimators as in Lazarus et al. (2018) due to the high dimensionality. Doing so may require a very large number of basis functions relative to the sample size to yield a full-rank covariance matrix. In practice, we find in simulations that inference based on (41) adequately adjusts for serial correlation in the errors, and systematically outperformed the HAR estimators that we considered.

4.3.2 Step-by-step Implementation of the Weak-Instruments Test

1. Using $\hat{\Sigma}_{v_H^{\perp}}$ in (41), construct $\hat{\Omega} = R'(\hat{\Sigma}_{v_H^{\perp}} \otimes I_H)R$, $\hat{S} = \left(\hat{\Omega}^{-\frac{1}{2}} \otimes I_H\right)\hat{\Sigma}_{v_H^{\perp}}^{\frac{1}{2}}$, and $\hat{\mathcal{S}} = \hat{S}\hat{S}'$.
2. Calculate the test statistic $g_{min} = N_z^{-1}\,\mathrm{mineval}\{\hat{\Omega}^{-\frac{1}{2}}R'(Y_H^{\perp}P_{Z^{\perp}}Y_H^{\perp\prime} \otimes I_H)R\hat{\Omega}^{-\frac{1}{2}}\}$
3. Obtain $\ell_{min}(\xi)$ using the numerical procedure outlined above and calculate the first three cumulants of the bounding limiting distribution of g_{min} under the null hypothesis

[8] To impose the VAR dynamics in the Generalized SP-IV formula (25), replace $y_H^{\perp}P_{Z^{\perp}}$ by $\hat{\Theta}_y^{VAR}(ZM_XZ'/T)^{-\frac{1}{2}}ZM_X$ and to construct $\check{Y}_H$ in the KLM statistic in (37), replace $Y_H^{\perp}P_{Z^{\perp}}$ by $\hat{\Theta}_Y^{VAR}(ZM_XZ'/T)^{-\frac{1}{2}}ZM_X$.

of a minimum eigenvalue of $\ell_{min}(\xi)$,

$$
\begin{aligned}
\kappa_1 &= N_z(1+\ell_{min}(\xi)) \ , \\
\kappa_2 &= 2\left(N_z\text{maxeval}\{R'(\hat{\mathcal{S}}^2 \otimes I_H)R\} + 2\ell_{min}(\xi)N_z\right) \ , \\
\kappa_3 &= 8\left(N_z\text{maxeval}\{R'(\hat{\mathcal{S}}^3 \otimes I_H)R\} + 3\ell_{min}(\xi)N_z\text{maxeval}\{R'(\hat{\mathcal{S}}^2 \otimes I_H)R\}\right) \ .
\end{aligned}
$$

Using $\nu = \kappa_2/\kappa_3$ and $\delta = 8\kappa_2\nu^2$, compute the critical value

$$
g^*_{min}(\xi,\alpha) = \frac{1}{N_z}\left(\frac{\chi^2_\delta(\alpha)-\delta}{4\nu} + \kappa_1\right) \ , \tag{42}
$$

where $\chi^2_\delta(\alpha)$ is the upper $\alpha\%$ percentile of a central χ^2 distribution with δ degrees of freedom.

Weak instruments are defined as instruments that generate a bias in $\hat{\beta}$ that in absolute value is ξ percent of the worst-case OLS bias or larger. The test rejects the null hypothesis of weak instruments when g_{min} exceeds the critical value $g^*_{min}(\xi,\alpha)$, with significance level of α. The procedure is generally straightforward, although the computation of $\ell_{min}(\xi)$ can be time-consuming for large H and K. As in Montiel-Olea and Pflueger (2013), a simpler alternative plug-in procedure is to replace $\ell_{min}(\xi)$ with $1/\xi$ in the cumulants above.

4.3.3 Constructing Confidence Sets for Strong Instruments

If the instruments are strong (as indicated by the first-stage test), asymptotically valid confidence sets can be constructed by inverting the usual Wald statistic. In particular, (24) provides the asymptotic variance for β, the elements of which can be replaced with their natural consistent estimators. Explicitly, the $1-\alpha$ confidence set for β is

$$
CS^W_\beta(\alpha) = \{b : T(\hat{\beta}-b)'V_\beta^{-1}(\hat{\beta}-b) < \chi^2_K(\alpha)\} \ . \tag{43}
$$

As usual, confidence sets for subsets of β can be formed by using selection matrices to take linear combinations of β, with comparison to limiting distributions with correspondingly reduced degrees of freedom.

4.3.4 Constructing Robust Confidence Sets

If the first-stage test fails to reject the null hypothesis of weak instruments, or if the econometrician prefers to use inference that will be valid regardless of instrument strength, confidence sets can be constructed by inverting either the AR or KLM statistics in (36) and (37) respectively. In particular, confidence sets for β are

$$
CS^{AR}_\beta(\alpha) = \{b : AR(b) < \chi^2_{HN_Z}(\alpha)\} \ ; \ CS^{KLM}_\beta(\alpha) = \{b : K(b) < \chi^2_K(\alpha)\} \ . \tag{44}
$$

These confidence sets provide valid inference for the full vector of structural parameters, β. If the econometrician is interested in a confidence set for only a subset of β, $\tilde{\beta}$ (with complement $\breve{\beta}$), confidence sets for $\tilde{\beta}$ can be defined as

(45)
$$CS^{AR}_{\tilde{\beta}}(\alpha) = \{\tilde{b} : \min_{\breve{b}} AR(\tilde{b}, \breve{b}) < \chi^2_{HN_Z}(\alpha)\} \ ; \ CS^{KLM}_{\tilde{\beta}}(\alpha) = \{\tilde{b} : \min_{\breve{b}} K(\tilde{b}, \breve{b}) < \chi^2_{K}(\alpha)\} \ .$$

Note that, as is the norm for robust inference, these confidence sets use the projection method, comparing the test statistics to critical values with degrees of freedom equal to that of the full-vector test. Reductions in degrees of freedom, like those for Wald tests on subsets of a parameter vector, only arise in special cases.

As in any weakly-identified setting, there are potential complications when constructing robust confidence sets. For instance, AR confidence sets may be infinite, if non-identification cannot be rejected, or empty, in the case of over-identifying restrictions that do not appear to be satisfied, see Andrews et al. (2019) for a discussion. Robust confidence sets will not, in general, be centered at the estimated values of the parameters. Since the KLM statistic is a quadratic form of the score, it will not only fail to reject parameter values associated with minima, but also with maxima and inflection points, meaning that confidence sets should be inspected to determine whether the included region contains these zeros as well, see Kleibergen (2002) for a discussion.

5 Performance of SP-IV in Model Simulations

We evaluate the performance of SP-IV using data generated from the workhorse macroeconomic model of Smets and Wouters (2007).[9] The objective is to estimate the parameters of the Hybrid New-Keynesian Phillips Curve (HNKPC) in (2), which is one of the model equations in a system of fourteen simultaneous dynamic equilibrium equations for the dynamics of key macroeconomic aggregates at a quarterly frequency. In Smets and Wouters (2007), the error term in (2) is the ARMA(1,1) process

(46)
$$u_t = \rho_u u_{t-1} + \epsilon^p_t - \mu_p \epsilon^p_{t-1} \ , \ \mid \rho_u \mid < 1 \ ,$$

where ϵ^p_t is an i.i.d. normally distributed price markup shock. Inverting the autoregressive term in (46) yields $u_t = \epsilon^p_t + \rho_u(1-\mu_p)\epsilon^p_{t-1} + \rho_u(\rho_u - \mu_p)\epsilon^p_{t-2} + \rho^2_u(\rho_u - \mu_p)\epsilon^p_{t-3} + \dots$, which makes clear that the error term in general depends on the entire history of price markup shocks $\epsilon^p_t, \epsilon^p_{t-1}, \epsilon^p_{t-2}, \dots$ The period t values of the endogenous variables in the model are functions of all current and lagged values of a 7×1 vector of shocks ϵ_t, of which ϵ^p_t is one

[9]The data is generated from the Smets and Wouters (2007) model using the Dynare replication code kindly provided by Johannes Pfeifer at https://sites.google.com/site/pfeiferecon/dynare.

element. Lagged values of standard endogenous macro variables are therefore in general not valid instruments. These variables either violate the lag exogeneity assumption, or else lose relevance if the data is first conditioned on predetermined variables to avoid the lag exogeneity requirement.

We assume that the econometrician cannot exploit the ARMA(1,1) error structure in (46). To achieve identification, we assume that a sample of N_z elements of the vector of shocks ϵ_t (other than ϵ_t^p) is observed. These shocks are mutually independent i.i.d. normal variables, and satisfy the necessary lead, contemporaneous, and lag exogeneity requirements for all estimators we consider. Using the true shocks as instruments is not a realistic assumption, as in real world applications researchers must confront the problem of empirically identifying these economic shocks in finite samples.[10] In this section, we focus on simulations with true shocks as instruments in order to level the playing field across estimators in this dimension.

We do not assume access to a full information set, that is a set of controls spanning the full history of all the shocks in the model. Instead, we use a more realistic set of controls containing four lags of seven endogenous variables: the short term interest rate, inflation, marginal cost, output, consumption, investment and the real wage. Price setters' inflation expectations π_{t+1}^e are assumed to be unobserved, and are replaced in (2) by realized future inflation π_{t+1}. This is a common approach in the literature when expectations appear in structural equations. Under rational expectations, as assumed in the Smets and Wouters (2007) model, the resulting measurement error only depends on future realizations of ϵ_t, which does not create any additional endogeneity problems since the instruments satisfy lead exogeneity.

We examine the performance of several consistent estimators of β. First, the single-equation 2SLS estimator regresses $\tilde{y}_t$ and $\tilde{Y}_t$ on the contemporaneous value and $H-1$ lags of the shocks in the first stage, and therefore implicitly relies on IRFs from a distributed lag specification. Second, we consider a single-equation 2SLS estimator that uses Almon shrinkage to estimate the IRFs. This estimator shrinks the number of instruments from HN_z to $3N_z$ by using quadratic approximations to the IRFs, which is what Barnichon and Mesters (2020) propose to avoid many-weak instrument problems. In their 2SLS-Almon procedure, the first stage consists of a regression $\tilde{y}_t$ and $\tilde{Y}_t$ on sums of the contemporaneous value and $H-1$ lags of the shocks, interacted with constant, linear and quadratic terms in the impulse response horizon h. Next, we consider two versions of SP-IV that obtain the forecast errors of the endogenous variables using LP, as described in Section 4.1. The two

[10] Wolf (2020) provides simulation results on the ability of various identification schemes to recover the true monetary policy shocks in the Smets and Wouters (2007) model.

versions differ in the controls: the first, SP-IV LP, uses only a constant, while the second, SP-IV LP-C, adds the set of lagged controls described above. Next, we consider SP-IV based on a VAR in the seven variables of the control set and with four lags. We implement this SP-IV VAR estimator using the IRFs that are directly obtained from inverting the estimated VAR model, as we recommend in Section 4.2. Finally, we also examine feasible GLS versions of the previous three SP-IV estimators as in Section 2.6. These are labeled FGSP-IV, and also correspond to the efficient GMM estimates.

We generate 5000 Monte Carlo samples for 12 different specifications, varying sample size, the number of horizons H, and the set of instruments. We evaluate sample sizes of $T = 250$ and $T = 500$ to assess the performance in small samples, and $T = 5000$ to verify the asymptotic properties of the estimators and inference procedures. We consider specifications with the first $H = 8$ or $H = 20$ horizons, corresponding to two and five years worth of leads (or lags in the case of single-equation 2SLS) of quarterly data, respectively. Finally, we consider $N_z = 1$, instrumenting with the monetary policy shock as in Barnichon and Mesters (2020), and a richer set of $N_z = 3$ instruments, adding the government spending and risk premium shocks. The tables with simulation results for $N_z = 3$ are in Appendix D for brevity. Appendix D also presents simulation results for the underlying IRF estimators.

5.1 Bias and Variance

Table 1 reports the mean estimates of $\beta = [\gamma_b, \gamma_f, \lambda]'$ for $N_z = 1$, with $H = 8$ and $H = 20$. The true values β are in the first row of each panel. As a benchmark for the bias of the various consistent estimators, the second row provides the simple OLS estimate from the regression of the demeaned variables, $\tilde{y}_t$ on $\tilde{Y}_t$. Unsurprisingly, the OLS estimates are strongly biased for all sample sizes because of simultaneity and measurement error in inflation expectations.

The first panel in Table 1, with results for $H = 8$, shows that the mean estimates for 2SLS and SP-IV LP (without controls) are very similar, and both reduce the bias considerably relative to OLS. Implementing SP-IV with controls, either using LP or a VAR, tends to further reduce the bias. Across all parameters, SP-IV VAR typically delivers the lowest bias, although SP-IV LP-C occasionally exhibits lower bias for λ. The FGSP-IV mean estimates are generally very similar to their SP-IV counterparts. Finally, the 2SLS-Almon estimator proposed in Barnichon and Mesters (2020) shows larger bias than OLS for γ_b and λ when $T = 250$, and only becomes competitive with the other estimators in larger samples.

