

DISCUSSION PAPER SERIES

DP13765

IDENTIFYING MODERN MACRO EQUATIONS WITH OLD SHOCKS

Régis Barnichon and Geert Mesters

MONETARY ECONOMICS AND FLUCTUATIONS

IDENTIFYING MODERN MACRO EQUATIONS WITH OLD SHOCKS

Régis Barnichon and Geert Mesters

Discussion Paper DP13765

Published 28 May 2019

Submitted 23 May 2019

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **MONETARY ECONOMICS AND FLUCTUATIONS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Régis Barnichon and Geert Mesters

IDENTIFYING MODERN MACRO EQUATIONS WITH OLD SHOCKS

Abstract

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, of Euler equations for consumption or output, or of monetary policy rules--- have typically relied on using pre-determined 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 (i) substantially under-estimate the slope of the Phillips curve and (ii) over-estimate the role of forward-looking inflation expectations.

JEL Classification: C14, C32, E32, E52

Keywords: Structural equations, instrumental variables, Impulse Responses, Robust inference

Régis Barnichon - regis.barnichon@sf.frb.org
Federal Reserve Bank of San Francisco and CEPR

Geert Mesters - geert.mesters@upf.edu
Universitat Pompeu Fabra

Acknowledgements

We thank Christian Brownlees, Luca Fornaro, Patrik Guggenberger, Oscar Jorda, Frank Kleibergen, Sylvain Leduc, Karel Mertens, Barbara Rossi, Adam Shapiro, Jim Stock, Mark Watson and seminar participants at the San Francisco 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 paper 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 acknowledge 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).

IDENTIFYING MODERN MACRO EQUATIONS WITH OLD SHOCKS*

Regis Barnichon^(a) and *Geert Mesters*^(b)

^(a) Federal Reserve Bank of San Francisco and CEPR

^(b) