IDENTIFYING MODERN MACRO EQUATIONS WITH OLD SHOCKS*

Regis Barnichon and Geert Mesters

Despite decades of research, the consistent estimation of structural forward-looking macroeconomic equations remains a formidable empirical challenge because of pervasive endogeneity issues. Prominent cases—the estimation of Phillips curves, Euler equations, or monetary policy rules—have typically relied on using predetermined variables as instruments, with mixed success. In this work, we propose a new approach that consists in using sequences of independently identified structural shocks as instrumental variables. Our approach is robust to weak instruments and is valid regardless of the shocks' variance contribution. We estimate a Phillips curve using monetary shocks as instruments and find that conventional methods substantially underestimate the slope of the Phillips curve. *JEL* Codes: C14, C32, E32, E52.

I. Introduction

The estimation of structural forward-looking macroeconomic equations is a central task of macroeconomic research. Prominent examples include the estimation of New Keynesian Phillips curves (e.g., Galí and Gertler 1999), Euler equations (e.g., Deaton 1992; Fuhrer and Rudebusch 2004), monetary policy rules (e.g., Clarida, Galí, and Gertler 2000) and consumption-based asset pricing equations (e.g., Campbell 2003).

Obtaining reliable estimates for the structural coefficients of forward-looking equations has proved challenging because of

* We thank Hie Joo Ahn, Edilean Aragón, Christian Brownlees, Luca Fornaro, Patrik Guggenberger, Òscar Jordà, Frank Kleibergen, Sylvain Leduc, Tassos Magdalinos, Karel Mertens, Lili Morel, Barbara Rossi, Adam Shapiro, Jim Stock, Mark Watson, and seminar participants at the San Fransisco Fed, Texas A&M, Tinbergen Institute, UCL, the 2018 NBER Summer Institute, the 2018 St Louis Fed Econometrics Workshop, and the 2018 Sciences Po Workshop on Empirical Monetary Economics for helpful comments. The views expressed in this article are the sole responsibility of the authors and do not necessarily reflect the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System. Mesters acknowledges support from the Spanish Ministry of Economy and Competitiveness, grant ECO2015-68136-P, and FEDER, EU, the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563), and Fundacion BBVA scientific research grant (PR16-DAT-0043).

The Quarterly Journal of Economics (2020), 2255–2298. doi:10.1093/qje/qjaa022.
Advance Access publication on June 16, 2020.

pervasive endogeneity issues. Take as an example the case of the Phillips curve, which postulates that inflation is determined by three main factors: expected future inflation, the output gap (the difference between the level of economic activity and its natural flexible-price level), and supply factors. All three factors lead to endogeneity-related biases: (i) inflation expectations are unobserved, (ii) the natural level of output (and thus the output gap) is unobserved, and (iii) supply shocks lead to confounding. Similar issues affect other macro equations like the Euler equations or monetary policy rules.

Going back at least to Frisch (1934) and Reiersol (1941), the literature has traditionally addressed endogeneity concerns in macro by using predetermined variables as instruments, that is, lags of observable macro variables as instruments. This approach, which was popularized by the seminal contributions of Hansen and Singleton (1982) and Hansen (1982), has had mixed success, however. Despite decades of research, estimates display both high sampling uncertainty and high specification uncertainty, as minor specification changes can lead to very different estimates (e.g., Yogo 2004; Kleibergen and Mavroeidis 2009; Mavroeidis 2010; Mavroeidis, Plagborg-Møller, and Stock 2014). A common explanation is that predetermined variables are weak, possibly even invalid, instruments.

In this work, we propose a new approach to estimate forward-looking macro equations. Our approach consists of projecting the structural equation of interest on the space spanned by the present and past values of some well-chosen structural shocks. Taking again the Phillips curve as an example, we show that independently identified aggregate demand shocks, for instance monetary policy shocks, can be used to identify the parameters of the Phillips curve. Intuitively, projecting inflation and unemployment on past monetary shocks can address the endogeneity issues by projecting out (i) the influence of supply shocks, (ii) the measurement error in expected future inflation, and (iii) the measurement error in the natural level of output.[1]

1. In a static AD/AS setting, the intuition is straightforward: aggregate shocks that shift the AD curve will allow us to trace out the AS curve, that is, identify the coefficient on the unemployment gap. In a dynamic setting, we will see that aggregate demand shocks can separately identify the coefficients on the unemployment gap and on inflation expectations as long as they have different dynamic effects on future inflation and the output gap.

Our approach amounts to an instrumental variable (IV) regression, where—and this is our key contribution—the set of instruments is a sequence of past structural shocks. For the Phillips curve, monetary policy shocks are appropriate instruments, but different structural shocks will be called for depending on the structural equation of interest. For instance, an aggregate demand relation like the intertemporal IS curve could be identified with aggregate supply shocks.

Using sequences of structural shocks as instruments has an intuitive interpretation as a "regression in impulse response space." By projecting a macro equation on a space spanned by some past structural shocks, our approach can be seen as a regression where the variables of the macro equation are replaced by their impulse responses to the structural shock. Identification then comes from variation across the horizons of the impulse responses, that is, as the regression of an impulse response on other impulse responses.

Because structural shocks are not necessarily strong instruments, we rely on weak-instrument robust methods for conducting inference, see Andrews, Stock, and Sun (forthcoming) for a recent review of the literature. Intuitively, in our setting the weak-IV robust approach amounts to inferring how the residual of the macro equation of interest, say, the Phillips curve, responds over time to an innovation in the structural shock, for instance, a monetary shock. For values of the Phillips curve parameters close to their true values, the impulse response of the residual to a monetary shock should not be different from zero. But for values away from the truth, the impulse response of the residual should be a combination of the impulse responses of inflation and unemployment (the variables of the Phillips curve equation) and be nonzero.

We exploit this impulse response interpretation to improve the power of weak-IV robust tests. If the responses of macro variables to structural shocks are smooth, as is typically believed, the impulse response of the equation residual should also be smooth and we can exploit this "smoothness" to reduce the noise in the weak-IV robust statistics. Specifically, we parameterize the residual impulse response as a quadratic polynomial function that reduces the number of instruments but does not affect the exogeneity of the instruments. Thanks to this dimension reduction, the model becomes just-identified, which allows us to rely on the AR (Anderson and Rubin 1949) statistic for inference, which is known to be the uniformly most accurate unbiased test in this setting (see Moreira 2009). Moreover, when the instruments are strong, the AR test is asymptotically efficient in the usual sense, and so does not sacrifice power relative to the conventional t-test based on the two-stage least squares (2SLS) estimator (see Andrews, Stock, and Sun forthcoming).

Equipped with our new approach, we revisit the literature on the New Keynesian Phillips curve, where we use Romer and Romer (2004) narrative monetary shocks as instruments to identify the structural coefficients over 1969–2007. We find that the coefficient on the forcing variable (the slope of the Phillips curve), measured by either the output gap or the unemployment rate, is significantly different from zero and substantially larger than when using predetermined variables as instruments. In contrast, the role of forward-looking inflation expectations is smaller than estimated with the standard approach. We then study the Phillips curve over the more recent period by using high-frequency identified monetary surprises (e.g., Kuttner 2001) as instruments over 1990–2017. Over that period, the slope of the Phillips curve is smaller but still significant, while forward-looking inflation expectations play a larger role.

Our approach for estimating structural equations bridges two large literatures: the literature on the estimation of structural equations using limited-information methods (see Mavroeidis, Plagborg-Møller, and Stock 2014) and the literature on the identification of macroeconomic shocks and their impulse responses (e.g., Ramey 2016; Stock and Watson 2016).

The use of structural shocks as instruments considerably broadens the scope of identification schemes when compared with using predetermined variables, that is, lags of macro variables, as instruments. Although some specific literatures have taken advantage of structural shocks for identification (see for instance Hall 1988b in the context of production function estimation), modern forward-looking structural equations such as the Phillips curve and the Euler equation have not been identified using structural shocks. Moreover, the key new insight, as derived from the impulse response intuition, is that sequences of current and past structural shocks need to be used to induce sufficient variation in the endogenous macro variables.

Although structural shocks are generally not observable, the recent literature has produced a variety of proxies for structural shocks, which are sufficient for conducting instrumental variable–based inference (e.g., Stock and Watson 2018). Such proxies have been derived using a variety of methods requiring different modeling assumptions. In addition to the monetary shocks already discussed, examples include oil price shocks (Hamilton 2003; Kilian 2008), TFP shocks (Fernald 2012), government spending shocks (Ramey and Zubairy 2018), and potentially many others. All these shocks and notably their lags can potentially be exploited for identifying different structural equations.[2] That being said, the use of proxies for the structural shocks introduces measurement error, which can reduce the power of the hypothesis tests and can cloud the impulse response interpretation (see Stock and Watson 2018).

The remainder of this article is organized as follows. In Section II we review the empirical issues faced by limited-information methods and discuss the traditional solution based on lagged instruments. Section III outlines the use of independently identified structural shocks for identification. The estimation methodology is developed in Section IV, and the empirical findings for the Phillips curve are presented in Section V. Section VI concludes.

II. Structural Equations and Endogeneity Issues

In this section we consider general forward-looking structural equations and discuss the different sources of endogeneity that are present in such equations. We then outline the predominant approach in the literature for conducting inference in this setting: using lagged observables as instrumental variables. Our exposition is brief and is merely intended to lay the ground for the next section, where we introduce our new approach. More details can be found in, for example, Mavroeidis (2005).

Consider the general forward-looking equation

$$y_t = \gamma_b y_{t-1} + \gamma_f \mathrm{E}_t(y_{t+1}) + \lambda x_t + e_t, \tag{1}$$

where y_t is the variable of interest that depends on its own lag, its expected value $\mathrm{E}_t(y_{t+1})$, the forcing variable x_t, and the disturbance e_t. The expectation $\mathrm{E}_t(\cdot)$ is taken with respect to the time t information set $\mathcal{F}_t$. The forcing variable x_t is typically not observable as it is often formulated in deviation from some

2. In our limited-information context, the most appealing shock proxies are identified with little to no additional restriction on the data-generating process. That being said, shocks derived from SVARs identified with exclusion or sign restrictions are also possible, depending on the researcher's tolerance for additional modeling restrictions.

natural rate. For example, when x_t is taken as the unemployment gap it depends on the natural flexible-price level, which is unobserved. The structural coefficients of interest are γ_b, γ_f, and λ. The estimation of these parameters is complicated due to a variety of endogeneity issues. To highlight the different sources of endogeneity we rewrite equation (1) as follows

$$y_t = \gamma_b y_{t-1} + \gamma_f y_{t+1} + \lambda \hat{x}_t + \underbrace{e_t - \gamma_f(y_{t+1} - \mathrm{E}_t(y_{t+1})) - \lambda(\hat{x}_t - x_t)}_{u_t}, \tag{2}$$

where $\hat{x}_t$ is an observable proxy for the forcing variable.[3] In this way the first three variables on the right-hand side of equation (2) are observable and u_t is the unobserved error term. Three potential sources of endogeneity in equation (2) can be distinguished.

(i) **Simultaneous-equation bias and confounding with the error term:** The error term may simultaneously affect y_t and $\hat{x}_t$ through a system of simultaneous equations, in which case we have $\mathrm{E}(\hat{x}_t u_t) \neq 0$.

(ii) **Measurement error in the forcing variable**: Since the forcing variable is unobserved and thus subject to measurement error, we have $\mathrm{E}(\hat{x}_t u_t) \neq 0$.

(iii) **Unobserved inflation expectations:** Since $\mathrm{E}_t(y_{t+1})$ is unobserved and thus subject to measurement error, we have $\mathrm{E}(y_{t+1} u_t) \neq 0$.

This collection of endogeneity problems implies that we cannot use ordinary least squares to consistently estimate the structural parameters in equation (2).

The traditional approach for handling the endogeneity problems is to treat y_{t-1} as predetermined and use lags of the observed macro variables as instruments. To illustrate, we let $z_t^l = (y_{t-2}, \hat{x}_{t-1})'$, and we discuss the conditions under which the three sources of endogeneity bias disappear when we use z_t^l as an instrument.

(i) $\mathrm{E}(e_t z_t^l) = 0$ since $\mathrm{E}_{t-1}(e_t) = 0$ provided that the error term e_t has no serial correlation.

3. Other observable proxies for the expectation term, such as expectation measures from surveys, can equally well be considered.

(ii) $\mathrm{E}((y_{t+1} - \mathrm{E}_t(y_{t+1}))z_t^l) = 0$ since $\mathrm{E}_t(y_{t+1} - \mathrm{E}_t(y_{t+1})) = 0$ under rational expectations and by applying the law of iterated expectations.

(iii) $\mathrm{E}((\hat{x}_t - x_t)z_t^l) = 0$ provided that the measurement error $\hat{x}_t - x_t$ has no serial correlation

This implies that $\mathrm{E}(u_t z_t^l) = 0$ and z_t^l satisfies the exogeneity condition. Moreover, the same can be shown for all z_{t-j}^l with $j \geqslant 0$.

Unfortunately, this approach faces challenges, as it is difficult to find lagged economic variables that are both exogenous and strongly correlated with expected future variables.

First, lagged macro instruments are typically weak instruments, which can lead to considerable sampling uncertainty and to sensitivity of parameter estimates to minor changes in specification choices, in the set of right-hand-side variables or in the sample period (e.g., Mavroeidis, Plagborg-Møller, and Stock 2014). Moreover, conventional inference methods for computing standard errors and confidence bounds break down when instruments are weak and robust methods need to be adopted, see Kleibergen and Mavroeidis (2009).