The second panel in Table 1, with results for $H = 20$, shows that the ranking of the estimators and general performance is overall similar to the first panel, although the bias is often slightly larger. This suggests that the additional horizons generally do not add much

Table 1: Mean parameter estimates, $N_z = 1$

$H = 8$	$T = 250$			$T = 500$			$T = 5000$		
Estimator	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
β	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
2SLS	0.27	0.51	0.01	0.24	0.60	0.01	0.17	0.83	0.04
2SLS-Almon	0.55	0.85	-0.15	0.16	0.76	0.09	0.16	0.79	0.02
SP-IV LP	0.26	0.50	0.01	0.23	0.60	0.01	0.17	0.83	0.04
SP-IV LP-C	0.29	0.64	0.04	0.24	0.74	0.05	0.16	0.84	0.05
SP-IV VAR	0.23	0.81	0.03	0.18	0.84	0.05	0.12	0.83	0.09
FGSP-IV LP	0.25	0.44	0.00	0.23	0.54	0.00	0.17	0.82	0.03
FGSP-IV LP-C	0.30	0.66	0.03	0.25	0.76	0.03	0.16	0.84	0.05
FGSP-IV VAR	0.22	0.82	0.03	0.16	0.85	0.05	0.11	0.84	0.09
$H = 20$	$T = 250$			$T = 500$			$T = 5000$		
Estimator	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
β	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
2SLS	0.39	0.53	0.00	0.36	0.61	0.00	0.23	0.80	0.01
2SLS-Almon	0.28	0.73	0.02	0.55	0.48	-0.05	0.20	0.83	0.02
SP-IV LP	0.38	0.53	0.01	0.35	0.61	0.00	0.23	0.80	0.01
SP-IV LP-C	0.41	0.55	0.02	0.37	0.63	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
FGSP-IV LP	0.42	0.54	0.00	0.39	0.61	-0.01	0.26	0.79	0.01
FGSP-IV LP-C	0.46	0.57	0.00	0.42	0.65	0.00	0.25	0.80	0.01
FGSP-IV VAR	0.28	0.81	0.01	0.23	0.84	0.02	0.16	0.83	0.06

Notes: The top row in each panel contains the true parameter values $\beta = [\gamma_b, \gamma_f, \lambda]'$ of (2) in the Smets and Wouters (2007) model. The other rows show mean estimates across 5000 Monte Carlo samples of size T and with $h = 0, \ldots, H-1$. All IV estimators use the monetary policy shock as the instrument. 2SLS-Almon is the estimator proposed in Barnichon and Mesters (2020). SP-IV is the estimator in (9) while FGSP-IV is the feasible generalized estimator in (25). LP and LP-C denote implementations based on local projections discussed in Section 4.1, without and with controls, respectively. VAR denotes the implementation with a vector autoregression discussed in Section 4.2.

Table 2: Standard deviation of parameter estimates, $N_z = 1$

$H = 8$	$T = 250$			$T = 500$			$T = 5000$		
Estimator	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
2SLS	0.26	0.33	0.21	0.24	0.31	0.21	0.13	0.08	0.09
2SLS-Almon	21.15	8.77	8.79	11.38	10.56	7.17	4.38	3.17	3.92
SP-IV LP	0.27	0.34	0.23	0.25	0.30	0.21	0.13	0.09	0.09
SP-IV LP-C	0.28	0.29	0.27	0.26	0.21	0.24	0.12	0.06	0.08
SP-IV VAR	0.31	0.36	0.29	0.30	0.24	0.26	0.14	0.06	0.09
FGSP-IV LP	0.33	0.46	0.25	0.27	0.41	0.22	0.12	0.09	0.09
FGSP-IV LP-C	0.35	0.35	0.31	0.31	0.23	0.28	0.12	0.06	0.08
FGSP-IV VAR	0.36	0.42	0.35	0.33	0.27	0.30	0.13	0.06	0.09
$H = 20$	$T = 250$			$T = 500$			$T = 5000$		
Estimator	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
2SLS	0.11	0.12	0.05	0.11	0.11	0.06	0.07	0.05	0.03
2SLS-Almon	5.48	4.87	1.65	13.01	11.78	4.48	1.79	0.65	0.63
SP-IV LP	0.12	0.13	0.07	0.11	0.11	0.06	0.07	0.05	0.03
SP-IV LP-C	0.09	0.11	0.06	0.09	0.10	0.06	0.08	0.05	0.04
SP-IV VAR	0.22	0.25	0.11	0.20	0.19	0.09	0.11	0.06	0.05
FGSP-IV LP	0.15	0.17	0.07	0.12	0.15	0.06	0.07	0.05	0.03
FGSP-IV LP-C	0.11	0.12	0.07	0.10	0.09	0.06	0.08	0.05	0.03
FGSP-IV VAR	0.24	0.29	0.16	0.22	0.20	0.12	0.12	0.06	0.06

Notes: Rows show standard deviations across 5000 Monte Carlo samples of size T and with $h = 0, \dots, H-1$. All IV estimators use the monetary policy shock as the instrument. 2SLS-Almon is the estimator proposed in Barnichon and Mesters (2020). SP-IV is the estimator in (9) while FGSP-IV is the feasible generalized estimator in (25). LP and LP-C denote implementations based on local projections discussed in Section 4.1, without and with controls, respectively. VAR denotes the implementation with a vector autoregression discussed in Section 4.2.

useful identifying variation, and mostly exacerbate problems of weak identification. For the specifications with $N_z = 3$, see Appendix D, the general performance and ranking of the estimators remains very similar, although 2SLS-Almon performs better than in Table 1 with more instruments. In addition, the FGSP-IV estimators consistently show somewhat greater bias than their SP-IV counterparts with additional instruments.

Table 2 reports the standard deviations of the various estimators for $N_z = 1$, with results for $H = 8$ in the first panel, and $H = 20$ in the second. As expected, variances are everywhere decreasing in sample size. Comparing the first and second panel reveals that variances are also decreasing in H, indicating that additional horizons reduce the variability of all estimators. For $N_z = 1$, we find that the 2SLS-Almon estimator of Barnichon and Mesters (2020) is unstable, with standard deviations that are two orders of magnitude larger than those of all other estimators. The variances of single-equation 2SLS and the two LP-based SP-IV estimators are overall roughly similar for both $H = 8$ and $H = 20$, while the variance of SP-IV VAR is systematically somewhat higher than that of 2SLS or

SP-IV with LP. While the three FGSP-IV estimators are asymptotically more efficient, with $N_z = 1$, this advantage does not materialize in the sample sizes considered, as all FGSP-IV variances slightly exceed those of their SP-IV counterparts.

Using additional instruments leads to lower variances overall, see the results for $N_z = 3$ in Appendix D. The 2SLS-Almon estimator performs much better after adding instruments, but still tends to have higher variance than the other estimators. With $N_z = 3$, 2SLS and the LP-based SP-IV estimators continue to have similar variances, and the variance of SP-IV VAR continues to be generally slightly higher than 2SLS or SP-IV with LP. Finally, with more instruments, there is some sporadic evidence of (small) efficiency gains of FGSP-IV relative to their SP-IV counterparts.

Based on the simulation results across all 12 specifications, we draw several conclusions. First, the SP-IV VAR estimator consistently performs the best in terms of bias, followed by SP-IV LP-C, and then SP-IV LP and single-equation 2SLS. However, in terms of variance, the latter three estimators are comparable, and perform better than SP-IV VAR. This pattern may be surprising given existing results on the variance-bias trade-off between VARs and LPs for the estimation of IRFs (e.g., Plagborg-Møller and Wolf (2021), Li et al. (2021)).[11] However, the various SP-IV estimators do not estimate IRFs, but relationships across IRFs, see Section 2.4. Biases and covariances across IRFs can have offsetting or reinforcing effects on the bias and variance of the SP-IV estimators.

Second, SP-IV LP (without controls) and single-equation 2SLS are very close substitutes, both in terms of bias and variance. This is not that surprising since neither include controls and, given instruments that satisfy lead, contemporaneous, and lag exogeneity, both estimate very similar IRFs in the first stage. As discussed in Section 2.3, 2SLS can be asymptotically more efficient than SP-IV LP or vice versa, depending on H and the properties of the error term. In our simulations, the differences between the variances of the 2SLS and SP-IV LP estimators are generally very small. The benefits of SP-IV emerge once the controls are included, either in the LP-C or VAR implementation. As the simulation results show, adding predetermined variables as controls can lead to meaningful reductions in bias by moderating the weak-instrument bias in smaller samples. As we explained in Section 2.3, adding these controls is generally not feasible in single-equation 2SLS with lagged shocks as instruments. Recall also that, in our simulation setup, the instruments z_t always satisfy lag exogeneity. Were the raw instruments to be correlated with past price markup shocks, and thus fail the lag exogeneity condition, there would be a further performance wedge between SP-IV (LP-C or VAR) and single-equation 2SLS and 2SLS-Almon due to the bias induced

[11] Typically, imposing VAR dynamics introduces bias in the IRFs but yields efficiency gains relative to the LP approach. Appendix D shows that this trade-off is also present for the IRFs in our simulations.

Table 3: Empirical size of nominal 5% tests, $N_z = 1$

	$H=8$			$H=20$		
	$T=250$	$T=500$	$T=5000$	$T=250$	$T=500$	$T=5000$
WALD 2SLS	13.10	11.20	12.50	65.20	59.80	42.30
AR 2SLS	12.60	9.30	4.60	38.00	20.90	5.50
AR 2SLS-Almon	4.20	2.00	0.10	8.60	7.00	2.10
WALD SP-IV LP	15.70	13.00	12.90	69.40	63.90	43.70
WALD SP-IV LP-C	13.10	11.40	7.80	72.90	63.40	29.50
WALD SP-IV VAR	7.80	6.90	5.50	33.00	27.40	13.40
AR SP-IV LP	5.80	5.50	4.70	9.60	6.90	4.90
AR SP-IV LP-C	6.40	5.60	4.90	11.40	7.60	5.00
AR SP-IV VAR	4.40	4.80	4.80	5.20	5.80	4.70
KLM SP-IV LP	5.60	5.90	5.10	8.00	6.40	4.60
KLM SP-IV LP-C	7.20	5.70	5.10	11.70	7.30	4.80
KLM SP-IV VAR	5.30	5.30	4.90	8.20	6.60	4.60

Notes: Empirical rejection rates of various 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 using the monetary policy shock as the instrument. The 2SLS Wald test uses a HAR variance matrix following Lazarus et al. (2018). AR 2SLS and AR 2SLS-Almon are the Anderson and Rubin (1949) tests in Barnichon and Mesters (2020). WALD SP-IV is based on (23) with $\hat{\Sigma}_{u_H^\perp}$ as in (41), AR is based on (36), and KLM is based on (37). LP and LP-C denote implementations based on local projections discussed in Section 4.1, without and with controls, respectively. VAR denotes the implementation with a vector autoregression discussed in Section 4.2.

by invalid instruments.

Third, the feasible GLS (or efficient GMM) versions of our estimators (FGSP-IV) do not improve performance in practice, at least not in realistic sample sizes and for our data generating process. The bias is comparable or worse than the corresponding SP-IV estimators, and the variance is often higher for small sample sizes. The fact that GLS does not provide efficiency gains (and may fare slightly worse) in small samples likely results from estimation error in the $H \times H$ weighting matrix, which itself depends on the estimates $\hat{\beta}$.

Finally, the 2SLS-Almon estimator tends to perform poorly with the monetary policy shock as the only instrument in data from the Smets and Wouters (2007) model, although it fares better when there are more instruments. While a quadratic approximation to the IRFs reduces the number of instruments, in our simulations the approach proposed by Barnichon and Mesters (2020) appears generally inferior to conventional 2SLS or the SP-IV estimators in terms of bias and variance.

5.2 Inference

Table 3 reports empirical rejection rates of various nominal 5% tests of the true values of the full parameter vector, $\beta = [\gamma_b, \gamma_f, \lambda]'$. The table shows the empirical sizes for the specifications with $N_z = 1$; the table for $N_z = 3$ is in Appendix D.

We consider a 2SLS HAR Wald test – using the HAR estimator proposed by Lazarus et al. (2018) – as well as the Anderson and Rubin (1949) tests used in Barnichon and Mesters (2020) for 2SLS and 2SLS-Almon. The first row in Table 3 shows that the Wald test for 2SLS exhibits moderate size distortions for $H = 8$ and $N_z = 1$. The size distortions, however, become very large when $H = 20$, or when $N_z = 3$, see Appendix D. The AR test for 2SLS is relatively well-sized in small samples when $H = 8$ and $N_z = 1$, but becomes very oversized for $H = 20$, or when $N_z = 3$. This is entirely in keeping with the many-weak-instruments problem discussed earlier, and consistent with the findings in Barnichon and Mesters (2020) based on their data generating process.

The third row in Table 3 shows the AR tests for 2SLS-Almon, which shrinks the number of instruments to avoid many-weak instruments problems. For $H = 8$, we find that the tests are consistently under-sized. For $H = 20$, the tests have small positive size distortions in small samples, and negative size distortions in large samples. Moreover, we find that – unlike the AR test for 2SLS – the empirical rejection rates of the 2SLS-Almon AR test consistently converge towards zero as the sample size T increases.

The next three rows in Table 3 consider Wald tests for the three SP-IV estimators with $\hat{\Sigma}_{u_H^\perp}$ as in (41). Just as the Wald test for 2SLS, these tests are valid only under strong identification. Consistent with identification being weak, the SP-IV Wald tests exhibit large size distortions in many cases. Just as for the 2SLS Wald test, the size distortions are relatively moderate for $H = 8$ and $N_z = 1$, but become very large for higher H or N_z.[12] Note that the size distortions decrease with T, as the first-stage relationships stay fixed across specifications and the concentration parameter rises.

The remaining rows of Table 3 consider our robust inference methods for SP-IV. The next three rows report the AR tests based on (36). The SP-IV AR tests are generally well-sized, although they do exhibit some over-rejection in small samples when $H = 20$ in the case of the LP estimators. When $H = 8$, the SP-IV AR tests are also slightly conservative for the VAR estimator in small samples. Just as the AR test for 2SLS (but not for 2SLS Almon), the rejection rates of the SP-IV AR tests all approach 5% as the sample size increases, reflecting asymptotic validity. The final three rows of Table 3 report the KLM tests based on (37). Just as the SP-IV AR tests, the KLM tests exhibit at most only relatively small size distortions across all specifications, see also Appendix D.

