Dynamic Identification Using System Projections and Instrumental Variables*

Daniel J. Lewis — University College London

Karel Mertens† — FRB Dallas, CEPR

November 29, 2022

Abstract

We propose System Projections with Instrumental Variables (SP-IV) to estimate dynamic structural relationships. SP-IV replaces lag sequences of instruments in traditional IV with lead sequences of endogenous variables. SP-IV permits identification over many time horizons and allows the inclusion of controls, which weakens exogeneity requirements and improves effective instrument strength. SP-IV also enables the estimation of structural relationships between impulse responses obtained from local projections or vector autoregressions. We provide a bias-based test for instrument strength and present inference procedures under strong and weak identification. SP-IV outperforms competing estimators of the Phillips Curve parameters in simulations. We estimate the Phillips Curve implied by the main business cycle shock of Angeletos et al. (2020), and find evidence for forward-looking behavior. The data is consistent with weak but also relatively strong cyclical connections between inflation and unemployment.

*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.

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

Equation (1) nests a very wide range of dynamic relationships of interest in macroeconomics. To illustrate the range of difficulties that can arise in the estimation of β, consider the specific example of the Hybrid New Keynesian Phillips Curve (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 the 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 natural level in the absence of 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*. In practice, both the output gap and price setters' inflation expectations are not observed directly and must be replaced with proxy measures. A second major source of endogeneity problems is *simultaneity*, since the error term typically includes structural disturbances 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 to address such challenges is to rely on dynamics for identification and use lagged economic variables as instrumental variables. In the New Keynesian 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 independent of any of the determinants of the instrumenting lagged macroeconomic variables. In other words, the shocks (and lags thereof) comprising the innovation to π_t must be unrelated to 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) recently proposed current and lagged values of direct measures of structural shocks from the literature as instrumental variables.[2] Instrument exogeneity in this case requires measures of economic shocks that are independent of both the contemporaneous and lagged macroeconomic influences on the error term u_t. In practice, however, the literature is rarely comfortable with imposing the strong assumption of (unconditional) lag exogeneity on available empirical measures of structural shocks, and typically avoids doing so by including a rich set of lagged macroeconomic controls. Unfortunately, including such controls in an IV regression with lagged shocks as instruments shrinks the explanatory power of the instrument set towards that of only the contemporaneous value of the instrumenting shocks, since lagged macro variables typically span the lagged shocks that generate them. The result of adding lagged control variables in standard IV specifications is therefore weaker or even under-identification.

In this paper, we propose a novel approach to estimating β that allows the inclusion of lagged macroeconomic variables as controls without weakening identification. Specifically, we replace the single equation (1) with an H-dimensional system in forecast errors, conditional on controls,

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

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

for leads of y_t and Y_t, where H is the number of leads. The forecast errors can be derived from a variety of forecasting models, and we consider vector autoregressive models (VARs) and local projections (LPs) because of their widespread use. We use only the contemporaneous value of the N_z instrumental variables to yield HN_z moment conditions, which we solve in closed form for β, yielding a restricted two-stage least squares (2SLS) estimator in the system of reduced form forecast errors. We refer to this estimator as System Projections with Instrumental Variables, or SP-IV.

The SP-IV approach has several conceptual advantages. First, like Barnichon and Mesters (2020), it can leverage existing identified shocks, but with suitable controls it requires only the weaker assumption of contemporaneous exogeneity, since it can be applied to forecast errors. Our approach is therefore better aligned with the identification assumptions that the literature is willing to make about empirical measures of economic shocks. 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 the effective strength of a given instrument, potentially reducing weak instruments bias. These advantages are due to our use of lead sequences of the endogenous variables y_t and Y_t, instead of lag sequences of the instrumental variables, as in existing approaches.

Barnichon and Mesters (2020) observe that single-equation 2SLS with lag sequences of shocks as instruments is equivalent to a regression of the impulse response function (IRF) of y_t on the IRFs of Y_t, where the IRFs are estimated by regressions of the endogenous variables on distributed lags (DL) of the shocks. We show that since SP-IV is based on similar moments (applied to forecast errors), it is also equivalent to a regression of the IRF of y_t on the IRFs of Y_t. However, an additional advantage of SP-IV is that these IRFs can be obtained using any valid impulse response estimator and identification scheme (e.g. internal rather than external instruments). SP-IV therefore allows the estimation of structural relationships across IRFs as they are estimated in practice, which is rarely with distributed lag specifications, but with vector autoregressions or local projections. The SP-IV inference methods we describe enable formal testing of hypotheses about structural relationships across IRFs in macroeconomic applications, while in empirical practice claims about such relationships have typically relied on more informal arguments (e.g., Angeletos et al. (2020)).

The inference methods presented in this paper make SP-IV of practical use for a wide range of settings. 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. Such a pre-test provides a convenient way to assess whether standard inference will be reliable, or to compare the identifying information contained in different specifications - perhaps across alternative sets of instruments. As instrumental variables are often weak in practice, we describe two weak-instrument-robust inference procedures, Anderson and Rubin's (1949) AR statistic and Kleibergen's (2005) KLM statistic.

To demonstrate the advantages of SP-IV, we conduct simulations estimating the parameters of the Phillips curve in (2) using data generated from the Smets and Wouters (2007) model and weak instruments. OLS in this setting is strongly biased. When the instrument is correlated with lagged inflation according to the estimated relationship in actual U.S. data, 2SLS is also prohibitively biased, while SP-IV including controls exhibits only mild biases due to instrument strength. When the true shocks are available as an instrument, SP-IV estimators with controls still exhibits smaller biases than single-equation 2SLS estimators. A VAR implementation of SP-IV has the lowest bias of all estimators we consider, while LP implementations have lower variances. Standard inference is badly over-sized for all IV estimators due to weak instruments. Both proposed robust inference procedures remain well-sized in realistic sample sizes, exhibiting small size distortions only when H is large relative to T. Single-equation 2SLS with lagged instruments shows more substantial over-rejections when H or N_z is large.

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. Based on IRFs to an MBC shock identified in a monthly VAR, SP-IV finds greater weight on future inflation than lagged inflation. Robust confidence sets for the slope are consistent with both very weak and fairly strong cyclical responses of inflation. The IRFs to the MBC shock do not necessarily support the conclusion in Angeletos et al. (2020) that inflation dynamics are disconnected from the business cycle.

Henceforth, I_N denotes the N-dimensional identity matrix, $\otimes$ the Kronecker product, $\text{Tr}(\cdot)$ the trace operator, $\text{vec}(\cdot)$ the vectorization operator, $\text{mineval}\{\cdot\}/\text{maxeval}\{\cdot\}$ the minimum/maximum eigenvalue, $E[X \mid Y]$ the conditional expectation of X given Y, $\xrightarrow{p}$ convergence in probability, $\xrightarrow{d}$ convergence in distribution, and projection matrix $P_U = U'(UU')^{-1}U$.