Second, using lagged macro variables as instruments requires that none of the components in the error term u_t are autocorrelated.[4] A potential way of avoiding this concern is to increase the lag length of the instruments. For instance, to use z_{t-4} instead of z_t as instruments. Unfortunately, this solution leads to a trade-off between the exogeneity condition and the relevance condition as increasing the lag length dramatically worsens the weak-instrument problem (Mavroeidis, Plagborg-Møller, and Stock 2014, 163).

III. Aggregate Structural Shocks as Instruments

In this section we show that sequences of (well-chosen) structural shocks are valid instruments to identify the coefficients in

4. This can happen if the disturbance e_t is autocorrelated, or if the measurement error in y_t or x_t are serially correlated. This problem is likely to be very relevant in practice. For instance, in the context of the Phillips curve, Zhang and Clovis (2010) show that the residual in the Galí and Gertler (1999) specification of the Phillips curve is serially correlated. This can happen with autocorrelation in cost-push shocks (Galí 2015) or with autocorrelation in the measurement error of the natural rates of inflation expectations (e.g., Coibion, Gorodnichenko, and Ulate 2017).

equations like equation (2). Let ε_t^i denote the mean zero structural shock of type i for time period t.[5] Depending on the application ε_t^i can be either a monetary, fiscal, technology, credit, oil price, or some other structural shock. The idea in this work is to use sequences of past structural shocks for identifying the coefficients in equation (2). To this extent define $\varepsilon_{t:t-H}^i \equiv (\varepsilon_t^i, \ldots, \varepsilon_{t-H}^i)'$.

The following two conditions must be verified for the structural shocks $\varepsilon_{t:t-H}^i$ to be characterized as valid instruments:

$$\text{(Exogeneity)} \qquad \mathrm{E}(\varepsilon_{t:t-H}^i u_t) = 0$$

$$\text{(Relevance)} \qquad \mathrm{E}\left(\varepsilon_{t:t-H}^i (y_{t-1}, y_{t+1}, \hat{x}_t)\right) \qquad \text{full column rank.}$$

The exogeneity and relevance conditions imply that the validity of the instruments depends on the structural equation of interest. For instance, aggregate demand shocks will typically be valid instruments to identify an aggregate supply equation, and aggregate supply shocks will be valid to identify an aggregate demand equation. We provide specific examples for important macro equations below, but first we discuss the intimate connection between the exogeneity and relevance conditions, and the identification of impulse response functions.

III.A. *Identification Using Structural Shocks: Intuition*

In this section, we provide some intuition by showing how our approach recasts the problem of identifying structural coefficients as a well-known problem in macroeconomics: the identification of impulse responses to aggregate structural shocks.

We start by rewriting the exogeneity and relevance conditions in terms of impulse responses to the structural shocks $\varepsilon_{t:t-H}^i$. To do this in a simple way, we assume for the moment that all variables are stationary, that the structural shocks are mutually uncorrelated and that the macro variables $(y_{t-1}, y_{t+1}, \hat{x}_t)$ and the residual u_t can be written as linear functions of the structural shocks. Under these assumptions, the exogeneity and relevance conditions can be restated as

$$\text{(Exogeneity)} \qquad \mathcal{R}_h^u = 0 \qquad \forall \quad h = 0, \ldots, H$$

$$\text{(Relevance)} \qquad \left[\mathcal{R}_{h-1}^y, \mathcal{R}_{h+1}^y, \mathcal{R}_h^{\hat{x}}\right]_{h=0}^H \qquad \text{linearly independent,}$$

5. We refer to Ramey (2016), Blanchard and Watson (1986), and Bernanke (1986) for more discussion regarding the definition of a structural shock.

where $\mathcal{R}_h^j$ is the impulse response of j_t, for $j = u, y, \hat{x}$, to the structural shock ε_{t-h}^i. We provide a formal derivation in the Online Appendix.

The reformulated exogeneity condition implies that the impulse response function of the residual u_t to the structural shock is equal to zero. Intuitively, when the macro parameters (λ, γ_f, γ_b) are set at their true values, the impulse response of the residual u_t should be zero (under correct specification). The reformulated relevance condition states that the impulse response of the observed forcing variable $\hat{x}$ and the impulse responses of past and future y are not linearly dependent.

Next, postmultiply the forward-looking equation (2) by ε_{t-h}^i, take the expectation, and use the exogeneity condition to obtain:[6]

$$\mathcal{R}_h^y = \gamma_b \mathcal{R}_{h-1}^y + \gamma_f \mathcal{R}_{h+1}^y + \lambda \mathcal{R}_h^{\hat{x}}, \qquad \forall \quad h = 0, \ldots, H. \tag{3}$$

Expression (3) implies that all the information needed to recover the coefficients of the structural equation is encoded in the impulse response functions to the structural shocks. In fact, we can identify the coefficients of the macro equation from an OLS regression—across h—of the impulse response of the outcome variable on its own lag and lead, and on the impulse response of the forcing variable, that is, from a regression in "impulse response space."[7] The relevance condition can then be seen as the no multicollinearity condition of OLS: the dynamics of the impulse

6. Consider $y_t \varepsilon_{t-h}^i = \gamma_b y_{t-1} \varepsilon_{t-h}^i + \gamma_f y_{t+1} \varepsilon_{t-h}^i + \lambda \hat{x}_t \varepsilon_{t-h}^i + u_t \varepsilon_{t-h}^i$. Now taking expectations on both sides $\mathrm{E}(y_t \varepsilon_{t-h}^i) = \gamma_b \mathrm{E}(y_{t-1} \varepsilon_{t-h}^i) + \gamma_f \mathrm{E}(y_{t+1} \varepsilon_{t-h}^i) + \lambda \mathrm{E}(\hat{x}_t \varepsilon_{t-h}^i) + \mathrm{E}(u_t \varepsilon_{t-h}^i)$. The last term is zero by the exogeneity assumption and the other expectations are the impulse responses of y_{t-1}, y_{t+1}, and $\hat{x}_t$ to ε_{t-h}^i.

7. Specifically, by minimizing the sum of squared residuals $\sum_{h=0}^{H} (\mathcal{R}_h^y - \gamma_b \mathcal{R}_{h-1}^y - \gamma_f \mathcal{R}_{h+1}^y - \lambda \mathcal{R}_h^{\hat{x}})^2$, we can find the structural coefficients that best fit equation (3) for any h. This is an OLS regression in "impulse response space," that is, a regression across the horizon h of the impulse responses. While the "regression in impulse response space" interpretation is helpful to get the intuition behind our IV approach, we do not advocate estimating the coefficients in this way in practice. While the approach is consistent, it is not efficient. In fact, it can be easily verified that the OLS estimates obtained from equation (3) after replacing $\mathcal{R}_h^y$ and $\mathcal{R}_h^{\hat{x}}$ by their sample counterparts are equivalent to computing the GMM estimator for the structural equation (1) with instruments $\{\varepsilon_t^i, \ldots, \varepsilon_{t-H}^i\}$ and with the GMM weighting matrix equal to the identity matrix. This choice is not efficient and not robust to many and weak instruments. Our preferred methodology is described in the estimation section.

responses of $(y_{t-1}, y_{t+1}, \hat{x}_t)$ have to be rich enough such that there exists a unique parameter vector $(\lambda,\ \gamma_f,\ \gamma_b)$ satisfying expression (3).

III.B. Identification Using Structural Shocks: Examples

To illustrate our approach we discuss three important structural equations: the Phillips curve, the Euler equation (for output or consumption) and the central bank's monetary policy rule. In each case, we argue that sequences of well-chosen structural shocks can form valid instruments under relatively mild assumptions.

1. The Phillips Curve. Consider the hybrid New Keynesian Phillips curve (e.g., Galí and Gertler 1999) given by

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f \mathrm{E}_t(\pi_{t+1}) + \lambda x_t + \varepsilon_t^s, \tag{4}$$

where π_t is inflation, the output gap $x_t = g_t - g_t^n$ depends on the natural level of output g_t^n, and ε_t^s denotes some (possibly autocorrelated) exogenous cost-push factors. The parameters of interest γ_b, γ_f, and λ are typically functions of deep structural parameters of an underlying model (see, e.g., Galí 2015). Notice that the Phillips curve fits naturally in our general framework, equation (1).

Rewriting equation (4) to highlight the endogeneity issues, we have

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f \pi_{t+1} + \lambda \hat{x}_t + \underbrace{\varepsilon_t^s - \gamma_f(\pi_{t+1} - \mathrm{E}_t(\pi_{t+1})) - \lambda(\hat{x}_t - x_t)}_{u_t}. \tag{5}$$

The Phillips curve includes all three sources of endogeneity discussed in Section II: (i) cost-push factors can simultaneously affect inflation and the forcing variable through the systematic response of monetary policy to inflation developments (Kareken and Solow 1963; McLeay and Tenreyro 2018), (ii) measurement error in the forcing variable since the natural level of output is unobserved, and (iii) unobserved inflation expectations.

We now argue that monetary shocks $\varepsilon_{t:t-H}^m$—deviations of the central bank from its typical behavior (e.g., Romer and Romer 2004; Cochrane 2004)—are valid instruments to identify the Phillips curve, that is, that they are both (i) exogenous and (ii) relevant.

Exogeneity: The exogeneity condition $E(\varepsilon^i_{t:t-H}u_t) = 0$ is satisfied if monetary shocks are orthogonal to (i) cost-push factors, (ii) measurement error in the output gap, and (iii) measurement error in inflation expectations. While the systematic response of monetary policy to inflation can create a correlation between the output gap and cost-push factors, monetary shocks are innovations to the systematic conduct of monetary policy (e.g., Galí 2015; McLeay and Tenreyro 2018), and should thus be orthogonal to cost-push factors and satisfy condition (i).[8] Condition (ii) holds under the assumption that money is neutral under flexible prices, a relatively mild and uncontroversial assumption.[9] Condition (iii) holds under rational expectations or provided that survey measures of inflation expectations are available and accurate up to some additive (and possibly autocorrelated) measurement error term.[10]

Relevance: Monetary shocks are relevant instruments if they affect inflation and the output gap. This implies that (in addition to the Phillips curve equation (4)), there must exist an underlying IS curve, that is, an equation linking the output gap to the level of interest rate (and thus to monetary shocks). Our approach does not rely on specifying any parametric IS curve, only that such a curve exists so that the policy rate affects the output gap. Since the existence of an IS curve is a cornerstone of most macro models, we view this condition as mild and uncontroversial. In addition, because the Phillips curve equation (4) involves three

8. This is true as long as monetary policy has no effect on aggregate supply. Although this is a commonly held assumption, some cost effects of monetary policy are possible. For instance, if firms need to finance wage payments or need to hold inventory, a higher interest rate can raise firms' real marginal costs, the so-called cost channel of monetary policy (e.g., Barth and Ramey 2001). In that case, the exogeneity condition (i) is no longer verified, and one should include the interest rate on the right-hand side (Ravenna and Walsh 2006) and instrument it with monetary shocks. Another example whereby monetary policy can have cost-push effects is when oil prices respond to U.S. monetary policy. In that case, one would need to add (and instrument) the oil price on the right-hand side. More generally, the set of valid instruments depends on the specification of the Phillips curve posited by the researcher. Here, we focus on the standard New Keynesian Phillips curve encountered in most empirical studies (e.g., Mavroeidis, Plagborg-Møller, and Stock 2014).

9. The exogeneity condition $E(\varepsilon^m_{t-j}(\hat{x}_t - x_t)) = 0$ is verified, if $E(\varepsilon^m_{t-j}(\hat{g}^n_t - g^n_t)) = 0$, which holds if monetary policy is neutral under flexible prices.

10. The exogeneity condition $E(\varepsilon^m_{t-j}(\pi_{t+1} - E_t\pi_{t+1})) = 0$ is satisfied under rational expectations, since the law of iterated expectations implies $E(\varepsilon^m_{t-j}(\pi_{t+1} - E_t\pi_{t+1})) = E(\varepsilon^m_{t-j}E_t(\pi_{t+1} - E_t\pi_{t+1})) = 0$. For departures from rational expectations, we can still obtain consistent estimates, as long as the survey measurement error term is orthogonal to monetary shocks, a relatively mild assumption.

endogenous variables (lagged inflation, future inflation, and the output gap), satisfying the rank condition requires that the first-stage predicted values of the endogenous variables are not linearly dependent. From the intuition in Section III.A, it follows that the relevance condition holds if and only if the impulse responses of lagged inflation, future inflation, and the output gap are not linear functions of one another. With a hybrid Phillips curve ($\gamma_b > 0$), this is ensured even if the output gap x_t follows only a basic i.i.d. process (see Appendix A for a formal derivation), so we view this condition as mild and uncontroversial. Naturally however, as emphasized in the literature (Kleibergen and Mavroeidis 2009), the rank condition is not sufficient for reliable estimation and inference because of the problem of weak instruments. We will come back to this point in the estimation section.

2. The Euler Equation. Consider a linearized Euler equation of the form

$$x_t = \gamma_b x_{t-1} + \gamma_f \mathrm{E}_t(x_{t+1}) - \lambda \left(i_t - \mathrm{E}_t(\pi_{t+1}) - r_t^n\right), \tag{6}$$

with r_t^n the real natural rate of interest and where x_t can be the (log) output gap as in the output Euler equation, or (log) aggregate consumption as in the consumption Euler equation.[11] This equation forms the basis of numerous empirical works on the dynamic IS curve underlying the New Keynesian model (e.g., Fuhrer and Rudebusch 2004), or on the elasticity of intertemporal substitution (e.g., Hall 1988a; Yogo 2004; Ascari, Magnusson, and Mavroeidis forthcoming).