[12] The size distortions are generally worse for the FGSP-IV estimators than their SP-IV counterparts. Results for FGSP-IV are omitted from Table 3 for brevity.

The main takeaway from Table 3 (and also D.3 in Appendix) is that for all SP-IV implementations the AR and KLM robust inference procedures are well-sized, showing only small size distortions when H is large relative to T. By leading the endogenous variables rather than lagging the instruments, SP-IV effectively mitigates the many-weak-instrument problems that plague the 2SLS AR test in small samples. The 2SLS-Almon shrinkage procedure of Barnichon and Mesters (2020) also reduces the number of instruments, but their associated AR test appears to be incorrectly sized asymptotically. Other approaches in the literature may also be able to address many-weak-instrument problems in time series settings, see Mikusheva (2021) for suggestions, but they do not necessarily have any of the other advantages of SP-IV. There may also be test statistics offering refinements over the AR and KLM, but many – for example, the CLC of Andrews (2016) – are only directly applicable to the Generalized SP-IV estimators.

6 Application to the New Keynesian Phillips Curve with U.S. Data

In this section, we use SP-IV to estimate the parameters of the Hybrid New Keynesian Phillips curve in (2) using U.S. data, and compare our results with those from OLS, 2SLS, and 2SLS with the Almon approximation. We consider the following specification for the dynamics of quarterly inflation at a monthly frequency,

$$\pi_t^{1q} = (1-\gamma_f)\pi_{t-3}^{1y} + \gamma_f \pi_{t+12}^{1y} + \lambda U_t + u_t \ , \tag{47}$$

where π_t^{1q} is the annualized percent change in the Core CPI from a quarter ago in month t, π_t^{1y} is the percent change in the Core CPI over the preceding year in month t, and U_t is the headline unemployment rate in month t. The variable definitions in terms of quarterly and annual lagged and future inflation rates and unemployment as the gap measure are the same as in Barnichon and Mesters (2020), but we estimate (47) using monthly data instead of quarterly data. As is common in the literature, e.g. Mavroeidis et al. (2014), the specification in (47) restricts the coefficient on lagged and future inflation to sum to one, $\gamma_b + \gamma_f = 1$. This parameter restriction imposes that there is no long run trade-off between unemployment and inflation.

As the instrumenting shock, we use a monthly version of the Angeletos et al. (2020) Main Business Cycle (MBC) Shock. Specifically, we estimate a monthly six-variable VAR using the annualized one-month percent change in the core CPI, the unemployment rate, the 12-month change in log industrial production, the 12-month percent change in the PPI for all commodities, the 3-month T-bill rate, and the 10-year Treasury rate. The effective sample period is 1979:M1 to 2018:M4 (472 monthly observations), and we use 6 lags in the VAR. The MBC shock is identified as the shock that – out of all orthogonal rotations of structural shocks – maximizes the contribution to the variation in the unemployment rate at horizons of 18 to 96 months in the frequency domain.

In principle, there is a range of economic shock measures that could be suitable to identify the parameters of (47). There are several reasons why we use the MBC shock, and not measures of monetary policy shocks as in Barnichon and Mesters (2020). The most useful instruments have strong predictive power for the endogenous variables, while still satisfying the exogeneity requirements. High frequency measures of monetary policy shocks are far too weak as predictors of unemployment and inflation to be useful in practice. The same is typically – though not always – the case for monetary shocks identified through timing restrictions or the narrative measures of Romer and Romer (2004). Moreover, contractionary policy shocks identified by these last two methods robustly generate puzzling expansionary effects in updated samples, raising questions about their interpretation and exogeneity.[13] We also show below that point estimates based on the Romer and Romer (2004) shocks give rise to unrealistic cyclical inflation dynamics.

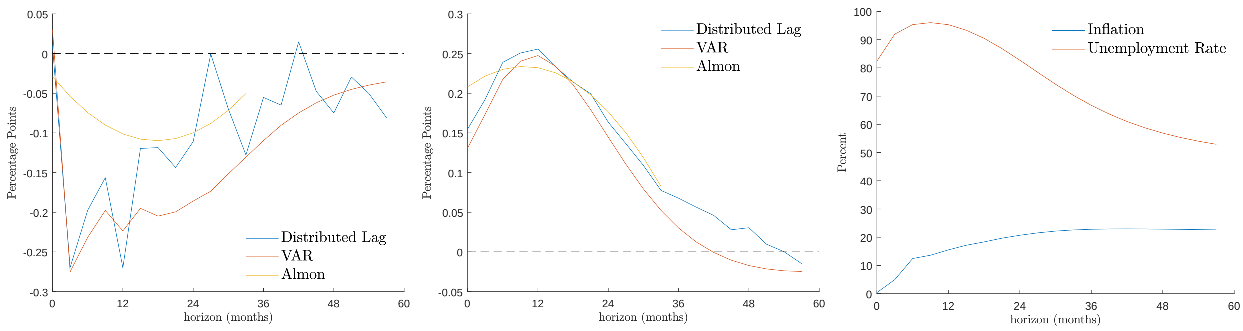
Angeletos et al. (2020) find that the MBC shock obtained by maximizing the contribution to cyclical unemployment fluctuations is interchangeable with shocks identified by maximizing the cyclical variance contribution to other major macro aggregates, such as GDP, consumption, investment, or hours worked. This interchangeability suggests a single main driver of business cycles with a common propagation mechanism. The authors argue that this main driver best fits the notion of an aggregate demand shock, making it potentially a good instrument for estimating the Phillips curve. Indeed, looking at the disconnect between the unemployment and inflation responses to the MBC shock, Angeletos et al. (2020) conclude that the Phillips curve must be overly flat, and suggest that demand-driven business cycles are perhaps not tied to nominal rigidities at all. Rather than relying on casual inspections of the IRFs, SP-IV allows for a formal investigation of such claims.

Our monthly version of the MBC shock produces IRFs that are very similar to those in Angeletos et al. (2020). The red lines in Figures 1a-1b plot the VAR-based IRFs of quarterly inflation π_t^{1q} and unemployment U_t following a one standard-deviation MBC shock, while Figure 1c shows the contributions of the MBC shock to the forecast error variance (FEV). As in Angeletos et al. (2020), the MBC shock looks like an aggregate demand shock, pushing unemployment higher and inflation lower. At the same time, the MBC shock explains a relatively small fraction of the FEV of inflation. On impact, the MBC shock explains essentially zero percent of the inflation FEV, and the contribution rises only to slightly above 20% after about two years. This is the apparent disconnect between inflation and the shock that explains most of the variance of unemployment at business cycle frequencies.

[13] See for instance Barakchian and Crowe (2013), Ramey (2016), or Miranda-Agrippino and Ricco (2021).

Figure 1: Impact of the Main Business Cycle Shock on Core CPI and Unemployment

(a) IRF Core CPI Inflation (b) IRF Unemployment Rate (c) FEV Contribution

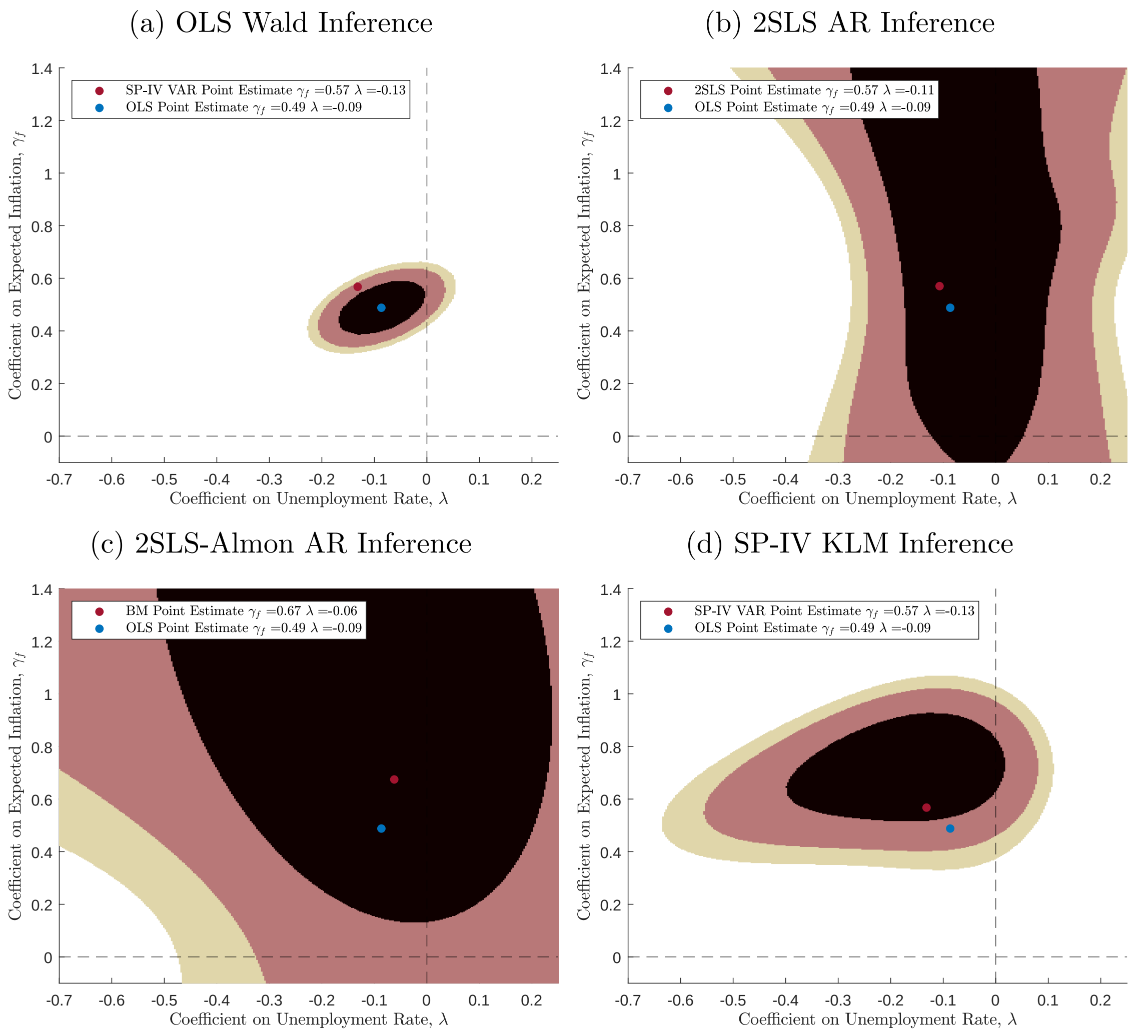


Notes: Inflation is the annualized inflation rate from a quarter ago (π_t^{1q}). Results in red are obtained from a six-variable VAR with six lags using the annualized one-month percent change in the core CPI, the unemployment rate, the 12-month change in log industrial production, the 12-month percent change in the PPI for all commodities, the 3-month T-bill rate, and the 10-year Treasury rate. The MBC shock maximizes the contribution to the variance of the unemployment rate at horizons of 18 to 96 months in the frequency domain, as in Angeletos et al. (2020). The effective sample period is 1979:M1 to 2018:M4. Blue lines show results from regressions on lag sequences of the MBC shock, whereas yellow lines show results from a quadratic (Almon) approximation.

By using the MBC shock as the instrument, the 2SLS, SP-IV and 2SLS-Almon estimators each produce estimates of the Phillips curve parameters as they are encoded in the IRFs associated with an MBC shock. The 2SLS estimator uses contemporaneous and lagged values the MBC shocks as the instrumental variables. The 2SLS-Almon estimator shrinks the number of instrumental variables down to three by using sums of the contemporaneous and lagged values interacted with constant, linear, and quadratic terms in the impulse response horizon h as the instrumental variables. The SP-IV estimator uses the contemporaneous MBC shock as a single instrument in a system of forecast errors.[14] In each case, the first stage leads to different estimators of the IRFs associated with the MBC shock. Figure 1a-1b show the three different IRFs that implicitly underlie the estimates of the parameters γ_f and λ in each case. The 2SLS estimator is built from the IRF coefficients obtained from regressions of π_t^{1q} (and π_{t-3}^{1y} and π_{t+12}^{1y}) and U_t on a distributed lag of the shock (blue lines). The 2SLS-Almon of Barnichon and Mesters (2020) is similar but uses quadratic approximations to those IRFs (yellow lines). SP-IV, in contrast, allows the direct use of the VAR-based IRFs (red lines). To make efficient use of the identifying information contained in the IRF dynamics, we use the coefficients in the first month of the first 12 quarters of the response horizons – that is at $h = 0, 3, 6, \ldots, 33$ – to construct each estimator. Figures 1a-1b show the first twelve IRF coefficients that are used in practice in the estimation, and also show the next eight quarters of the VAR and distributed lag IRF coefficients to visualize the full dynamics following an MBC shock.

[14]The forecast errors of π_t^{1q},π_{t-3}^{1y} and π_{t+12}^{1y} are straightforward to obtain from the VAR based on the forecast errors of monthly inflation π_t^{1m}.

Figure 2: OLS and SP-IV Confidence Sets for Estimates of Hybrid NK Phillips Curve Parameters



Notes: Figures show point estimates and 68%, 90% and 95% confidence sets. The OLS sets are based on the HAR Wald statistic. The 2SLS and 2SLS-Almon sets are those in Barnichon and Mesters (2020). The SP-IV sets are based on the KLM statistic described in Sections 3.2 and 4.3.4.

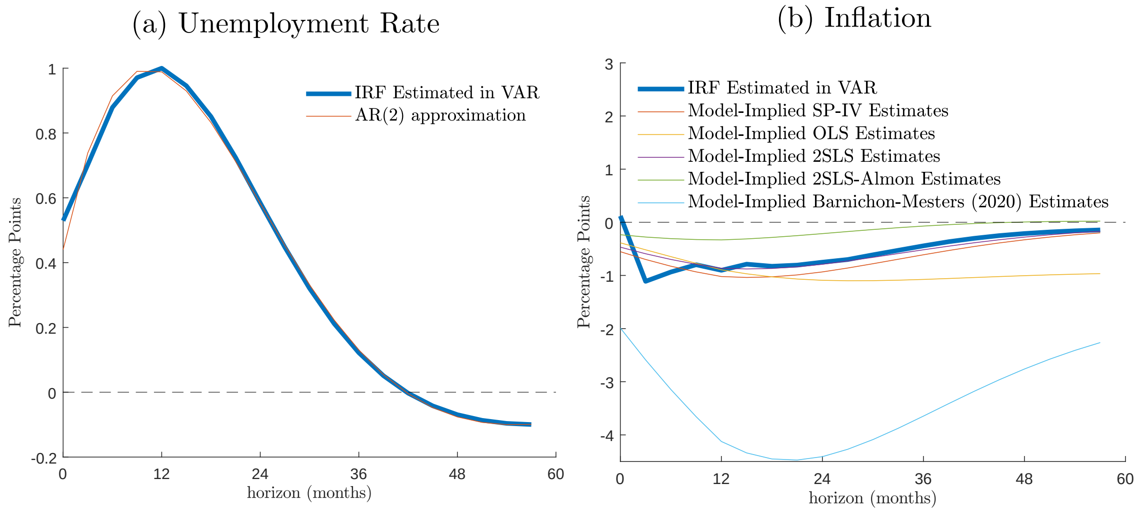
Figure 2 shows the different estimates of γ_f and λ, together with 68%, 90% and 95% confidence sets. The point estimates of γ_f, the weight on future inflation, are 0.57 for both 2SLS and SP-IV, which is larger than the OLS estimate of 0.49 and smaller than the 2SLS-Almon estimate of 0.67. The 2SLS-Almon point estimate for the coefficient on unemployment, λ, is the smallest in absolute value ($\lambda = -0.06$) while the OLS, 2SLS and SP-IV estimates each successively imply steeper slopes of the Phillips curve ($\lambda = -0.09$ to -0.11 and -0.13 respectively). Before discussing the implications of these estimates for the cyclical dynamics of inflation, we first discuss the inference results. The OLS confidence set shown in Figure 2a is based on the HAR Wald statistic, using the variance estimator of Lazarus et al. (2018). This set is relatively tight, but of course potentially misleading because of endogeneity bias. For 2SLS and 2SLS-Almon, the sets shown in Figures 2b-2c are the AR sets used in Barnichon and Mesters (2020). Neither reject weights on future inflation as low as zero or as high as one at the 5% or 10% significance levels, or are able to rule out a wide range of possible Phillips curve slopes. The 90% set for 2SLS includes values of λ as high as 0.2 and as low -0.3, and the 95% set includes an even wider range for λ. As is well known, and was also evident in our simulations earlier, AR-based inference for 2SLS can be very unreliable in the face of many weak-instruments problems. The shrinkage to three instruments under the 2SLS-Almon can avoid these problems, as shown in Barnichon and Mesters (2020), but our simulations also showed rejection rates tending to zero with sample size.

Turning to the SP-IV inference, we first apply the testing procedure in Section 3.1 to assess instrument strength. For the SP-IV VAR, we find a value for the g_{min} test statistic in (35) of 7.76. The 5% critical value associated with the hypothesis that the bias does not exceed 10% percent of the worst-case OLS bias is 21.88. Hence, we cannot reject that the MBC shock is a weak instrument, and must therefore rely on one of the robust inference procedures.[15] Figure 2d shows robust confidence sets for the SP-IV estimator based on the KLM statistic. The simulation evidence in Section 5 demonstrated the good performance of SP-IV VAR with KLM inference in small and large samples in both absolute terms and compared with the other approaches, particularly when $N_z = 1$ and H is not too large. We therefore view the KLM sets in Figure 2d, where $N_z = 1$ and $H = 12$, as the most reliable for inference. Compared with the 2SLS approaches, inference for SP-IV is much sharper for the weight on future inflation, with the confidence set ruling out values of γ_f that are meaningfully below 0.4 or above 1. At the same time, the KLM sets also do not rule a wide range of possible Phillips curve slopes, with values of λ ranging from -0.5 to slightly greater than zero within the 90% set.

To gain insight on the implications for cyclical inflation dynamics, we consider a back-of-the-envelope calculation by embedding the estimated Phillips curves in a simple dynamic rational expectations economic model in which unemployment follows an exogenous $AR(2)$ process, $U_t = \rho_1 U_{t-1} + \rho_2 U_{t-2} + \epsilon_t^{MBC}$. We set ρ_1 and ρ_2 to minimize the Euclidean distance between the IRFs of U_t to ϵ_t^{MBC} in the model and the VAR. Figure 3a shows that the resulting $AR(2)$ approximation of the unemployment dynamics closely matches the IRF of unemployment from the VAR. To assess the inflation dynamics implied by the various estimates of γ_f and λ, we consider the impact of a recessionary shock ϵ_t^{MBC} with a peak response of the unemployment rate of one percentage point. Figure 3b shows the responses of inflation based on Phillips curves parametrized by the OLS, SP-IV, 2SLS and 2SLS-Almon point estimates reported in Figure 2. The Phillips curves with the 2SLS and SP-IV parameters both provide good approximations to inflation dynamics estimated by the VAR

[15] A similar first-stage test of instrument strength that allows for serially correlated errors and multiple endogenous regressors is currently not available for single-equation 2SLS.

Figure 3: Empirical and Model Inflation Dynamics After a Recessionary Shock



Notes: Figures show model-based and VAR-estimated responses to a recessionary shock with a peak unemployment response of one percentage point. The model consists of the Phillips curve $\pi_t^{1q} = (1-\gamma_f)\pi_{t-3}^{1y} + E_t\gamma_f\pi_{t+12}^{1y} + \lambda U_t + u_t$, the AR(2) process $U_t = \rho_1 U_{t-1} + \rho_2 U_{t-2} + \epsilon_t^{MBC}$ and assumes rational expectations.

at all horizons, including those beyond the initial twelve (at three-month intervals) that are used in the estimation. The OLS estimates also lead to a reasonable approximation at horizons up to a year, but do not capture the dynamics at longer horizons very well. The 2SLS-Almon estimates considerably understate the inflation response to a recessionary shock at all horizons, as the simple quadratic function in practice provides a poor approximation to inflation response after an MBC shock, see Figure 1a. Finally, the figures also show the inflation dynamics implied by the estimates of $\gamma_f = 0.53$ and $\lambda = -0.45$ reported in Barnichon and Mesters (2020) based on 2SLS-Almon and responses to the Romer and Romer (2004) monetary policy shocks. These estimates imply a much stronger disinflationary response than suggested by the empirical inflation response to a recessionary MBC shock. The magnitude of these responses to a shock increasing the unemployment rate by one percentage point – a decrease in inflation by up to 4 percentage points – strikes us as implausible, and cause us to doubt that the responses to Romer and Romer (2004) shocks contain reliable information about the relationship between inflation and unemployment.

In the above application to the Phillips curve, single-equation 2SLS and SP-IV provide similar point estimates, and both lead to realistic cyclical inflation dynamics based on a Phillips curve with just two parameters. However, only the SP-IV inference is robust to the many-weak-instruments problems that arise when using identifying information across a relatively large number of time horizons. In other applications the 2SLS and SP-IV estimates may also differ more substantially, as the IRFs from VARs or LPs are in practice not always in such close agreement with those from distributed lag specifications without controls. The weaker exogeneity requirements of SP-IV can also be important, and a key advantage of SP-IV is that it enables researchers to fit structural equations directly to the IRFs obtained from methods that are preferred in practice, such as VARs or LPs. Moreover, the SP-IV methodology provides a way to formally test claims about structural relationships embedded in these empirical impulse responses. On the basis of the estimated inflation dynamics following the MBC shock of Angeletos et al. (2020), our SP-IV inference points towards a likely greater weight on future inflation than on lagged inflation, and the confidence sets are consistent with a wide range of possible cyclical responses of inflation, both weak and relatively strong. The evidence based on MBC shocks therefore does not necessarily support the conclusion that inflation dynamics are disconnected from the business cycle, or that the Hybrid New Keynesian Phillips curve is of little use to model these dynamics.

7 Concluding Remarks and Future Research

We conclude by discussing several other potential interesting applications, and some avenues for future research. SP-IV should be useful for estimating a wide variety of structural relationships in macroeconomics, such as Euler equations for consumption or investment, the wage Phillips curve, monetary or fiscal policy rules, or aggregate production functions. SP-IV can also be used more broadly to conduct inference on ratios (or other relationships) of impulse response coefficients, such as Okun coefficients, sacrifice ratios, multipliers, etc. conditonal on economic shocks.

In this paper, we have taken the selection of horizons as given. Future work can study methods for the optimal selection of the horizons. If $h = 0, ..., H-1$ indexes cross-sectional groups rather than time horizons, then this paper also describes instrumental variables in the cross-section with heterogeneity in the first stage. Our methodology could be extended to panel data settings, and be potentially useful in applications that commonly rely on lagged variables as instruments, such as the estimation of production functions in industrial organization, see Wooldridge (2009). Finally, the derivation of the first-stage test for instrument strength in this paper naturally follows the same steps required for a first-stage test for single-equation 2SLS under heteroskedasticity and autocorrelation with respect to the instruments, therefore extending Montiel-Olea and Pflueger (2013) to the case of multiple endogenous regressors. We plan to pursue these and other avenues in future research.

References

Alzer, Horst (1997). "On Some Inequalities for the Gamma and Psi Functions". In: *Mathematics of Computation* 66.217, pp. 373–389. DOI: 10.1090/S0025-5718-97-00807-7.

Anderson, T. W. and Herman Rubin (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations". In: *The Annals of Mathematical Statistics* 20.1, pp. 46 –63. DOI: 10.1214/aoms/1177730090.

Andrews, Isaiah (2016). "Conditional Linear Combination Tests for Weakly Identified Models". In: *Econometrica* 84.6, pp. 2155–2182. DOI: 10.3982/ECTA12407.

Andrews, Isaiah, James H. Stock, and Liyang Sun (2019). "Weak Instruments in Instrumental Variables Regression: Theory and Practice". In: *Annual Review of Economics* 11.1, pp. 727–753. DOI: 10.1146/annurev-economics-080218-025643.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas (2020). "Business-Cycle Anatomy". In: *American Economic Review* 110.10, pp. 3030–70. DOI: 10.1257/aer.20181174.

Barakchian, S. Mahdi and Christopher Crowe (2013). "Monetary policy matters: Evidence from new shocks data". In: *Journal of Monetary Economics* 60.8, pp. 950–966. DOI: 10.1016/j.jmoneco.2013.09.

Barnichon, Regis and Geert Mesters (2020). "Identifying Modern Macro Equations with Old Shocks". In: *The Quarterly Journal of Economics* 135.4, pp. 2255–2298. DOI: 10.1093/qje/qjaa022.

Bekker, Paul A. (1994). "Alternative Approximations to the Distributions of Instrumental Variable Estimators". In: *Econometrica* 62.3, pp. 657–681. DOI: 10.2307/2951662.

Cragg, John G. and Stephen G. Donald (1993). "Testing Identifiability and Specification in Instrumental Variable Models". In: *Econometric Theory* 9.2, pp. 222–240. DOI: 10.1017/S0266466600007519.

Forni, Mario and Luca Gambetti (2014). "Sufficient information in structural VARs". In: *Journal of Monetary Economics* 66, pp. 124–136. DOI: 10.1016/j.jmoneco.2014.04.005.

Galí, Jordi and Luca Gambetti (2020). "Has the U.S. Wage Phillips Curve Flattened? A Semi-Structural Exploration". In: ed. by Gonzalo Castex, Jordi Galí, and Diego Saravia. Vol. 27. Changing Inflation Dynamics, Evolving Monetary Policy. Central Bank of Chile.

Galí, Jordi and Mark Gertler (1999). "Inflation dynamics: A structural econometric analysis". In: *Journal of Monetary Economics* 44.2, pp. 195–222. DOI: 10.1016/S0304-3932(99)00023-9.

Imhof, J. P. (Dec. 1961). "Computing the distribution of quadratic forms in normal variables". In: *Biometrika* 48.3-4, pp. 419–426. DOI: 10.1093/biomet/48.3-4.419.

Jordà, Òscar (2005). "Estimation and Inference of Impulse Responses by Local Projections". In: *American Economic Review* 95.1, pp. 161–182. DOI: 10.1257/0002828053828518.

Kilian, Lutz and Helmut Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge University Press. DOI: 10.1017/9781108164818.

Kleibergen, Frank (2002). "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression". In: *Econometrica* 70.5, pp. 1781–1803. DOI: 10.1111/1468-0262.00353.

Kleibergen, Frank (2005). “Testing Parameters in GMM Without Assuming that They Are Identified”. In: *Econometrica* 73.4, pp. 1103–1123. DOI: 10.1111/j.1468-0262.2005.00610.x.