1 System Projections with Instrumental Variables

We begin by reformulating the dynamic relationship of interest in (1) in terms of forecast errors. Taking h-horizon leads and subtracting the expectation conditional on an N_x-dimensional vector of predictors X_{t-1} (including a constant) yields

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

where $y^{\perp}_{t+h} = y_{t+h} - E[y_{t+h} \mid X_{t-1}]$, $Y^{\perp}_{t+h} = Y_{t+h} - E[Y_{t+h} \mid X_{t-1}]$, and $u^{\perp}_{t+h} = u_{t+h} - E[u_{t+h} \mid X_{t-1}]$. Let z_t denote an $N_z \times 1$ vector of instrumental variables, and define $z^{\perp}_t = z_t - E[z_t \mid X_{t-1}]$. 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 lags $z_t, z_{t-1}, \ldots$ of the instruments z_t. Instead of the usual approach of imposing orthogonality between z_{t-h} and u_t for various $h \geq 0$, we impose

$$E[u^{\perp}_{t+h} z^{\perp}_t] = 0 \ ; \ h = 0, \ldots, H-1 \ . \tag{4}$$

Without conditioning on X_{t-1}, the orthogonality conditions in (4) are equivalent to imposing orthogonality between z_{t-h} and u_t under stationarity. The key departure approaches including lags 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 predictors X_{t-1}. As we show below, the use of forecast errors leads to at least three key advantages. Crucially, the same advantages do not arise from simply including X_{t-1} in equation (1), and proceeding with 2SLS with $z_t, \ldots, z_{t-H+1}$ as instruments and X_{t-1} as included regressors. To arrive at an estimator that *will* realize the benefits of conditioning on X_{t-1}, we next formulate the GMM problem associated with the orthogonality conditions in (4).

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 $y^{\perp}_{H,t}$ and $u^{\perp}_{H,t}$ denote the $H \times 1$ vectors in which the $(h+1)$-th element is $y^{\perp}_{t+h}$ and $u^{\perp}_{t+h}$ respectively. Let $Y^{\perp}_{H,t}$ denote the $HK \times 1$ vector stacking the $H \times 1$ vectors $Y^{k,\perp}_{H,t}$, where Y^k_t is the k-th variable in Y_t. Using this notation, the moment conditions are

$$E[u^{\perp}_{H,t}(\beta) \otimes z^{\perp}_t] = 0 \ , \tag{5}$$

where $u^{\perp}_{H,t}(b) \equiv y^{\perp}_{H,t} - (b' \otimes I_H) Y^{\perp}_{H,t}$, the truth is $b = \beta$, and $E\left[u^{\perp}_{H,t}(\beta) u^{\perp}_{H,t}(\beta)'\right] = \Sigma_{u^{\perp}_H}$.

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 (LP) and vector autoregressions (VARs), both of which are widely used in applied macroeconomics.[3] The moment conditions for this step are

$$E\left[\left[y^{\perp\prime}_{H,t}(\zeta), Y^{\perp\prime}_{H,t}(\zeta), z^{\perp\prime}_t(\zeta)\right]' \otimes X_{t-1}\right] = 0, \tag{6}$$

where $y^{\perp}_{H,t}(d)$, $Y^{\perp}_{H,t}(d)$, $z^{\perp}_t(d)$ are functions of parameters d that depend on the forecasting model chosen, and the true value of d is ζ.

The moments in (5) and (6) can be stacked in a moment function $f(y_{H,t}, Y_{H,t}, z_t, X_{t-1}; b, d)$ with $E[f(y_{H,t}, Y_{H,t}, z_t, X_{t-1}; \beta, \zeta)] = 0$. Let $W_t = (y^{\perp\prime}_{H,t},\ Y^{\perp\prime}_{H,t},\ z^{\perp\prime}_t,\ X'_{t-1})'$. The associated GMM objective function is

$$F_T(b,d) = \frac{1}{T}\left(\sum_{t=1}^{T} f(W_t; b, d)\right)' \Phi(b,d) \left(\sum_{t=1}^{T} f(W_t; b, d)\right) \ , \tag{7}$$

[3] 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, April).

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. We make an additional assumption to ensure 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.*

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 $z_t^{\perp}$.[4] The resulting solution to (7) for β in population is

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

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 $y_{H,t}^{\perp}$, $Y_{H,t}^{\perp}$, and $z_t^{\perp}$ respectively. The natural sample analog of (8) is

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

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

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

[4] Alternatively, the efficient GMM estimator uses the weighting matrix $\Phi_s(\beta,\zeta) = (\Sigma_{u_H^{\perp}}^{-1} \otimes Q^{-1})$. This estimator naturally leads to a 'Generalized Least Squares' version of SP-IV, see Appendix B.

after projection on the instruments $Z^{\perp}$. Thus, $\hat{\beta}$ is also the restricted 2SLS estimator in the system of equations (10). We thus refer to $\hat{\beta}$ as the **S**ystem of equations after **P**rojection on **I**nstrumental **V**ariables (SP-IV) estimator. The only restrictions in (10) are those implied by (1).

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

$$\hat{\Theta}_Y = \frac{Y_H^{\perp} Z^{\perp\prime}}{T} \left(\frac{Z^{\perp} Z^{\perp\prime}}{T} \right)^{-\frac{1}{2}} \quad ; \quad \hat{\Theta}_y = \frac{y_H^{\perp} Z^{\perp\prime}}{T} \left(\frac{Z^{\perp} Z^{\perp\prime}}{T} \right)^{-\frac{1}{2}} , \tag{11}$$

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

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

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

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

which shows $\hat{\beta}$ is interpreted as the slope in OLS regression of $\hat{\mathbf{\Theta}}_y$ on $\hat{\mathbf{\Theta}}_Y$, or coefficients in 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 impulse response functions using instruments satisfying the exogeneity conditions. 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 in (2), for example, 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 SP-IV estimator is obtained regressing the IRF of the outcome variable y_t on the IRFs of the endogenous variables Y_t. For the Phillips curve, 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 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 can estimates 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 base SP-IV on a subset of horizons rather than all $h = 0, \dots, H-1$, and SP-IV can be based on any subset of horizons.

1.3 The Difference Between SP-IV and Regular IV

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

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

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 H 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 interpretdee] as a regression in impulse response space. In 2SLS the regression uses IRFs estimated by 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 LP or VAR specifications in which the h-step ahead forecasts of y_t and Y_t given z_t can be conditioned on a set of additional predictors, X_{t-1}.