Rewriting the Euler equation to highlight the endogeneity issues gives

$$\hat{x}_t = \gamma_b \hat{x}_{t-1} + \gamma_f \hat{x}_{t+1} - \lambda\left(i_t - \pi_{t+1}\right) + u_t, \tag{7}$$

where the residual u_t captures endogeneity bias from (i) confounding from movements in the real rate of interest (e.g., from productivity shocks; Galí 2015), (ii) measurement error in the

11. Compared with the conventional Euler equation implied by the baseline New Keynesian model (e.g., Galí 2015), specification (6) features the lag of the output gap as an explanatory variable. This added persistence can arise with habit formation in consumption; see Fuhrer (2000), for instance.

output gap, and (iii) unobserved inflation expectations and output gap expectations.[12]

Again, monetary shocks are good candidates for valid instruments to identify equation (7). The reasons are similar to the case of the Phillips curve, and we do not repeat them. The only difference is that the confounding factors are no longer cost-push shocks, but instead shocks to the natural real rate of interest, for instance, productivity shocks (e.g., Galí 2015). Again, the common assumption that monetary policy is neutral under flexible prices implies that monetary shocks are orthogonal to movements in the natural rate of interest, which means that monetary shocks satisfy the exogeneity condition for the Euler equation as well.

Another set of possible candidates for exogenous instruments are cost-push shocks. These shocks are relevant instruments as long as there exist some underlying Phillips curve and monetary rule with rich enough dynamics (that need not be specified), such that the impulse responses to a cost-push shock of the three endogenous variables in the Euler equation—inflation, the output gap, and the nominal interest rate—are not linear functions of one another.

3. The Monetary Policy Rule. The final example that we discuss is a simplified version of the interest rate rule from Clarida, Galí, and Gertler (2000) and Mavroeidis (2010) that is given by

$$i_t = \gamma_b i_{t-1} + \gamma_f \mathrm{E}_t(\pi_{t+1}) + \lambda x_t + \varepsilon_t^m, \tag{8}$$

where i_t denotes the nominal interest rate, x_t the output gap, and ε_t^m the monetary policy shock.

We rewrite equation (8) in terms of the observables to obtain

$$i_t = \gamma_b i_{t-1} + \gamma_f \pi_{t+1} + \lambda \hat{x}_t + u_t. \tag{9}$$

12. The residual u_t satisfies

$$u_t = \lambda r_t^n - \lambda(\pi_{t+1} - \mathrm{E}_t(\pi_{t+1})) - \gamma_f(\hat{x}_{t+1} - \mathrm{E}_t(x_{t+1})) + \sum_{j=0,1}(-\gamma_b)^j(\hat{x}_{t-j} - x_{t-j}).$$

Equation (6) admits the general form discussed in Section II, but with one additional source of endogeneity compared to the Phillips curve: because the left-hand side variable in equation (6) is the unobserved variable x_t, serially correlated measurement error in x_t will imply $\mathrm{E}(\hat{x}_{t-1}u_t) \neq 0$.

The sources of endogeneity bias in equation (9) are confounding from monetary shocks, unobserved inflation expectations, and measurement error in the output gap.[13] In this case, productivity shocks are valid instruments as long as there exist some underlying Phillips curve and IS curve with rich enough dynamics (that need not be specified), such that the impulse responses of inflation and the output gap to those shocks are not linear functions of one another.

IV. Estimation Methodology

In this section we discuss inference for the parameters of the general forward-looking model (2) using structural shocks as instruments. For ease of exposition consider the following compact model representation

$$y_t = w_t'\delta + u_t, \tag{10}$$

where

$$w_t = (y_{t-1}, y_{t+1}, \hat{x}_t)' \qquad \text{and} \qquad \delta = (\gamma_b, \gamma_f, \lambda)'.$$

Although structural shocks are typically not observed, the literature has produced a variety of proxies for structural shocks, which are sufficient for conducting IV-based inference (e.g., Stock and Watson 2018). To distinguish between the structural shocks and their proxies, we denote the latter by ξ_t^i and work under the assumption that ξ_t^i correlates only with ε_t^i and not with other structural shocks. Hence, the identification arguments of the previous section are assumed to hold when we replace $\varepsilon_{t:t-H}^i$ with $\xi_{t:t-H}^i$.

IV.A. Naive Moment Estimators

Given the sequence of proxies $\xi_{t:t-H}^i$, a straightforward approach for estimating δ is to use method of moment estimators. In general, following the textbook treatment of White (2000), we can consider estimators of the form

$$\hat{\delta}^{IV} = \left(S_{\xi w}'\hat{\Omega}_\xi S_{\xi w}\right)^{-1} S_{\xi w}'\hat{\Omega}_\xi s_{\xi y}, \tag{11}$$

13. The residual is given by $u_t = \varepsilon_t^m + \gamma_f(\mathrm{E}_t(\pi_{t+1}) - \pi_{t+1}) + \lambda(x_t - \hat{x}_t)$.

where $S_{\xi w} = \frac{1}{n}\sum_{t=1}^{n} \xi_{t:t-H}^{i} w_t'$, $s_{\xi y} = \frac{1}{n}\sum_{t=1}^{n} \xi_{t:t-H}^{i} y_t$, and $\hat{\Omega}_\xi$ is some positive-definite weight matrix. A set of general assumptions under which $\sqrt{n}(\hat{\delta}^{IV} - \delta_0)$ converges to a normal distribution is given in White (2000; see Theorem 5.23). Based on such normal limiting approximations we may conduct hypothesis tests and construct confidence intervals.[14]

This naive approach suffers from two problems however: weak instruments and many instruments.

First, structural shocks need not explain a large share of the variance of macro variables (e.g., Gorodnichenko and Lee forthcoming; Plagborg-Møller and Wolf 2018), which implies that in such cases the shocks are weak instruments. Consequently, the conventional normal limiting distribution of the moment estimator $\hat{\delta}^{IV}$ provides a poor description of the finite sample behavior of the estimator (e.g. Staiger and Stock 1997).

Second, we typically want to consider the number of structural shocks between $H = 12$ and $H = 20$ for quarterly data because this is the horizon for which macroeconomic impulse responses are typically found to be significantly different from zero. When the number of instruments used is large relative to the sample size, we face a many instruments problem, and again the traditional asymptotic approximation for the moment estimator $\hat{\delta}^{IV}$ provides a poor description of its finite sample behavior (e.g., Bekker 1994). Moreover, with many instruments, tests based on conventional weak-instrument robust statistics have poor power and size properties; see Andrews and Stock (2007).

IV.B. Inference with the Almon-Restricted AR Statistic

Our preferred inference approach follows the weak-instrument robust literature (e.g., Andrews, Stock, and Sun 2019) by considering test statistics for which the limiting distribution does not depend on the strength of the instruments. In addition, we exploit the impulse response intuition from Section III.A to reduce the number of effective instruments, thus avoiding the many instruments problem.

14. A special case of this naive approach is a two-step approach where in the first step the structural impulse responses of w_t to the structural shock proxies $\xi_{t:t-H}^{i}$ are estimated using SVAR-IV or LP-IV (see Mertens and Ravn 2013; Stock and Watson 2018), and in the second step the estimated impulse responses are regressed on each other based on equation (3).

Consider testing the hypothesis H_0: $\delta = \delta_0$. From the exogeneity condition $\mathrm{E}(\xi^i_{t:t-H} u_t) = 0$, we can test H_0 based on the distributed lag model

$$y_t - w_t' \delta_0 = \theta' \xi^i_{t:t-H} + \eta_t, \tag{12}$$

where θ is the $(H + 1) \times 1$ impulse response of the macro-equation residual u_t to the proxies $\xi^i_{t:t-H}$, and η_t is a disturbance term.[15] Under H_0 the exogeneity condition implies that the impulse response θ is zero. So a test for H_0: $\delta = \delta_0$ can be implemented by testing $\theta = 0$. Intuitively, for values of the macro parameters close to their true values, the impulse response of the residual $u_t = y_t - w_t' \delta_0$ to the structural shock proxies should not be different from zero. Conversely, for values away from the truth, the impulse response of the residual should be a combination of the impulse responses of y_t (the left-hand side variable of the macro equation) and the impulse responses of $\hat{x}_t$ and future and past y_t (the right-hand side variables of the macro equation), and thus be nonzero.[16]

This approach to conducting inference goes back to Anderson and Rubin (1949), and we can test H_0: $\delta = \delta_0$ by testing $\theta = 0$ using an Anderson-Rubin (AR) type statistic

$$AR[\delta_0] = \hat{\theta}' \hat{\Sigma}_\theta^{-1} \hat{\theta}, \tag{13}$$

where $\hat{\theta}$ is the OLS estimate for θ based on equation (12) and $\hat{\Sigma}_\theta$ denotes any heteroskedasticity and serial-correlation robust estimator for the variance of $\hat{\theta}$.

The important feature of such an AR-type statistic is that its limiting distribution does not depend on the strength of the instruments (e.g., Staiger and Stock 1997).[17]

15. Note that we changed the impulse response notation from $\mathcal{R}$ to θ to highlight that this is the impulse response to the proxies for the structural shocks instead of the structural shocks themselves.

16. For instance, if the parameters of the Phillips curve equation are set to zero, the impulse response of the residual corresponds to the impulse response of inflation, the left-hand variable of the Phillips curve equation.

17. In the homoskedastic case under random sampling the AR test statistic is equivalent to the F-statistic of the regression of $y_t - w_t' \delta_0$ on $\xi^i_{t:t-H}$. More general forms that allow for, among others, dependent data can be found in, for example, Stock and Wright (2000). Other popular test statistics for H_0: $\delta = \delta_0$ include the Lagrange multiplier (LM) statistic of Kleibergen (2002) and the conditional likelihood ratio statistic of Moreira (2003).

Unfortunately, hypothesis tests based on the standard AR-statistic have poor power and size properties when, as in our setting, the number of instruments is large relative to the sample size (Andrews, Stock, and Sun 2019). To reduce the dimension of the problem, we exploit the fact that the impulse responses of macro variables are typically believed to be smooth functions, and we draw from Almon (1965) and parameterize the elements of the impulse response θ as a polynomial function

$$\theta_h = a + bh + ch^2, \qquad \text{for} \quad h = 0, \ldots, H, \tag{14}$$

where a, b, and c are the polynomial coefficients. While alternative basis functions for θ_h can also be considered, polynomial basis functions are attractive in our setting because the resulting estimation problem remains linear.

With the Almon parametrization we reduce the number of instruments to three with

$$z_t^i = \left(\sum_{h=0}^{H} \xi_{t-h}^i, \sum_{h=0}^{H} h\xi_{t-h}^i, \sum_{h=0}^{H} h^2 \xi_{t-h}^i \right)', \tag{15}$$

and we can rewrite our distributed lag model (12) as

$$y_t - w_t'\delta_0 = \theta_a' z_t^i + \eta_t, \tag{16}$$

where the parameters to estimate are the three Almon polynomial coefficients $\theta_a = (a, b, c)'$.

Notice that our new set of instruments z_t^i is merely a deterministic linear function of the exogenous structural shocks and hence z_t^i inherits the exogeneity properties of $\xi_{t:t-H}^i$, that is, we have $\mathrm{E}(z_t^i(y_t - w_t'\delta_0)) = 0$ under H_0. This implies that our approach remains valid even if the true impulse responses are not smooth functions and a quadratic polynomial provides a poor approximation. In such cases the Almon restriction will only impose a cost in terms of lower power.

With the number of instruments equal to the number of endogenous variables,[18] we can construct an "Almon-restricted"

18. In such just-identified settings Chernozhukov, Hansen, and Jansson (2009) have shown that the Anderson and Rubin (1949) statistic for testing H_0: $\delta = \delta_0$ is admissible. Intuitively, this means that we can be robust to weak instruments without sacrificing power. Moreover, Moreira (2009) shows that the AR test is uniformly most accurate unbiased in this setting.

AR statistic given by

$$AR_a[\delta_0] = \hat{\theta}_a' \hat{\Sigma}_{\theta_a}^{-1} \hat{\theta}, \tag{17}$$

where

$$\hat{\theta}_a = \left(\sum_{t=H+1}^{n} z_t^i z_t^{i'} \right)^{-1} \sum_{t=H+1}^{n} z_t^i (y_t - w_t' \delta_0), \quad \hat{\Sigma}_{\theta_a} = \left(\sum_{t=H+1}^{n} z_t^i z_t^{i'} \right)^{-1} \hat{s}_u^2,$$

and $\hat{s}_u^2$ is any consistent estimate for the long-run variance of $u_t = y_t - w_t'\delta_0$. The form of the variance estimate $\hat{\Sigma}_{\theta_a}$ is motivated by our asymptotic theory in which we let the number of instruments increase with the sample size, for example, $\frac{H}{n} \to c \in (0, 1)$ as $n \to \infty$.

When the structural shocks are strictly exogenous, that is, $\mathrm{E}(u_t \xi_s^i) = 0$ for all s, t, we show in Appendix B that the Almon-restricted AR statistic converges in distribution to a $\chi^2(3)$ under mild regularity conditions, in particular allowing for autocorrelation and heteroskedasticity in both the macro-equation residual u_t and in the shock proxy ξ_t^i. Confidence sets for δ are then computed by inverting the AR_a statistic for different values of $\delta_0 \in \mathfrak{D} \subset \mathbb{R}^3$. We provide a detailed implementation guide in the Online Appendix.