Kleibergen, Frank, Sophocles Mavroeidis, Fabio Canova, John C. Chao, Norman R. Swanson, Jean-Marie Dufour, Anna Mikusheva, Jonathan H. Wright, Motohiro Yogo, Eric Zivot, and Saraswata Chaudhuri (2009). “Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve”. In: *Journal of Business & Economic Statistics* 27.3, pp. 293–339. DOI: 10.1198/jbes.2009.08280.

Lazarus, Eben, Daniel J. Lewis, James H. Stock, and Mark W. Watson (2018). “HAR Inference: Recommendations for Practice”. In: *Journal of Business & Economic Statistics* 36.4, pp. 541–559. DOI: 10.1080/07350015.2018.1506926.

Li, Dake, Mikkel Plagborg-Moller, and Christian K. Wolf (Apr. 2021). *Local Projections vs. VARs: Lessons From Thousands of DGPs*. Papers 2104.00655. arXiv.org.

Mathai, Arak (Jan. 1980). “Moments of the trace of a noncentral Wishart matrix”. In: *Communications in Statistics - Theory and Methods* 9, pp. 795–801. DOI: 10.1080/03610928008827921.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H. Stock (2014). “Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve”. In: *Journal of Economic Literature* 52.1, pp. 124–88. DOI: 10.1257/jel.52.1.124.

McLeay, Michael and Silvana Tenreyro (2019). “Optimal Inflation and the Identification of the Phillips Curve”. In: *NBER Macroeconomics Annual* 34. DOI: https://doi.org/10.1086/707181.

Mertens, Karel and Morten O. Ravn (2013). “The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States”. In: *American Economic Review* 103.4, pp. 1212–47. DOI: 10.1257/aer.103.4.1212.

Mikusheva, Anna (Feb. 2021). *Many Weak Instruments in Time Series Econometrics*. World Congress of Econometric Society. MIT.

Miranda-Agrippino, Silvia and Giovanni Ricco (2019). *Identification with External Instruments in Structural VARs under Partial Invertibility*. Discussion Paper 13853. Centre for Economic Policy Research. URL: cepr.org/active/publications/discussion_papers/dp.php?dpno=13853.

— (2021). “The Transmission of Monetary Policy Shocks”. In: *American Economic Journal: Macroeconomics* 13.3, pp. 74–107. DOI: 10.1257/mac.20180124.

Montiel-Olea, José Luis and Carolin Pflueger (2013). “A Robust Test for Weak Instruments”. In: *Journal of Business & Economic Statistics* 31.3, pp. 358–369. DOI: 10.1080/00401706.2013.806694.

Montiel Olea, José Luis and Mikkel Plagborg-Møller (2021). “Local Projection Inference is Simpler and More Robust Than You Think”. In: *Econometrica* 89.4, pp. 1789–1823. URL: 10.3982/ECTA18756.

Muirhead, Robb J. (1982). *Aspects of Multivariate Statistical Theory*. Hoboken, New Jersey: John Wiley and Sons Ltd. DOI: 10.1002/9780470316559.

Nason, James M. and Gregor W. Smith (2008). "Identifying the New Keynesian Phillips Curve". In: *Journal of Applied Econometrics* 23.5, pp. 525–551. DOI: 10.1002/jae.1011.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708. URL: 10.2307/1913610.

Plagborg-Møller, Mikkel and Christian K. Wolf (2021). "Local Projections and VARs Estimate the Same Impulse Responses". In: *Econometrica* 89.2, pp. 955–980. DOI: 10.3982/ECTA17813.

Ramey, Valerie (2016). "Chapter 2 - Macroeconomic Shocks and Their Propagation". In: ed. by John B. Taylor and Harald Uhlig. Vol. 2. Handbook of Macroeconomics. Elsevier, pp. 71–162. DOI: 10.1016/bs.hesmac.2016.03.003.

Romer, Christina D. and David H. Romer (2004). "A New Measure of Monetary Shocks: Derivation and Implications". In: *American Economic Review* 94.4, pp. 1055–1084. DOI: 10.1257/0002828042002651.

Rothenberg, T. J. and C. T. Leenders (1964). "Efficient Estimation of Simultaneous Equation Systems". In: *Econometrica* 32.1/2, pp. 57–76. DOI: 10.2307/1913734.

Smets, Frank and Rafael Wouters (2007). "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach". In: *American Economic Review* 97.3, pp. 586–606. DOI: 10.1257/aer.97.3.586.

Stock, James and Motohiro Yogo (2005). "Testing for Weak Instruments in Linear IV Regression". In: *Identification and Inference for Econometric Models*. Ed. by Donald W.K. Andrews. New York: Cambridge University Press, pp. 80–108. DOI: 10.1017/CBO9780511614491.

Stock, James H. (2008). "Lecture 7: Recent Developments in Structural VAR Modeling". In: Presented at the National Bureau of Economic Research Summer Institute Minicourse: What's New in Econometrics: Time Series, Cambridge, MA, July 15. URL: http://www.nber.org/minicourse_2008.html.

Stock, James H. and Mark W. Watson (2012). "Disentangling the Channels of the 2007-2009 Recession". In: *Brookings Papers on Economic Activity, Spring 2012*, pp. 81–135. DOI: 10.1353/eca.2012.0005.

— (May 2018). "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments". In: *The Economic Journal* 128.610, pp. 917–948. DOI: 10.1111/ecoj.12593.

Stock, James H. and Jonathan H. Wright (2000). "GMM with Weak Identification". In: *Econometrica* 68.5, pp. 1055–1096. DOI: 10.1111/1468-0262.00151.

Wolf, Christian K. (2020). “SVAR (Mis)identification and the Real Effects of Monetary Policy Shocks”. In: *American Economic Journal: Macroeconomics* 12.4, pp. 1–32. DOI: 10.1257/mac.20180328.

Wooldridge, Jeffrey M. (2009). “On estimating firm-level production functions using proxy variables to control for unobservables”. In: *Economics Letters* 104.3, pp. 112–114. ISSN: 0165-1765. DOI: https://doi.org/10.1016/j.econlet.2009.04.026.

A Relative Efficiency of $\hat{\beta}$ with an AR(1) Error Term

Recall that $\tilde{x}_t = x_t - E[x_t]$ and $\bar{x}_t = x_t - E[x_t \mid \mathcal{I}_{t-1}^{full}]$ where $\mathcal{I}_{t-1}^{full}$ is a full information set for $[y_t, Y_t']'$. Define

- $\hat{\tilde{\beta}}$, as in (9) with $y_t^{\perp} = \tilde{y}_t$, $Y_t^{\perp} = \tilde{Y}_t$, $Z^{\perp} = \tilde{z}_t$.
- $\hat{\bar{\beta}}$, as in (9) with $y_t^{\perp} = \bar{y}_t$, $Y_t^{\perp} = \bar{Y}_t$, $Z^{\perp} = \bar{z}_t$.
- $\hat{\tilde{\beta}}_{SE}$, the standard single-equation 2SLS estimator of $\tilde{y}_t$ on $\tilde{Y}_t$ using $\tilde{z}_{t-h},\ h = 0, \ldots, H-1$ as instruments.
- $\hat{\beta}_{SE}^{*}$, the standard single-equation 2SLS estimator of $\tilde{y}_t$ on $\tilde{Y}_t$ using $\bar{z}_{t-h},\ h = 0, \ldots, H-1$ as instruments.

and assume that the exogeneity requirements are met in all cases.

We consider $\hat{\beta}_j$ asymptotically more efficient than $\hat{\beta}_i$ if $aVar(\hat{\beta}_i) - aVar(\hat{\beta}_j)$ is positive semi-definite, as in e.g. Rothenberg and Leenders (1964). The asymptotic variances for the estimators above are given by

$$\begin{aligned}
aVar(\hat{\tilde{\beta}}) &= (\boldsymbol{\Theta}_Y'\boldsymbol{\Theta}_Y)^{-1}\boldsymbol{\Theta}_Y' \left(I_{N_z} \otimes Var(\tilde{u}_{H,t})\right) \boldsymbol{\Theta}_Y (\boldsymbol{\Theta}_Y'\boldsymbol{\Theta}_Y)^{-1} \ , \\
aVar(\hat{\bar{\beta}}) &= (\boldsymbol{\Theta}_Y'\boldsymbol{\Theta}_Y)^{-1}\boldsymbol{\Theta}_Y' \left(I_{N_z} \otimes Var(\bar{u}_{H,t})\right) \boldsymbol{\Theta}_Y (\boldsymbol{\Theta}_Y'\boldsymbol{\Theta}_Y)^{-1} \ , \\
aVar(\hat{\tilde{\beta}}_{SE}) = aVar(\hat{\beta}_{SE}^{*}) &= (\boldsymbol{\Theta}_Y'\boldsymbol{\Theta}_Y)^{-1} Var(\tilde{u}_t) \ .
\end{aligned}$$

Consider the special case in which $\tilde{u}_t$ follows an AR(1)-process:

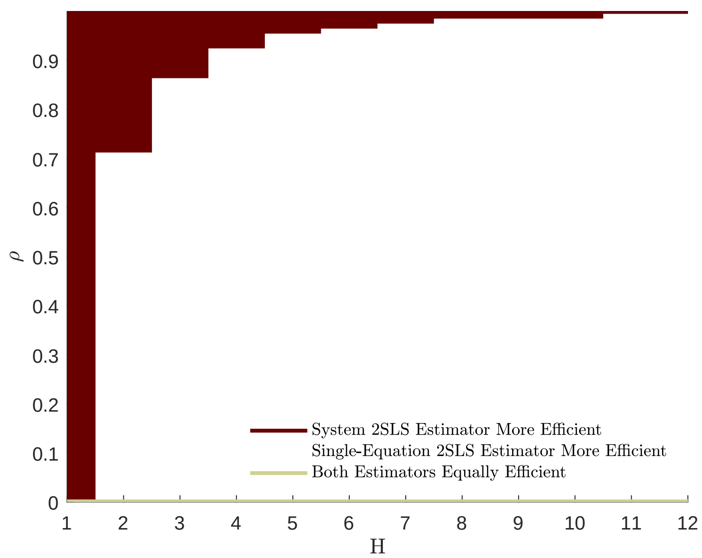
$$\tilde{u}_t = \rho_u \tilde{u}_{t-1} + \upsilon_t \quad ; \quad 0 \leq \rho_u < 1 \quad ; \quad \upsilon_t \sim N(0, \sigma_\upsilon^2) \ . \tag{A.1}$$

Proposition 1. *If $\tilde{u}_t$ follows an AR(1)-process given by (A.1), $\hat{\tilde{\beta}}_{SE}$ is asymptotically more efficient than $\hat{\tilde{\beta}}$ whenever $\rho_u > 0$ and $H > 1$.*

Proof. It suffices to show that $aVar(\hat{\tilde{\beta}}) - aVar(\hat{\tilde{\beta}}_{SE})$ is positive definite. This will be the case as long as $Var(\tilde{u}_t) = \sigma_\upsilon^2/(1-\rho_u^2) < \text{maxeval}\, Var(\tilde{u}_{H,t})$. $Var(\tilde{u}_{H,t})$ is a matrix with $\rho_u^{|h-v|}\sigma_\upsilon^2/(1-\rho_u^2)$ in the h-row and v-th column. When $\rho_u > 0$, by the Perron-Frobenius theorem this matrix has a unique positive dominant eigenvalue that is bounded from below by the minimum row sum. The minimum row sum is $(\sum_{h=0}^{H-1} \rho_u^h)\sigma_\upsilon^2/(1-\rho_u^2)$ which is strictly larger than $Var(\tilde{u}_t)$ when $\rho_u > 0$ and $H > 1$. Therefore, $\text{maxeval}\, Var(\tilde{u}_{H,t}) > Var(\tilde{u}_t)$ when $\rho_u > 0$ and $H > 1$. □

Proposition 2. *If $\tilde{u}_t$ follows an AR(1)-process given by (A.1), $\hat{\bar{\beta}}$ is asymptotically more efficient than $\hat{\beta}_{SE}^{*}$ over a range of combinations of H and ρ_u, where H is sufficiently low and ρ_u is sufficiently large. $\hat{\bar{\beta}}$ is asymptotically as efficient than $\hat{\beta}_{SE}^{*}$ when $\rho_u = 0$.*

FIGURE A.1: Asymptotic Efficiency of System Estimator $\hat{\hat{\beta}}$ vs. Single-Equation Estimator $\hat{\beta}^*_{SE}$



Notes: Results are for an AR(1) error term with persistence ρ_u. H is horizon length.

