Federal Reserve Bank of Dallas

Dynamic Identification Using System Projections and Instrumental Variables

Daniel J. Lewis and Karel Mertens

Working Paper 2204 **March 2022 (Revised January 2023)**

Research Department

https://doi.org/10.24149/wp2204r1

Dynamic Identification Using System Projections and Instrumental Variables*

Daniel J. Lewis† and Karel Mertens‡

March 22, 2022
Revised: January 4, 2023

Abstract

We propose System Projections on Instrumental Variables (SP-IV) to estimate dynamic structural relationships using impulse responses obtained from local projections or vector autoregressions. SP-IV replaces lag sequences of instruments in traditional IV with lead sequences of endogenous variables. By allowing the inclusion of lagged variables as controls, SP-IV weakens exogeneity requirements and can improve efficiency and effective instrument strength relative to 2SLS. We provide inference procedures under strong and weak identification, and show that SP-IV outperforms conventional IV estimators of Phillips Curve parameters in simulations. We estimate the Phillips Curve implied by the main business cycle shock of Angeletos et al. (2020), and find that the impulse response estimates are 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 Bank of Dallas or the Federal Reserve System. We are grateful to Kirill Borusyak, Xu Cheng, Andrew Chesher, Richard Crump, Marco Del Negro, Dennis Kristensen, Geert Mesters, José Luis Montiel Olea, Mikkel Plagborg-Møller, Andrei Zeleneev, and seminar participants at UPenn, Penn State, University of Houston, UCL, Johns Hopkins, Boston University, Padova, Bologna, NBER/NSF Time Series, and the Federal Reserve Banks of Kansas City and New York.

†Daniel J. Lewis, University College London.

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

This paper studies the estimation of β in structural time series equations of the form

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

where y_t is a 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 β contains the K structural parameters of interest. The explanatory variables Y_t may contain contemporaneous variables, but also lagged variables or agents' expectations of future variables that may not be measured well by the econometrician. We are interested in applications where $E[Y_t u_t] \neq 0$, such that standard regression techniques yield inconsistent estimates of β due to endogeneity.

Equation (1) nests a wide range of dynamic relationships of interest in macroeconomics. Consider the example of the Hybrid New Keynesian Phillips Curve (henceforth, the "Phillips Curve"),

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

where π_t denotes inflation, π^e_{t+1} is a measure of price setters' period t expectation of inflation in $t+1$, and gap_t is an output gap measure (the deviation of actual economic activity from the level without price rigidities). Equation (2) maps into the more general problem in (1) with $y_t = \pi_t$, $Y_t = [\pi_{t-1}, \pi^e_{t+1}, gap_t]'$ and $\beta = [\gamma_b, \gamma_f, \lambda]'$. The estimation of $\beta = [\gamma_b, \gamma_f, \lambda]'$ 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 general source of endogeneity is *measurement error*, as in practice the output gap and inflation expectations must be replaced with proxy measures. A second source of endogeneity problems is *simultaneity*, since the error term typically includes structural shocks that also influence the endogenous variables in Y_t. Many theoretical dynamic relationships include expectations and other endogenous explanatory variables, and therefore face similar problems.

A common approach in the literature is to rely on dynamics for identification and use lagged variables as instrumental variables. In the Phillips

Curve 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 that the error term u_t is uncorrelated with any of the determinants of the instrumenting lagged macroeconomic variables. In other words, the shocks (and lags thereof) comprising the error term u_t must be uncorrelated with the shocks generating lags of gap_t and π_t. There is no general 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) propose current and lagged values of direct measures of structural shocks believed to be uncorrelated with u_t as instrumental variables. In practice, however, the literature is rarely comfortable with imposing the strong assumption of unconditional lag exogeneity on available empirical measures of structural shocks, such as empirical measures of monetary policy surprises, and typically avoids doing so in other contexts by including a rich set of lagged macroeconomic controls. Unfortunately, including such controls in both stages of 2SLS with lagged economic shocks as instruments shrinks their explanatory power towards that of only the contemporaneous value of the instrumenting shocks, resulting in weaker or even under-identification.

In this paper, we propose a novel approach to estimating β that allows the inclusion of lagged variables as controls without weakening identification. Specifically, we replace the single equation (1) with an H-dimensional system of structural equations in forecast errors 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, including vector autoregressive models (VARs) and local projections (LPs). The contemporaneous values of the N_z instrumental variables generate HN_z moment conditions, which we solve in closed form for β, yielding a restricted IV estimator in the system of reduced form forecast errors. We refer to this estimator as System Projections on Instrumental Variables, or SP-IV.

[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.

We show that SP-IV is equivalent to a regression of the impulse response function (IRF) of y_t on the IRFs of Y_t, where the IRFs can be obtained from a VAR, LPs or any other valid impulse response estimator. 2SLS with lag sequences of shocks as instruments is also equivalent to a regression with IRFs, but the implicit IRFs are estimated by regressions of the endogenous variables on distributed lags (DL) of the shocks. An advantage of SP-IV is that it estimates structural relationships across IRFs as they are estimated in practice, which is rarely with DL specifications but instead with VARs or LPs. In addition, SP-IV works with any LP or VAR identification scheme and does not necessarily require a direct measure of an economic shock.

The SP-IV approach has several other advantages over the conventional 2SLS approach with DL instruments. First, like Barnichon and Mesters (2020), it can leverage existing shock measures, but with suitable controls it requires only the weaker assumptions of contemporaneous and lead exogeneity, compared to contemporaneous, lead, and lag exogeneity for 2SLS. Second, the use of forecast errors instead of raw variables can improve efficiency in estimating β. Third, similar efficiency gains in the first stage increase effective instrument strength and mitigate weak instrument problems.

We develop inference methods for SP-IV that enable the formal testing of hypotheses about structural relationships across IRFs in macroeconomic applications. We describe 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. As instrumental variables are often weak in practice, we describe weak instrument robust inference procedures based on the Anderson and Rubin (1949) statistic and Kleibergen's (2005) KLM statistic.

We demonstrate the better performance of SP-IV in simulations estimating the Phillips curve parameters with weak instruments using data generated from the Smets and Wouters (2007) model. When the instrument is lag endogenous, 2SLS is prohibitively biased, but SP-IV with suitable controls is not. When the instruments are valid for both estimators, SP-IV with controls still exhibits considerably smaller bias than 2SLS. A

VAR implementation of SP-IV has the lowest bias of all estimators we consider, while LP implementations have lower variances. The robust inference procedures for SP-IV remain well-sized in realistic sample sizes and exhibit smaller size distortions when HN_z is large than traditional 2SLS with DL instruments.

As an empirical application, we estimate the Phillips curve in US data using the Main Business Cycle (MBC) shock of Angeletos et al. (2020) as an instrument. Identified as the shock that maximally explains the cyclical variation in unemployment, Angeletos et al. (2020) conclude from its muted impact on inflation that the Phillips curve must be very flat. SP-IV enables a formal econometric assessment of claims about structural relationships between IRFs that accounts for the sampling error in the IRF estimates. We find that robust confidence sets for the slope of the Phillips curve are consistent with weak but also fairly strong cyclical connections between inflation and economic activity. After properly accounting for estimation uncertainty, the evidence from IRFs to an MBC shock does not necessarily provide strong support for inflation dynamics that are disconnected from the business cycle.

Henceforth, I_N denotes the N-dimensional identity matrix, $\otimes$ 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, $E[Y \mid X]$ the conditional expectation of Y given X, $\xrightarrow{p}$ convergence in probability, $\xrightarrow{d}$ convergence in distribution, and $P_U = U'(UU')^{-1}U$ the projection matrix.

1 System Projections on Instrumental Variables

We first reformulate the dynamic relationship in (1) in terms of forecast errors. Taking h-horizon leads and subtracting the expectation conditional on an $N_x \times 1$ vector of predictors X_{t-1} (including a constant) yields

$$y_t^{\perp}(h) = \beta' Y_t^{\perp}(h) + u_t^{\perp}(h) , \tag{3}$$

where $y_t^{\perp}(h) = y_{t+h} - E[y_{t+h} \mid X_{t-1}]$, $Y_t^{\perp}(h) = Y_{t+h} - E[Y_{t+h} \mid X_{t-1}]$, and $u_t^{\perp}(h) = u_{t+h} - E[u_{t+h} \mid X_{t-1}]$. Let z_t denote an $N_z \times 1$ vector of instrumental variables, and define $z_t^{\perp} = z_t - E[z_t \mid X_{t-1}]$. As explained in the introduction, we focus on applications that rely on dynamics for identification, exploiting orthogonality conditions between the error term u_t and $z_t, z_{t-1}, \ldots$ Instead of the usual approach of imposing orthogonality between z_{t-h} and u_t for various $h \geq 0$, we impose

$$E[u_t^{\perp}(h) z_t^{\perp}] = 0 \; ; \; h = 0, \ldots, H-1 \; . \tag{4}$$

Without conditioning on X_{t-1} and under stationarity, the orthogonality conditions in (4) are equivalent to imposing orthogonality between z_{t-h} and u_t. The key departure from 2SLS with a distributed lag of z_t as instruments is that the moments in (4) are not in terms of the unconditional data but in terms of forecast errors conditional on the predictors X_{t-1}.

1.1 The Generalized Method of Moments Problem

The conditions in (4) provide 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 $y_t^{\perp}(h)$ or $u_t^{\perp}(h)$ respectively. Let $Y_{H,t}^{\perp}$ denote the $HK \times 1$ vector stacking the $H \times 1$ vectors $Y_{H,t}^{k,\perp}$, where Y_t^k is the k-th variable in Y_t. Using this notation, the moment conditions are

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

where $u_{H,t}^{\perp}(b) \equiv y_{H,t}^{\perp} - (b' \otimes I_H) Y_{H,t}^{\perp}$ and the truth is $b = \beta$.

The moment conditions in (5) can be augmented to account for the estimation of the forecast errors. 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 includes local projections and vector autoregressions, both of which are widely used in applied macroeconomics.[2] The moment conditions for this 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

[2]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).

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 $f(y_{H,t}, Y_{H,t}, z_t, X_{t-1}; b, d)$ with $E[f(y_{H,t}, Y_{H,t}, z_t, X_{t-1}; \beta, \zeta)] = 0$. Let $W_t = [y_{H,t}^{\perp\prime},\ Y_{H,t}^{\perp\prime},\ z_t^{\perp\prime},\ X_{t-1}^{\prime}]^{\prime}$. The associated GMM objective function is

$$F_T(b,d) = \frac{1}{T}\left(\sum_{t=1}^{T} f(W_t; b, d)\right)^{\prime} \Phi(b,d) \left(\sum_{t=1}^{T} f(W_t; b, d)\right) , \tag{7}$$

where $\Phi(b,d)$ is a positive definite weighting matrix. The forecasting step and the structural estimation step are separable for estimation purposes since b does not enter (6) and the Jacobian of (5) with respect to d is zero in expectation at ζ. This means we can henceforth take the forecasts as given and focus on the structural estimation step. An additional assumption ensures that estimation error in the forecast errors is asymptotically negligible for inference on the structural parameters:

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.*

In what follows, we suppress the dependence of the forecast errors on the consistently estimated ζ.

1.2 The SP-IV Estimator

Let $\Phi_s(b,d)$ denote the block in the weighting matrix $\Phi(b,d)$ corresponding to the identifying moments in (5). Our baseline estimator uses $\Phi_s(b,d) = I_H \otimes Q^{-1}$, where $Q = E[z_t^{\perp} z_t^{\perp\prime}]$, to standardize and orthogonalize $z_t^{\perp}$.[3] The resulting solution to (7) for β in population is

$$\begin{aligned} \beta = & \left(R^{\prime}(E[Y_{H,t}^{\perp} z_t^{\perp\prime}] Q^{-1} E[Y_{H,t}^{\perp} z_t^{\perp\prime}]^{\prime} \otimes I_H) R\right)^{-1} \\ & \times R^{\prime} \operatorname{vec}(E[y_{H,t}^{\perp} z_t^{\perp\prime}] Q^{-1} E[Y_{H,t}^{\perp} z_t^{\perp\prime}]^{\prime}) , \end{aligned} \tag{8}$$

where $R \equiv I_K \otimes \operatorname{vec}(I_H)$. 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 $y_{H,t}^{\perp}$,

[3] Efficient GMM uses $\Phi_s(\beta,\zeta) = (\Sigma_{u_H^{\perp}}^{-1} \otimes Q^{-1})$, where $\Sigma_{u_H^{\perp}} = E\left[u_{H,t}^{\perp}(\beta) u_{H,t}^{\perp}(\beta)^{\prime}\right]$. Appendix B presents the resulting GLS version of SP-IV, as well as the CUE SP-IV estimator.

$Y^{\perp}_{H,t}$, and $z^{\perp}_t$ respectively. The natural sample analog of (8) is

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

which minimizes (7) with respect to b, using the sample weighting matrix, $I_H \otimes (Z^{\perp}Z^{\perp\prime}/T)^{-1}$. That minimization problem is equivalent to minimizing $\operatorname{Tr}(u^{\perp}_H P_{Z^{\perp}} u^{\perp\prime}_H)$, or the sum of squared residuals in the system

$$(10) \qquad y^{\perp}_H \;=\; (\beta' \otimes I_H)Y^{\perp}_H + u^{\perp}_H \ ,$$

after projection on the instruments $z^{\perp}_t$. Thus, $\hat{\beta}$ is also the restricted 2SLS estimator in the system of equations in (10), where the only restrictions are those implied by (1). For this reason, we refer to $\hat{\beta}$ as the **S**ystem **P**rojections on **I**nstrumental **V**ariables (SP-IV) estimator.

The SP-IV estimator has a useful interpretation in terms of the impulse response functions of y_t and Y_t to innovations in the instruments z_t. Consider the following IRF estimates,

$$(11) \qquad \hat{\Theta}_Y = \frac{Y^{\perp}_H Z^{\perp\prime}}{T}\left(\frac{Z^{\perp}Z^{\perp\prime}}{T}\right)^{-\frac{1}{2}} \quad ; \quad \hat{\Theta}_y = \frac{y^{\perp}_H Z^{\perp\prime}}{T}\left(\frac{Z^{\perp}Z^{\perp\prime}}{T}\right)^{-\frac{1}{2}} \ ,$$

which are OLS coefficients from regressing $Y^{\perp}_{H,t}$ and $y^{\perp}_{H,t}$ on standardized innovations to the instruments, $\left(Z^{\perp}Z^{\perp\prime}/T\right)^{-\frac{1}{2}} z^{\perp}_t$. Using $\hat{\Theta}_y$, construct the $HN_z \times 1$ vector $\hat{\mathbf{\Theta}}_y$ stacking the N_z vectors of IRF coefficients of y_t. Construct the $HN_z \times K$ matrix $\hat{\mathbf{\Theta}}_Y$ similarly stacking $\hat{\Theta}_Y$. Formally,

$$(12) \qquad \begin{aligned} \hat{\mathbf{\Theta}}_Y &= ((Z^{\perp}Z^{\perp\prime}/T)^{-\frac{1}{2}} Z^{\perp} \otimes I_H/T)\mathbf{Y}^{\perp}_H \ ; \\ \hat{\mathbf{\Theta}}_y &= ((Z^{\perp}Z^{\perp\prime}/T)^{-\frac{1}{2}} Z^{\perp} \otimes I_H/T)\mathbf{y}^{\perp}_H \ , \end{aligned}$$

where $\mathbf{y}^{\perp}_H = \operatorname{vec}\left(y^{\perp}_H\right)$ is $TH \times 1$ and $\mathbf{Y}^{\perp}_H = [\operatorname{vec}(Y^{\perp}_{H,1}), \ldots, \operatorname{vec}(Y^{\perp}_{H,K})]$ is $TH \times K$. Then the SP-IV estimator $\hat{\beta}$ in (9) can be expressed as

$$(13) \qquad \begin{aligned} \hat{\beta} &= (\mathbf{Y}^{\perp\prime}_H (P_{Z^{\perp}} \otimes I_H)\mathbf{Y}^{\perp}_H)^{-1}\mathbf{Y}^{\perp\prime}_H (P_{Z^{\perp}} \otimes I_H)\mathbf{y}^{\perp}_H \ , \\ &= (\hat{\mathbf{\Theta}}'_Y \hat{\mathbf{\Theta}}_Y)^{-1}\hat{\mathbf{\Theta}}'_Y \hat{\mathbf{\Theta}}_y \ , \end{aligned}$$

which shows $\hat{\beta}$ is the slope in OLS regression of $\hat{\mathbf{\Theta}}_y$ on $\hat{\mathbf{\Theta}}_Y$, i.e. the coefficients in a regression of IRFs of y_t and Y_t to z_t conditional on X_{t-1}.

The expression for $\hat{\beta}$ in (13) presents a two-step procedure for implementing SP-IV. The first step consists of estimating IRFs using instruments satisfying the exogeneity conditions. In the second step, the SP-IV estimator is obtained by regressing the IRF of the outcome variable y_t on the IRFs of the endogenous variables Y_t. To theoretically justify the moment conditions in (4), it will often be natural to choose instruments leading to impulse responses to interpretable economic shocks, such as monetary policy shocks, government spending shocks, etc. For the Phillips curve example in (2), the first step estimates IRFs of inflation π_t and the slack measure gap_t to a monetary policy shock (or other aggregate demand shocks orthogonal to the cost-push term, u_t). In the second step, the IRF of π_t is regressed on the IRF of gap_t as well as the IRFs of lagged and expected future inflation, π_{t-1} and π^e_{t+1}. The latter are obtained by lagging and leading the IRF of π_t by one horizon. Appendix A gives practical details on the implementation using LPs or VARs.

A large literature studies the identification of economic shocks presenting potential instruments for SP-IV, see Ramey (2016) or Kilian and Lütkepohl (2017) for surveys. Essentially any valid strategy for identifying structural IRFs based on LPs or VARs can be used in conjunction with SP-IV provided the underlying shocks satisfy the exogeneity conditions (5). SP-IV produces the coefficients in structural economic relationships that best fit the IRFs of the variables in those relationships in response to shocks chosen by the econometrician. SP-IV actually only requires IRFs to an identified rotation of economic shocks that satisfy the exogeneity conditions. In other words, the shocks and their associated IRFs need not necessarily be separately identified. In practice, it is also possible to perform SP-IV with a subset of horizons rather than all $h = 0, \dots, H-1$.

1.3 The Difference Between SP-IV and 2SLS

The standard approach for estimating β in (1) with $z_t, ..., z_{t-H+1}$ as instruments exploits the HN_z orthogonality conditions

$$E[u_t z_{t-h}] = 0 \quad ; \quad h = 0, \dots, H-1 \ . \tag{14}$$

In a 2SLS implementation, the first stage consists of regressing the endogenous variables Y_t on the lag sequence $z_t, ..., z_{t-H+1}$, and the second stage consists of regressing y_t on the predicted values. When z_t consists of measures of economic shocks, the first stage estimates the IRF coefficients of Y_t to the shocks z_t using a DL model. Barnichon and Mesters (2020) observe that, after similarly estimating the IRF of y_t, the 2SLS estimates equal the estimates in OLS regression of the IRF of y_t on the IRFs of Y_t. The 2SLS estimator with lagged shocks as instruments therefore can – like SP-IV – be interpreted as a regression with IRFs. In 2SLS, the regression uses IRFs estimated by single equation DL models, i.e. regressions of y_t and Y_t on $z_t, ..., z_{t-H+1}$ without additional controls. In contrast, in SP-IV the IRFs can be obtained from LPs or VARs in which the h-step ahead forecasts of y_t and Y_t given z_t are conditioned on a set of additional predictors, X_{t-1}. In the literature, IRFs are typically estimated using LPs or VARs, not DL models. An advantage of SP-IV is therefore that it estimates structural relationships across IRFs as they are estimated in practice. Whereas the IRFs for 2SLS rely on the availability of external measures of economic shocks, the IRFs for SP-IV can also exploit internal instruments generated from recursivity assumptions, or any other short or long-run covariance restrictions. Thus, SP-IV greatly expands the options for identification.

The ability to accommodate controls also yields three further theoretical advantages of SP-IV. To exposit these advantages, we adopt the usual impulse-propagation paradigm to express y_t and Y_t in terms of current and past realizations of "structural shocks", 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 the error term can be expressed as a linear combination of current and past shock realizations:

$$u_t = \mu_0' \epsilon_t + \mu_1' \epsilon_{t-1} + \mu_2' \epsilon_{t-2} + \ldots . \tag{15}$$

We also assume stationarity throughout.

1. Weaker Exogeneity Requirements With suitably chosen predictors X_{t-1}, SP-IV has weaker exogeneity requirements than 2SLS:

Proposition 1. *The exogeneity condition for 2SLS with lags of z_t is*

$$\mu_l' E[\epsilon_{t+h-l} z_t'] = 0 \quad ; \quad l = 0, \ldots, \infty \quad ; \quad h = 0, \ldots, H-1. \tag{16}$$

If X_{t-1} spans all $\epsilon_{j,t-l}$ such that $\mu_{j,l} = 0$ in (15), SP-IV including X_{t-1} as controls requires only

$$\mu_l' E[\epsilon_{t+h} z_t'] = 0 \quad ; \quad h = 0, \ldots, H-1 \tag{17}$$

Proof. The 2SLS result follows from substituting (15) in (14) and stationarity. The SP-IV result follows similarly, orthogonalizing (15) to X_{t-1}. □

Following Stock and Watson (2018), we denote the conditions in (16) with $l > h$ as *lag exogeneity*, with $l = h$ as *contemporeneous exogeneity*, and with $l < h$ as *lead exogeneity*. 2SLS requires all three forms of exogeneity to hold.[4] In contrast, SP-IV requires only contemporaneous and lead exogeneity, since conditioning on X_{t-1} eliminates the influence of all past values of ϵ_t on $u_t^{\perp}(h)$. With a sufficiently rich set of predictors, the exogeneity conditions on z_t are thus substantially weaker, echoing an argument of Stock and Watson's (2018) for including controls in LP-IV. Even if X_{t-1} does not fully span the shocks in practice, it can still reduce the scope of the exogeneity condition.

Consider the Phillips Curve example in (2). As instruments, Barnichon and Mesters (2020) consider a DL of Romer and Romer's (2004) measure of monetary policy surprises, z_t^{RR}, which are the residuals in a regression of the intended funds rate change at FOMC meetings on the current rate and Greenbook forecasts of output growth and inflation. Assume no measurement error and that the error term in (2) is just an exogenous cost-push shock following $u_t = \rho_u u_{t-1} + \upsilon_t$, with $0 \leq \rho_u < 1$, and υ_t is white noise. Unless $\rho_u = 0$, u_t depends on υ_t, and all lags $\upsilon_{t-1}, \upsilon_{t-2}, \ldots$. If z_t^{RR} is uncorrelated with υ_t, its leads up to $H-1$, and all of its lags, it

[4] The conditions in (16) are sufficient but not strictly necessary, as exogeneity requires only $\sum_{l=0}^{\infty} \mu_l' E[\epsilon_{t+h-l} z_t'] = 0$. However, there are no theoretical reasons to expect this knife-edge case.

satisfies the exogeneity requirements for 2SLS estimation of the Phillips curve with $z_t^{RR}, ..., z_{t-H+1}^{RR}$ as instruments. Suppose, however, that the regression generating z_t^{RR} is misspecified by omitting one or more lags of inflation. In that case, z_t^{RR} generally still depends on lags of υ_t, and the lag exogeneity requirement for 2SLS is not satisfied. However, by including lags of inflation amongst predictors X_{t-1}, the exogeneity requirements for SP-IV remain satisfied as long as contemporaneous and lead exogeneity hold. We return to this example later in the simulations of Section 3.