Finally, note that the Almon restriction can also be used to reduce the number of instruments for the naive moment estimator of Section IV.A. In particular, we can consider the Almon-restricted moment estimator

$$\hat{\delta}_a^{IV} = S_{zw}^{-1} s_{zy}, \tag{18}$$

where $S_{zw} = \frac{1}{n}\sum_{t=H+1}^{n} z_t^i w_t'$ and $s_{zy} = \frac{1}{n}\sum_{t=H+1}^{n} z_t^i y_t$. This simple IV estimator does not suffer from the many instrument problem, thanks to the Almon restriction, but it is not robust to weak instruments. Therefore our preferred approach is based on the $AR_a[\delta_0]$ statistic, which is robust to weak instruments and does not suffer from the many instruments problem.

IV.C. Inference with the Subset Almon-Restricted AR Statistic

Often we are interested in conducting inference on a subset of parameters. For instance, a researcher may only be interested in getting a confidence interval for the effect of the forcing variable.

To conduct subset inference we partition the parameters δ as follows $\delta = (\beta', \alpha')'$. The subset hypothesis of interest is given by H_0: $\beta = \beta_0$, and we may regard the parameters α as nuisance parameters. To test the null hypothesis, without assuming strong identification, we propose a subset version of the Almon-restricted AR statistic, building on Stock and Wright (2000), Kleibergen and Mavroeidis (2009), and Guggenberger et al. (2012).

In particular, we consider

$$AR_{a,s}[\beta_0] = \min_{\alpha \in \mathbb{R}^{\dim(\alpha)}} AR_a[(\beta_0', \alpha')']. \tag{19}$$

We show in Appendix B that $AR_{a,s}[\beta_0]$ is upper bounded by a chi-squared random variable with degrees of freedom equal to the dimension of β.[19] To compute the subset AR statistic we minimize $AR_a[(\beta_0', \alpha')']$ with respect to α and subsequently compare $AR_{a,s}[\beta_0]$ with the critical values of the $\chi^2(\dim(\beta))$ distribution. Again, see the detailed implementation guide in the Online Appendix for more details.

In certain applications it may be desirable to use more shock instruments when compared with the number of endogenous variables. To make our approach suited for such settings Appendix C generalizes our methodology to cover structural equations with an arbitrary number of structural parameters and multiple Almon-restricted structural shock instruments. For these settings we may continue to use the (subset) Almon-restricted AR statistic as long as the effective number of instruments is at least as large as the number of endogenous variables.

IV.D. Summary of the Simulation Study

In this section we briefly discuss the findings from a simulation study that we conducted to assess the finite sample performance of our proposed methodology. A full description of the simulation study is presented in Appendix D.

We simulated data from model (1) where the forcing variable followed an AR(2) process. The structural shocks were chosen such that their variance contributions mimic the recent empirical findings for monetary policy shocks (e.g., Plagborg-Møller and Wolf 2018; Gorodnichenko and Lee forthcoming), and notably the fact that monetary shocks may account for a relatively small

19. Note that if we assume that α is strongly identified, we have that $AR_{a,s}[\beta_0] \xrightarrow{d} \chi^2(\dim(\beta))$, see Stock and Wright (2000). When identification is weak, the $\chi^2(\dim(\beta))$ distribution provides merely an upper bound, implying that inference based on the subset statistic is conservative (e.g., Guggenberger et al. 2012).

share of the variance of macro variables. Based on this data-generating process, we compared the standard Wald test (based on the 2SLS estimator in equation (11)), the Wald test computed with Almon-restricted instruments (based on the IV moment estimator with Almon restriction equation (18)), the standard AR test equation (13), and the Almon-restricted AR_a test equation (17). We vary $H = 20$, 40 to investigate the sensitivity of the methodology to different choices for H.

We compared the empirical rejection frequencies of these tests and found that only the AR_a test has correct size. All other tests severely overreject. For the standard Wald test this is caused by both many and weak instruments, for the Almon-restricted Wald test this is caused only by weak instruments and for the standard AR test this is caused by the use of many instruments relative to the sample size. Importantly, our proposed Almon-restricted AR_a test has correct size regardless of the strength of the instruments and the value of H.

For the subset Almon-restricted $AR_{a,s}$ test we find that if the instruments are strong the size of the subset test is correct. When the instruments are weak the subset statistic is conservative. These findings hold for all combinations of H and n considered and correspond with the asymptotic theory outlined in Appendix B. Additional simulation results are provided in the Online Appendix.

V. The U.S. Phillips Curve

In this section we illustrate our approach by estimating the New Keynesian Phillips curve using quarterly data for the United States. We consider a standard hybrid Phillips curve of the form

$$\pi_t = \gamma_b \pi_{t-1}^4 + \gamma_f \mathrm{E}_t\left(\pi_{t+4}^4\right) + \lambda x_t + \varepsilon_t^s, \quad (20)$$

with π_t (annualized) quarter-to-quarter inflation and $\pi_{t-1}^4 = \frac{1}{4}(\pi_{t-1} + \pi_{t-2} + \pi_{t-3} + \pi_{t-4})$ average inflation over the past year.

In Section III.B we showed that one can identify the parameters of the Phillips curve equation (20) by using monetary policy shocks as instrumental variables. To operationalize the use of monetary shocks for identification we rely on two different proxies for monetary policy shocks. Our baseline estimates are based on the Romer and Romer (2004) narrative measure of exogenous monetary policy changes, which has the advantage of covering the longest sample period thanks to Tenreyro and Thwaites (2016)'s extension of the Romer and Romer series (1969–2007). As an alternative, we also rely on the recent high-frequency identification (HFI) approach, pioneered by Kuttner (2001) and Gürkaynak, Sack, and Swanson (2005), and use surprises in futures/bond prices around Federal Open Market Committee (FOMC) announcements as proxies for monetary shocks.

Before presenting our results, we note that these monetary shock proxies have limitations, both in terms of the validity of the exogeneity condition and in terms of the instrument strength. Regarding the exogeneity condition, Romer and Romer (2004) identify monetary shocks holding constant the staff's Greenbook forecasts for output and inflation, but one concern is that policy makers respond to information beyond what is in the Greenbook. If this response is in reaction to cost-push factors, the exogeneity condition could be violated. For HFI surprises, the limitation comes from a possible Federal Reserve information effect, whereby an FOMC announcement releases some information that was known by the Federal Reserve but not by private agents (Romer and Romer 2000; Nakamura and Steinsson 2017). If some of the Fed informational advantage is related to cost-push factors, the exogeneity condition could be violated. In terms of instrument strength, if monetary policy has been set more systematically in the post-1990 period (see Ramey 2016; McLeay and Tenreyro 2018), this would leave only a limited amount of true exogenous variation to identify the Phillips curve over that period. While the asymptotic distribution of our test statistics does not depend on the strength of the instruments, the power of our tests will be lower when the instruments are weaker.

V.A. *Identification from Romer and Romer Monetary Shocks, 1969–2007*

We first present our results based on using the Romer and Romer monetary shocks as instruments with $H = 20$ over 1969–2007. For our baseline results, we measure inflation from changes in the PCE price level excluding food and energy prices (core PCE), and as forcing variable we use detrended unemployment or the detrended real GDP gap, with the underlying trend estimated from an HP-filter with $\lambda^{hp} = 1{,}600$. We later consider alternative specifications.

TABLE I
THE PHILLIPS CURVE, 1969–2007, RR ID

	Unrestricted		Restricted	
γ_b	0.51	[0.11, 1.02]		
γ_f	0.53	[0.07, 0.89]	0.53	[0.11, 0.88]
λ_U	−0.42	[−1.61, −0.05]	−0.45	[−1.57, −0.07]
γ_b	0.62	[0.18, 3.31]		
γ_f	0.42	[−2.05, 0.83]	0.40	[−1.62, 0.82]
λ_Y	0.28	[0.03, 2.95]	0.31	[0.05, 2.53]

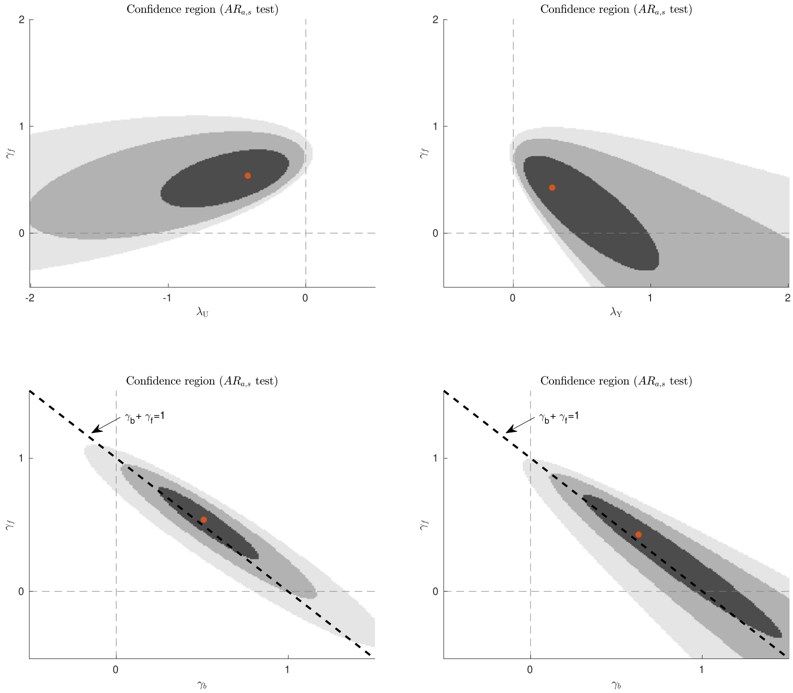
Notes. The table reports the parameter estimates and weak-IV robust confidence intervals for the U.S. Phillips curve (1969–2007). We show the Almon-restricted IV point estimates based on the Romer and Romer (2004) shocks as instruments ($H = 20$) and the $AR_{a,s}$ based 95% confidence bounds. The forcing variable is the unemployment gap λ_U or the output gap λ_Y.

In Table I we show the results for the Phillips curve coefficients γ_b, γ_f, and λ. We report the Almon-restricted IV point estimates equation (18) for the individual parameters γ_b, γ_f, and λ, and we use the subset $AR_{a,s}$ statistic, as in equation (19), to obtain the weak-IV robust confidence intervals. Finally, we complement our study by reporting the same set of estimates computed under the restriction that $\gamma_b + \gamma_f = 1$, a restriction that is often imposed in empirical studies and is consistent with the existence of a vertical long-run Phillips curve.

The slope of the Phillips curve (λ) is significantly different from zero, regardless of the forcing variable, whereas lagged inflation (γ_b) and expected future inflation (γ_f) are equally important in determining inflation. In fact, the coefficient on lagged inflation is always positive and significant, indicating that the hybrid Phillips curve is preferable to the strictly forward-looking Phillips curve.

To better capture the interaction between the coefficient estimates, Figure I shows two-dimensional confidence regions. The top row shows the two-dimensional confidence regions for γ_f and λ, obtained by using the subset $AR_{a,s}$ statistic, where only lagged inflation was integrated out.[20] Overall, we can exclude zero for the slope of the Phillips curve for most values of γ_f, but we have difficulty rejecting combinations of a large (absolute) slope and a small (in absolute value) coefficient on expected future inflation.

20. Formally, in the notation of the subset statistic equation (19) we take $\alpha = \gamma_b$ and construct the confidence set for $\beta = (\gamma_f, \lambda)'$ by inverting the subset AR–based test $\beta = 0$ for different values of β.



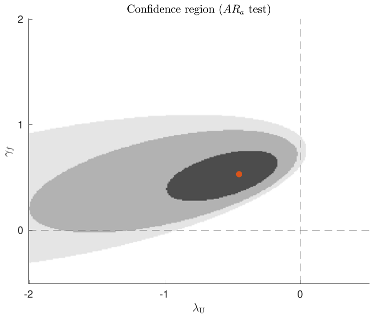
(A) Unemployment gap (B) Output gap

FIGURE I

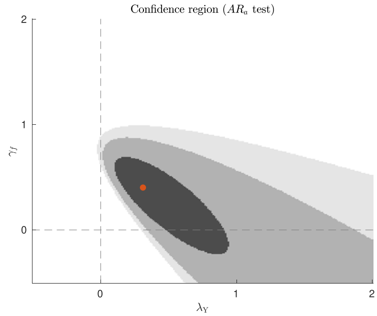
The Phillips Curve, 1969–2007, RR id

Robust confidence sets for the Phillips curve coefficients are obtained by inverting the $AR_{a,s}$ statistic. Top row: 95%, 90%, and 68% confidence sets for λ (the slope of the Phillips curve) and γ_f (the loading on inflation expectations). Bottom row: confidence sets for γ_f and γ_b (the loading on expected and lagged inflation). The dashed line depicts the $\gamma_f + \gamma_b = 1$ set. The estimation is based on using the Romer-Romer (RR) monetary shocks as instruments for 1969–2007. The black (red online) dot is the Almon-restricted IV estimate. Specification with the unemployment gap (left column) or the output gap (right column) as the forcing variable.