Proof. $\hat{\hat{\beta}}$ is asymptotically more efficient than $\hat{\beta}^*_{SE}$ if $aVar(\hat{\beta}^*_{SE}) - aVar(\hat{\hat{\beta}})$ is positive definite, i.e. as long as $Var(\tilde{u}_t) = \sigma_v^2/(1-\rho_u^2) \geq \text{maxeval}\, Var(\bar{u}_{H,t})$. $\hat{\hat{\beta}}$ is as efficient as $\hat{\beta}^*_{SE}$ if $Var(\tilde{u}_t) = \text{maxeval}\, Var(\bar{u}_{H,t})$. $Var(\bar{u}_{H,t})$ is a matrix with $\sum_{j=1}^{\min\{h,v\}} \sigma_v^2 \rho_u^{h+v-2j}$ in the h-th row and v-th column. When $\rho_u = 0$, $\text{maxeval}\, Var(\bar{u}_{H,t}) = \sigma_v^2 = Var(\tilde{u}_t)$ for all H. For $\rho_u > 0$, Figure A.1 shows the region in (H, ρ_u)-space in which $\text{maxeval}\, Var(\bar{u}_{H,t}) < \sigma_v^2/(1-\rho_u^2) = Var(\tilde{u}_t)$ (this range does not depend on σ_v^2). □

B Bounding Limiting Distributions of g_{min}

B.1 Upper Bounds for the Cumulants of g_{min}

A Single Endogenous Variable When $K = 1$, $R'(W \otimes I_H)R = \text{Tr}(W)$ is a scalar. The trace of a noncentral Wishart $W \sim \mathcal{W}(N_z, \mathcal{S}, D)$ is a linear combination of noncentral χ^2 variables. While there is no tractable formula for its probability distribution, Mathai (1980) provides an analytical expression for the n-th order cumulant of $\text{Tr}(W)$,

$$\kappa_n = 2^{n-1}(n-1)!\,(N_z \,\text{Tr}(\mathcal{S}^n) + n\,\text{Tr}(\mathcal{S}^n D)) \ . \tag{B.1}$$

The mean is $\kappa_1 = N_z(1+\ell)$, where $\ell = \text{Tr}(\mathcal{S}D)/N_z$ is the concentration parameter. The higher order cumulants are bounded by

$$\kappa_n \leq 2^{n-1}(n-1)!\left(N_z \,\text{Tr}(\mathcal{S}^n) + nN_z\ell\,\text{Tr}(\mathcal{S}^{n-1})\right) \ . \tag{B.2}$$

We show in Appendix B.2 that an approximating distribution based on Imhof (1961) that matches the mean and the upper bounds for the second and third cumulants is conservative in the right tail relative to the distribution with strictly smaller cumulants. As a result, the approximating distribution at the upper bounds with $\ell = \ell_{min}(\xi)$ provides conservative critical values for testing the null hypothesis that the instruments are weak using g_{min}.

Multiple Endogenous Variables When $K > 1$, $g_{min}N_z$ is distributed as the minimum eigenvalue of a matrix with elements that are traces of the $H \times H$ subpartitions of a non-central Wishart matrix $W \sim \mathcal{W}(N_z, \mathcal{S}, D)$. Analogous to Stock and Yogo (2005), we use the distribution of $\gamma' R'(W \otimes I_H)R\gamma \geq \text{mineval}\{R'(W \otimes I_H)R\}$ as a bounding distribution, where γ is the eigenvector associated with the minimum eigenvalue of $R'(\mathcal{S}D \otimes I_H)R$ and $\gamma'\gamma = 1$. We extend the results in Mathai (1980) to obtain an analytical expression for the n-th order cumulant of the distribution of $\gamma' R'(W \otimes I_H)R\gamma$,

$$\kappa_n = 2^{n-1}(n-1)!\left(N_z \operatorname{Tr}\left(((\gamma\gamma' \otimes I_H)\mathcal{S})^n\right) + n \operatorname{Tr}(((\gamma\gamma' \otimes I_H)\mathcal{S})^n D)\right) . \tag{B.3}$$

For the mean, we have the upper bound

$$\kappa_1 = N_z \operatorname{Tr}((\gamma\gamma' \otimes I_H)\mathcal{S}) + N_z\ell \leq N_z(1+\ell) , \tag{B.4}$$

where $\ell = \text{mineval}\{R'(\mathcal{S}D \otimes I_H)R\}/N_z = \operatorname{Tr}\left((\gamma\gamma' \otimes I_H)\mathcal{S}D\right)/N_z$ is the minimum eigenvalue of the concentration matrix. For the higher-order cumulants, we have the bounds

$$\begin{aligned}
\kappa_n &= 2^{n-1}(n-1)!\left(N_z \operatorname{Tr}(((\gamma\gamma' \otimes I_H)\mathcal{S})^n) + n \operatorname{Tr}(((\gamma\gamma' \otimes I_H)\mathcal{S})^n D\right) && \text{(B.5)}\\
&\leq 2^{n-1}(n-1)!\Big(N_z \text{maxeval}\{R'(\mathcal{S}^n \otimes I_H)R\} && \text{(B.6)}\\
&\qquad\qquad + nN_z\ell\,\text{maxeval}\{R'(\mathcal{S}^{n-1} \otimes I_H)R\}\Big) .
\end{aligned}$$

Appendix B.2 shows that the Imhof (1961) approximating distribution with the first three cumulants at the upper bounds is conservative in the right tail relative to the distribution with strictly smaller cumulants. Therefore, the approximating distribution at $\ell = \ell_{min}(\xi)$ yields conservative critical values for testing the null that the instruments are weak using $g_{min}N_z$.

B.2 Conservative Imhof Approximations

In the special cases where $H = 1$ or $\mathcal{S}$ is diagonal, the bounds on the cumulants in B.1 are those of a noncentral χ^2 with noncentrality $N_z\ell$ and N_z degrees of freedom. In the general case, we use the approximation in Imhof (1961) for the cdf of quadratic forms in normal variables,

$$\text{(B.7)} \quad \Pr(\mathcal{F} < x) \approx \Pr(\chi_h^2 < (x-\kappa_1)4\omega + h) = \int_{\kappa_1 - h(4\omega)^{-1}}^{x} \phi(z)dz \quad \text{, where}$$

$$h = 8\kappa_2\omega^2 \quad ; \quad \omega = \kappa_2/\kappa_3 \quad ; \quad \phi(z) = \left(1 + \frac{z-\kappa_1}{2\kappa_2\omega}\right)^{h/2-1} e^{-\frac{h}{2}\left(1+\frac{z-\kappa_1}{2\kappa_2\omega}\right)} \frac{(h/2)^{h/2-1}\omega}{2^{h/2-2}\Gamma(h/2)} \ .$$

This approximation matches the first three central moments of the true distribution of $\mathcal{F}$. The pdf $\phi(z)$ has a mode at $z^m = \kappa_1 - (2\omega)^{-1}$ if $h \geq 2$, and at zero otherwise.

The critical value associated with the upper α-percentile is implicitly defined by $\alpha = \int_{x(\alpha)}^{\infty} \phi(z)dz$. To find the largest possible critical value among all possible distributions, we solve the following optimization problem:

$$\text{(B.8)} \qquad \max_{\kappa_1,\kappa_2,\kappa_3} \ x(\alpha) \ \text{ s.t. } \ \kappa_n \leq \bar{\kappa}_n \text{ for } n = 1,2,3 \ .$$

The Kuhn-Tucker conditions are[1]

$$\text{(B.9)} \qquad \int_{x(\alpha)}^{\infty} \frac{\partial \phi(z)}{\partial \kappa_n} dz = \mu_n,$$

together with $\mu_n \geq 0, \ n = 1,2,3$, the constraints and the complementary slackness conditions, where μ_n are the multipliers times $\phi(x(\alpha)) > 0$. The partial derivatives are

$$\text{(B.10)} \qquad \frac{\partial \phi(z)}{\partial \kappa_1} = \frac{1 + (z-\kappa_1)2\omega}{2\kappa_2\omega} \left(1 + \frac{z-\kappa_1}{2\kappa_2\omega}\right)^{-1} \phi(z) \ ,$$

$$\text{(B.11)} \qquad \frac{\partial \phi(z)}{\partial \kappa_2} = \frac{\phi(z)}{\kappa_2} G_1\left((z-\kappa_1)4\omega + h\right) \ ,$$

$$\text{(B.12)} \qquad \frac{\partial \phi(z)}{\partial \kappa_3} = \frac{\phi(z)}{\kappa_3} G_2\left((z-\kappa_1)4\omega + h\right) \ ,$$

where

$$\text{(B.13)} \qquad G_1(y) = -\frac{1}{2}\left(y - 2h(h-2)/y + h\right) + 3/2(\ln(y/2) - \psi(h/2))h \ ,$$

$$\text{(B.14)} \qquad G_2(y) = \frac{1}{2}\left(y - h(h-2)/y\right) - (\ln(y/2) - \psi(h/2))h \ ,$$

and $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function (the logarithmic derivative of the gamma function $\Gamma(x)$). From Alzer (1997) (equation 2.2), we know that

$$\text{(B.15)} \qquad 1/h < \ln(h/2) - \psi(h/2) < 2/h \ .$$

[1] This follows from the implicit function theorem and Leibniz's rule: $1 = -\phi(x(\alpha))\frac{\partial x(\alpha)}{\partial y} + \int_{x(\alpha)}^{\infty} \frac{\partial \phi(z)}{\partial y} dz \Rightarrow \frac{\partial x(\alpha)}{\partial y} = \int_{x(\alpha)}^{\infty} \frac{\partial \phi(z)}{\partial y} dz / \phi(x(\alpha))$ with $\phi(x(\alpha)) > 0$ for $\alpha \in (0,1)$.

For $n = 1$, the LHS of (B.9) is always positive to the right of the mode, which means the constraint on the mean ($n = 1$) is always binding. The Alzer bounds imply that in the right tail of any optimal distribution, the LHS of (B.9) is always strictly positive for $n = 2, 3$, which means that the constraints for $n = 1, 2$ are also binding as long as α is sufficiently small. In other words, the Imhof approximation matching the upper bounds for the cumulants is a conservative approximation for the right tail of true distribution of $g_{min} N_z$.

C Limiting Distributions of Robust Test Statistics

C.1 The AR Statistic with Multiple Outcome Variables

By the cyclic property of the trace,

$$\operatorname{Tr}\left(u_H^{\perp}(b) P_{Z^{\perp}} u_H^{\perp}(b)' \left(u_H^{\perp}(b) M_{Z^{\perp}} u_H^{\perp}(b)'\right)^{-1}\right) =$$
$$\operatorname{Tr}\left(\left(u_H^{\perp}(b) M_{Z^{\perp}} u_H^{\perp}(b)'\right)^{-1/2} u_H^{\perp}(b) P_{Z^{\perp}} u_H^{\perp}(b)' \left(u_H^{\perp}(b) M_{Z^{\perp}} u_H^{\perp}(b)'\right)^{-1/2}\right) .$$

Note that $u_H^{\perp}(b) Z^{\perp\prime}(Z^{\perp} Z^{\perp\prime})^{-1} = \hat{\psi}(b)$, where ψ is the regression coefficient in the regression of $u_H^{\perp}(b)$ on $Z^{\perp}$. Let $e(b) = u_H^{\perp}(b) - \psi(b) Z^{\perp}$. Using the idempotence of $P_{Z^{\perp}}$ it follows that, under the null hypothesis (from which it follows that $\psi(\beta) = 0$),

$$\begin{aligned} u_H^{\perp}(\beta) P_{Z^{\perp}} u_H^{\perp}(\beta)' &= \left(\psi(\beta) Z^{\perp} + e(\beta)\right) P_{Z^{\perp}} \left(\psi(\beta) Z^{\perp} + e(\beta)\right)' \\ &= e(\beta) Z^{\perp\prime}(Z^{\perp} Z^{\perp\prime})^{-1} Z^{\perp} Z^{\perp\prime}(Z^{\perp} Z^{\perp\prime})^{-1} Z^{\perp} e(\beta)' \\ &= \left(e(\beta) Z^{\perp\prime}(Z^{\perp} Z^{\perp\prime})^{-1/2}\right) \left(e(\beta) Z^{\perp\prime}(Z^{\perp} Z^{\perp\prime})^{-1/2}\right)' . \end{aligned}$$

Observe that under Assumptions 2 and 3,

$$\text{(C.1)}\quad T^{1/2} \operatorname{vec}\left(\left(u_H^{\perp}(\beta) M_{Z^{\perp}} u_H^{\perp}(\beta)'\right)^{-1/2} \left(e(\beta) Z^{\perp\prime}(Z^{\perp} Z^{\perp\prime})^{-1/2}\right)\right) \xrightarrow{d} N(0, I_{HN_z}) .$$

This implies that the diagonal entries in $T u_H^{\perp}(\beta) P_{Z^{\perp}} u_H^{\perp}(\beta)' \left(u_H^{\perp}(\beta) M_{Z^{\perp}} u_H^{\perp}(\beta)'\right)^{-1}$ converge in distribution to sums of N_z squared independent standard normal random variables, so each diagonal element converges in distribution to a $\chi^2_{N_z}$ random variable. Taking the trace of this matrix takes the sum of those H diagonal elements, which converges in distribution to the sum of H independent $\chi^2_{N_z}$ random variables, which is itself a $\chi^2_{HN_z}$ random variable. We include a degrees of freedom adjustment in the expression in the text.