The assumption that a set of variables X_{t-1} spans the history of shocks ϵ_t determining u_t resembles the invertibility assumption in VARs and the practice of including lagged controls in LPs to avoid lag exogeneity requirements (Ramey 2016; Stock and Watson 2018). Here though, the assumption is weaker than that needed to estimate dynamic causal effects using LPs of VARs: X_{t-1} must span the shocks included in the error term u_t in the structural equation of interest, rather than the history of all shocks driving y_t and Y_t jointly.[5] Nevertheless, a richer set of predictors offers better insurance against violations of the lag exogeneity assumption.

It is not possible to circumvent the lag exogeneity requirement of 2SLS by first projecting z_t on X_{t-1} and using the residuals, $z_t^{\perp}, ..., z_{t-H+1}^{\perp}$ as the instrumental variables in 2SLS. This is the implicit procedure, for example, when a shock is first identified in a VAR or LPs with X_{t-1} as controls, and a DL of that shock is then used as the instruments in 2SLS. u_t must still be orthogonal to all lags of the identified shock, which will not generally hold. Even if such a weaker form of lag exogeneity is plausible, this procedure does not have any of the other advantages of SP-IV.

2. Efficiency Gains The second advantage of SP-IV is that conditioning on predictors X_{t-1} can lead to asymptotic efficiency gains relative to 2SLS with lagged instruments. Whether SP-IV improves efficiency depends on the data generating process (DGP) driving u_t and the informativeness of the predictors X_{t-1}. Intuitively, SP-IV is more efficient than

[5]For the Phillips curve, z_t^{RR} could still be contaminated by other demand shocks after conditioning on X_{t-1}, and the IRFs identified with z_t^{RR} in VARs or LPs with X_{t-1} as controls may therefore not represent the causal effects of monetary policy shocks; nevertheless, as long as X_{t-1} eliminates the influence of $\upsilon_{t-1}, \upsilon_{t-2}, \ldots$, z_t^{RR} remains a valid instrument for SP-IV including on X_{t-1}.

2SLS if the variances of forecast errors $u_t^{\perp}(h)$ at $h = 0, ..., H-1$ are small relative to the variance of the error term u_t. The ranking of estimators depends on the DGP. We consider an AR(1) illustrative model,

$$u_t = \rho_u u_{t-1} + \upsilon_t, 0 \leq \rho_u < 1, E[\upsilon_t] = 0, E[\upsilon_t^2] = \sigma_\upsilon^2, E[\upsilon_t \upsilon_s] = 0, s \neq t. \tag{18}$$

Let $\Sigma_{u_H^{\perp}}$ denote the covariance of $u_{H,t}^{\perp}$.

Proposition 2.

(i) If u_t is i.i.d., SP-IV is asymptotically as efficient as 2SLS.

(ii) If u_t follows the AR(1) process in (18) and X_{t-1} is empty or otherwise uninformative for u_t, then $u_{H,t}^{\perp} = u_{H,t}$ and $\hat{\beta}_{2SLS}$ is asymptotically more efficient than $\hat{\beta}$ whenever $\rho_u > 0$ and $H > 1$.

(iii) $\hat{\beta}$ is asymptotically more efficient than $\hat{\beta}_{2SLS}$ if $\sigma_u^2 > \mathrm{maxeval}(\Sigma_{u_H^{\perp}})$. If u_t follows the AR(1) process in (18) and X_{t-1} spans past shocks, then the condition becomes $\frac{\sigma_\upsilon^2}{1-\rho_u^2} > \mathrm{maxeval}\left(\Sigma_{u_H^{\perp}}\right)$, where the (h,i) entry of $\Sigma_{u_H^{\perp}}$ is given by $\sum_{j=1}^{\min\{h,i\}} \sigma_\upsilon^2 \rho_u^{h+i-2j}$.

Proof. See Appendix C. □

When u_t is i.i.d., the errors in both estimators are identical in population since X_{t-1} does not predict $u_t, \dots u_{t+H-1}$ and forecast errors do not accumulate over $h = 0, \dots, H-1$; so too are their asymptotic variances. Otherwise, the ranking depends on the DGP. Under (18), if X_{t-1} has no predictive power but u_t is persistent, then 2SLS dominates SP-IV. However, when X_{t-1} spans the influence of $\upsilon_{t-1}, \upsilon_{t-2}, \dots$ on the errors, SP-IV is asymptotically more efficient as long as ρ_u is sufficiently large and the forecast horizon H is not too large, see the Figure in Appendix C. Efficiency gains from using SP-IV are more likely more generally when u_t is predictable and the maximum forecast horizon, H, is moderate.

3. Stronger Identification The ability to condition on X_{t-1} in SP-IV can also improve the effective strength of the instruments. Weak instruments lead to bias in 2SLS estimators and make conventional inference methods invalid. In many time series applications, instruments are weak, while the endogenous variables can be highly persistent, and thus predictable. Let ω_t be the error term in the first stage of 2SLS with variance σ^2_ω. The $H \times 1$ vector of errors in the SP-IV first stage regression of $Y_{H,t}$ on z_t (but no additional predictors) is $v_{H,t}$, with covariance Σ_{v_H}. The errors in the SP-IV first stage regression of $Y_{H,t}$ on z_t and the additional predictors X_{t-1} is $v^{\perp}_{H,t}$, with covariance $\Sigma_{v^{\perp}_H}$.

Proposition 3. *When $K = 1$, for a given z_t and H:*

(i) Unless X_{t-1} is completely irrelevant, the concentration parameter for SP-IV conditional on X_{t-1} is larger than for SP-IV without controls.

(ii) If $\mathrm{Tr}(\Sigma_{v^{\perp}_H})/H < \sigma^2_\omega$, the concentration parameter for SP-IV is larger than that for 2SLS;

Proof. See Appendix D. □

Part(i) in proposition 3 states that when the predictors have explanatory power for the endogenous regressors, their inclusion in SP-IV increases the effective strength of the instruments as measured by the concentration parameter, and conditioning on X_{t-1} therefore decreases bias. Part(ii) in proposition 3 states that the effective instrument strength can increase also relative to 2SLS, depending on the persistence and predictability of the errors as well as on H. As the predictability of the endogenous variables diminishes with the forecast horizon H, the advantage of conditioning on lagged variables can be outweighed by the recency of z_t for Y_t in 2SLS. When $K > 1$, instrument strength depends on the entire eigenstructure of the first stage parameters (and that of $\Sigma_{v^{\perp}_H}$), such that a fully general result is not readily available. Intuitively, however, conditioning on X_{t-1} should similarly strengthen the instruments when X_{t-1} has explanatory power.

In contrast to SP-IV, adding X_{t-1} as additional regressors in both stages of 2SLS with $z_t, \ldots, z_{t-H+1}$ as instruments weakens identification and may introduce endogeneity. As an extreme case, suppose conditioning on X_{t-1} eliminates the influence of all past realizations of the structural shocks, ϵ_t, on Y_t and z_t. Including X_{t-1} as additional regressors then implies only the contemporaneous instruments z_t remain relevant, since X_{t-1}

spans all lags of z_t; by construction, all z_{t-h} for $h > 0$ are uncorrelated with $Y_t^\perp$ and are completely irrelevant as instruments. Identification can no longer exploit information from the dynamic relationship between z_t and Y_t. Moreover, when $N_z < K$, dropping these lags results in under-identification. In less extreme cases, when X_{t-1} does not span all past shocks, residualizing u_t with respect to X_{t-1} subtracts some linear combination of the lags of all past shocks. Doing so will generally induce correlation with the lagged shocks that are spanned by X_{t-1}, including those proxied by $z_{t-1}, \ldots, z_{t-H+1}$. Thus, lags of an instrument that were previously exogenous with respect to u_t may become endogenous if the shocks proxied by $z_{t-1}, \ldots z_{t-H+1}$ partially determine the controls in X_{t-1}.

1.4 Consistency of the SP-IV Estimator

Consider the following high-level assumptions on covariances.

Assumption 2. *The following probability limits and rank condition hold:*

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

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

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

(2.d) $\quad R'(\Theta_Y \Theta_Y' \otimes I_H)R$ *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 sample weighting matrix. Condition 2.b states that the covariance between $Y_H^\perp$ and $Z^\perp$ is consistently estimated. The population covariance $\Theta_Y Q^{\frac{1}{2}}$ is a rotation of Θ_Y, a matrix containing the impulse response coefficients of $Y_t^\perp$ to $z_t^\perp$, after standardization. Condition 2.c is the exogeneity condition. Finally, the rank condition 2.d is sufficient for the existence of a unique solution to the moment conditions (5), and ensures that the denominator of the closed form solution (8) is full rank; with the definition of Θ_Y, it implies that the instruments are relevant. 2.b and 2.d jointly imply that the instruments are strong, an assumption we relax in Section 2.

Assumption 2 resembles the usual (strong) IV assumptions, see for instance Stock and Yogo (2005). Condition 2.d does not require there to be at least as many instruments as endogenous regressors, $N_z \geq K$. Since $\operatorname{rank}(R'(\Theta_Y \Theta_Y' \otimes I_H)R) = \min\{K, H \operatorname{rank}(\Theta_Y \Theta_Y')\}$, the order condition is $HN_z \geq K$, since there are HN_z moment conditions in (5). Adding leads of y_t and Y_t makes up for $N_z < K$ just as adding lags of z_t does in 2SLS. Proposition 4 states the consistency result for the SP-IV estimator in (9).

Proposition 4. *Under Assumptions 1 and 2, $\hat{\beta} \xrightarrow{p} \beta$.*

Proof. Both terms in (9) converge by the stated assumptions, and the result follows from the continuous mapping theorem. □

2 Inference for SP-IV

2.1 Inference under Strong Instruments

When the instruments are strong, under the conditions in Assumption 1-2, inference for SP-IV can proceed analogously to standard 2SLS. With a further high-level assumption, the limiting distribution of $\hat{\beta}$ follows:

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

Proposition 5. *Under Assumptions 1-3,*

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_\beta) \ , \tag{19}$$

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

Proof. The result is immediate, after rearranging (9), from Proposition 4, the stated assumptions, and the continuous mapping theorem. □

V_β can be estimated by replacing $\Sigma_{u_H^{\perp}}$ with a consistent estimate, and $\Theta_Y \Theta_Y'$ with $Y_H^{\perp} P_{Z^{\perp}} Y_H^{\perp\prime}$. Inference can be based on standard Wald tests.[6]

[6]Given the model (6) and Assumption 1, estimation error in $Y_H^{\perp}$ etc. does not impact the asymptotic variance of $\hat{\beta}$. This is a Frisch-Waugh result; the expected Jacobian of the moments is block-diagonal since derivatives of second-stage moments with respect to first-stage parameters are products of controls X_{t-1} and forecast errors $y_{H,t}^{\perp}, Y_{H,t}^{\perp}, z_t^{\perp}$, which are orthogonal by construction.

A natural consistent estimator is

$$\hat{\Sigma}_{u_H^\perp} = \hat{u}_H^\perp \hat{u}_H^{\perp\prime}/(T - N_x - K). \tag{20}$$

Including adequate lags in X_{t-1} obviates the need for an autocorrelation robust estimate by eliminating autocorrelation in both $u_t^\perp$ and $z_t^\perp$. Any mechanical correlation between $u_t^\perp(0)$ and $u_{t-h}^\perp(h)$, say, drops out of $\text{var}(u_{H,t}^\perp \otimes z_t^\perp)$, since when $z_t^\perp$ is serially uncorrelated, so too is $u_{H,t}^\perp \otimes z_t^\perp$.[7] This is not the case for 2SLS, which generally requires autocorrelation robust methods due to mechanical autocorrelation in a lag sequence of z_t.

2.2 A Test for Weak Instruments

In many applications, the available instruments may be weak. If so, Wald inference will be invalid, leading to empirical rejection rates that generally exceed nominal levels. In the Online Appendix, we derive a bias-based test of instrument strength for SP-IV that is analogous to the popular Stock and Yogo (2005) bias-based test of weak instruments for standard 2SLS. We consider a Nagar approximation of the bias under weak instrument asymptotics, as in Montiel-Olea and Pflueger (2013) and Lewis and Mertens (2022). Like Stock and Yogo (2005) and Lewis and Mertens (2022), we use a weighted ℓ_2-norm of the bias to accommodate multiple endogenous regressors ($K>1$). Weak instruments are defined as those for which the bias in $\hat{\beta}$ is at least τ percent of a worst-case benchmark under weak instrument asymptotics. The test statistic is similar to that of Cragg and Donald (1993), and the test rejects the null hypothesis of weak instruments when the statistic exceeds the level-α critical value of a bounding distribution. The test nests the Stock and Yogo (2005) test when $H = 1$.

2.3 Weak Instrument Robust Inference for SP-IV

We describe two robust test statistics for SP-IV.

[7] The argument is analogous to that of Montiel Olea and Plagborg-Møller (2021) for LP with instrumental variables.

AR Statistic The "S-statistic" of Stock and Wright (2000) extends the AR statistic to the GMM setting. For SP-IV, the statistic and its limiting distribution under the null hypothesis are defined as

$$
\begin{aligned}
(21)\qquad & AR(b) = (T - d_{AR}) \operatorname{Tr}\left(u_H^{\perp}(b) P_{Z^{\perp}} u_H^{\perp}(b)' \left(u_H^{\perp}(b) M_{Z^{\perp}} u_H^{\perp}(b)' \right)^{-1} \right), \\
& AR(\beta) \xrightarrow{d} \chi^2_{HN_z} ,
\end{aligned}
$$

where $M_{Z^{\perp}} = I_T - P_{Z^{\perp}}$ is the residualizing matrix and $d_{AR} = N_z + N_x$ is a degrees of freedom correction. Rather than the moment covariance matrix, we use the normalizing matrix typically used with the AR statistic, asymptotically equivalent under the null hypothesis.

KLM Statistic The AR statistic can have poor power when there are over-identifying restrictions. This is the case when $HN_z > K$, i.e. when the number of IRF coefficients exceeds the number of endogenous regressors. As this may often be the case, we consider the Kleibergen (2005) KLM statistic, which can improve power (Andrews et al. 2019).

Following Kleibergen (2005),

$$
\begin{aligned}
(22)\qquad K(b) &= (T - d_K) \operatorname{vec}\left(\Xi^{-1} u_H^{\perp}(b) \check{Y}_H' \right)' R \\
&\quad \times \left(R' (\check{Y}_H \check{Y}_H' \otimes \Xi^{-1} u_H^{\perp}(b) u_H^{\perp\prime}(b) \Xi^{-1}) R \right)^{-1} \\
&\quad \times R' \operatorname{vec}\left(\Xi^{-1} u_H^{\perp}(b) \check{Y}_H' \right)' , \\
K(\beta) &\xrightarrow{d} \chi^2_K ,
\end{aligned}
$$

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 $Z^{\perp}$, $\Xi = u_H^{\perp}(b) M_{Z^{\perp}} u_H^{\perp\prime}(b)$, $\check{v}_H^{\perp} = v_H^{\perp} M_{Z^{\perp}}$, $\check{u}_H^{\perp}(b) = u_H^{\perp}(b) M_{Z^{\perp}}$, and $d_K = N_z + N_x$ is a degrees of freedom correction. Intuitively, 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 $(Z^{\perp} Z^{\perp\prime})^{-1/2} Z^{\perp}$. Our formulation differs from Kleibergen (2005) only by the replacement of $u_H^{\perp}$ and $v_H^{\perp}$ with $\check{u}_H^{\perp}$ and $\check{v}_H^{\perp}$. This choice is consistent with the IV statistic in Kleibergen (2002) and asymptotically equivalent to the form in Kleibergen (2005) under the null.

3 Performance of SP-IV in Model Simulations

In this section, we demonstrate the performance improvements offered by SP-IV in simulations. The objective in all simulations is to estimate the parameters of the Phillips Curve in (2) using data generated from the macroeconomic model of Smets and Wouters (2007) (SW).[8] The Phillips Curve in (2) is one of the equations in the SW model within a system of fourteen simultaneous equations for the dynamics of key macroeconomic aggregates. An important feature of the estimated SW model is that the shocks underlying the error term u_t in the Phillips curve explain a very large fraction of the variance of inflation. This means that, in realistic sample sizes, the weak instrument problem is generally severe. Moreover, the error term u_t is persistent, as are most of the macro aggregates generated by the model. Both features make the estimation of the Phillips curve parameters challenging. Conventional IV methods tend to perform poorly, and our simulation setup is therefore an ideal laboratory to evaluate the potential improvements offered by SP-IV.

As mentioned in the introduction, using a sequence of lagged endogenous variables as instruments – as in Galí and Gertler (1999) and subsequent literature – is not valid for identification in this setting. In the SW model, the error term in (2) is the ARMA(1,1) process

$$u_t = \rho_u u_{t-1} + \epsilon_t^p - \mu_p \epsilon_{t-1}^p \ , \ \rho_u = 0.99, \ \mu_p = 0.83 \tag{23}$$

where ϵ_t^p is an i.i.d. normally distributed price markup shock.[9] Inverting the autoregressive term in (23) yields $u_t = \epsilon_t^p + \rho_u(1-\mu_p)\epsilon_{t-1}^p + \rho_u(\rho_u - \mu_p)\epsilon_{t-2}^p + \rho_u^2(\rho_u - \mu_p)\epsilon_{t-3}^p + \ldots$, which shows that the error term u_t generally depends on the entire history of price markup shocks $\epsilon_t^p, \epsilon_{t-1}^p, \epsilon_{t-2}^p, \ldots$. The period t values of the endogenous model variables are functions of all current and lagged values of a 7×1 shock vector ϵ_t, including ϵ_t^p. Lagged values of these endogenous variables therefore violate the lag exogeneity requirement, and lose relevance if the data is first conditioned on prede-

[8] The data is generated from the SW model using the Dynare replication code kindly provided by Johannes Pfeifer at https://sites.google.com/site/pfeiferecon/dynare.

[9] We assume that the econometrician cannot exploit the ARMA(1,1) error structure in (23).

termined variables to avoid the lag exogeneity requirement.

Because lagged endogenous variables are not valid instruments, we consider a measure of the monetary policy shock as z_t, as in Barnichon and Mesters (2020). We present two sets of simulations. In the first, we use a measure of monetary policy shocks that violates the lag exogeneity requirement in an arguably realistic manner; the SP-IV estimator – unlike the 2SLS estimator – remains consistent. In the second, we use the true model monetary policy shock as the instrument to level the playing field across estimators and compare the small sample performance of 2SLS and SP-IV when both are consistent. Additional simulation results with multiple shocks as instruments and the generalized (or efficient GMM) version of SP-IV are available in the Online Appendix.

We do not assume that the econometrician possesses a set of controls spanning the full history of model shocks. Instead, we use a realistic set of controls, four lags of seven endogenous model variables: the short-term interest rate, inflation, marginal cost, output, consumption, investment, and the real wage. Inflation expectations π^e_{t+1} are unobserved and are replaced in (2) by realized future inflation π_{t+1}, as is typical in the literature when expectations appear in structural equations. Under rational expectations – as assumed in the SW model – the resulting measurement error depends only on future realizations of the model shocks, which does not create any additional endogeneity problems given that all instruments satisfy lead exogeneity.

3.1 Simulations with Violations of Lag Exogeneity

Our first set of simulations demonstrates how SP-IV can help ensure exogeneity by conditioning on lagged macroeconomic variables. We are motivated by the identification of the Phillips Curve, for example, with monetary policy shock measures like those constructed by Romer and Romer (2004), or based on high-frequency changes in Fed Funds futures as in Kuttner (2001). A practical concern with such measures is that, despite careful construction, they may still contain a meaningful predictable component (Barakchian and Crowe 2013; Bauer and Swanson 2022; Coibion 2012; Miranda-Agrippino and Ricco 2021; Ramey 2016). Consequently, researchers identifying IRFs using these measures typically include various lagged macro variables as controls in their models. However, when the same measures are used as instruments in structural equations via 2SLS – as in Barnichon and Mesters (2020) for example – estimation proceeds without controls.

To illustrate the potential implications of excluding controls, we simulate "Romer and Romer (2004) instruments" that consist of the true monetary policy shocks in the SW model, augmented with a linear function of inflation over the past four quarters. We estimate the coefficients on lagged inflation by regressing the actual Romer and Romer (2004) measures on four lags of the log change in the GDP deflator (the inflation measure used to estimate the SW model) over the 1969-2004 sample. The resulting instruments have non-zero covariances with lagged inflation that are calibrated to the U.S. data (with an R^2 of 0.08), and therefore violate the lag exogeneity requirement. However, the simulated instruments are exogenous conditional on suitable controls. In our simulations, we consider both LP and VAR implementations of SP-IV using four lags of the previously described conditioning set, X_{t-1}.

The left panel in Table 1 reports mean estimates of $\beta = [\gamma_b, \gamma_f, \lambda]'$ across 5000 Monte Carlo samples. We consider specifications with horizons of $H = 8$ and $H = 20$ quarters. To focus on the violation of the exogeneity requirements, Table 1 considers a long sample $T = 5000$, to minimize small-sample features. The true model parameters are shown in the first row, with OLS estimates in the second. The remaining rows report results for 2SLS with H lags of the monetary policy instrument, 2SLS conditioning on X_{t-1} (2SLS-C), the SP-IV based on LP without controls (SP-IV LP), and LP and VAR implementations of SP-IV (SP-IV LP-C and SP-IV VAR) conditioning on X_{t-1}.

Unsurprisingly, the OLS estimates are severely biased because of endogeneity, pointing incorrectly to a completely flat Phillips curve. Because of the violation of lag exogeneity, the 2SLS estimates are also strongly biased. The average estimate of λ even has the wrong sign for both $H = 8$ and $H = 20$. The 2SLS-C estimates show smaller bias for γ_b and γ_f, but the average estimate of λ continues to have the wrong sign for both

Table 1: Results with Lag Endogenous Instrument, $T = 5000$

	Mean Estimates		
	γ_b	γ_f	λ
True Value	0.15	0.85	0.05
OLS	0.48	0.48	0.00
$H = 8$			
2SLS	0.26	0.58	-0.09
2SLS-C	0.19	0.87	-0.06
SP-IV LP	0.26	0.60	-0.08
SP-IV LP-C	0.16	0.84	0.05
SP-IV VAR	0.12	0.83	0.09
$H = 20$			
2SLS	0.24	0.76	-0.02
2SLS-C	0.21	0.84	-0.05
SP-IV LP	0.24	0.75	-0.02
SP-IV LP-C	0.23	0.81	0.02
SP-IV VAR	0.17	0.83	0.05

Empirical Size of Nominal 5% Tests		
	$H = 8$	$H = 20$
WALD 2SLS	52.10	95.80
WALD 2SLS-C	81.30	87.80
WALD SP-IV LP	59.10	95.90
WALD SP-IV LP-C	7.80	29.70
WALD SP-IV VAR	5.50	13.10
AR 2SLS	73.10	71.40
AR 2SLS-C	71.00	55.10
AR SP-IV LP	69.50	57.00
AR SP-IV LP-C	4.90	5.00
AR SP-IV VAR	4.90	4.80
KLM 2SLS	83.20	86.50
KLM 2SLS-C	30.90	21.30
KLM SP-IV LP	82.50	76.40
KLM SP-IV LP-C	5.30	4.80
KLM SP-IV VAR	5.00	4.60

Notes: In the left panel, the top row reports the true Smets and Wouters (2007) model parameters, and the remaining rows the means estimates across 5000 Monte Carlo samples. All IV estimators use $h = 0, ..., H-1$ and the lag endogenous monetary policy instruments described in the text. SP-IV LP and LP-C denote implementations based on local projections without and with X_{t-1} (described in the text) as controls, respectively. SP-IV VAR denotes implementation with a vector autoregression for X_t with four lags. Robust tests for 2SLS use a Newey-West HAR variance matrix with Sun (2014) fixed-b critical values; inference procedures for SP-IV are described in Section 2.

$H = 8$ and $H = 20$. While conditioning may remove dependence on past inflation, doing so weakens the relevance of the instruments and induces endogeneity of its own, as explained in Section 1.3. The next row shows the SP-IV estimator without controls X_{t-1}; it is also biased because, like 2SLS, it requires lag exogeneity to hold. The bias is almost identical to that of 2SLS since they exploit similar moments for identification. The next two rows show SP-IV estimators that condition on X_{t-1} using either LPs or a VAR. Both procedures produce mean estimates with the correct sign and values that are much closer to the truth. The reason for the smaller bias is the conditioning step, which helps eliminate the persistent influence of past cost-push shocks that leads to a violation of the lag exogeneity requirement. While the SP-IV LP-C and SP-IV VAR estimators have much smaller bias, some bias remains. This residual bias arises either because X_{t-1} does not fully span the history of cost-push shocks, because the IRFs are misspecified, or because weak instrument bias remains even as $T = 5000$.

The right panel of Table 1 reports empirical rejection rates for a nominal 5% test that β equals the true value for the various SP-IV inference procedures described in Section 2 and analogous HAR procedures for 2SLS. When exogeneity fails, rejection rates will not match nominal levels. Every test associated with estimators for which exogeneity is violated (OLS, 2SLS, 2SLS-C, and SP-IV LP) is badly oversized. Conversely, for the SP-IV estimators that condition on X_{t-1}, (SP-IV LP-C and SP-IV VAR), the robust AR and KLM tests, defined in (21) and (22) respectively, exhibit empirical rejection rates very close to 5%, again demonstrating that the conditioning step adequately protects against the violation of lag exogeneity. The Wald tests for these estimators remain somewhat oversized, especially when $H = 20$. The fact that robust inference procedures effectively control size indicates that the residual bias is primarily related to the weakness of the instruments, even in a relatively large sample.

The results in Table 1 illustrate the advantage of weakening the exogeneity condition by using SP-IV with lagged controls instead of 2SLS. The same controls cannot be included in 2SLS specifications, because doing so renders the lagged instruments irrelevant. The results in Table 1 also show that instrument weakness is a concern, even at $T = 5000$. Next, we demonstrate the additional advantages of SP-IV in simulations with smaller samples, and therefore more severe weak instrument problems.

3.2 Small Sample Performance

Given the limited role of monetary policy shocks for inflation dynamics in the SW model, estimating the parameters of the Phillips curve using monetary policy shocks as instruments is especially challenging in small samples. The main goal of the next simulations is to show how the conditioning step in SP-IV can not only weaken exogeneity requirements but also substantially alleviate weak instrument problems. To level the playing field across estimators, we now assume that the econometrician has the true monetary policy shocks as instruments. This assumption is unrealistic but permits a fair comparison between the various estimators as the exogeneity requirement is now satisfied for all 2SLS and SP-IV estimators. We consider a sample of $T = 250$ quarters, a best-case in most macro applications, roughly corresponding to the postwar period, but also report results for $T = 500$ and $T = 5000$ to verify the asymptotic properties of the estimators and inference procedures.[10] Results for $N_z = 3$ in the Online Appendix are qualitatively similar.

Bias. Table 2 reports the mean estimates of $\beta = [\gamma_b, \gamma_f, \lambda]'$ for the various samples sizes. The first two rows report the true model parameters and OLS results. As expected, OLS is severely biased regardless of T due to endogeneity. The other rows show the results for the various 2SLS and SP-IV estimators with $H = 8$ or $H = 20$ quarters.

As the first row under $H = 8$ in Table 2 shows, 2SLS produces estimates that on average are closer to the true parameter values than OLS. Because the instruments are now also lag exogenous, the 2SLS estimates also converge to the truth as the sample size grows. However, despite the use of valid instruments, there remains considerable bias in realistic samples with $T = 250$. The Phillips Curve slope, λ, is estimated to be much flatter on average than in the model: 0.01 compared to 0.05. The backward and forward-looking inflation terms are also heavily misweighted, with γ_f too low on average, and γ_b too high. The next row, labeled 2SLS-C, shows that adding the lagged controls X_{t-1} to 2SLS does not mitigate the small sample problems; rather, the bias is worse than for 2SLS without controls, except for γ_f in some specifications. Moreover, the bias for λ grows even worse as T increases. This is because conditioning not only weakens the relevance of the lag instruments but also renders them endogenous when the controls do not span all past shocks. The next row

[10]In further simulations, available on request, we consider Barnichon and Mesters's (2020) 2SLS estimator with Almon shrinkage. The performance is poor, with bias highly variable over H, T, and parameters, and standard deviations one to two orders of magnitude larger than the other estimators. No inference procedure controls size well across all specifications. In contrast, we find that SP-IV LP-C and VAR perform very well when tested on Barnichon and Mesters's (2020) DGP.

TABLE 2: MEAN PARAMETER ESTIMATES

	$T = 250$			$T = 500$			$T = 5000$		
	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
True Value	0.15	0.85	0.05	0.15	0.85	0.05	0.15	0.85	0.05
OLS	0.47	0.47	0.00	0.48	0.48	0.00	0.48	0.48	0.00
$H = 8$									
2SLS	0.27	0.51	0.01	0.23	0.60	0.01	0.17	0.83	0.04
2SLS-C	0.31	0.68	-0.06	0.27	0.76	-0.04	0.18	0.89	-0.11
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
$H = 20$									
2SLS	0.39	0.53	0.01	0.36	0.61	0.00	0.23	0.80	0.01
2SLS-C	0.37	0.57	-0.07	0.33	0.64	-0.06	0.21	0.85	-0.08
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	0.40	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

Notes: The top row gives the true parameter values in the Smets and Wouters (2007) model. The others report the mean estimates across 5000 Monte Carlo samples. All IV estimators are based on $h = 0, ..., H-1$ and use true model shocks as instruments. 2SLS-C denotes the 2SLS estimator including X_{t-1} (described in the text) as controls. SP-IV LP and LP-C denote implementations based on local projections without and with X_{t-1}, respectively. SP-IV VAR denotes implementation with a vector autoregression for X_t with four lags.

shows that, without controls, the bias of SP-IV is almost identical to that of 2SLS for all T. This is again unsurprising, as in this case, both exploit essentially the same identifying moments.

The next two rows under $H = 8$ illustrate the possible bias reductions when using the LP-C or VAR implementations of SP-IV, both of which condition on X_{t-1}. For the LP-C implementation, the estimates of λ average 0.04 in samples with $T = 250$, which is much closer to the true value of 0.05 than for 2SLS. The forward-looking coefficient in the Phillips Curve, γ_f, is also considerably closer to the truth, and the bias in the backward-looking coefficient, γ_b, is only marginally worse. The VAR implementation of SP-IV also delivers substantial bias improvements in all three coefficients relative to 2SLS, although the improvement for λ is slightly smaller than for the LP-C implementation. Taken together, the reductions in small sample bias by adopting SP-IV LP-C or SP-IV VAR are substantial. These reductions are also economically meaningful, as the average differences in parameter estimates have considerable implications for inflation dynamics and the inflation-output gap trade-off. As discussed in Section 1.3, the improvements relative to 2SLS arise because the conditioning step amplifies the signal provided by the monetary policy shock instruments, which is generally weak in the SW DGP.

The improvements in small sample performance of SP-IV relative to 2SLS depend on the choice of H. Including additional horizons can add useful identifying variation. On the other hand, the endogenous variables become harder to predict at longer horizons. The results in Table 2 for $H = 20$ show that the relative performance of the estimators is qualitatively the same as for $H = 8$. Quantitatively, however, the reductions in bias under the LP-C or VAR implementations of SP-IV are smaller than they are for $H = 8$. In general, as predicted in Section 1.3, the advantages of SP-IV over 2SLS diminish as the number of lags included as instruments in 2SLS – which is also the maximum forecast horizon in SP-IV – grows larger.

Variance. Table 3 reports the standard deviations of the various estimators. We omit OLS and 2SLS-C given their poor performance in terms of bias. Section 1.3 showed that SP-IV can be asymptotically more efficient than 2SLS after conditioning on controls when H is not too large and the error term u_t is a sufficiently persistent AR(1) process. While the error term in our simulations is the ARMA(1,1) process in (23), similar efficiency gains can arise. Table 3 indeed shows efficiency gains for $T = 5000$. For $H = 8$, the standard deviations of the SP-IV LP-C estimates are uniformly smaller than those of the 2SLS estimates. For the VAR implementation, the standard deviation is smaller for estimates of γ_f, and roughly similar to 2SLS for the other two parameters. Consistent with the theory, the relative efficiency of SP-IV disappears for larger H, as can be seen for $H = 20$ and $T = 5000$ in the bottom panel. Also consistent with the theory is that the conditioning step is essential to re-

Table 3: Standard deviation of parameter estimates

	$T = 250$			$T = 500$			$T = 5000$		
	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
$H = 8$									
2SLS	0.26	0.33	0.21	0.24	0.30	0.20	0.13	0.09	0.09
SP-IV LP	0.28	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	0.37	0.28	0.30	0.25	0.26	0.14	0.06	0.09
$H = 20$									
2SLS	0.11	0.12	0.06	0.10	0.11	0.06	0.07	0.05	0.03
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	0.21	0.25	0.10	0.20	0.19	0.09	0.11	0.06	0.05

Notes: The table shows standard deviations of the estimates across 5000 Monte Carlo samples from the Smets and Wouters (2007) model. All IV estimators are based on $h = 0, ..., H-1$ and use true model shocks as instruments. SP-IV LP and LP-C denote implementations based on local projections without and with X_{t-1} (described in the text) as controls, respectively. SP-IV VAR denotes implementation with a vector autoregression for X_t with four lags.

alize any efficiency gains: the SP-IV estimates that do no condition on X_{t-1}, in the second row of each panel, have similar or larger variance than 2SLS. In smaller samples, the LP-C implementation of SP-IV has similar variance to 2SLS, with most standard deviations being somewhat smaller than 2SLS, and some only slightly larger. The standard deviations of the VAR implementation of the SP-IV, on the other hand, are systematically somewhat greater than those of 2SLS with $T = 250$ or $T = 500$.

At least for the DGP considered here, the LP-C implementation of SP-IV consistently generates lower bias than 2SLS, while it has similar or smaller variance. The VAR implementation yields further reductions in bias in our setting, but generally also has slightly higher variance. That the VAR implementation has smaller bias but greater variance may be surprising given the bias-variance trade-off between VARs and LPs for the estimation of IRFs.[11] However, the SP-IV estimators are not IRFs, but relationships between IRFs. Biases and covariances across IRFs can

[11] Typically, imposing VAR dynamics introduces bias in the IRFs but yields efficiency gains relative to the LP approach, see Plagborg-Møller and Wolf (2021), Li et al. (2021). In the Online Appendix, we show that this trade-off is also present for the IRFs in our simulations.

have offsetting or reinforcing effects on the bias and variance of the SP-IV estimators. The relative bias-variance properties of the LP-C and VAR implementations of SP-IV are likely application-specific.

Finally, all standard deviations in Table 3 are decreasing in H, indicating that additional horizons reduce the variability of all estimators. Given our bias results, this implies a bias-variance trade-off when choosing the maximum horizon H for SP-IV: larger H provides smaller bias improvements relative to 2SLS, but generates less variable estimates.

Inference. Given that monetary policy shocks are weak instruments, a key question is how severe size distortions are using standard Wald inference, and how well the weak instrument robust procedures control size in practice. It is well known that robust procedures may still perform poorly when the number of instruments is large (Bekker 1994). Barnichon and Mesters (2020), for example, report severe size distortions for AR inference for 2SLS with long lag sequences of instrumenting shocks. Since SP-IV uses HN_z moments, it potentially faces the same theoretical "many moments" problem as 2SLS with HN_z instruments (Han and Phillips 2006; Newey and Windmeijer 2009).

Table 4 reports empirical rejection rates for nominal 5% tests of the true values of the full parameter vector, $\beta = [\gamma_b, \gamma_f, \lambda]'$ for sample sizes of $T = 250, 500$ and 5000. We also report results for $N_z = 3$ with $H = 20$, see the Online Appendix for details. Any size distortions will generally decrease with T, since the first-stage relationships remain fixed. Identification strength therefore improves with T.

Wald tests for 2SLS exhibit meaningful size distortions for $H = 8$, with empirical rejection rates substantially above the nominal 5%. The size distortions become much larger for $H = 20$ and $N_z = 3$. These distortions are not surprising given the weakness of the instruments and demonstrate the need for robust inference procedures. For $H = 8$ and $N_z = 1$, the 2SLS AR test is relatively well-sized in small samples. Indicative of many instrument problems, the 2SLS AR test becomes noticeably oversized in small samples when $H = 20$, dramatically so for $N_z = 3$. The 2SLS KLM test controls size better but can be conservative, potentially due to the use

Table 4: Empirical size of nominal 5% tests

	$H=8$, $N_z=1$			$H=20$, $N_z=1$			$H=20$, $N_z=3$		
$T=$	250	500	5000	250	500	5000	250	500	5000
WALD 2SLS	8.6	7.3	9.3	65.1	58.0	40.9	100.0	99.8	92.5
WALD 2SLS-C	22.1	22.3	91.3	63.7	56.3	92.7	99.9	99.1	96.9
WALD SP-IV LP	15.2	12.7	12.9	68.5	64.1	43.8	99.9	99.9	91.7
WALD SP-IV LP-C	13.0	11.6	7.8	72.0	63.5	29.7	100.0	99.7	76.6
WALD SP-IV VAR	7.8	7.1	5.5	32.2	27.4	13.1	86.1	76.5	53.8
AR 2SLS	6.6	6.0	4.1	11.9	8.1	4.2	55.3	25.7	3.9
AR 2SLS-C	8.8	11.9	84.9	10.3	9.0	67.3	32.4	16.0	44.2
AR SP-IV LP	5.7	5.7	4.6	9.7	7.1	4.9	14.6	8.7	4.9
AR SP-IV LP-C	6.7	5.8	4.9	11.4	7.7	5.0	17.3	9.9	5.2
AR SP-IV VAR	4.6	4.6	4.9	5.3	5.8	4.8	6.3	5.9	4.7
KLM 2SLS	3.9	3.5	3.7	6.0	5.5	3.9	0.5	12.7	4.9
KLM 2SLS-C	3.8	5.2	36.8	2.9	2.7	20.2	1.2	3.1	7.3
KLM SP-IV LP	5.6	5.5	4.8	8.3	6.2	4.7	8.2	6.3	5.3
KLM SP-IV LP-C	6.9	6.0	5.3	11.9	7.1	4.8	11.3	7.9	5.1
KLM SP-IV VAR	5.4	5.1	5.0	8.1	6.4	4.6	10.7	8.6	5.3

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

of only approximate fixed-b critical values for HAR inference. The same tests for 2SLS-C all have unit size asymptotically since the instruments are not generally exogenous after conditioning.

Just like 2SLS, the SP-IV Wald size distortions for $H = 8$ are substantial and very large for $H = 20$. The SP-IV AR tests are well-sized overall. For the LP and LP-C implementations, they do over-reject in small samples when $H = 20$, but not to the extent seen for 2SLS. The SP-IV KLM tests in the final three rows are also generally well-sized. Just like the AR tests, the KLM tests exhibit some over-rejection in small samples when $H = 20$. Overall, however, the size distortions of the robust SP-IV tests with large HN_z are considerably milder than those for robust 2SLS tests, especially for the VAR implementation. A likely reason that the robust

SP-IV procedures control size distortions better is that the 2SLS errors exhibit mechanical serial correlation due to the lag sequence of z_t, requiring HAR covariance estimates. Since 2SLS requires an autocorrelation robust covariance matrix, the number of parameters to be estimated increases much more quickly with HN_z than for SP-IV covariances. Because of many-instrument problems, we nevertheless recommend avoiding very large HN_z also when using SP-IV.

In practice, there is no need to use all horizons for identification. Researchers can, for example, select impulse response horizons at lower frequencies than that of the time series (e.g. quarterly horizons in monthly data, annual horizons in quarterly data, etc.), especially since adjacent horizons do not necessarily contain much independent identifying information for typical shapes of IRFs. Further refinements are also possible to address any remaining many instrument problems, see for example Mikusheva (2021) for suggestions. In the context of 2SLS with DLs of shocks as instruments, Barnichon and Mesters (2020) propose quadratic approximations to the IRFs to avoid many instrument problems, and similar approximations are possible with SP-IV. Other test statistics could possibly be adapted to SP-IV and offer improvements over the AR and KLM tests, for example, those based on Moreira (2003) or Andrews (2016). Given the relatively good performance of our test statistics in the simulations, we leave such extensions for future work.

4 Application to the Phillips Curve with U.S. Data

In this section, we use SP-IV to estimate the parameters of the Phillips curve in (2) using U.S. data and compare the results with 2SLS. We consider the following specification for quarterly inflation,

$$\pi_t^{1q} = (1-\gamma_f)\pi_{t-3}^{1y} + \gamma_f \pi_{t+12}^{1y} + \lambda U_t + u_t \ , \tag{24}$$

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 specification and variable definitions are exactly as in Barnichon and Mesters (2020), but we use monthly data from 1979:M1 to 2018:M4 (472 monthly observations) instead of quarterly data. As is common in the literature, e.g. Mavroeidis et al. (2014), (24) restricts the coefficients on lagged and future inflation to sum to unity, $\gamma_b + \gamma_f = 1$, which imposes that there is no long-run trade-off between unemployment and inflation. The restriction is implemented by rewriting the structural equation as $\pi_t^{1q} - \pi_{t-3}^{1y} = \gamma_f(\pi_{t+12}^{1y} - \pi_{t-3}^{1y}) + \lambda U_t + u_t$. We consider a maximum forecast horizon of 3 years (36 months). To make efficient use of the identifying information in the IRF dynamics and mitigate many-instrument problems, we only use the coefficients in the first month of each of the first 12 quarters of the response horizons – that is, $h = 0, 3, 6, \ldots, 33$. We consider identification with a single economic shock, such that for both SP-IV and 2SLS there are twelve identifying moments. We use the VAR implementation of SP-IV, using a VAR with six lags in the following standard monthly macro variables as controls: 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 Treasury rate, and the 10-year Treasury rate.

As the instrument, we use a monthly version of the Angeletos et al. (2020) Main Business Cycle (MBC) Shock, identified within our monthly VAR by maximizing the contribution to cyclical unemployment fluctuations in the frequency domain. Angeletos et al. (2020) find that the resulting shock 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. Empirically, this propagation mechanism best fits the notion of an aggregate demand shock, making the MBC shock a plausible instrument for estimating the Phillips curve.

We choose the MBC shock for two main reasons. The first reason is that it serves as a good illustration of how SP-IV can be useful when interpreting empirical IRFs. Observing the disconnect between the unemployment and inflation impulse responses to the MBC shock, a key conclusion in Angeletos et al. (2020) is that the Phillips curve must be nearly completely flat. Rather than relying on informal visual inspections of IRFs, SP-IV allows a formal econometric investigation of the Phillips curve relationship embedded in the VAR-based IRFs. The second reason is that the MBC shock is likely the strongest available instrument for identifying the Phillips curve. By construction, the MBC shock is highly predictive for unemployment fluctuations over business cycle horizons. We find that the MBC shock also has some strength for inflation at horizons up to three years. The first two columns in the first row of Table 5 show test statistics and critical values of the weak instruments test for SP-IV, along with those for 2SLS (without controls) based on the HAR first-stage test of Lewis and Mertens (2022). For illustrative purposes, the other columns in Table 5 report results for each endogenous regressor separately. The test statistic is 7.3 for SP-IV, with a critical value of 22.1 for the null of at most 10 percent bias at the 5% level. The test statistics are 17.9 and 7.8, respectively, for unemployment and inflation separately (critical values of 21.7 and 18.5). The test statistics for 2SLS are all much lower relative to similar critical values of around 20, which illustrates how including the additional predictors in SP-IV amplifies the signal of the instrument relative to 2SLS. Despite this amplification, the MBC shock is still judged to be weak at conventional tolerance levels according to the SP-IV first-stage test. The MBC shock is, nevertheless, by far the strongest instrument across all candidate shocks that we explored.

In principle, many other shock measures could be used to identify the parameters of (24), including monetary policy shocks as in Barnichon and Mesters (2020). Table 5 reports the first stage test results for various popular monetary policy shock measures.[12] The main takeaway is that, at least in the sample that we consider, each of the monetary policy shocks is far too weak as an instrument to be useful for identifying the Phillips curve in practice. A few have some strength for inflation separately in the first stage of 2SLS, but none do for both endogenous regressors jointly, which is what matters for identification. Moreover, any hint of instrument strength disappears entirely after including lagged macroeconomic

[12] The SP-IV IRFs are identified in an "internal instrument" VAR, i.e. by adding the shocks to the other six macro variables in the VAR (see, e.g., Plagborg-Møller and Wolf (2021)).

Table 5: First-Stage Test Results

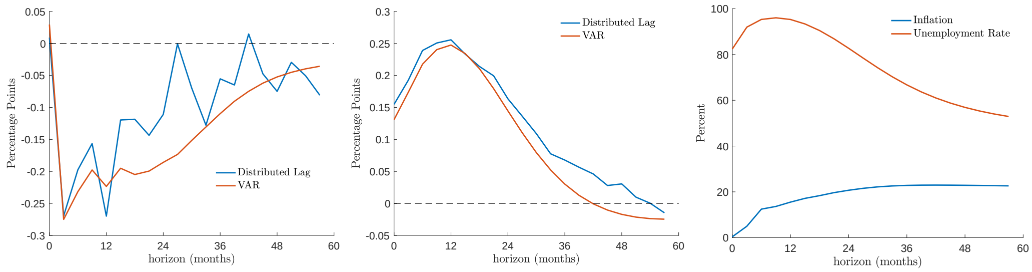
	π, U jointly				U separately				π separately			
	2SLS		SPIV		2SLS		SPIV		2SLS		SPIV	
	g	cv	g	cv	g	cv	g	cv	g	cv	g	cv
MBC	1.9	21.8	7.3	22.1	4.0	20.0	17.9	21.7	2.1	20.0	7.8	18.5
Monetary Policy Shock Measures:												
RR	0.3	20.3	0.5	20.9	0.3	17.5	0.7	20.1	1.3	17.8	1.1	18.0
GK	0.5	21.7	0.1	22.3	0.9	21.1	0.1	22.0	7.9	18.5	0.1	19.1
MAR	0.3	22.9	0.0	22.3	0.9	20.3	0.9	22.0	0.3	18.9	0.0	18.7
JK	0.4	20.0	0.1	22.3	1.0	18.3	0.4	22.0	2.3	19.0	0.1	19.1
SWA	0.8	18.9	0.1	17.2	0.8	16.9	0.3	17.1	3.3	16.5	0.2	15.0
BC	1.2	20.9	0.0	22.3	1.9	19.6	1.6	22.0	6.8	18.7	0.0	19.2
NS	1.1	20.1	0.0	22.4	1.1	19.3	0.1	22.1	6.7	17.4	0.1	18.9

Notes: The table reports test results for the null hypothesis of weak instrument bias less than or equal to 10% of the worst-case benchmark. g is the test statistic, cv is the 5% critical value. U and π are the endogenous regressors in the restricted equation, i.e. U_t and $\pi^{1y}_{t+12} - \pi^{1y}_{t-3}$. For 2SLS, results are for the HAR test of Lewis and Mertens (2022). For SP-IV, the test is described in the Online Appendix. MBC is the main business cycle shock of Angeletos et al. (2020). The monetary policy shock measures are Romer and Romer (2004) (RR), Gertler and Karadi (2015) (GK), Miranda-Agrippino and Ricco (2021) (MAR), Jarociński and Karadi (2020) (JK), Swanson (2021) (SWA), Barakchian and Crowe (2013) (BC), Nakamura and Steinsson (2018) (NS). RR, GK and BC are updated versions from Ramey (2016); the other series are from the original sources.

variables as controls in SP-IV. That none of the 2SLS/SP-IV first-stage tests with monetary policy shocks comes close to rejecting the null of weak instruments is not surprising, as it reflects the broadly held view that monetary disturbances are relatively unimportant as drivers of inflation and economic activity. We also considered several other plausible demand shock measures, such as the credit spread shock of Gilchrist and Zakrajšek (2012) and the Bloom (2009) uncertainty shock, but none are

Figure 1: Impact of the MBC Shock on Inflation and Unemployment

(a) IRF Inflation (b) IRF Unemployment (c) FEV Contribution



Notes: Inflation is the annualized Core CPI inflation rate from a quarter ago (π_t^{1q}). IRFs in red in (a) and (b) are from a 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 Treasury rate, and the 10-year Treasury rate. The estimation sample is 1979:M1 to 2018:M4. Blue lines in (a) and (b) show results from the DL regressions in the first stage of 2SLS. Panel (c) shows the FEV contributions of the MBC shock in the VAR.

nearly as strong as the MBC shock in the post-1979 sample. Using multiple shocks could improve identification strength but creates potential many-instrument problems. To the extent the MBC shock indeed collects a range of demand disturbances that satisfy the exogeneity requirements, it is by far the most informative available instrument for the identification of the Phillips curve.

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 VAR-based IRFs of π_t^{1q} and U_t to a one standard-deviation MBC shock. The figures show the first twelve IRF coefficients that are used in the estimation, and also show the next eight quarters to visualize the full dynamics. As in Angeletos et al. (2020), the MBC shock looks like an aggregate demand shock, driving unemployment higher and inflation lower. At the same time, the MBC shock explains a relatively small fraction of the forecast error contribution (FEVD) of inflation, nearly zero on impact and only 20% after two years, see Figure 1c. This finding illustrates the apparent "disconnect" between inflation and the shock that explains most of the cyclical variation in unemployment, see also Del Negro et al. (2020). As previously explained, both 2SLS and SP-IV estimates can be expressed as the coefficients in regressions of IRFs. In SP-IV, these IRFs are the red lines in Figures 1a-1b.[13] The 2SLS estimator instead uses the IRFs obtained from DL regressions of π_t^{1q} (and π_{t-3}^{1y} and π_{t+12}^{1y}) and U_t on the current and lagged values of the MBC shock. For illustration, these IRFs are shown in blue in Figures 1a-1b.

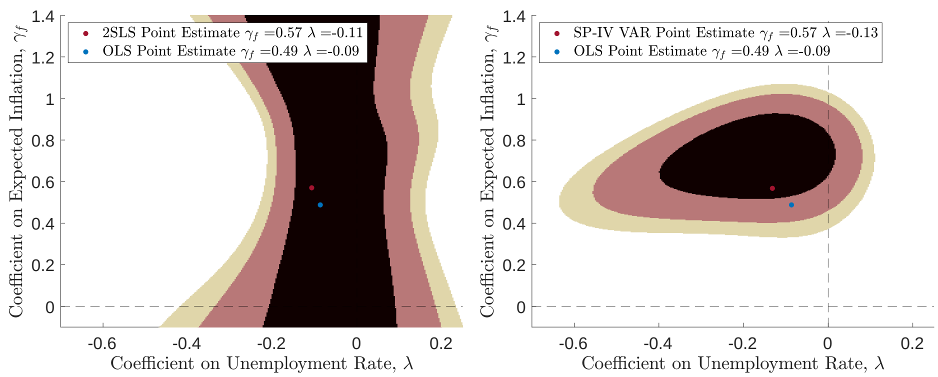
Figure 2 displays the estimates of γ_f and λ, together with 68%, 90% and 95% confidence sets. Since neither of the first-stage tests in Table 5 rejects the null of weak instruments, the confidence sets are both based on the KLM statistic. The point estimates of γ_f, the weight on future inflation, are 0.57 for both 2SLS and SP-IV. The slope estimates are also close, $\lambda = -0.13$ in SP-IV versus $\lambda = -0.11$ in 2SLS, and have the expected negative sign since unemployment is the gap measure. The similarity in point estimates is not too surprising, given that we use the VAR-identified MBC shock to construct the 2SLS instruments. The inference results, on the other hand, are much less similar. The 2SLS confidence sets do not reject any plausible values of γ^f, nor do they rule out a wide range of possible values of λ. The 90% set includes values of λ as high as 0.2 and as low -0.4, and the 95% set includes an even wider range for λ. Compared with 2SLS, 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 SP-IV sets also do not rule out 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. We attribute the relatively more informative SP-IV confidence sets to the greater effective strength of the instruments, as discussed in Section 1.3. Our earlier simulation results also showed that KLM inference for SP-IV was more reliable overall than for 2SLS.

As to the inflation-activity disconnect, our robust inference results act as a warning against drawing strong conclusions from informal comparisons of IRF point estimates. When judging relationships across IRFs, it is important to take into account that these estimates are inevitably uncertain. SP-IV estimates the posited relationships between IRFs from VARs or LPs formally and allows inference that is robust to the distor-

[13]The IRF of $\pi_{t+12}^{1y} - \pi_{t-3}^{1y}$ is straightforward to construct from the IRF of π_t^{1q}.

Figure 2: 2SLS and SP-IV Confidence Sets for Estimates of Phillips Curve Parameters

(a) 2SLS (b) SP-IV



Notes: Figures show point estimates and 68%, 90% and 95% confidence sets based on the KLM statistic described in Section 2.3 and a HAR version for 2SLS.

tions caused by sampling error in the IRF estimates. The confidence sets in Figure 1b, for example, are consistent with weak but also relatively strong cyclical connections between inflation and unemployment. The business cycle anatomy of Angeletos et al. (2020), therefore, does not provide strong evidence that inflation and activity have been largely disconnected in the post-1979 sample.

5 Concluding Remarks and Future Research

While we focused mainly on the (challenging) problem of estimating the parameters of the inflation Phillips curve, 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, and aggregate production functions. SP-IV can be used more broadly to conduct inference on ratios (or other relationships) of impulse response coefficients, such as Okun coefficients, sacrifice ratios, multipliers, etc., conditional on economic shocks. Our methodology could be extended to panel data settings and should be more generally useful in applications that commonly rely on lagged variables as instruments, such as the estimation of production functions in industrial organization. SP-IV could also be used in cross-sectional applications. If

$h = 0, ..., H-1$ indexes cross-sectional groups rather than time horizons, then SP-IV amounts to instrumental variables in the cross-section with heterogeneity in the first stage coefficients. Future work can also develop methods to optimally select the horizons/groups used for identification. We plan to pursue these and other avenues in future research.

References

Anderson, T. W., & Rubin, H. (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics*, *20*(1), 46 –63.

Andrews, I. (2016). Conditional linear combination tests for weakly identified models. *Econometrica*, *84*(6), 2155–2182.

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, *11*(1), 727–753.

Angeletos, G.-M., Collard, F., & Dellas, H. (2020). Business-cycle anatomy. *American Economic Review*, *110*(10), 3030–70.

Barakchian, S. M., & Crowe, C. (2013). Monetary policy matters: Evidence from new shocks data. *Journal of Monetary Economics*, *60*(8), 950–966.

Barnichon, R., & Mesters, G. (2020). Identifying Modern Macro Equations with Old Shocks. *The Quarterly Journal of Economics*, *135*(4), 2255–2298.

Bauer, M. D., & Swanson, E. T. (2022). *A reassessment of monetary policy surprises and high-frequency identification* (Working Paper No. 29939). National Bureau of Economic Research.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, *62*(3), 657–681.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3).

Coibion, O. (2012). Are the effects of monetary policy shocks big or small? *American Economic Journal: Macroeconomics*, *4*(2), 1–32.

Cragg, J. G., & Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, *9*(2), 222–240.

Del Negro, M., Lenza, M., Primiceri, G. E., & Tambalotti, A. (2020). What's up with the Phillips Curve?

Galí, J., & Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, *44*(2), 195–222.

Gertler, M., & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, *7*(1), 44–76.

Gilchrist, S., & Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, *102*(4), 1692–1720.

Han, C., & Phillips, P. C. B. (2006). Gmm with many moment conditions. *Econometrica*, *74*(1), 147–192.

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

Jordà, (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, *95*(1), 161–182.

Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press.

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

Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. *Econometrica*, *73*(4), 1103–1123.

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of Monetary Economics*, *47*(3), 523–544.

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

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

Mavroeidis, S., Plagborg-Møller, M., & Stock, J. H. (2014). Empirical evidence on inflation expectations in the new keynesian phillips curve. *Journal of Economic Literature*, *52*(1), 124–88.

McLeay, M., & Tenreyro, S. (2019). Optimal inflation and the identification of the phillips curve. *NBER Macroeconomics Annual*, *34*.

Mertens, K., & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review*, *103*(4), 1212–47.

Mikusheva, A. (2021). *Many weak instruments in time series econometrics* (World Congress of Econometric Society). MIT.

Miranda-Agrippino, S., & Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*, *13*(3), 74–107.

Montiel-Olea, J. L., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, *31*(3), 358–369.

Montiel Olea, J. L., & Plagborg-Møller, M. (2021). Local projection inference is simpler and more robust than you think. *Econometrica*, *89*(4), 1789–1823.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, *71*(4), 1027–1048.

Nakamura, E., & Steinsson, J. (2018). High-Frequency Identification of Monetary Non-Neutrality: The Information Effect*. *The Quarterly Journal of Economics*, *133*(3), 1283–1330.

Newey, W. K., & Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, *77*(3), 687–719.

Plagborg-Møller, M., & Wolf, C. K. (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, *89*(2), 955–980.

Plagborg-Møller, M., & Wolf, C. K. (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, *89*(2), 955–980.

Ramey, V. Chapter 2 - macroeconomic shocks and their propagation (J. B. Taylor & H. Uhlig, Eds.). In: ed. by Taylor, J. B., & Uhlig, H. Vol. 2. Handbook of Macroeconomics. Elsevier, 2016, pp. 71–162.

Romer, C. D., & Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, *94*(4), 1055–1084.

Rothenberg, T. J., & Leenders, C. T. (1964). Efficient estimation of simultaneous equation systems. *Econometrica*, *32*(1/2), 57–76.

Smets, F., & Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, *97*(3), 586–606.

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

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

Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, *128*(610), 917–948.

Stock, J. H., & Wright, J. H. (2000). Gmm with weak identification. *Econometrica*, *68*(5), 1055–1096.

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

Swanson, E. T. (2021). Measuring the effects of federal reserve forward guidance and asset purchases on financial markets. *Journal of Monetary Economics*, *118*, 32–53.

Appendix

A Practical Implementation of SPIV with LPs or VARs

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 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). X_t can include not only current values, but also lags of y_t, Y_t, Z_t, or any other time series.

Local Projections Define the $N_x \times T$ matrix X with controls X_{t-1} in the t-th column, 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

$$\text{(A.1)} \qquad y_H^{\perp} = y_H M_X \quad , \quad Y_H^{\perp} = Y_H M_X \quad , \quad Z^{\perp} = Z M_X \ ,$$

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 $\hat{\mathbf{\Theta}}_y$ and $\hat{\mathbf{\Theta}}_Y$ and subsequently constructing the SP-IV estimator using the alternative expression for $\hat{\beta}$ in (13). When $Z^{\perp}$ are measures of economic shocks, 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 Y_t^k 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. using the reduced form projections of the outcome variables on z_t and X_{t-1}.

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,

$$\text{(A.2)} \qquad X_t = AX_{t-1} + e_t \ .$$

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 (in "companion form"). 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

$$\text{(A.3)} \qquad X_t^{\perp}(h) = \sum_{j=0}^{h} \hat{A}^{h-j} \hat{e}_{t+j} \quad , \quad \hat{e}_t = X_t - \hat{A} X_{t-1} \ .$$

The appropriate selection of elements in $X_t^{\perp}(h)$ leads to $y_H^{\perp}$, $Y_H^{\perp}$ 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 D in $e_t = D\epsilon_t$, allowing the estimation of IRFs that are interpretable as dynamic causal effects of the associated economic shocks in ϵ_t. If $\hat{\epsilon}_t^{1:Nz}$ are the N_z identified shocks in the structural VAR, it is possible to use $z_t^{\perp} = \hat{\epsilon}_t^{1:Nz}$ to form $Z^{\perp}$ and use these shock estimates for identification in the SP-IV estimator. This procedure also nests identification with "external instruments", which can be directly included in the VAR and combined with zero restrictions in D as proposed by Plagborg-Møller and Wolf (2021), or used indirectly as instruments to identify columns in D as in the "proxy SVAR" or "SVAR-IV" approach (Mertens and Ravn 2013; Stock and Watson 2012; Stock and Watson 2018). Note that (11), or equivalently (12), are consistent estimators of the IRFs associated with $\hat{\epsilon}_t^{1:Nz}$. In finite samples, however, these IRF estimates will not be numerically identical to those obtained from $\hat{\Theta}_{X,h}^{VAR} = \hat{A}^h D^{1:Nz}$, $h = 0, \ldots, H-1$, where $D^{1:Nz}$ denotes the first N_z columns of D. The reason is that the restrictions implied by the VAR dynamics are imposed on the reduced form forecast errors, but (11) or (12) do not impose the same VAR dynamics on the IRFs. Our preferred implementation of SP-IV with structural VARs is instead to select the elements corresponding to y_t and Y_t in $\hat{\Theta}_{X,h}^{VAR}$ 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 (13). 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 $y_H^{\perp} P_{Z^{\perp}} Y_H^{\perp}$ by $\hat{\Theta}_y^{VAR} \hat{\Theta}_Y^{VAR\prime}$ and $Y_H^{\perp} P_{Z^{\perp}} Y_H^{\perp}$ by $\hat{\Theta}_Y^{VAR} \hat{\Theta}_Y^{VAR\prime}$, where $\hat{\Theta}_Y^{VAR}$ is the $HK \times N_z$ matrix stacking the K blocks of the VAR IRF coefficients of Y_t, and $\hat{\Theta}_y^{VAR}$ contains the

$H \times N_z$ VAR IRF coefficients of y_t.[14] When comfortable imposing VAR dynamics, it makes sense to impose these restrictions consistently, and we therefore recommend this second implementation for SP-IV with VARs.

B Generalized and CUE SP-IV

Using the weighting matrix $\Phi_s(\beta,\zeta) = (\Sigma_{u_H^\perp}^{-1} \otimes Q^{-1})$, where $\Sigma_{u_H^\perp}$ is the covariance of $u_{H,t}^\perp$, leads to the efficient GMM estimator of β. This estimator is also the "Generalized Least Squares" version of SP-IV minimizing $\mathrm{Tr}\left((u_H^\perp P_{Z^\perp} u_H^{\perp\prime})\Sigma_{u_H^\perp}^{-1}\right)$. Given $\Sigma_{u_H^\perp}$, the closed form generalized SP-IV estimator is

(B.1)

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

For inference, 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,

$$\text{(B.2)} \quad \sqrt{T}(\hat{\beta}_G - \beta) \xrightarrow{d} N(0, V_{\beta_G}) \quad , \quad V_{\beta_G} = \left(R' \left(\Theta_Y \Theta_Y' \otimes \Sigma_{u_H^\perp}^{-1}\right) R\right)^{-1} .$$

The Generalized SP-IV estimator is feasible replacing $\Sigma_{u_H^\perp}$ with a consistent estimator like the one in Section 2.1, using a two-step or iterated procedure. Alternatively, the CUE estimator minimizes the AR statistic in (21) with respect to b. The KLM statistic in (22) is zero at the CUE estimator, so both AR and KLM confidence sets contain the CUE.

C Proof of Proposition 2

Proof. The asymptotic variance of the SP-IV estimator in (9) is

$$\text{(C.1)} \qquad aVar(\hat{\beta}) = (\boldsymbol{\Theta}_Y'\boldsymbol{\Theta}_Y)^{-1}\boldsymbol{\Theta}_Y' \left(I_{N_z} \otimes \mathrm{var}(u_{H,t}^\perp)\right) \boldsymbol{\Theta}_Y (\boldsymbol{\Theta}_Y'\boldsymbol{\Theta}_Y)^{-1} \ ,$$

[14] To impose the VAR dynamics in the Generalized SP-IV formula (B.1), 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 (22), replace $Y_H^\perp P_{Z^\perp}$ by $\hat{\Theta}_Y^{VAR}(ZM_XZ'/T)^{-\frac{1}{2}}ZM_X$.

The asymptotic variance of the 2SLS estimator is

$$aVar(\hat{\beta}_{2SLS}) = (\boldsymbol{\Theta}_Y'\boldsymbol{\Theta}_Y)^{-1}\,\mathrm{var}(u_t)\ . \tag{C.2}$$

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 (Rothenberg and Leenders 1964).

If u_t is i.i.d., then it is unpredictable and $E[u_t^2] = E[u_t(h)^2]\ \forall h$ and $E[u_t(s)u_t(h)] = 0, s \neq h$, so $\mathrm{var}(u_{H,t}^{\perp}) = \mathrm{var}(u_t)I_H$, and part $i)$ follows.

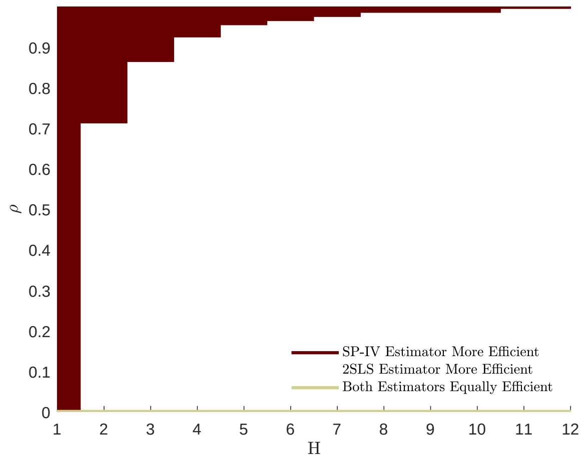
Suppose that X_{t-1} is only a constant, or uninformative; then $u_{H,t}^{\perp} = u_{H,t}$. $aVar(\hat{\beta})-aVar(\hat{\beta}_{2SLS})$ will be positive definite as long as $\mathrm{var}(u_t) = \sigma_\upsilon^2/(1-\rho_u^2) < \mathrm{maxeval}\,(\mathrm{var}(u_{H,t}))$. $\mathrm{var}(u_{H,t})$ is a matrix with h,i entry $\rho_u^{|h-i|}\sigma_\upsilon^2/(1-\rho_u^2)$. 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 $\mathrm{var}(u_t)$ when $\rho_u > 0$ and $H > 1$. Therefore, $\mathrm{maxeval}\,\mathrm{var}(u_{H,t}) > \mathrm{var}(u_t)$ when $\rho_u > 0, H > 1$, completing part $ii)$.

Finally, $aVar(\hat{\beta}_{2SLS})-aVar(\hat{\beta})$ is positive definite if $\mathrm{var}(u_t) = \sigma_\upsilon^2/(1-\rho_u^2) > \mathrm{maxeval}\left(\mathrm{var}(u_{H,t}^{\perp})\right)$, giving the first part of (iii) . If X_{t-1} spans the full history of υ_t up to $t-1$, $u_t^{\perp}(h) = \sum_{j=0}^{h}\rho_u^j\upsilon_{t+h-j}$, and the condition specializes to $\sigma_\upsilon^2/(1-\rho_u^2) > \mathrm{maxeval}\,\mathrm{var}(u_{H,t}^{\perp})$, where the h,i entry of $\mathrm{var}(u_{H,t}^{\perp})$ is $\sum_{j=1}^{\min\{h,i\}}\sigma_\upsilon^2\rho_u^{h+v-2j}$, as stated in the proposition. □

D Proof of Proposition 3

Proof. Consider the weak instruments asymptotic embedding $\Theta_Y = C/\sqrt{T}$ where C is a $HK \times N_z$ fixed matrix. When $K = 1$, the concentration parameter for 2SLS is $\mathrm{Tr}(CC')/(HN_z)/\sigma_\omega^2$, where σ_ω^2 is the variance of the first stage error term. For SP-IV without conditioning on X_{t-1}, the concentration parameter is $\mathrm{Tr}(CC')/(N_z\,\mathrm{Tr}(\Sigma_{v_H}))$, see Definition 1 in the Online Appendix. For SP-IV with conditioning on X_{t-1}, the concentration parameter is $\mathrm{Tr}(CC')/(N_z\,\mathrm{Tr}(\Sigma_{v_H^{\perp}}))$, see Definition 1 in the Online Appendix. $\mathrm{Tr}(\Sigma_{v_H})$ is larger than $\mathrm{Tr}(\Sigma_{v_H^{\perp}})$ unless X_{t-1} is completely irrelevant for predicting $Y_{t+h}, h = 0, ..., H-1$. Parts (i) and (ii) follow from the expressions for the concentration parameters.

Figure C.1: Asymptotic Efficiency of SP-IV and 2SLS



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

$\square$