The bottom row of Figure I shows the confidence sets for (γ_b, γ_f), that is, after differentiating out the forcing variable. Our results support a vertical long-run Phillips curve, as the confidence sets for (γ_b, γ_f) lie on the $\gamma_b + \gamma_f = 1$ line. In fact, consistent with that result, imposing the common restriction $\gamma_b + \gamma_f = 1$ (e.g., Kleibergen and Mavroeidis 2009) barely changes our IV point estimates and confidence sets for (λ, γ_f), except that the sets become slightly smaller (Figure II and Table I). Again



(A) Unemployment gap



(B) Output gap

FIGURE II

The Phillips Curve, 1969–2007, RR id, $\gamma_f + \gamma_b = 1$

Robust confidence sets for the Phillips curve coefficients are obtained by inverting the AR_a test. 95%, 90%, and 68% confidence sets for λ (the slope of the Phillips curve) and γ_f (the loading on inflation expectations). The estimation is based on using the Romer-Romer (RR) monetary shocks as instruments for 1969–2007. The black (red online) dot is the Almon-restricted IV estimate. The specification imposes $\gamma_f + \gamma_b = 1$ with the unemployment gap (left column) or the output gap (right column) as the forcing variable.

however, we have a hard time discarding large (absolute) values for λ when $|\gamma_f|$ is small.

1. Intuition. To get some intuition behind this last result and more generally better understand how we construct our confidence sets from the impulse responses of the residual, Figure III displays the heatmap of the AR_a statistic for our restricted ($\gamma_f + \gamma_b = 1$) estimates based on using the unemployment gap as the forcing variable. Intuitively, the AR_a statistic can be seen as an F-test of overall significance for the impulse response of the Phillips curve residual to a monetary shock. Darker (bluer) values indicate values of the AR_a statistic close to zero—impulse responses of the residual close to zero—and thus more "likely" parameter values. For values away from the truth, the impulse response of the residual should be a combination of the impulse responses of inflation and the unemployment gap and thus be nonzero.

To illustrate how the impulse response of the residual changes with parameter values, the bottom panel of Figure III plots the impulse responses of the residual for nine different values of (λ, γ_f), first unsmoothed (in blue) and then smoothed with an Almon restriction (in red; color version available online). The

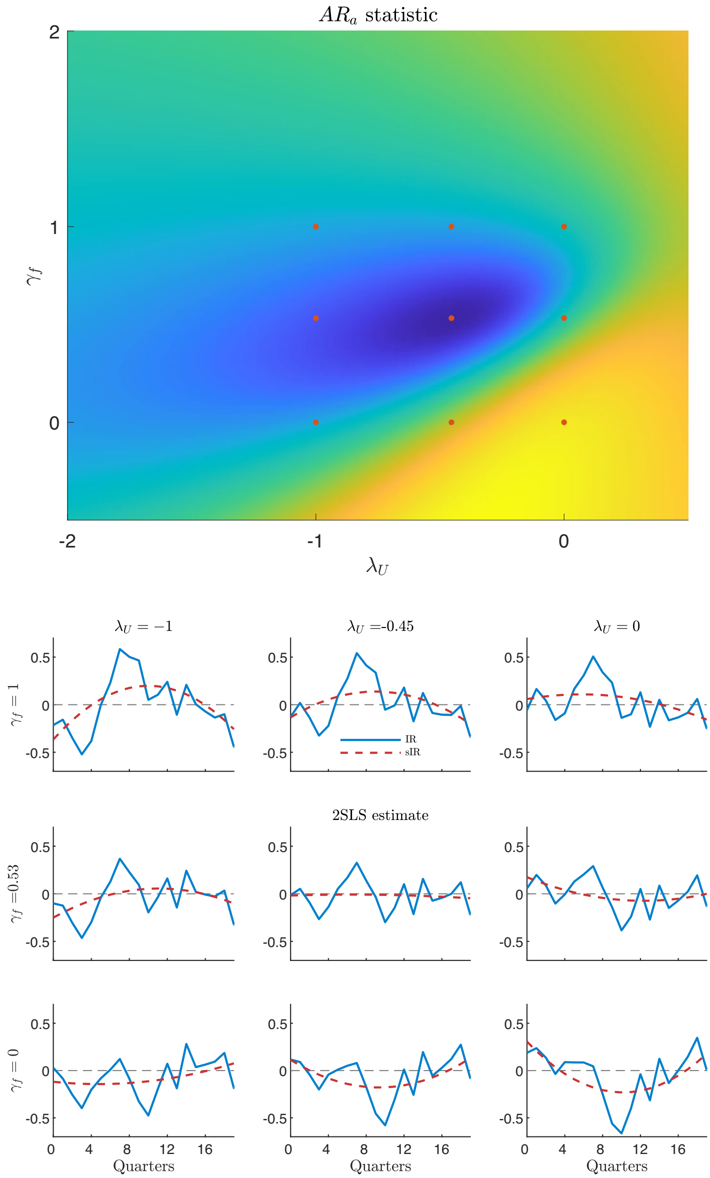


FIGURE III

The Phillips Curve, 1969–2007, RR id, $\gamma_f + \gamma_b = 1$

Top panel: Heatmap of the Almon AR statistic (AR_a) across the parameter space of λ (the slope of the Phillips curve) and γ_f (the loading on inflation expectations). The estimation is based on using the Romer-Romer (RR) monetary shocks

FIGURE III (CONTINUED)

as instruments over 1969–2007. The black (red online) dots denote the parameter values corresponding to the nine impulse responses plotted in the bottom panel, with the center dot corresponding to the Almon-restricted IV estimate. Bottom panel: Impulse responses (IR in solid blue) of the Phillips curve residual for different values of λ and γ_f. The impulses responses smoothed with an Almon restriction (sIR) are reported in dashed red lines.

small red dots in the top panel of Figure III denote the different parameter values corresponding to the nine impulse responses. For λ and γ_f at their IV estimates (center red dot in top panel), the impulse response of the residual is close to zero, consistent with the idea that the point estimates are close to their true values. As we move away from these values, the impulse response of the residual becomes a combination of the impulse responses of inflation and unemployment. For instance, with $\lambda = 0$ and $\gamma_f = 0$ (impulse response in the right-bottom panel), one can show that the residual is simply $\Delta\pi_t$. Because $\Delta\pi_t$ decreases following a positive (i.e., contractionary) monetary shock, this allows us to discard this parameter pair. As we decrease λ however (moving to the impulse response in the left-bottom panel), the residual becomes a (weighted) sum of $\Delta\pi_t$ and u_t, two variables that move in the opposite direction following a monetary shock. With the impulse responses of $\Delta\pi_t$ and u_t partially canceling each other out, it becomes difficult to reject H_0, that is, difficult to reject combinations of a large (absolute) slope $|\lambda|$ and a small (absolute) γ_f.

2. Comparison with Traditional Methods. To put our results in the context of the literature, we also estimated the New Keynesian Phillips curve in the traditional way, using lagged macro variables as instruments. Our implementation follows Kleibergen and Mavroeidis (2009), and we use four lags of inflation and the forcing variable.

In addition, to more systematically explore how our estimates differ from those based on the traditional approach, we repeated our estimation procedure using different inflation measures and different gap measures. Specifically, we considered five popular measures of inflation: core PCE, PCE, core CPI, CPI, and the GDP deflator. For the unemployment gap, we considered the raw unemployment rate, the CBO unemployment gap, unemployment detrended with an HP-filter with $\lambda^{hp} = 1{,}600$, and unemployment

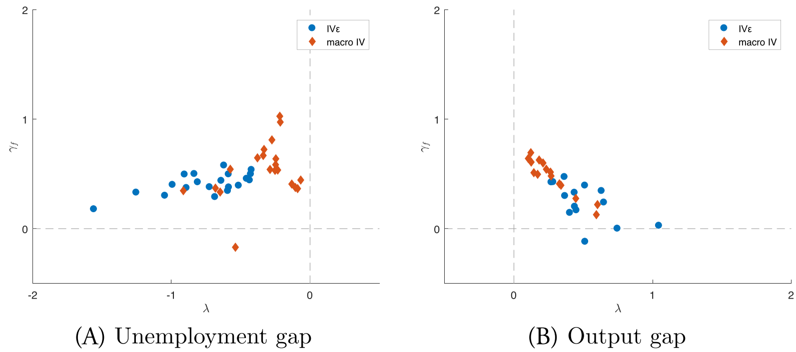


FIGURE IV

Point Estimates Using Shocks or Lagged Macro Variables as Instruments, 1969–2007

Phillips curve IV estimates for λ (the slope of the Phillips curve) and γ_f (the loading on inflation expectations) for different inflation and gap measures. The instruments are the Romer and Romer (2004) shocks (IVε, blue circles) or lagged macro variables (macro IV, red diamonds).

detrended with a smoother HP-filter with $\lambda^{hp} = 10^5$. For the output gap, we considered the CBO output gap, the output gap from an HP-filter with $\lambda^{hp} = 1{,}600$ and the output gap from an HP-filter with $\lambda^{hp} = 10^5$.

Figure IV reports the IV point estimates for the different combinations of inflation and gap measures. Two main conclusions emerge. First, our estimates for the slope of the Phillips curve are substantially larger (in absolute value) than the estimates based on using lagged macro variables as instruments. This finding is in line with what one would expect if the "lagged macro-instruments" violate the exclusion restriction because of serial correlation in the cost-push factors (Mavroeidis, Plagborg-Møller, and Stock 2014) or because of serial correlation in the measurement error in the forcing variable.[21] Second, our estimates for the coefficient on expected future inflation are substantially smaller than the estimates based on using lagged macro variables as instruments. This indicates that earlier methods have

21. Confounding with supply factors will lead to a downward bias in the lagged macro-instrument estimates, because supply shocks lead to a positive correlation between inflation and the unemployment gap. Measurement error in the forcing variable will also lead to downward bias coming from attenuation.

tended to overestimate the role of forward-looking inflation expectations.

V.B. Identification from HFI Monetary Surprises, 1990–2017

Our results based on the full 1969–2007 sample mix very different policy regimes. In fact, a number of Phillips curve–based studies have suggested substantial changes in the persistence of inflation as well as in the magnitude of the inflation-unemployment trade-off; from the close to unit-root behavior of inflation in the 1970s (e.g., King and Watson 1994) to the flattening of the Phillips curve in the post-1990 period (e.g., Ball and Mazumder 2011; Blanchard 2016).

In this section, we use HFI monetary surprises—changes in bond/futures prices around FOMC announcements—to estimate the Phillips curve over the more recent 1990–2017 period, a period with a relatively stable policy regime. As the instruments, we take the sum of the three-month-ahead monthly Fed Funds futures, which capture variations in the Fed Funds futures prior to the zero-lower-bound period (see Gertler and Karadi 2015), and surprises to the 10-year yield, which capture interest rate variations from slope policies in the post-2007 period (see Eberly, Stock, and Wright 2019).[22] Given the short sampling period, we impose the restriction $\gamma_f + \gamma_b = 1$.

Table II displays the Almon-restricted IV point estimates for γ_f and λ along with the weak-IV robust confidence intervals derived from the subset $AR_{a,s}$ statistic. Similarly to Figure II, Figure V plots the confidence sets for γ_f and λ.

Before we contrast our HFI results based on the more recent 1990–2017 period with our results based on the 1969–2007 Romer and Romer (RR) monetary shocks, we note that comparing estimates across different identification schemes (HFI versus RR) can be challenging. As we saw earlier, HFI and RR instruments have potential imperfections. Because these imperfections are different for HFI and for RR, differences in results across identification schemes could be caused by differences in

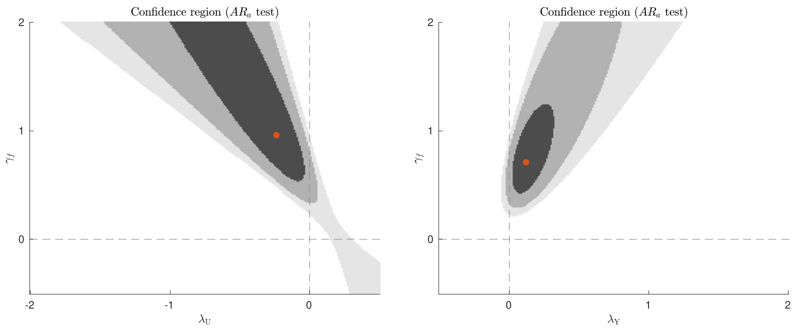
22. Intuitively, since the relevant interest rate for economic decisions is a longer-term yield like the 10-year yield, our goal is to capture as much exogenous variation in the 10-year yield as possible. While taking the sum of FF4 and 10-year yield surprises is a crude way to capture exogenous variations in the 10-year yield over the 1990–2017 period, a regression of the 10-year yield on these two surprises show that both terms enter significantly and with roughly equal coefficients.

TABLE II

THE PHILLIPS CURVE, 1990–2017, HFI ID

γ_f	0.96	$[-\infty\ _{0.50},\ \ \infty\ _{14.88}]$
λ_U	−0.24	$[-\infty\ _{-6.72},\ \ \infty\ _{-0.02}]$
γ_f	0.71	$[\ 0.34\ _{0.40},\ \ 1.86\ _{1.38}]$
λ_Y	0.12	$[-0.01\ _{0.01},\ \ 0.56\ _{0.37}]$

Notes. The table reports the parameter estimates and weak-IV robust confidence intervals for the U.S. Phillips curve (1990–2017). We show the Almon-restricted IV point estimates based on the high-frequency identified (HFI) monetary surprises as instruments, the $AR_{a,s}$-based 95% confidence bounds, and in subscript the $AR_{a,s}$-based 90% confidence bounds. The forcing variable is the unemployment gap λ_U or the output gap λ_Y.