C.2 Derivation and Limiting Distribution of the KLM Statistic $K(\beta)$

To derive our test statistic, $K(b)$, we differentiate $AR(b) = T \operatorname{Tr}\left(u_H^{\perp}(b) P_{Z^{\perp}} u_H^{\perp\prime}(b) \Xi^{-1}\right)$, with respect to b (where Ξ denotes $u_H^{\perp}(b) M_{Z^{\perp}} u_H^{\perp\prime}(b)$), as in Kleibergen (2002). The (re-scaled) score is

$$-\frac{1}{2T}\frac{\partial AR(b)}{\partial b} = R' \operatorname{vec}\left(\Xi^{-1}u_H^{\perp}P_{Z^{\perp}}Y_H^{\perp\prime} - \Xi^{-1}u_H^{\perp}\left(P_{Z^{\perp}}u_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}\check{u}_H^{\perp}\check{v}_H^{\perp\prime}\right)\right) \tag{C.2}$$
$$= R' \operatorname{vec}(\Xi^{-1}u_H^{\perp}\check{Y}_H') \ .$$

where the dependence of $u_H^{\perp}$ on b is suppressed for brevity. Henceforth, we maintain the null hypothesis and let $u_H^{\perp} = u_H^{\perp}(\beta)$ throughout. Consider the transformation $\eta = v_H^{\perp} - \Sigma'_{u_H^{\perp}v_H^{\perp}}\Sigma^{-1}_{u_H^{\perp}}u_H^{\perp} = Y_H^{\perp} - \Theta_Y Q^{-1/2}Z^{\perp} - \Sigma'_{u_H^{\perp}v_H^{\perp}}\Sigma^{-1}_{u_H^{\perp}}u_H^{\perp}$. By construction, η is orthogonal to $u_H^{\perp}$. In particular, $\operatorname{var}(\eta) \equiv \Sigma_{\eta} = \Sigma_{v_H^{\perp}} - \Sigma'_{u_H^{\perp}v_H^{\perp}}\Sigma^{-1}_{u_H^{\perp}}\Sigma_{u_H^{\perp}v_H^{\perp}}$ such that $\operatorname{var}\left(\begin{bmatrix}u_H^{\perp\prime}, & \eta'\end{bmatrix}'\right) \equiv \check{\Sigma} = \begin{bmatrix}\Sigma_{u_H^{\perp}} & 0 \\ 0 & \Sigma_{\eta}\end{bmatrix}$, and under Assumption 4.b,

$$T^{-1/2}\left(\operatorname{vec}(Z^{\perp}u_H^{\perp\prime}), \operatorname{vec}(Z^{\perp}\eta')\right) \xrightarrow{d} (\operatorname{vec}(\Psi_{Zu}), \operatorname{vec}(\Psi_{Z\eta})) \sim \mathcal{N}(0, Q \otimes \check{\Sigma}). \tag{C.3}$$

The two summations, $\operatorname{vec}(Z^{\perp}u_H^{\perp\prime})$ and $\operatorname{vec}(Z^{\perp}\eta')$, are therefore asymptotically independent. By Assumption 4.a, under the null hypothesis (so $u_H^{\perp}$ and $v_H^{\perp}$ are orthogonal to $Z^{\perp}$), $\Xi/T = \check{u}_H^{\perp}\check{u}_H^{\perp\prime}/T \xrightarrow{p} \Sigma_{u_H^{\perp}}$ and $u_H^{\perp}M_{Z^{\perp}}v_H^{\perp\prime}/T = \check{u}_H^{\perp}\check{v}_H^{\perp\prime}/T \xrightarrow{p} \Sigma_{u_H^{\perp}v_H^{\perp}}$. Given the $\sqrt{T}$-consistency of these estimators,

$$T^{-1/2}\operatorname{vec}\left(Z^{\perp}\left[Y_H^{\perp} - \Theta_Y Q^{-1/2}Z^{\perp} - \check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}u_H^{\perp}\right]'\right) \tag{C.4}$$
$$= T^{-1/2}\operatorname{vec}\left(Z^{\perp}\left[Y_H^{\perp} - \Theta_Y Q^{-1/2}Z^{\perp} - \Sigma'_{u_H^{\perp}v_H^{\perp}}\Sigma^{-1}_{u_H^{\perp}}u_H^{\perp}\right]'\right)$$
$$- T^{-1/2}\operatorname{vec}\left(Z^{\perp}\left[\left(\check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1} - \Sigma'_{u_H^{\perp}v_H^{\perp}}\Sigma^{-1}_{u_H^{\perp}}\right)u_H^{\perp}\right]'\right)$$
$$\xrightarrow{d} \operatorname{vec}(\Psi_{Z\eta}) - 0 = \operatorname{vec}(\Psi_{Z\eta}),$$

where the second term converges in probability to zero. Therefore, $T^{-1/2}\operatorname{vec}(Z^{\perp}\eta')$ and $T^{-1/2}\operatorname{vec}(Z^{\perp}\left[Y_H^{\perp} - \Theta_Y Q^{-1/2}Z^{\perp} - \check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}u_H^{\perp}\right]')$ have the same limiting behavior.

We use (C.3)-(C.4) to construct the limiting distribution of $K(\beta)$ under the cases of strong, weak, and non-identification. We proceed under Assumptions 2.a-2.c and Assumption 4, and assume that all coefficients in Θ_Y are of the same order.

First, we consider the case of strong identification, implying that Θ_Y is a fixed nonzero matrix. In that case, $T^{-1/2}\operatorname{vec}(Z^{\perp}\left[Y_H^{\perp} - \check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}u_H^{\perp}\right]' - Z^{\perp}Z^{\perp\prime}Q^{-1/2}\Theta_Y') \xrightarrow{d} \operatorname{vec}(\Psi_{Z\eta})$, such that

$$
\begin{aligned}
T^{-1/2}u_H^{\perp}\check{Y}_H' &= T^{-1/2}u_H^{\perp}P_{Z^{\perp}}\left[Y_H^{\perp}-\check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}u_H^{\perp}\right]' \\
&= T^{-1/2}u_H^{\perp}P_{Z^{\perp}}\left[Y_H^{\perp}-\Theta_Y Q^{-1/2}Z^{\perp}-\check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}u_H^{\perp}\right]' + T^{-1/2}u_H^{\perp}P_{Z^{\perp}}Z^{\perp\prime}Q^{-1/2}\Theta_Y' \\
&\xrightarrow{d} \Psi_{Zu}'Q^{-1/2}\Theta_Y'.
\end{aligned} \tag{C.5}
$$

The first term vanishes since it is equal to

$$
T^{-1/2}\left(T^{-1/2}u_H^{\perp}Z^{\perp\prime}\right)\left(Z^{\perp}Z^{\perp\prime}/T\right)^{-1}\left(T^{-1/2}Z^{\perp}\left[Y_H^{\perp}-\Theta_Y Q^{-1/2}Z^{\perp}-\check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}u_H^{\perp}\right]'\right) \xrightarrow{p} 0\,.
$$

Therefore, $T^{1/2}\operatorname{vec}(\Xi^{-1}u_H^{\perp}\check{Y}_H') \xrightarrow{d} \operatorname{vec}(\Sigma_{u_H^{\perp}}^{-1}\Psi_{Zu}'Q^{-1/2}\Theta_Y') \sim \mathcal{N}(0,\Theta_Y\Theta_Y'\otimes\Sigma_{u_H^{\perp}}^{-1})$, such that

$$
T^{1/2}R'\operatorname{vec}(\Xi^{-1}u_H^{\perp}\check{Y}_H')\left(R'(\check{Y}_H\check{Y}_H'\otimes\Xi^{-1}u_H^{\perp}u_H^{\perp\prime}\Xi^{-1})R\right)^{-1/2} \xrightarrow{d} \mathcal{N}(0,I_K)\ , \tag{C.6}
$$

since $R'\left(\check{Y}_H\check{Y}_H'\otimes\Xi^{-1}u_H^{\perp}u_H^{\perp\prime}\Xi^{-1}\right)R \xrightarrow{p} R'\left(\Theta_Y\Theta_Y'\otimes\Sigma_{u_H^{\perp}}^{-1}\right)R$, which under the additional Assumption 2.d$'$, is invertible. Finally, it follows that $K(\beta)\xrightarrow{d}\chi_K^2$.

Next, we consider the case of weak identification, $\Theta_Y = C/\sqrt{T}$, where C is a fixed full rank $KH\times N_z$ matrix. In this case, $T^{-1/2}Z^{\perp}\left[Y_H^{\perp}-\check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}u_H^{\perp}\right]' \xrightarrow{d} \Psi_{Z\eta}+Q^{1/2}C'$, which implies that

$$
u_H^{\perp}\check{Y}_H' = u_H^{\perp}P_{Z^{\perp}}\left[Y_H^{\perp}-\check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}u_H^{\perp}\right]' \xrightarrow{d} \Psi_{Zu}'Q^{-1}(\Psi_{Z\eta}+Q^{1/2}C')\ , \tag{C.7}
$$

$$
\begin{aligned}
\check{Y}_H\check{Y}_H' &= \left[Y_H^{\perp}-\check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}u_H^{\perp}\right]P_{Z^{\perp}}\left[Y_H^{\perp}-\check{v}_H^{\perp}\check{u}_H^{\perp\prime}(\check{u}_H^{\perp}\check{u}_H^{\perp\prime})^{-1}u_H^{\perp}\right]' \\
&\xrightarrow{d} (\Psi_{Z\eta}+Q^{1/2}C')'Q^{-1}(\Psi_{Z\eta}+Q^{1/2}C').
\end{aligned} \tag{C.8}
$$

Consequently,

$$
T\operatorname{vec}(\Xi^{-1}u_H^{\perp}\check{Y}_H) \xrightarrow{d} \operatorname{vec}(\Sigma_{u_H^{\perp}}^{-1}\Psi_{Zu}'Q^{-1}(\Psi_{Z\eta}+Q^{1/2}C'))\ . \tag{C.9}
$$

As a result of the independence of Ψ_{Zu} and $\Psi_{Z\eta}$, the conditional distribution of $\operatorname{vec}(\Sigma_{u_H^{\perp}}^{-1}\Psi_{Zu}'Q^{-1}(\Psi_{Z\eta}+C'))$, given $\Psi_{Z\eta}$, is

$$
\mathcal{N}(0,(\Psi_{Z\eta}+Q^{1/2}C')'Q^{-1}(\Psi_{Z\eta}+Q^{1/2}C')\otimes\Sigma_{u_H^{\perp}}^{-1})\ . \tag{C.10}
$$

Therefore, the conditional distribution given $\Psi_{Z\eta}$ of

$$
\operatorname{vec}\left(\Sigma_{u_H^{\perp}}^{-1}\Psi_{Zu}'Q^{-1}(\Psi_{Z\eta}+C')\right)\left(\Psi_{Z\eta}+Q^{1/2}C')'Q^{-1}(\Psi_{Z\eta}+Q^{1/2}C')\otimes\Sigma_{u_H^{\perp}}^{-1}\right)^{-1/2} \tag{C.11}
$$

is $\mathcal{N}(0, I_K)$ and does not depend on $\Psi_{Z\eta}$, which implies that this random variable is also unconditionally distributed $\mathcal{N}(0, I_K)$. It follows that,

$$T^{1/2} R' \operatorname{vec}(\Xi^{-1} u_H^{\perp} \check{Y}_H) \left(R'(\check{Y}_H \check{Y}_H' \otimes \Xi^{-1} u_H^{\perp} u_H^{\perp\prime} \Xi^{-1})R\right)^{-1/2} \xrightarrow{d} \mathcal{N}(0, I_K) \ , \tag{C.12}$$

since $\check{Y}_H \check{Y}_H' \xrightarrow{d} (\Psi_{Z\eta} + Q^{1/2}C')' Q^{-1} (\Psi_{Z\eta} + Q^{1/2}C')$ and $T\ \Xi^{-1} u_H^{\perp} u_H^{\perp\prime} \Xi^{-1} \xrightarrow{p} \Sigma_{u_H^{\perp}}$ (note that the limit of the variance in the denominator is $O(1/T)$ times that in (C.10) since the numerator is $T^{-1/2}$ times that in (C.9)). This implies that $K(\beta) \xrightarrow{d} \chi^2_K$.

Finally, we consider the case where the model is completely unidentified, and $\Theta_Y = 0$. As a result, $T^{-1/2} Z^{\perp} \left[Y_H^{\perp} - \check{v}_H^{\perp} \check{u}_H^{\perp\prime} (\check{u}_H^{\perp} \check{u}_H^{\perp\prime})^{-1} u_H^{\perp}\right]' \xrightarrow{d} \Psi_{Z\eta}$ such that

(C.13)

$$u_H^{\perp} \check{Y}_H' = u_H^{\perp} P_{Z^{\perp}} \left[Y_H^{\perp} - \check{v}_H^{\perp} \check{u}_H^{\perp\prime} (\check{u}_H^{\perp} \check{u}_H^{\perp\prime})^{-1} u_H^{\perp}\right]' \xrightarrow{d} \Psi_{Zu}' Q^{-1} \Psi_{Z\eta}$$

(C.14)

$$\check{Y}_H \check{Y}_H' = \left[Y_H^{\perp} - \check{v}_H^{\perp} \check{u}_H^{\perp\prime} (\check{u}_H^{\perp} \check{u}_H^{\perp\prime})^{-1} u_H^{\perp}\right] P_{Z^{\perp}} \left[Y_H^{\perp} - \check{v}_H^{\perp} \check{u}_H^{\perp\prime} (\check{u}_H^{\perp} \check{u}_H^{\perp\prime})^{-1} u_H^{\perp}\right]' \xrightarrow{d} \Psi_{Z\eta}' Q^{-1} \Psi_{Z\eta},$$

It follows that $T \operatorname{vec}(\Xi^{-1} u_H^{\perp} \check{Y}_H) \xrightarrow{d} \operatorname{vec}(\Sigma_{u_H^{\perp}}^{-1} \Psi_{Zu}' Q^{-1} \Psi_{Z\eta})$. Using the same reasoning as under the weak identification case, the distribution of $\operatorname{vec}(\Sigma_{u_H^{\perp}}^{-1} \Psi_{Zu}' Q^{-1} \Psi_{Z\eta})$, conditional on $\Psi_{Z\eta}$, is $\mathcal{N}(0, \Psi_{Z\eta}' Q^{-1} \Psi_{Z\eta} \otimes \Sigma_{u_H^{\perp}}^{-1})$. This implies that, unconditionally,

$$T^{1/2} R' \operatorname{vec}(\Xi^{-1} u_H^{\perp} \check{Y}_H) \left(R'(\check{Y}_H \check{Y}_H' \otimes \Xi^{-1} u_H^{\perp} u_H^{\perp\prime} \Xi^{-1})R\right)^{-1/2} \xrightarrow{d} \mathcal{N}(0, I_K), \tag{C.15}$$