A preliminary advantage of SP-IV is that it estimates structural relationships across IRFs as they are estimated in practice – using LPs or VARs, not DL models. The 2SLS approach also requires external measures of economic shocks. Such external measures can be used for identification in LPs and VARs (Jordà 2005; Stock and Watson 2012; Mertens and Ravn 2013), but IRFs for SP-IV can also exploit internal instruments (recursivity assumptions) or other short or long-run restrictions on forecast error variances. Thus, SP-IV greatly expands the econometrician's options.

SP-IV's ability to accommodate controls yields three theoretical advantages. To exposit these advantages, we adopt the Slutsky-Frisch paragdigm to express y_t and Y_t in terms of current and past realization of '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 additionally assume stationarity throughout. Let ω_t be the error term in the first stage of 2SLS using $z_t, z_{t-1}, \ldots, z_{t-H+1}$, with variance σ_ω^2. Denote the $H \times 1$ vector of errors in the SP-IV first stage regression of $Y_{H,t}$ on z_t as $v_{H,t}$, with covariance Σ_{v_H}, and, in the first stage regression of $Y_{H,t}^\perp$ on $z_t^\perp$ (conditional on controls X_{t-1}) as $v_{H,t}^\perp$, $\Sigma_{v_H^\perp}$.

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

Theorem 1. *The exogeneity condition for 2SLS with lags of z_t requires*

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

If X_{t-1} spans all $\epsilon_{t-l} : \mu_l \neq 0$ 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}. □

The 2SLS conditions require that z_t is uncorrelated with those past, contemporaneous, or future shocks entering u_t. Following Stock and Watson (2018), we denote 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 types of exogeneity conditions to hold.[5] 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+h}. With a sufficiently rich set of predictors, the exogeneity conditions on z_t are thus substantially weaker. 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 satisfies the exogeneity requirements for 2SLS estimation of the Phillips

[5] 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.

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 echos the invertibility assumption in VARs and the practice of including lagged controls in LPs to avoid lag exogeneity requirements (Stock and Watson 2018; Ramey 2016). 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. [6] In practice, a richer set of predictors offers better insurance against violations of the lag exogeneity assumption.

Finally, 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 will realize any of the other advantages of SP-IV.[7]

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 this is the case 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 2SLS if the vari-

[6] In 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 $\upsilon_{t-1}, \upsilon_{t-2}, \ldots$, however, z_t^{RR} remains a valid instrument for SP-IV including on X_{t-1}.

[7] Another drawback is such a procedure requires modifications to standard IV inference and pretest formulas and software, which assume that both the IV stages contain the same set of controls.

ances of forecast errors $u^{\perp}_{t+h}$ at $h = 0, ..., H-1$ are small relative to the variance of the error term u_t. The ranking of estimators depends on the data generating process. We consider an AR(1) illustrative model,

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

Theorem 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 $u^{\perp}_{H,t} = u_{H,t}$ 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 $\sigma_u^2 > \text{maxeval}(\Sigma_{u_H^{\perp}})$. If u_t follows the AR(1) process in (18) and X_{t-1} spans past shocks, then the condition becomes $\frac{\sigma_\upsilon^2}{1-\rho_u^2} > \text{maxeval}\left(\Sigma_{u_H^{\perp}}\right)$, where the h, i entry of $\Sigma_{u_H^{\perp}}$ is given by $\sum_{j=1}^{\min\{h,i\}} \sigma_\upsilon^2 \rho_u^{h+i-2j}$.

When u_t is i.i.d., the errors in both estimators are identical in population since X_{t-1} does not predict $u_t, \ldots u_{t+H-1}$ and forecast errors do not accumulate over $h = 0, \ldots, H-1$; 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 $\upsilon_{t-1}, \upsilon_{t-2}, \ldots$ on errors, SP-IV is asymptotically more efficient as long as ρ_u is sufficiently large and the forecast horizon H is not too large. Figure C.1 in the Online Appendix shows the region where SP-IV is more efficient. The greater the persistence, the more predictive power X_{t-1} contains, but the longer the horizon, the harder it becomes to predict u_t, so improvements due to SP-IV are more likely when u_t is highly persistent – often true in macroeconomic applications – and the length of the 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 bias 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.

Theorem 3. *When $K = 1$, for a given set of instruments and horizons,*

(i) If $\mathrm{Tr}(\Sigma_{v_H^\perp})/H < \sigma_\omega^2$, the concentration parameter for SP-IV is larger than that for 2SLS;

(ii) 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.

We derive the concentration matrix for SP-IV and prove the Theorem 3 in the Online Appendix, where we also show that, as in Stock and Yogo (2005) and Lewis and Mertens (2022), the worst case weak instruments bias is decreasing in its minimum eigenvalue. Thus, Theorem 3 shows that when the controls have explanatory power for the endogenous regressors their inclusion increases the effective strength of the instruments in SP-IV, relative to SP-IV without controls and 2SLS, and therefore decreases bias. When $K > 1$, the concentration parameter depends on the entire eigenstructure of the first stage parameters (and that of $\Sigma_{v_H^\perp}$), so a fully general result is not possible, although intuitively it will hold for many DGPs. Whether the effective instrument strength increases relative to 2SLS is application specific, and depends on the persistence and predictability of the errors as well as H, just as efficiency did in the previous section. As 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.

In contrast, adding X_{t-1} as additional regressors in both stages of 2SLS with $z_t, \dots, z_{t-H+1}$ as instruments weakens identification. 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 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 completely irrelevant 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, the inclusion of X_{t-1} will still shrink the explanatory power of $z_t, \dots, z_{t-H+1}$ towards that of z_t.

1.4 Consistency of the SP-IV Estimator

We conclude the discussion of the SP-IV estimator by presenting conditions for its consistency for β in (1). We make the following high-level assumptions on covariances.

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

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

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

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

(2.d) $R'(\Theta_Y\Theta_Y' \otimes I_H)R$ *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 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 definiton 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.

The conditions in Assumption 2 resemble the usual (strong) IV assumptions, see for instance Stock and Yogo (2005). However, condition 2.d does not require there to be at least as many instruments as endogenous regressors, $N_z \geq K$. Since $\text{rank}(R'(\Theta_Y\Theta_Y' \otimes I_H)R) = \min\{K, H\,\text{rank}(\Theta_Y\Theta_Y')\}$, the order condition is $HN_z \geq K$, since there are HN_z moment conditions in (5). Adding leads can thus make up for $N_z < K$ just as adding lags does so in single-equation 2SLS with lag sequences as instruments. Proposition 1 states the consistency result for the SP-IV estimator in (9).