(A) Unemployment gap (B) Output gap

FIGURE V

The Phillips Curve, 1990–2017, HFI id, $\gamma_f + \gamma_b = 1$

Robust confidence sets for the Phillips curve coefficients obtained by inverting the AR_a test. 95%, 90%, and 68% confidence sets for λ (the slope of the Phillips curve) and γ_f (the loading on inflation expectations). The estimation is based on using the high-frequency identified (HFI) monetary surprises as instruments for 1990–2017. The black (red online) dot is the Almon-restricted IV estimate. The specification imposes $\gamma_f + \gamma_b = 1$ with the unemployment gap (left) or the output gap (right) as the forcing variable.

imperfections and not by genuine changes in the underlying Phillips curve.

With this caveat in mind, we note two main differences. First, in terms of point estimates, the slope of the Phillips curve is substantially smaller with the HFI identification scheme, about half as large but still marginally significant, whereas the coefficient on expected future inflation is larger. In terms of confidence sets, the sets obtained with HFI instruments are markedly different from those obtained with the RR instruments, notably in terms of their shape and main orientation. Specifically, while the confidence sets in Figure II clearly exclude large values for γ_f, the opposite holds in Figure V, where the confidence sets are in positive territory for γ_f and in fact cannot exclude large values for γ_f. Although only suggestive, these results are consistent with a change in the main determinants of inflation since 1990, with forward-looking inflation expectations playing a larger role, and slack playing a smaller role.

VI. Conclusion

In this article, we used sequences of structural shocks as instrumental variables to address endogeneity issues and obtain consistent estimates of forward-looking structural equations including the Phillips curve, the dynamic IS curve, and the interest rate rule. We showed that the Anderson-Rubin statistic can be used to conduct inference in a powerful way that is robust to the weak-instruments problem. In our empirical work we have shown that the methodology is able to give new insights into the Phillips curve literature.

Looking beyond the current article, the impulse response interpretation associated with using sequences of structural shocks allows for further methodological developments. While we propose one refinement based on parameterizing the residual impulse response as a polynomial function, using structural shocks as instruments allows us to exploit many other features of impulse response functions. Examples include (i) combining different types of structural shocks (for instance, different types of aggregate demand shocks) so as to also exploit variation across impulse responses to improve inference, (ii) exploiting nonlinearities in the impulse responses to structural shocks and (iii) exploiting time variation in the impulse responses to shocks (e.g., Magnusson and Mavroeidis 2014).

Moreover, while the present article focuses on estimating linear equations, shocks, instead of predetermined variables, can also be used as instruments to estimate nonlinear forward-looking equations, which is of high relevance for the asset pricing literature (Hansen and Singleton 1982).

APPENDIX A: THE RANK CONDITION FOR A FORWARD-LOOKING STRUCTURAL EQUATION

Consider the general forward-looking structural equation

$$y_t = \gamma_b y_{t-1} + \gamma_f \mathrm{E}_t y_{t+1} + \lambda x_t + e_t, \tag{21}$$

and for tractability assume that the forcing variable follows an AR(1)

$$x_t = \rho x_{t-1} + \varepsilon_t + \nu e_t, \tag{22}$$

with e_t and ε_t some i.i.d. shocks, and γ_b, γ_f, λ, ρ, and ν parameters of the model.

PROPOSITION 1. The model characterized by equations (21) and (22) can be identified using the sequence of shocks $z_t = \varepsilon_{t:\,t-3}$ as instruments if and only if $\gamma_b \neq 0$ and $\delta_1 \neq -\rho - \rho(\rho+1)$ with δ_1 the stable root of the second-order difference equation (21).

Proof. Solving for x_t and y_t, we get

$$\begin{cases} x_t = \sum_{j=0}^{\infty} \rho^j \left(\varepsilon_{t-j} + \nu e_{t-j}\right) \\ y_t = \delta_1 y_{t-1} + \frac{\lambda}{\delta_2 \gamma_f} \sum_{j=0}^{\infty} \left(\frac{1}{\delta_2}\right)^j \mathrm{E}_t x_{t+j} \end{cases},$$

where δ_1 and δ_2 are the stable and unstable roots of the second-order difference equation given by equation (21).[23]

Some simple algebra for $z_t = \varepsilon_{t:\,t-3}$ then gives

$$\Gamma = \mathrm{E}\left(w_t z_t'\right) = \begin{pmatrix} 1 & \rho & \rho^2 \\ \delta_1\kappa + \rho\kappa & \delta_1(\delta_1\kappa + \rho\kappa) + \rho^2\kappa & \delta_1\kappa\left(\rho^2 + \rho\delta_1 + \delta_1^2\right) + \rho^3\kappa \\ 0 & \kappa & \delta_1\kappa + \rho\kappa \end{pmatrix}$$

with $\kappa = E(\pi_t \varepsilon_t) = \frac{\lambda}{\delta_2 \gamma_f (1-\rho/\delta_2)} \neq 0$.[24] $\det \Gamma = \kappa \delta_1^2(\rho + \delta_1 + \rho(\rho+1))$, so that the rank condition is satisfied if $\delta_1 \neq 0$, that is, if $\gamma_b \neq 0$. □

Although based on a simple DGP for the output gap, Proposition 1 shows that a necessary condition for our approach to be valid is that past inflation helps determine future inflation, that is, that inflation cannot be strictly forward looking ($\gamma_b \neq 0$). We can relax this assumption at the expense of assuming more

23. We have $\delta_1 = \frac{1-\sqrt{1-4\gamma_b\gamma_f}}{2\gamma_f}$ and $\delta_2 = \frac{1+\sqrt{1-4\gamma_b\gamma_f}}{2\gamma_f}$.

24. This follows from the recursion $E\pi_t\varepsilon_{t-j}^m = \delta_1 E\pi_t\varepsilon_{t-j+1}^m + \rho^j\kappa$, for $j > 0$.

elaborate dynamics for the forcing variable. In particular, γ_b can be equal to zero if the forcing variable follows an AR(2) process.

APPENDIX B: ASYMPTOTIC THEORY

We discuss an asymptotic theory for the Almon-restricted AR statistic AR_a and its subset counterpart $AR_{a,s}$ in which we allow the number of lags H to increase with the sample size, for example, $\frac{H}{n} \to c \in (0, 1)$ as $n \to \infty$. This is important because H corresponds to the number of lagged structural shocks included, and since we typically want to allow for $H \approx 20$ to capture sufficient variation in the endogenous variables, our theory needs to reflect that H is proportional to n (see Richardson and Stock 1989; Valkanov 2003 for similar arguments).

The AR_a and $AR_{a,s}$ statistics both depend on the long-run variance estimate $\hat{s}_u^2$ which we assume to be the form $\hat{s}_u^2 = \frac{1}{n-H} \sum_{t=H+1}^{n} \sum_{s=H+1}^{n} \hat{u}_t \hat{u}_s \kappa(\frac{t-s}{b_n})$, where $\hat{u}_t = (y_t - w_t'\delta) - z_t^{i'}\hat{\theta}_a$ and the kernel function $\kappa()$ has bandwidth parameter b_n which is increasing in n.[25] The exact assumptions for $\kappa()$ are spelled out below, but include the standard Newey and West (1987) approach and many others.

The limiting distributions of $AR_a[\delta_0]$ and $AR_{a,s}[\beta_0]$ can be characterized in terms of the behavior of the partial sums of the disturbances and structural shock proxies. To ensure the applicability of a functional central limit theorem, we impose mild moment and dependence assumptions. Our dependence assumptions rely on the concept of near epoch dependent (NED) stochastic processes for which we use the following definition (Davidson 1994, Definition 17.1; see also Gallant and White 1987).

DEFINITION 1. A sequence of integrable random vectors $\{X_t\}$ is L_2-NED on a stochastic sequence $\{V_t\}$ on probability space $(\Omega, \mathcal{F}, P)$ if for $m \geqslant 0$

$$\|X_t - \mathrm{E}(X_t|\mathcal{F}_{t-m}^{t+m})\|_2 < d_t \nu_m,$$

where $\mathcal{F}_s^t = \sigma(V_s, \ldots, V_t) \subset \mathcal{F}$, $t \geqslant s$, d_t is a sequence of constants, and $\nu_m \to 0$ as $m \to \infty$.

25. Alternatively, we can also directly impose H_0 and consider $s_u^2 = \frac{1}{n-H} \sum_{t=H+1}^{n} \sum_{s=H+1}^{n} u_t u_s \kappa\left(\frac{t-s}{b_n}\right)$, where $u_t = y_t - w_t'\delta_0$. These variance estimates are asymptotically equivalent as proven in Lemma 5 in the Online Appendix.

We will say that the sequence is L_2-NED of size $-s$ when $\nu_m = O(m^{-s-\varepsilon})$ for some $\varepsilon > 0$. Using this definition we impose the following assumptions.

ASSUMPTION 1. The observations $\{y_t, w_t\}$ are generated by the linear IV model

$$\begin{aligned} y_t &= w_t'\delta + u_t \\ &= w_{\beta,t}'\beta + w_{\alpha,t}'\alpha + u_t \end{aligned}$$

$$\underbrace{\begin{pmatrix} w_{\beta,t} \\ w_{\alpha,t} \end{pmatrix}}_{w_t} = \underbrace{\begin{pmatrix} \Pi_\beta' \\ \Pi_\alpha' \end{pmatrix}}_{\Pi'} z_t^i + \underbrace{\begin{pmatrix} v_{\beta,t} \\ v_{\alpha,t} \end{pmatrix}}_{v_t}, \qquad t = H+1, \ldots, n,$$

where $w_t = (w_{\beta,t}', w_{\alpha,t}')'$ and $\delta = (\beta', \alpha')'$ are $m \times 1$, with $m = 3$; β, $w_{\beta,t}$, and $v_{\beta,t}$ are $m_\beta \times 1$; α, $w_{\alpha,t}$, and $v_{\alpha,t}$ are $m_\alpha \times 1$; $m = m_\alpha + m_\beta$; Π is $3 \times m$; Π_α is $3 \times m_\alpha$; Π_β is $3 \times m_\beta$; $z_t^i = \left(\sum_{h=0}^H \xi_{t-h}^i, \sum_{h=0}^H h\xi_{t-h}^i, \sum_{h=0}^H h^2\xi_{t-h}^i\right)'$; and let $\eta_t = (\xi_t^i, u_t, v_t')'$. We assume that

(i) for all t, s we have (a) $\mathrm{E}(\eta_t) = 0$, (b) $\mathrm{E}(u_t\xi_s^i) = 0$, and (c) $\mathrm{E}(v_t\xi_s^i) = 0$,

(ii) for some $r > 2$ and finite constant Δ we have $\sup_t \|\eta_t\|_{2r} \leqslant \Delta$,

(iii) η_t is L_2-NED of size $-\frac{r-1}{r-2}$ with $d_t = 1$ on V_t, where $\{V_t\}$ is an α-mixing process of size $-\frac{r}{r-2}$,

(iv) for integers $p, q \geqslant 0$ we have uniformly in n and H, with $H < n$, that

$$\begin{aligned} \omega_{\xi,p,n,H}^2 &= \mathrm{Var}\left(\sum_{t=H+1}^n t^p \xi_t^i\right) \\ &= \omega_{\xi,p}^2 (n-H)^{2p+1} + o\big((n-H)^{2p+1}\big) \\ \Omega_{uv,q,n,H} &= \mathrm{Var}\left(\sum_{t=H+1}^n t^q \left(u_t, v_t'\right)'\right) \\ &= \Omega_{uv,q}(n-H)^{2q+1} + o\big((n-H)^{2q+1}\big), \end{aligned}$$

with finite $\omega_{\xi,p}^2 > 0$ and $\Omega_{uv,q} \succ 0$.

(v) $b_n = o(n)$ and $\kappa(\cdot) \in \mathcal{K}$ where

$$\mathcal{K} = \Big\{ \kappa(\cdot) : \mathbb{R} \to [-1, 1],\ \kappa(0) = 1,$$
$$\kappa(x) = \kappa(-x)\ \forall x \in \mathbb{R}, \int_{-\infty}^{\infty} |\kappa(x)| \mathrm{d}x < \infty,$$
$$\int_{-\infty}^{\infty} \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} \kappa(x) e^{ivx} \mathrm{d}x \right| \mathrm{d}v < \infty,$$
$$\kappa(\cdot) \text{ is continuous at 0 and at all but a finite number of points.} \Big\}$$

Note that no assumptions are placed on the matrix Π, which leaves the strength of the instruments z_t^i unrestricted. The first assumption imposes that the shocks are mean zero and more important that the structural shock proxies are uncorrelated, at all leads and lags, with u_t and v_t. Note that these conditions correspond to the definition of a structural shock in Ramey (2016) and are the same as in the LP-IV and SVAR-IV literature (see condition LP-IV in Stock and Watson 2018, 924). Parts (ii) and (iii) of the assumption impose mild restrictions on the dependence, heterogeneity, and moments of η_t. Importantly, they allow for serial correlation and heteroskedasticity in the structural shock proxies ξ_t^i and the error term u_t, which is deemed important in Alloza, Gonzalo, and Sanz (2019) and Zhang and Clovis (2010), respectively. Part (iv) defines the convergence rate of the long-run variance, which is standard apart from the additional rescaling to account for the fact that the standard deviations are proportional to t^p (see also Wooldridge and White 1988, example 2.12). In our setting, this form of explosive variance is caused by the polynomial instruments z_t^i. Part (v) allows for a rich class of kernel functions for the estimation of $\hat{s}_u^2$. In particular, the class includes the Bartlett, Parzen, quadratic spectral, and Tukey-Hanning kernels, see de Jong and Davidson (2000). Also, the assumption bounds the bandwidth parameter at a rate that is similar as in Andrews (1991) and de Jong and Davidson (2000).