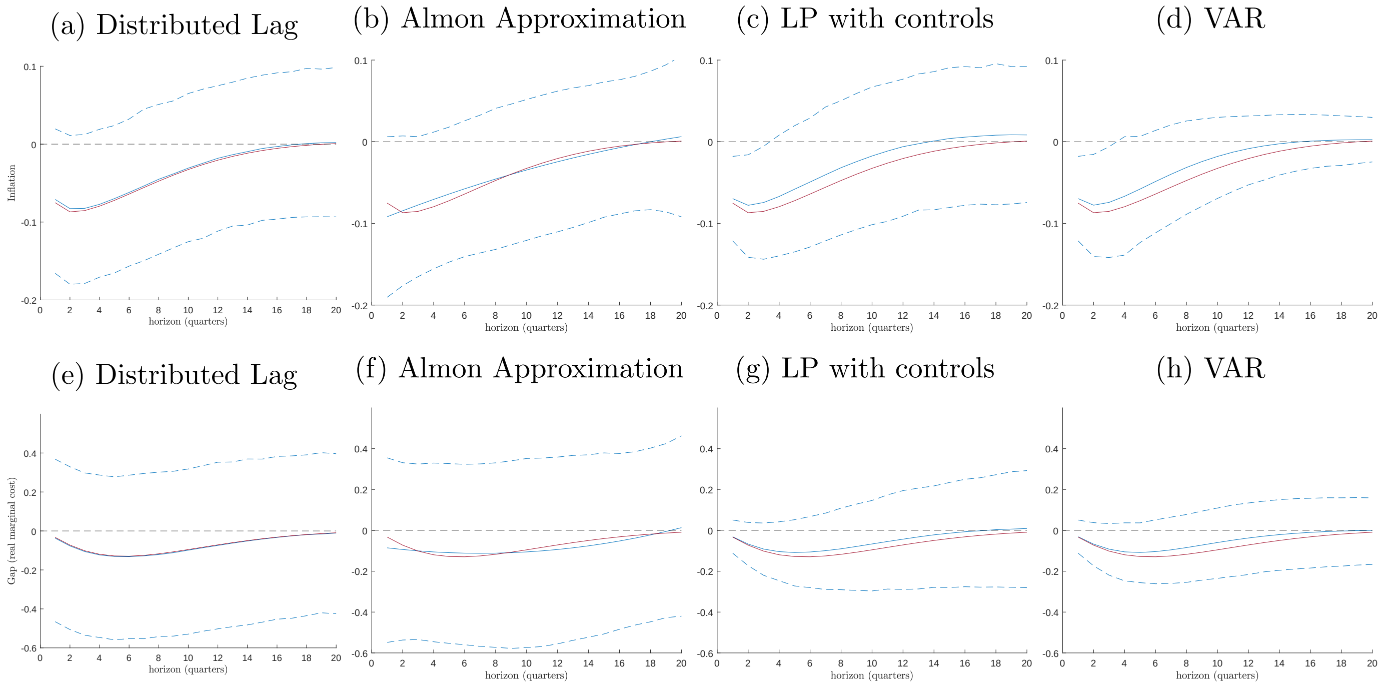
since $\check{Y}_H \check{Y}_H' \xrightarrow{d} \Psi_{Z\eta}' Q^{-1} \Psi_{Z\eta}$ and $T\ \Xi^{-1} u_H^{\perp} u_H^{\perp\prime} \Xi^{-1} \xrightarrow{p} \Sigma_{u_H^{\perp}}$. As a result, $K(\beta) \xrightarrow{d} \chi^2_K$.

D Model Simulations: Additional Results

D.1 IRF Estimates in the Simulations

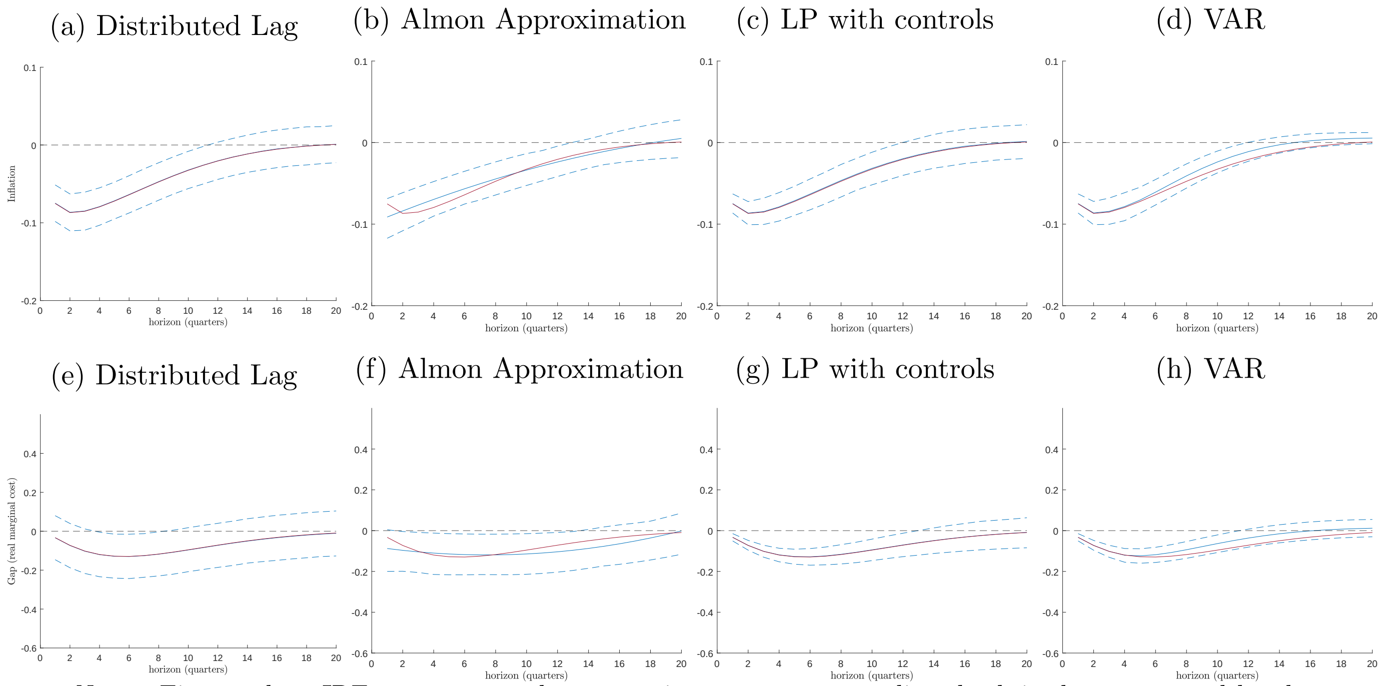
Figures D.1 and D.2 show the mean IRFs, together with 2.5% and 97.5% percentiles, across simulations from the Smets and Wouters (2007) model. The figures show IRFs estimated using a distributed lag specification, the quadratic (Almon) approximation as in Barnichon and Mesters (2020), local projections with the predetermined control variables described in the main text, and a VAR in the variables of the control set described in the main text. For brevity, we only show the IRFs associated with the monetary policy shock for $H = 20$ and $T = 250, 5000$. Results for the other specifications are available on request.

Figure D.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 D.2: True and Estimated IRFs in Simulations, Large Sample (T=5000)



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.

TABLE D.1: MEAN PARAMETER ESTIMATES, $N_z = 3$

$H = 8$ Estimator	$T = 250$ γ_b	γ_f	λ	$T = 500$ γ_b	γ_f	λ	$T = 5000$ γ_b	γ_f	λ
β	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
2SLS	0.40	0.55	0.01	0.36	0.63	0.01	0.23	0.82	0.02
2SLS-Almon	0.40	0.67	0.00	0.34	0.73	0.00	0.22	0.84	0.01
SP-IV LP	0.39	0.56	0.01	0.36	0.63	0.01	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
FGSP-IV LP	0.43	0.58	0.00	0.39	0.64	0.00	0.24	0.81	0.01
FGSP-IV LP-C	0.47	0.57	0.01	0.42	0.64	0.01	0.22	0.81	0.03
FGSP-IV VAR	0.40	0.67	0.00	0.34	0.74	0.01	0.22	0.81	0.03
$H = 20$ Estimator	$T = 250$ γ_b	γ_f	λ	$T = 500$ γ_b	γ_f	λ	$T = 5000$ γ_b	γ_f	λ
β	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
2SLS	0.45	0.51	0.00	0.43	0.55	0.00	0.28	0.76	0.01
2SLS-Almon	0.44	0.61	0.00	0.39	0.67	0.00	0.27	0.80	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	0.55	0.01	0.27	0.76	0.02
SP-IV VAR	0.35	0.69	0.01	0.31	0.75	0.01	0.23	0.82	0.02
FGSP-IV LP	0.49	0.54	0.00	0.47	0.57	0.00	0.32	0.74	0.00
FGSP-IV LP-C	0.51	0.53	0.00	0.49	0.56	0.00	0.31	0.74	0.01
FGSP-IV VAR	0.42	0.64	0.00	0.36	0.71	0.00	0.27	0.79	0.01

Notes: The top row in each panel contains the true parameter values $\beta = [\gamma_b, \gamma_f, \lambda]'$ of (2) in the Smets and Wouters (2007) model. The other rows show mean estimates across 5000 Monte Carlo samples of size T and with $h = 0, \ldots, H-1$. All IV estimators use the monetary policy shock, government spending shock and the risk premium shocks as instruments. 2SLS-Almon is the estimator proposed in Barnichon and Mesters (2020). SP-IV is the estimator in (9) while FGSP-IV is the feasible generalized estimator in (25). LP and LP-C denote implementations based on local projections discussed in Section 4.1, without and with controls, respectively. VAR denotes the implementation with a vector autoregression discussed in Section 4.2.

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

In this section, we present simulation results for specifications using three instruments. A brief summary of these results (and comparison to the $N_z = 1$ results) is provided in the main text. In general, the results are quite similar to those for a single instrument.

Table D.2: Standard deviation of parameter estimates, $N_z = 3$

$H = 8$	$T = 250$			$T = 500$			$T = 5000$		
Estimator	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
2SLS	0.09	0.09	0.03	0.08	0.09	0.03	0.06	0.05	0.02
2SLS-Almon	0.13	0.13	0.03	0.11	0.11	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
FGSP-IV LP	0.11	0.12	0.04	0.09	0.11	0.03	0.06	0.05	0.02
FGSP-IV LP-C	0.10	0.10	0.05	0.09	0.08	0.05	0.06	0.04	0.03
FGSP-IV VAR	0.11	0.12	0.05	0.10	0.11	0.05	0.06	0.05	0.03
$H = 20$	$T = 250$			$T = 500$			$T = 5000$		
Estimator	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
2SLS	0.04	0.04	0.01	0.04	0.04	0.01	0.04	0.04	0.01
2SLS-Almon	0.12	0.11	0.01	0.10	0.10	0.01	0.05	0.05	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	0.10	0.02	0.09	0.10	0.02	0.05	0.04	0.02
FGSP-IV LP	0.03	0.03	0.01	0.03	0.03	0.01	0.04	0.04	0.01
FGSP-IV LP-C	0.03	0.03	0.01	0.03	0.03	0.01	0.04	0.03	0.01
FGSP-IV VAR	0.07	0.09	0.02	0.08	0.09	0.02	0.05	0.04	0.02

Notes: Rows show standard deviations across 5000 Monte Carlo samples of size T and with $h = 0, \ldots, H-1$. All IV estimators use the monetary policy shock, government spending shock and the risk premium shocks as instruments. 2SLS-Almon is the estimator proposed in Barnichon and Mesters (2020). SP-IV is the estimator in (9) while FGSP-IV is the feasible generalized estimator in (25). LP and LP-C denote implementations based on local projections discussed in Section 4.1, without and with controls, respectively. VAR denotes the implementation with a vector autoregression discussed in Section 4.2.

Table D.3: 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	80.70	77.50	56.00	99.80	99.70	91.60
AR 2SLS	50.90	27.90	5.70	99.80	90.60	10.60
AR 2SLS-Almon	3.50	1.30	0.00	8.00	6.30	1.30
WALD SP-IV LP	82.10	78.70	56.50	99.90	99.90	91.60
WALD SP-IV LP-C	73.50	60.40	17.50	100.00	99.70	76.60
WALD SP-IV VAR	39.20	28.30	13.10	86.10	76.40	54.20
AR SP-IV LP	7.10	6.60	5.10	14.40	8.50	4.90
AR SP-IV LP-C	7.30	6.20	5.10	17.40	9.80	5.20
AR SP-IV VAR	3.90	4.90	4.90	6.10	5.80	4.70
KLM SP-IV LP	6.40	5.30	4.80	7.90	6.60	5.30
KLM SP-IV LP-C	7.70	6.10	4.90	11.40	7.80	5.10
KLM SP-IV VAR	7.00	7.00	4.90	10.70	8.50	5.40

Notes: Empirical rejection rates of various 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 using the monetary policy shock, government spending and risk premium shocks as the instruments. The 2SLS Wald test uses a HAR variance matrix following Lazarus et al. (2018). AR 2SLS and AR 2SLS-Almon are the Anderson and Rubin (1949) tests in Barnichon and Mesters (2020). WALD SP-IV is based on (23) with $\hat{\Sigma}_{u_H^\perp}$ as in (41), AR is based on (36), and KLM is based on (37). LP and LP-C denote implementations based on local projections discussed in Section 4.1, without and with controls, respectively. VAR denotes the implementation with a vector autoregression discussed in Section 4.2.