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

Proof. Both terms in (9) converge by the 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 2, inference for SP-IV can proceed analogously to standard 2SLS. With a further high-level assumption, we can derive a limiting distribution of $\hat{\beta}$:

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 2. *Under Assumptions 1-3,*

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

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

Proof. The result is immediate after rearranging (9) from Proposition 1, the stated assumptions, and 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.[8] A natural consistent estimator is $\hat{\Sigma}_{u_H^{\perp}} = \hat{u}_H^{\perp}\hat{u}_H^{\perp\prime}/(T-N_x-K)$. A choice of X_{t-1} including adequate lags obviates the need for an autocorrelation robust estimate by eliminating autocorrelation in both $u_t^{\perp}$ and $z_t^{\perp}$. Any mechanical correlation between $u_t^{\perp}$ and $u_{(t-h)+h}^{\perp}$, say, drops out of $\operatorname{var}(u_{H,t}^{\perp} \otimes z_t^{\perp})$, since when $z_t^{\perp}$ is serially uncorrelated, so too is $u_{H,t}^{\perp} \otimes z_t^{\perp}$. This is not the case for 2SLS, which generally requires HAR methods.

2.2 A Test for Weak Instruments

Available instruments may be weak in many applications. Then, Wald inference will be invalid, leading to empirical rejection rates that generally

[8] 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.

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, extending the test in Lewis and Mertens (2022). We consider a Nagar approximation of the bias under weak instrument asymptotics, as in Montiel-Olea and Pflueger (2013). Like Stock and Yogo (2005) and Lewis and Mertens (2022), we use a weighted ℓ_2-norm of the worst-case bias relative to a benchmark to accommodate cases with more than one endogenous regressor ($K>1$). Weak instruments are defined as those for which the bias in $\hat{\beta}$ is τ percent of the benchmark or larger 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 Online Appendix provides a step-by-step description of the testing procedure, implemented in the accompanying Matlab Code.

2.3 Weak-Instrument Robust Inference for SP-IV

We describe two test statistics for SP-IV, robust to weak instruments.

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

$$(20) \qquad AR(b) = (T - d_{AR}) \operatorname{Tr}\left(u_H^{\perp}(b) P_{Z^{\perp}} u_H^{\perp}(b)' \left(u_H^{\perp}(b) M_{Z^{\perp}} u_H^{\perp}(b)' \right)^{-1} \right),$$
$$AR(\beta) \xrightarrow{d} \chi^2_{HN_z} \ ,$$

where $M_{Z^{\perp}} = I_T - P_{Z^{\perp}}$ is the residualizing matrix , $d_{AR} = N_z + N_x$ 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 $HN_z > K$: when the number of IRF coefficients exceeds the number of endogenous regressors.

As this is likely in practice, we consider the KLM statistic proposed in Kleibergen (2002), which can improve power (e.g., Andrews et al. (2019)).

Since SP-IV is a GMM estimator, following (Kleibergen 2005),

$$
\begin{aligned}
K(b) &= (T - d_K)\,\mathrm{vec}\left(\Xi^{-1} u_H^{\perp}(b) \check{Y}_H'\right)' R \\
&\quad \times \left(R'(\check{Y}_H \check{Y}_H' \otimes \Xi^{-1} u_H^{\perp}(b) u_H^{\perp\prime}(b) \Xi^{-1}) R\right)^{-1} \\
&\quad \times R'\,\mathrm{vec}\left(\Xi^{-1} u_H^{\perp}(b) \check{Y}_H'\right)' \,, \\
K(\beta) &\xrightarrow{d} \chi^2_K \,,
\end{aligned}
\tag{21}
$$

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

We now evaluate the performance of SP-IV in simulations to demonstrate its practical advantages relative to 2SLS.

The objective in all simulations is to estimate the parameters of the Phillips Curve in (2) using data generated from the workhorse macroeconomic model of Smets and Wouters (2007) (hence SW).[9] 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 Smets and Wouters (2007) 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

[9] 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.

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 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

$$u_t = \rho_u u_{t-1} + \epsilon_t^p - \mu_p \epsilon_{t-1}^p \ , \ \rho_u = 0.99, \ \mu_p = 0.83 \tag{22}$$

where ϵ_t^p is an i.i.d. normally distributed price markup shock.[10] Inverting the autoregressive term in (22) 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 either violate the lag exogeneity requirement, but lose relevance if the data is first conditioned on predetermined variables to avoid the lag exogeneity requirement.

Because lagged endogenous variables are not valid instruments, we consider a measure of the monetary policy shocks 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. We conduct further simulations with multiple model shocks as instruments and discuss the performance of the generalized (or efficient GMM) version of SP-IV, with these additional

[10]We assume that the econometrician cannot exploit the ARMA(1,1) error structure in (22).

results largely relegated to the Online Appendix.

In our simulations, 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 assumed to be 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 we use satisfy lead exogeneity.

3.1 Simulations with Violations of Lag Exogeneity

Our first set of simulations demonstrates how SP-IV provides can help ensure exogeneity by conditioning on lagged macroeconomic variables. We are motivated by identification of Phillips curve parameters, 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 around FOMC meetings as in Kuttner (2001). A practical concern with such measures is that, despite careful construction, they may still contain a meaningful predictable component (Ramey 2016; Coibion 2012; Barakchian and Crowe 2013; Miranda-Agrippino and Ricco 2021; Bauer and Swanson 2022). Consequently, researchers identifying IRFs to monetary policy shocks 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 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

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			
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			
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	48.60	94.30
WALD 2SLS-C		
AR 2SLS	73.10	71.40
AR 2SLS-C		
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 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 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 remaining rows the means estimates across 5000 Monte Carlo samples. All IV estimators are based on $h = 0, ..., H - 1$ and use 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. In the right panel, tests for 2SLS use a HAR variance matrix following Lazarus et al. (2021); inference procedures for SP-IV are described in Section 2.

log change in the GDP deflator, used to estimate the SW model, over the sample 1969-2004. 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 for 2SLSor the SP-IV estimator without controls. However, the simulated instruments are exogenous conditional are 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 thes second. The remaining rows report results for 2SLS regression with H lags of the monetary policy instrument, 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}, for specifications with $H = 8$ and $H = 20$.

Unsurprisingly, 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 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. In contrast, both procedures produce mean estimates with the correct sign and 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 miss-specified, or because weak instruments 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 various SP-IV inference procedures described in Section 2 and analogous heteroskedasticity and autocorrelation robust (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 and SP-IV LP) is badly oversized. Conversely, for the estimators that condition on X_{t-1}, (SP-IV LP-C and SP-IV VAR), the robust AR and KLM tests, defined in (20) and (21) 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 test for these estimators remains somewhat oversized, especially when $H = 20$. These results indicate that the residual bias is primarily related to the weakness of the instruments, even in a relatively large sample, since robust inference procedures effectively control size.

The results in Table 1 illustrate the advantage of weakening the exogeneity condition by using SP-IV with lagged controls instead of 2SLS, as we argued in Section 1.3. 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 weak-instrument problems are a serious practical concern in our setting, 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

Our next set of simulations investigates the relative performance of SP-IV and 2SLS in more realistic sample sizes. Given the limited role of monetary policy shocks for inflation dynamics in the Smets and Wouters (2007) model, estimating the parameters of the Phillips curve using monetary policy shocks as instruments is especially challenging in small samples. The main goal of these 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 scenario 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.

Bias. We first discuss the bias of the estimators. 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,

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	0.35	0.67	0.00	0.29	0.75	0.00	0.22	0.87	-0.06
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	0.40	0.57	0.00	0.37	0.64	0.00	0.25	0.83	-0.04
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 the lag endogenous monetary policy instruments described in the text. 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.

OLS is severely biased regardless of T due to endogeneity. The next five rows show the results for various 2SLS and SP-IV estimators with $H = 8$ quarters, and the bottom five rows show results with $H = 20$ quarters.

We focus first on the performance of the IV estimators for the specifications with $H = 8$. 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 exogenous in this exercise, 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 sample, $T = 250$. The Phillips Curve slope, λ, is estimated 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 for two of the three parameters for $T = 250$. Moreover, the bias for λ grows *worse* as T grows larger, with the mean estimate having the wrong sign for $T = 5000$. The next row shows that, without controls, the bias of SP-IV is almost identical to that of 2SLS for all T. This is again unsurprising, as both exploit essentially the same identifying moments.

The next two rows 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 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 or the magnitude of 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 very weak in the SW DGP. Similar strengthening of identification is not possible by conditioning on X_{t-1} in 2SLS. As discussed, such conditioning on lagged variables tends to magnify rather than mitigate the weak-instrument problems of 2SLS with lagged shocks as instruments, as the 2SLS-C results show.

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. However, the advantage of including controls in SP-IV diminishes with the forecast horizon H since predibtability based

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 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.

on time t information falls. The final panel in Table 2 shows results for $H = 20$ instead of $H = 8$. The relative performance of the estimators is qualitatively unchanged. 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, 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. We next evaluate the variance of the estimators across Monte Carlo samples. Table 3 reports standard deviations of the various estimators, with results for $H = 8$ quarters in the top panel and $H = 20$ in the bottom. We omit OLS and 2SLS-C for brevity, given their poor performance in terms of bias.

We showed earlier in Section 1.3 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 (22), similar efficiency gains can arise. Table 3 indeed shows demonstrates 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 realize 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 larger 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 higher variance. That the VAR implementation has smaller bias but greater variance may be surprising given conventional wisdom on 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 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, the standard deviations for all estimators in Table 3 are decreasing in H, indicating that additional horizons generally reduce the variability of all estimators. Given our bias results, this means there is also a bias-variance trade-off when choosing the maximum horizon H for SP-IV, since larger H provides smaller bias improvements relative to 2SLS,

[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, April). In the Online Appendix, we show that this trade-off is also present for the IRFs in our simulations.

while also generating more efficient estimates.

Inference. Lastly, we evaluate various inference procedures. Given the monetary policy shocks are weak instruments, one key question is how severe size distortions are using standard Wald inference, which is only valid under strong identification. The other main question is how well the weak-instrument-robust procedures control size in practice. It is well known that these 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). As before, we consider Wald, AR and KLM tests for SP-IV and HAR versions for 2SLS.

Table 4 reports empirical rejection rates for nominal 5% tests of the true values of the full parameter vector, $\beta = [\gamma_b, \gamma_f, \lambda]'$. We again consider sample sizes of $T = 250, 500$ and 5000. We also reports results for $N_z = 3$ with $H = 20$, see Section B.2 for details. Any size distortions will generally decrease with T, since we keep the first-stage relationships fixed across specifications, and identification strength therefore improves with T.

The first row in Table 4 shows that Wald tests for 2SLS exhibit meaningful size distortions for $H = 8$, with empirical rejection rates substantially above the nominal level, 5%. The size distortions become very significant for $H = 20$, with rejection rates as high as 65 percent for $T = 250$. These distortions are unsurprising given the weakness of the monetary policy shocks as instruments, and demonstrates the necessity of robust inference procedures. For the specifications with $H = 8$, the 2SLS AR test in the second row is relatively well-sized in small samples. However, for the specifications with $H = 20$ the 2SLS AR test becomes noticeably oversized in small samples, dramatically so for $N_z = 3$, which is symptomatic of many-weak-instrument problems.

The next three rows in Table 4 consider Wald tests for the three SP-IV estimators. Just like 2SLS, the Wald size distortions for $H = 8$ are

Table 4: Empirical size of nominal 5% tests

$T =$	$H = 8,\ N_z = 1$ 250	500	5000	$H = 20,\ N_z = 1$ 250	500	5000	$H = 20,\ N_z = 3$ 250	500	5000
WALD 2SLS	13.3	11.1	12.6	65.4	59.7	42.6	99.9	99.7	91.6
AR 2SLS	6.8	6.1	4.1	12.1	8.1	4.3	56.0	26.0	3.9
WALD:									
SP-IV LP	15.2	12.7	12.9	68.5	64.1	43.8	99.9	99.9	91.7
SP-IV LP-C	13.0	11.6	7.8	72.0	63.5	29.7	100.0	99.7	76.6
SP-IV VAR	7.8	7.1	5.5	32.2	27.4	13.1	86.1	76.5	53.8
AR:									
SP-IV LP	5.7	5.7	4.6	9.7	7.1	4.9	14.6	8.7	4.9
SP-IV LP-C	6.7	5.8	4.9	11.4	7.7	5.0	17.3	9.9	5.2
SP-IV VAR	4.6	4.6	4.9	5.3	5.8	4.8	6.3	5.9	4.7
KLM:									
SP-IV LP	5.6	5.5	4.8	8.3	6.2	4.7	8.2	6.3	5.3
SP-IV LP-C	6.9	6.0	5.3	11.9	7.1	4.8	11.3	7.9	5.1
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 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. Tests for 2SLS use a HAR variance matrix following Lazarus et al. (2021); inference procedures for SP-IV are described in Section 2.

substantial, and they are very large for $H = 20$. The remaining rows consider the AR and KLM tests for SP-IV. The SP-IV AR tests are well-sized overall. However, 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 KLM tests in the final three rows are also generally well-sized. Just as the SP-IV AR tests, however, the KLM tests exhibit some over-rejection in small samples when $H = 20$. The size distortions of the robust SP-IV tests due to "many moments" problems are milder than those for robust 2SLS tests, especially those for the VAR implementation.

The simulation results in Table 4 confirm the severity of the weak instruments problems associated with standard Wald inference. Our proposed robust inference procedures for SP-IV generally control size distortions induced by weak instruments, at least as well as similar ones for

2SLS. That said, just as with 2SLS, it is important to avoid specifications with very large HN_z when the instruments are weak.

Fortunately, there is no need to use all horizons for identification in practice. 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-weak-instrument problems, see for example Mikusheva (2021, February) 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-weak 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 eq2 using U.S. data and compare our results with those from 2SLS. We consider the following specification for quarterly inflation at a monthly frequency,

$$\pi_t^{1q} = (1-\gamma_f)\pi_{t-3}^{1y} + \gamma_f \pi_{t+12}^{1y} + \lambda U_t + u_t \ , \tag{23}$$

where π_t^{1q} is the annualized percent change in the Core CPI from a quarter ago in month t, π_t^{1y} is the percent change in the Core CPI over the preceding year in month t, and U_t is the headline unemployment rate in month t. The variable definitions in terms of quarterly and annual lagged and future inflation rates and unemployment as the gap measure are identical to those in Barnichon and Mesters (2020), but we estimate (23) using monthly data instead of quarterly data. As is common in the literature (e.g., Mavroeidis et al. (2014)), (23) restricts the coefficient on lagged and future inflation to sum to one, $\gamma_b + \gamma_f = 1$. This restriction implies there is no long run trade-off between unemployment and inflation.

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

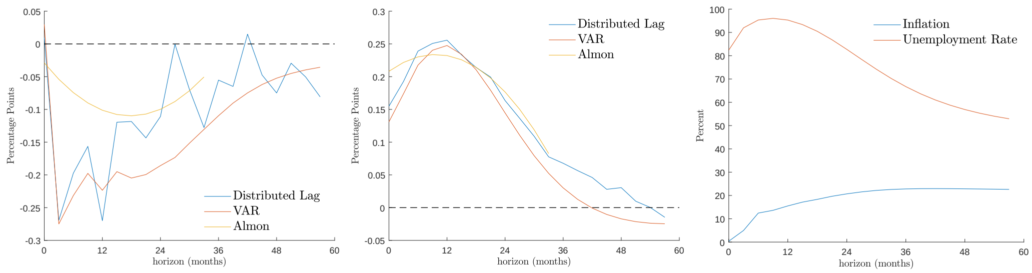
In principle, there is a range of economic shock measures that could be used to identify the parameters of (23), including monetary policy shocks as in Barnichon and Mesters (2020). However, the most useful instruments have strong predictive power for the endogenous variables, while still satisfying the exogeneity requirements. High frequency monetary policy shocks are far too weak predictors of unemployment and inflation to be useful in practice. The same is often the case for monetary shocks identified through timing restrictions or the narrative measures of Romer and Romer (2004). Moreover, contractionary policy shocks identified by these last two methods robustly generate puzzling expansionary effects in updated samples, calling into question their interpretation and contemporaneous exogeneity.[12]

Angeletos et al. (2020) find that the MBC shock obtained by maximizing the contribution to cyclical unemployment fluctuations is interchangeable with shocks identified by maximizing the cyclical variance contribution to other major macro aggregates, such as GDP, consumption, investment, or hours worked. This interchangeability suggests a single main driver of business cycles with a common propagation mechanism.

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

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}). Results in red are obtained from a six-variable monthly 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 MBC shock maximizes the contribution to the variance of the unemployment rate at horizons of 18 to 96 months in the frequency domain, as in Angeletos et al. (2020). The effective sample period is 1979:M1 to 2018:M4. Blue lines show results from regressions on lag sequences of the MBC shock, as in 2SLS.

The authors argue this main driver best fits the notion of an aggregate demand shock, making it potentially a good instrument for estimating the Phillips curve. Indeed, observing the disconnect between the unemployment and inflation responses to the MBC shock, Angeletos et al. (2020) conclude that the Phillips curve must be nearly flat, and suggest that demand-driven business cycles are perhaps not tied to nominal rigidities at all. Rather than relying on casual inspections of the IRFs, SP-IV provides a formal investigation of such claims.

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

This illustrates the apparent disconnect between inflation and the shock explaining most variation in unemployment at business cycle frequencies.

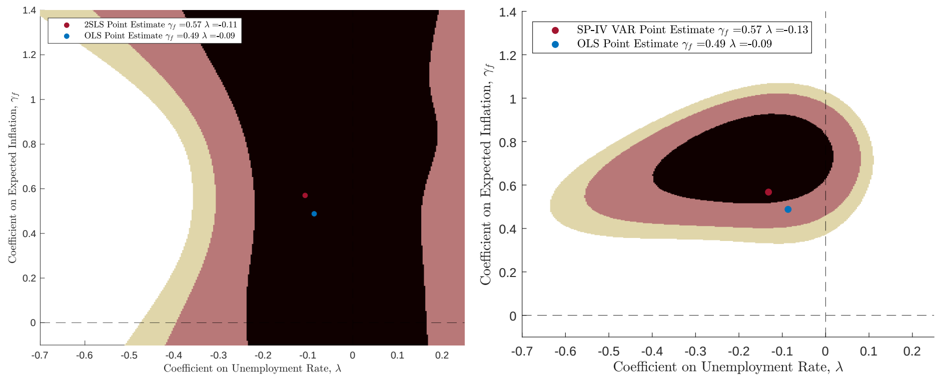
By using the MBC shock as the instrument the 2SLS and SP-IV estimators each produce estimates of the Phillips curve parameters connecting the IRFs associated with an MBC shock. The 2SLS estimator uses contemporaneous and lagged values the MBC shocks as the instrumental variables. The SP-IV estimator uses the contemporaneous MBC shock as a single instrument in a system of forecast errors - in particular, the forecast errors implied by the underlying VAR.[13] Each first stage leads to different estimators of the IRFs associated with the MBC shock. Figure 1a-1b shows the different IRFs that implicitly underlie the estimates of the parameters γ_f and λ in each case. The 2SLS estimator is built from the IRF coefficients obtained from regressions of π_t^{1q} (and π_{t-3}^{1y} and π_{t+12}^{1y}) and U_t on a DL of the shock (blue lines). SP-IV, in contrast, allows the direct use of the VAR-based IRFs (red lines). To make efficient use of the identifying information contained in the IRF dynamics, we use the coefficients in the first month of the first 12 quarters of the response horizons – that is at $h = 0, 3, 6, \ldots, 33$ – to construct each estimator. Figures 1a-1b show the first twelve IRF coefficients that are used in practice in the estimation, and also show the next eight quarters of the VAR and DL IRF coefficients to visualize the full dynamics following an MBC shock.

Figure 2 displays the estimates of γ_f and λ, together with 68%, 90% and 95% confidence sets. The point estimates of γ_f, the weight on future inflation, are 0.57 for both 2SLS and SP-IV. SP-IV estimates a steeper slope of the Phillips curve ($\lambda = -0.13$ to -0.11). However, the inference results are far less similar. For 2SLS, the confidence sets shown in Figures 2a are the AR sets used in Barnichon and Mesters (2020). They do not reject weights on future inflation as low as zero or as high as one at any level, nor are they able to rule out a wide range of possible Phillips curve slopes. The 90% set includes values of λ as high as 0.2 and as low -0.3, and the 95% set includes an even wider range for λ. As is well known, and evident in our simulations earlier, robust inference for 2SLS can be

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

Figure 2: 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.

unreliable in the face of many weak instruments problems.

Turning to inference for SP-IV, we first apply the first stage test in Section 2.2 to assess instrument strength. The test statistic is 7.76, while the 5% critical value associated with the null hypothesis that the bias does not exceed 10% percent of the worst-case bias is 21.88. Hence, we cannot reject that the MBC shock is a weak instrument, and therefore conduct robust inference. Figure 2b shows robust confidence sets for the SP-IV estimator based on the KLM statistic. The simulation evidence in Section 3 established favourable performance of SP-IV VAR with KLM inference in small and large samples in both absolute terms and compared with the other approaches, particularly when $N_z = 1$ and H is not too large, as is the case here. We therefore view the KLM sets in Figure 2b, where $N_z = 1$ and $H = 12$, as the most reliable for inference. Compared with the 2SLS approaches, inference for SP-IV is much sharper for the weight on future inflation, with the confidence set ruling out values of γ_f that are meaningfully below 0.4 or above 1. At the same time, the KLM sets also do not rule a wide range of possible Phillips curve slopes, with values of λ ranging from -0.5 to slightly greater than zero within the 90% set. The more informative confidence sets likely result from the greater effective strength of the instruments discussed in Section 1.3.

In the above application to the Phillips curve, single-equation 2SLS and SP-IV yield similar point estimates. However, only for SP-IV was robust inference well-sized in simulations across all specifications. In other applications the 2SLS and SP-IV estimates may differ more substantially, as in practice IRFs from VARs or LPs do not always agree so closely with those from DL specifications without controls. Moreover, the weaker exogeneity requirements of SP-IV can also be important, since the IRFs estimated by the DL 2SLS model may not always be identified. A key advantage of SP-IV is that it enabled us to fit structural equations directly to the IRFs obtained from VARs (or LPs), a method preferred by researchers in practice. The SP-IV methodology provides a way to formally test claims about structural relationships encoded in these empirical impulse responses. On the basis of the estimated inflation dynamics to the MBC shock of Angeletos et al. (2020), our SP-IV inference suggests a greater weight on future inflation than on lagged inflation, and the confidence sets are consistent with a wide range of possible cyclical responses of inflation, both weak and relatively strong. The evidence therefore does not necessarily support the conclusion that inflation dynamics are disconnected from the business cycle, as spurred by MBC shocks, or that the Phillips curve is of little use to model these dynamics.

5 Concluding Remarks and Future Research

We conclude by discussing several other potentially interesting applications and avenues for future research. SP-IV should be useful for estimating a wide variety of structural relationships in macroeconomics, such as Euler equations for consumption or investment, the wage Phillips curve, monetary or fiscal policy rules, 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.

In this paper, we have taken the selection of horizons as given. Future work can develop methods to optimally select the horizons. If $h =$

$0, ..., H-1$ indexes cross-sectional groups rather than time horizons, then this paper also describes instrumental variables in the cross-section with heterogeneity in the first stage. Our methodology could be extended to panel data settings, and be potentially useful in applications that commonly rely on lagged variables as instruments, such as the estimation of production functions in industrial organization, see Wooldridge (2009). We plan to pursue these and other avenues in future research.

References

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

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*.

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

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

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

Galí, J., & Gambetti, L. Has the u.s. wage phillips curve flattened? a semi-structural exploration (G. Castex, J. Galí, & D. Saravia, Eds.). In: ed. by Castex, G., Galí, J., & Saravia, D. Vol. 27. Changing Inflation Dynamics, Evolving Monetary Policy. Central Bank of Chile, 2020.

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.

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.

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

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.

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

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

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

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.

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

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

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

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.

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

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.

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

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

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

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

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.

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.

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

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

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

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.

Lazarus, E., Lewis, D. J., & Stock, J. H. (2021). The size-power tradeoff in har inference. *Econometrica*, *89*(5), 2497–2516.

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

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

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

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

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

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

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, *104*(3), 112–114.

Stock, J. H. Lecture 7: Recent developments in structural var modeling. In: Presented at the National Bureau of Economic Research Summer Institute Minicourse: What's New in Econometrics: Time Series, Cambridge, MA, July 15, 2008.

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

Appendix

A Practical Implementation of SPIV with LPs or VARs

This section describes the implementations of SP-IV with Jordà (2005) local projections or vector autoregressive models (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 $k = 1, \dots, K$, 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, and the projection matrix $P_X = X'(XX')^{-1}X$ and residualizing matrix $M_X = I_T - P_X$. Using a direct forecasting approach, the forecast errors after projection on X_{t-1} are given by

$$y_H^\perp = y_H M_X \quad , \quad Y_H^\perp = Y_H M_X \quad , \quad Z^\perp = Z M_X \ , \tag{A.1}$$

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 $\mathcal{Y}_t$ and controls X_{t-1} using z_t as instruments, a procedure often referred to as 'LP-IV'. Such IRFs can be used for identification in the SP-IV estimator exactly as described above, i.e. 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, which for the present development only we assume does not include a constant, and that X_t evolves according to a VAR,

$$X_t = AX_{t-1} + e_t \ . \tag{A.2}$$

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

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

The appropriate selection of elements in $X^{\perp}_{t+h}$ leads to $y^{\perp}_H$, $Y^{\perp}_H$ and $Z^{\perp}$, which can be used to obtain the SP-IV estimator $\hat{\beta}$ in (9). 'Structural' VARs are VARs in which researchers make assumptions to identify columns of B in $e_t = B\epsilon_t$, allowing the estimation of IRFs that are interpretable as dynamic causal effects of the associated economic shocks in ϵ_t. If $\hat{\epsilon}_t^{1:Nz}$ 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 B, or used indirectly as instruments to identify columns in B as in the 'proxy SVAR' or 'SVAR-IV' approach (Stock and Watson 2018; Stock and Watson 2012; Mertens and Ravn 2013; Stock 2008). 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 the structural VAR impulse responses obtained from $\hat{\Theta}_{X,h}^{VAR} = \hat{A}^h B^{1:Nz}$, $h = 0, \ldots, H-1$, where $B^{1:Nz}$ denotes the first N_z columns of B. The reason is that the restrictions implied by the VAR dynamics are imposed on the reduced form forecast errors, but (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 in practical applications of SP-IV with VAR-based IRFs.

[14] To impose the VAR dynamics in the Generalized SP-IV formula (B.1), replace $y_H^{\perp} P_{Z^{\perp}}$ by $\hat{\Theta}_y^{VAR} (Z M_X Z'/T)^{-\frac{1}{2}} Z M_X$ and to construct $\check{Y}_H$ in the KLM statistic in (21), replace $Y_H^{\perp} P_{Z^{\perp}}$ by $\hat{\Theta}_Y^{VAR} (Z M_X Z'/T)^{-\frac{1}{2}} Z M_X$.

B Generalized SP-IV

Using the weighting matrix $\Phi_s(\beta,\zeta) = (\Sigma^{-1}_{u^\perp_H} \otimes Q^{-1})$ leads to the efficient GMM estimator of β. This estimator is also the 'Generalized Least Squares' version of SP-IV minimizing $\mathrm{Tr}\left((u^\perp_H P_{Z^\perp} u^{\perp\prime}_H)\Sigma^{-1}_{u^\perp_H}\right)$. Given $\Sigma_{u^\perp_H}$, the closed form generalized SP-IV estimator is

(B.1)
$$\hat{\beta}_G = \left(R'\left(Y^\perp_H P_{Z^\perp} Y^{\perp\prime}_H \otimes \Sigma^{-1}_{u^\perp_H}\right)R\right)^{-1} R'\left(Y^\perp_H P_{Z^\perp} \otimes \Sigma^{-1}_{u^\perp_H}\right)\mathrm{vec}(y^\perp_H P_{Z^\perp}) \; .$$

For inference, we replace Assumption 2.d by

Assumption 2.d′. *$R'(\Theta_Y \Theta'_Y \otimes \Sigma^{-1}_{u^\perp_H})R$ is a fixed matrix with full rank.*

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

$$\text{(B.2)} \quad \sqrt{T}(\hat{\beta}_G - \beta) \xrightarrow{d} N(0, V_{\beta_G}) \;\; , \;\; V_{\beta_G} = \left(R'\left(\Theta_Y \Theta'_Y \otimes \Sigma^{-1}_{u^\perp_H}\right)R\right)^{-1} .$$

The Generalized SP-IV estimator is feasible replacing $\Sigma_{u^\perp_H}$ 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 (20) with respect to b. The KLM statistic in (21) is zero at the CUE estimator, so both AR and KLM confidence sets contain the CUE.

C Proof of Theorem 2

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

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

The asymptotic variance of the 2SLS estimator is

$$\text{(C.2)} \qquad aVar(\hat{\beta}_{2SLS}) = (\mathbf{\Theta}'_Y \mathbf{\Theta}_Y)^{-1} \mathrm{var}(u_t) \; .$$

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 $E[u^2_t] = E[u^2_{t+h}]$ $\forall h$ and $E[u_{t+s}u_{t+h}] = 0, s \neq h$, so $\mathrm{var}(u^\perp_{H,t}) = \mathrm{var}(u_t) I_H$, and part i) follows.

Suppose that X_{t-1} is empty, or uninformative; then $u^{\perp}_{H,t} = u_{H,t}$. $aVar(\hat{\beta}) - aVar(\hat{\beta}_{2SLS})$ will be positive definite as long as $\text{var}(u_t) = \sigma_v^2/(1-\rho_u^2) < \text{maxeval}\left(\text{var}(u_{H,t})\right)$. $\text{var}(u_{H,t})$ is a matrix with h,i entry $\rho_u^{|h-i|}\sigma_v^2/(1-\rho_u^2)$. 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 $(\sum_{h=0}^{H-1}\rho_u^h)\sigma_v^2/(1-\rho_u^2)$ which is strictly larger than $\text{var}(u_t)$ when $\rho_u > 0$ and $H > 1$. Therefore, $\text{maxeval}\,\text{var}(u_{H,t}) > \text{var}(u_t)$ when $\rho_u > 0, H > 1$, completing part *ii*).

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

D Proof of Theorem 3

Proof. The concentration parameter for a given SP-IV specification is derived in the Online Appendix, equation (A.11), $\Lambda = (HN_z)^{-1}R'_{K,H}\left((\tilde{\mathcal{S}}CC'\tilde{\mathcal{S}})\otimes I_H\right)R_{K,H}$, where $\tilde{\mathcal{S}} = ((R'_{K,H}(\mathbf{W}_2\otimes I_H)R_{K,H}/H)^{-\frac{1}{2}}$, $R_{K,H} = I_K\otimes\text{vec}(I_H)$, C is the location parameter in the weak instruments asymptotic embedding for Θ_Y in Assumption 4, and $\mathbf{W}_2$ is the covariance of $z_t\otimes v^{\perp}_{H,t}$. For 2SLS, Z_{2SLS} contains lags $0,\dots,H-1$ of Z, so, under stationarity, $C_{2SLS} = \text{vec}(C)$; the same formula then applies, setting $H=1$. It follows that $\Lambda^{2SLS} = C_{2SLS}C'_{2SLS}/\left(HN_z\sigma_\omega^2\right) = \text{Tr}(CC')/\left(HN_z\sigma_\omega^2\right)$. For SP-IV conditional on X_{t-1}, $\tilde{\mathcal{S}} = (\text{Tr}(\Sigma_{v_H^{\perp}})/H)^{-\frac{1}{2}}I_H$, so $\Lambda^{\perp} = \text{Tr}(CC')/\left(N_z\,\text{Tr}(\Sigma_{v_H^{\perp}})\right)$, and the condition in (i) follows. Finally, without controls, $\Lambda^{\emptyset} = \text{Tr}(CC')/\left(N_z\,\text{Tr}(\Sigma_{v_H})\right)$, which is smaller than $\Lambda^{\perp}$ unless $\text{Tr}(\Sigma_{v_H}) = \text{Tr}(\Sigma_{v_H^{\perp}})$, which occurs if and only if X_{t-1} is completely irrelevant, yielding (ii). □