To formalize the result for the subset statistic we follow Guggenberger et al. (2012) and define the parameter space Φ for the parameters $(\alpha, \Pi_\alpha, \Pi_\beta, F)$, where β is omitted as it is fixed under the subset hypothesis H_0: $\beta = \beta_0$ and F summarizes the distribution of the shocks $\{\eta_t\}$, with $\eta_t = (\xi_t^i, u_t, v_t')'$. We define Φ, under H_0: $\beta = \beta_0$, as follows

$$\Phi = \big\{\phi = (\alpha, \Pi_\alpha, \Pi_\beta, F) : \alpha \in \mathbb{R}^{m_\alpha}, \Pi_\alpha \in \mathbb{R}^{3\times m_\alpha}, \Pi_\beta \in \mathbb{R}^{3\times m_\beta}, \\ F \text{ satisfies Assumptions 1.i–1.iv}\big\}.$$

The asymptotic size of the subset AR statistic is defined as

$$\text{AsySz}_{AR_{a,s}} = \limsup_{n\to\infty,\, \frac{H}{n}\to c\in(0,1)} \sup_{\phi\in\Phi} \mathbb{P}_\phi\big(AR_{a,s}[\beta_0] > \chi^2_{1-\alpha}(m_\beta)\big),$$

where $\mathbb{P}_\phi$ denotes the probability of an event when the null data-generating process is pinned down by $\phi \in \Phi$ and $\chi^2_{1-\alpha}(m_\beta)$ denotes the $1-\alpha$ critical value of the χ^2 distribution with m_β degrees of freedom.

Give these definitions we have the following result.

THEOREM 1. Let Assumption 1 hold. Under H_0: $\delta = \delta_0$ for $\frac{H}{n} \to c \in (0, 1)$ as $n \to \infty$ we have that

$$AR_a[\delta_0] \xrightarrow{d} \chi^2(3),$$

and also, under H_0: $\beta = \beta_0$, we have that

$$\text{AsySz}_{AR_{a,s}} = \alpha.$$

The proof of this result is deferred to the Online Appendix. Intuitively, the first result in Theorem 1 is very similar to Park and Phillips (1988; see their theorem 5.4), where it is shown that under a strict exogeneity assumption, the Wald statistic defined by a regression with nonstationary explanatory variables has a χ^2 limit. The differences in our setting are caused by the nonstandard integration limits and the explosive variances, but the intuition for the result is similar. The second result in Theorem 1 follows similarly as in Guggenberger et al. (2012), where the key insight is that for $\frac{H}{n} \to c \in (0, 1)$ the limiting distribution of the appropriately scaled sums $\sum_{t=H+1}^{n} z_t^i(u_t, v_t')'$ converges to a normally distributed random vector whose variance, conditional on the instruments, has a Kroneker product structure. The latter is a key requirement for the second result in Theorem 1 and hinges crucially on the strict exogeneity of the instruments.

Appendix C: General Structural Equations

In general, the structural macro equation of interest may not have three coefficients, or the researcher may want to use multiple sequences of structural shock proxies. To outline our methodology for this more general case let w_t be an arbitrary $L \times 1$ vector of endogenous variables and let z_t denote the $\dim(z) \times 1$ vector of structural-shock polynomial instruments. For instance, if $\{\xi_t^1\}$ and $\{\xi_t^2\}$ are two sequences of structural shocks we may consider $z_t = \left(\sum_{h=0}^{H} \xi_{t-h}^1, \sum_{h=0}^{H} h\xi_{t-h}^1, \sum_{h=0}^{H} h^2\xi_{t-h}^1, \sum_{h=0}^{H} \xi_{t-h}^2, \sum_{h=0}^{H} h\xi_{t-h}^2,\right.$ $\left.\sum_{h=0}^{H} h^2\xi_{t-h}^2\right)'$. We require that $\dim(z_t) \geqslant L$ and may compute the AR_a statistic for testing H_0: $\delta = \delta_0$ similarly as in equation (17) with z_t replacing z_t^i. In this case we have that $AR_a[\delta_0] \xrightarrow{d} \chi^2(\dim(z_t))$ when $\frac{H}{n} \to c \in (0, 1)$ as $n \to \infty$. Furthermore, if we are interested in testing the subset hypothesis H_0: $\beta = \beta_0$ given $\delta = (\alpha', \beta')'$ we consider the subset Almon-AR statistic $AR_{a,s}[\beta_0]$. Under similar assumptions as in the previous section we then have that the limiting distribution of the $AR_{a,s}[\beta_0]$ statistic is upper bounded by a χ^2 random variable with $\dim(z_t) - \dim(\alpha)$ degrees of freedom. Note that in our baseline Theorem 1, with exact identification, we have that $\dim(z_t) - \dim(\alpha) = \dim(\beta)$.

In overidentified settings the degrees of freedom increase proportionally to the number of instruments. Hence it might be advantageous to rely on alternative weak-instrument robust statistics, such as the conditional likelihood ratio statistic; see Andrews, Stock, and Sun (forthcoming) for more discussion.

Appendix D: Simulation Evidence

In this section we discuss the results from a simulation study that is designed to evaluate the finite sample performance of the methodology. We concern ourselves with testing the hypothesis H_0: $\delta = \delta_0$ and the subset hypothesis H_0: $\lambda = \lambda_0$ using different methods based on using structural shocks as instruments. The Online Appendix provides additional simulation results for different data-generating processes.

Simulation Design

We consider the following data-generating process

$$\begin{aligned} y_t &= \gamma_b y_{t-1} + \gamma_f \mathrm{E}_t(y_{t+1}) + \lambda x_t + e_t \\ x_t &= \rho_1 x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t^i + \nu e_t, \end{aligned} \tag{23}$$

where the forcing variable x_t follows an AR(2) process. Model (23) has two shocks: e_t and ε_t^i. We assume, without loss of generality, that ξ_t^i, our instrument for ε_t^i, satisfies $\varepsilon_t^i = \xi_t^i$. Furthermore, we emphasize that although model (23) is highly stylized, it includes all the elements that are required to evaluate our methodology. The choice of an AR(2) process is motivated by the time series properties of the output and unemployment gaps.

The following parameter configurations are considered. For the structural equation we fix $\lambda = 0.4$, $\gamma_b = 0.6$, and $\gamma_f = 0.3$. These parameters are close to our empirical findings for the Phillips curve. For the forcing variable we match ρ_1 and ρ_2 to the fitted values that are obtained from considering the unemployment gap: $\rho_1 = 1.2$ and $\rho_2 = -0.4$. We fix $\nu = -1$ to mimic the intuition that cost-push shocks should increase inflation and reduce output.

To consider realistic values for the structural shock variances we match the configuration of the shocks to the recent findings for monetary policy shocks from Gorodnichenko and Lee (forthcoming), Plagborg-Møller and Wolf (2018), and Caldara and Herbst (2019). Using different methodologies, they find that monetary shocks are able to explain only a small portion of the variance observed in output and inflation. For instance, Gorodnichenko and Lee (forthcoming) find that at least between 10% and 20% of the fluctuations in output are driven by monetary policy shocks and about 10% of the fluctuations in inflation.[26] Similarly, Plagborg-Møller and Wolf (2018) find that, under weaker assumptions, the monetary policy shocks can explain at most 30% of the variation in output and 8% of the variation in inflation, but cannot reject zero influence of monetary policy shocks.

To match these numbers we proceed as follows. The shocks are generated from $\varepsilon_t^i \sim N(0, \sigma_i^2)$, with standard deviation $\sigma_i =$ 0.1, 0.25, 0.5, 1, and $e_t = \rho e_{t-1} + \sqrt{1-\rho^2}\zeta_t$ with $\zeta_t \sim N(0,\ 1)$. This implies that we can distinguish between different scenarios. When $\sigma_i = 0.1$ the structural-shock instrument explains approximately 1% of the variance in the outcome variable y_t and 2% of the variance in the forcing variable x_t. These percentages increase when we increase σ_i. In Table A.I we provide the details. The last scenario where $\sigma_i = 1$ is perhaps overoptimistic as the structural shock explains over 50% of the variation, but scenarios where $\sigma_i = 0,\ 1,\ 0.25,\ 0.5$ all correspond to empirical findings

26. When using local projection methods, they find substantially larger influences of the monetary shocks.

TABLE A.I

SIMULATION DESIGN: VARIANCE DECOMPOSITION FOR STRUCTURAL SHOCKS

σ_i^2	V(y)	V(x)
0.10	1%	2%
0.25	6%	11%
0.50	20%	30%
1.00	50%	67%

Notes. The table reports the details for the different simulation designs considered. We show the average percentage of variance explained by the structural shock in the variables y_t and x_t, respectively. The remainder of the variance is explained by the shock e_t. See Appendix D for more details.

for monetary policy shocks, for example, Gorodnichenko and Lee (forthcoming), Plagborg-Møller and Wolf (2018), and Caldara and Herbst (2019). The parameter ρ allows for serial correlation in the disturbance e_t and we consider the values $\rho = 0$ or $\rho = 0.5$.

For each combination of parameter values and sample sizes $n = 200, 500$ we simulate 5,000 data sets and for each data set we test the hypotheses H_0: $\delta = \delta_0$ and H_0: $\lambda = \lambda_0$ using the methodology outlined in Section IV. The choice for λ is arbitrary and similar results can be obtained for subset tests for γ_b and γ_f. For the hypothesis H_0: $\delta = \delta_0$ we consider the standard Wald test based on the 2SLS estimator,[27] the standard Wald test based on the Almon-restricted IV estimator equation (18), the standard AR test given in equation (13), and our preferred Almon (1965) restricted AR_a test as defined in equation (17). For the subset hypothesis H_0: $\lambda = \lambda_0$ we consider the $AR_{a,s}$ statistic. All tests are implemented using $H = 20$ or $H = 40$ shocks as instruments. Note that for the Almon-restricted Wald test, the AR_a test, and the $AR_{a,s}$ test the effective number of instruments remains three regardless of the value of H. We vary the value of H to investigate the influence of the persistence in the Almon-restricted instruments.

Results

We report the average rejection frequencies ($\alpha = 0.05$ level) for the different test statistics for H_0: $\delta = \delta_0$ in Table A.II. We find the following patterns. First, the standard Wald statistic based on the normal limiting distribution of the 2SLS estimator is severely oversized when the strength of the instruments is small. This

27. That is we consider $\hat{\delta}^{IV}$ as in equation (11), where the weighting matrix is taken as $S_{\xi\xi}^{-1}$ where $S_{\xi\xi} = \frac{1}{n}\sum_{t=1}^{n} \xi_{t:t-H}\xi_{t:t-H}^{'}$. Different choices for the weighting matrix do not change the conclusions below.

TABLE A.II
SIMULATION RESULTS: REJECTION FREQUENCIES

n	H	σ_i^2	ρ	IV-ε	IV$_a$-ε	AR	AR_a	$AR_{a,s}$
200	20	0.10	0.0	0.528	0.352	0.590	0.057	0.007
200	20	0.25	0.0	0.369	0.352	0.586	0.066	0.026
200	20	0.50	0.0	0.140	0.212	0.991	0.056	0.049
200	20	1.00	0.0	0.001	0.048	0.993	0.067	0.057
200	40	0.10	0.0	0.773	0.370	0.990	0.048	0.000
200	40	0.25	0.0	0.574	0.336	0.990	0.051	0.005
200	40	0.50	0.0	0.140	0.192	0.097	0.058	0.012
200	40	1.00	0.0	0.024	0.060	0.094	0.052	0.027
500	20	0.10	0.0	0.507	0.426	0.259	0.055	0.013
500	20	0.25	0.0	0.346	0.382	0.262	0.060	0.039
500	20	0.50	0.0	0.047	0.242	0.260	0.061	0.059
500	20	1.00	0.0	0.000	0.060	0.250	0.059	0.057
500	40	0.10	0.0	0.732	0.444	0.722	0.052	0.003
500	40	0.25	0.0	0.518	0.398	0.732	0.048	0.012
500	40	0.50	0.0	0.072	0.245	0.716	0.050	0.033
500	40	1.00	0.0	0.000	0.052	0.708	0.052	0.042
200	20	0.10	0.5	0.781	0.560	0.530	0.042	0.009
200	20	0.25	0.5	0.694	0.567	0.534	0.040	0.016
200	20	0.50	0.5	0.500	0.508	0.533	0.044	0.038
200	20	1.00	0.5	0.108	0.315	0.538	0.041	0.046
200	40	0.10	0.5	0.948	0.578	0.981	0.055	0.000
200	40	0.25	0.5	0.915	0.586	0.981	0.051	0.003
200	40	0.50	0.5	0.745	0.515	0.980	0.059	0.011
200	40	1.00	0.5	0.160	0.318	0.980	0.060	0.028
500	20	0.10	0.5	0.739	0.589	0.216	0.039	0.009
500	20	0.25	0.5	0.669	0.610	0.219	0.037	0.026
500	20	0.50	0.5	0.386	0.527	0.216	0.039	0.041
500	20	1.00	0.5	0.042	0.319	0.227	0.040	0.047
500	40	0.10	0.5	0.930	0.651	0.629	0.052	0.001
500	40	0.25	0.5	0.896	0.655	0.635	0.052	0.008
500	40	0.50	0.5	0.655	0.561	0.642	0.049	0.028
500	40	1.00	0.5	0.061	0.350	0.659	0.060	0.044

Notes. The table reports the empirical rejection frequencies for H_0: $\delta = \delta_0$ and (in the rightmost column) H_0: $\lambda = \lambda_0$, both with level $\alpha = 0.05$. For the IV-ε estimator these correspond to the Wald statistic based on the limiting distribution of the 2SLS estimator equation (11) with H instruments. The IV$_a$-ε corresponds to the Wald statistic based on the limiting distribution of the Almon-restricted 2SLS estimator equation (18). The AR column corresponds to the test based on the Anderson-Rubin statistic that was computed using H structural shocks as instruments. The AR_a column corresponds to the test based on the Anderson-Rubin statistic with Almon restriction as defined in equation (17). The $AR_{a,s}$ column corresponds to the test based on the subset Anderson-Rubin statistic with Almon restriction as defined in equation (19).

holds for both the Almon-restricted Wald test and the unrestricted version that uses H instruments. The empirical rejection frequency is much larger when compared with the nominal size when the variance of the structural shocks is relatively small, for example, $\sigma_i = 0.1, 0.25, 0.5$. The Almon-restricted version performs slightly better because it only suffers from the weak-instruments problem and not from the many instruments problem. The unrestricted Wald test is unreliable across all specifications.

Furthermore, the conventional AR statistic (denoted by AR) based on H structural shocks is severely oversized. This corresponds to the theoretical derivations of Andrews and Stock (2007), who show that the AR test is only correctly sized when $\frac{H^3}{n} \to 0$. This is clearly not the case in the current setting where $H = 20, 40$ and $n = 200, 500$.

In contrast, Table A.II clearly shows that the AR test with Almon restriction is always correctly sized. That is, for any combination of n, H, σ_i^2, and ρ the empirical rejection frequency is close to the nominal $\alpha = 0.05$ level. This indicates that the AR_a test with Almon restriction can be used for empirical work.

The average rejection frequencies for the subset statistic for H_0: $\lambda = \lambda_0$ are shown in the rightmost column of Table A.II. We find that the subset $AR_{a,s}$ statistic has rejection frequency close to 0.05 for strong instruments, that is, $\sigma_i = 1$. When the instruments are weak, the $AR_{a,s}$ statistic is conservative, having rejection frequencies that are smaller then 0.05. This is in line with our asymptotic theory which shows that the $AR_{a,s}$ statistic is asymptotically upper bounded by a $\chi^2(1)$ random variable. Note that when H increases the effective strength of the instruments goes down, because in the underlying model the influence of the structural shocks dies out exponentially fast. This implies that distant shocks do not explain much variance in the endogenous variables, thus making the Almon-type instruments weaker and leading to a more conservative subset test.

In the Online Appendix we show a number of additional results. First, we consider scenarios with different forms of heteroskedasticity and serial correlation in the structural shocks u_t. The results for these cases are the same as in Table A.II. Second, in a recent paper, Eberly, Stock, and Wright (2019) adopt the methodology of this article and extend it by considering an alternative way of reducing the number of instruments by an exponential weighted moving average approach. In the Online Appendix we discuss the results from a simulation study that compares the different approaches. We find both methods excellently control the size of the Anderson-Rubin statistic and do not differ much in power.

Federal Reserve Bank of San Francisco
Universitat Pompeu Fabra, Barcelona Graduate School of Economics, and Vrije Universiteit Amsterdam

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online.

Data Availability

Data and code replicating tables and figures in this article can be found in Barnichon and Mesters (2020), in the Harvard Dataverse, doi: 10.7910/DVN/2RR5SC.

References

Alloza, Mario, Jesus Gonzalo, and Carlos Sanz, "Dynamic Effects of Persistent Shocks," Banco de España Working Paper no. 1944, 2019.

Almon, Shirley, "The Distributed Lag between Capital Appropriations and Expenditures," *Econometrica*, 33 (1965), 178–196.

Anderson, Theodore W., and Herman Rubin, "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20 (1949), 46–63.

Andrews, Donald W. K., "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59 (1991), 817–858.

Andrews, Donald W. K., and James H. Stock, "Testing with Many Weak Instruments," *Journal of Econometrics*, 138 (2007), 24–46.

Andrews, Isaiah, James Stock, and Liyang Sun, "Weak Instruments in IV Regression: Theory and Practice," *Annual Review of Economics* (forthcoming).

Ascari, Guido, Leandro M. Magnusson, and Sophocles Mavroeidis, "Empirical Evidence on the Euler Equation for Consumption in the US," *Journal of Monetary Economics*, forthcoming.

Ball, Laurence, and Sandeep Mazumder, "Inflation Dynamics and the Great Recession," *Brookings Papers on Economic Activity*, 42 (2011), 337–405.

Barnichon, Regis, and Geert Mesters, "Replication Data for: 'Identifying Modern Macro Equations with Old Shocks'," (2020), Harvard Dataverse, doi: 10.7910/DVN/2RR5SC.

Barth, Marvin J. III, and Valerie A. Ramey, "The Cost Channel of Monetary Transmission," *NBER Macroeconomics Annual*, 16 (2001), 199–240.

Bekker, Paul A., "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62 (1994), 657–681.

Bernanke, Ben S., "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25 (1986), 49–99.

Blanchard, Olivier, "The Phillips Curve: Back to the '60s?," *American Economic Review*, 106 (2016), 31–34.

Blanchard, Olivier, and Mark W. Watson, "Are Business Cycles All Alike?" in *The American Business Cycle: Continuity and Change* (Cambridge, MA: National Bureau of Economic Research, 1986), 123–180.

Caldara, Dario, and Edward Herbst, “Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs,” *American Economic Journal: Macroeconomics*, 11 (2010), 157–192.
Campbell, John Y., “Consumption-Based Asset Pricing,” *Handbook of the Economics of Finance*, 1 (2003), 803–887.
Chernozhukov, Victor, Christian Hansen, and Michael Jansson, “Admissible Invariant Similar Tests for Instrumental Variable Regression,” *Econometric Theory*, 25 (2009), 806–818.
Clarida, Richard, Jordi Galí, and Mark Gertler, “Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory,” *Quarterly Journal of Economics*, 115 (2000), 147–180.
Cochrane, John, “Comments on ‘A New Measure of Monetary Shocks: Derivation and Implications’ by Romer and Romer,” NBER EFG Meeting, Technical Report, 2004.
Coibion, Olivier, Yuriy Gorodnichenko, and Mauricio Ulate, “The Cyclical Sensitivity in Estimates of Potential Output,” NBER Working Paper No. 23580, 2017.
Davidson, James, *Stochastic Limit Theory* (Oxford: Oxford University Press, 1994).
de Jong, Robert M., and James Davidson, “Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices,” *Econometrica*, 68 (2000), 407–423.
Deaton, Angus, *Understanding Consumption* (Oxford: Oxford University Press, 1992).
Eberly, Janice C., James H. Stock, and Jonathan H. Wright, “The Federal Reserve’s Current Framework for Monetary Policy: A Review and Assessment,” Harvard University Working Paper, 2019.
Fernald, John G., “A Quarterly, Utilization-Adjusted Series on Total Factor Productivity,” Federal Reserve Bank of San Francisco, Working Paper Series 2012-19, 2012.
Frisch, Ragnar, “Statistical Confluence Analysis by Means of Complete Regression Systems,” University Institute of Economics, Oslo, publication no. 5, 1934.
Fuhrer, Jeffrey C., “Habit Formation in Consumption and its Implications for Monetary-Policy Models,” *American Economic Review*, 90 (2000), 367–390.
Fuhrer, Jeffrey C., and Glenn D. Rudebusch, “Estimating the Euler Equation for Output,” *Journal of Monetary Economics*, 51 (2004), 1133–1153.
Galí, Jordi, *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and its Applications* (Princeton, NJ: Princeton University Press, 2015).
Galí, Jordi, and Mark Gertler, “Inflation Dynamics: A Structural Econometric Analysis,” *Journal of Monetary Economics*, 44 (1999), 195–222.
Gallant, A. Ronald, and Halbert L. White Jr., *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models* (Oxford: Basil Blackwell, 1987).
Gertler, Mark, and Peter Karadi, “Monetary Policy Surprises, Credit Costs, and Economic Activity,” *AEJ: Macroeconomics*, 7 (2015), 44–76.
Gorodnichenko, Yuriy, and Byoungchan Lee, “A Note on Variance Decomposition with Local Projections,” *Journal of Business and Economic Statistics* (forthcoming).
Guggenberger, Patrik, Frank Kleibergen, Sophocles Mavroeidis, and Linchun Chen, “On the Asymptotic Sizes of Subset Anderson-Rubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression,” *Econometrica*, 80 (2012), 2649–2666.
Gürkaynak, S., Refet, Brian Sack, and Eric Swanson, “The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models,” *American Economic Review*, 95 (2005), 425–436.
Hall, Robert E., “Intertemporal Substitution in Consumption,” *Journal of Political Economy*, 96 (1988a), 339–357.
———, “The Relation between Price and Marginal Cost in U.S. Industry,” *Journal of Political Economy*, 96 (1988b), 921–947.
Hamilton, James D., “What Is an Oil Shock?,” *Journal of Econometrics*, 113 (2003), 363–398.

Hansen, Lars Peter, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50 (1982), 1029–1054.
Hansen, Lars Peter, and Kenneth J. Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50 (1982), 1269–1286.
Kareken, John, and Robert M. Solow, "Lags in Monetary Policy," *Stabilization Policies* (1963), 14–96.
Kilian, Lutz, "Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?," *Review of Economics and Statistics*, 90 (2008), 216–240.
King, Robert G., and Mark W. Watson, "The Post-War U.S. Phillips Curve: A Revisionist Econometric History," *Carnegie-Rochester Conference Series on Public Policy*, 41 (1994), 157–219.
Kleibergen, Frank, "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70 (2002), 1781–1803.
Kleibergen, Frank, and Sophocles Mavroeidis, "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business and Economic Statistics*, 27 (2009), 293–311.
Kuttner, Kenneth N., "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market," *Journal of Monetary Economics*, 47 (2001), 523–544.
Magnusson, Leandro, and Sophocles Mavroeidis, "Identification Using Stability Restrictions," *Econometrica*, 82 (2014), 1799–1851.
Mavroeidis, Sophocles, "Identification Issues in Forward-Looking Models Estimated by GMM, with an Application to the Phillips Curve," *Journal of Money, Credit and Banking*, 37 (2005), 421–448.
———, "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence," *American Economic Review*, 100 (2010), 491–503.
Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H. Stock, "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve," *Journal of Economic Literature*, 52 (2014), 124–188.
McLeay, Michael, and Silvana Tenreyro, "Optimal Inflation and the Identification of the Phillips Curve," CFM Discussion Paper 1815, 2018.
Mertens, Karel, and Morten O. Ravn, "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *American Economic Review*, 103 (2013), 1212–1247.
Moreira, Marcelo J., "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71 (2003), 1027–1048.
———, "Tests with Correct Size when Instruments Can Be Arbitrarily Weak," *Journal of Econometrics*, 152 (2009), 131–140.
Nakamura, Emi, and Jón Steinsson, "Identification in Macroeconomics," NBER Working Paper No. 23968, 2017.
Newey, Whitney K., and Kenneth D. West, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55 (1987), 703–708.
Park, Joon Y., and Peter C. B. Phillips, "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4 (1988), 468–497.
Plagborg-Møller, Mikkel, and Christian K. Wolf, "Instrumental Variable Identification of Dynamic Variance Decompositions," Princeton University Working Paper, 2018.
Ramey, Valerie, "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, J. B. Taylor and H. Uhlig, eds. (Amsterdam: Elsevier, 2016).
Ramey, Valerie A., and Sarah Zubairy, "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data," *Journal of Political Economy*, 126 (2018), 850–901.
Ravenna, Federico, and Carl E. Walsh, "Optimal Monetary Policy with the Cost Channel," *Journal of Monetary Economics*, 53 (2006), 199–216.

Reiersol, Olav, "Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis," *Econometrica*, 9 (1941), 1–24.
Richardson, Matthew, and James H. Stock, "Drawing Inferences from Statistics Based on Multiyear Asset Returns," *Journal of Financial Economics*, 25 (1989), 323–348.
Romer, Christina D., and David H. Romer, "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review*, 90 (2000), 429–457.
———, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94 (2004), 1055–1084.
Staiger, Douglas, and James H. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65 (1997), 557–586.
Stock, James H., and Mark W. Watson, "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," in *Handbook of Macroeconomics*, vol. 2 (Amsterdam: Elsevier, 2016), 415–525.
———, "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *Economic Journal*, 128 (2018), 917–948.
Stock, James H., and Jonathan H. Wright, "GMM with Weak Identification," *Econometrica*, 68 (2000), 1055–1096.
Tenreyro, Silvana, and Gregory Thwaites, "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions," *American Economic Journal: Macroeconomics*, 8 (2016), 43–74.
Valkanov, Rossen, "Long-Horizon Regressions: Theoretical Results and Applications," *Journal of Financial Economics*, 68 (2003), 201–232.
White, Halbert L. Jr, *Asymptotic Theory for Econometrians*, 2nd ed. (San Diego, CA: Academic Press, 2000).
Wooldridge, Jeffrey M., and Halbert L. White Jr, "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory*, 4 (1988), 210–230.
Yogo, Motohiro, "Estimating the Elasticity of Intertemporal Substitution When Instruments are Weak," *Review of Economics and Statistics*, 86 (2004), 797–810.
Zhang, Chengsi, and Joel Clovis, "The New Keynesian Phillips Curve of Rational Expectations: A Serial Correlation Extension," *Journal of Applied Economics*, 13 (2010), 159–179.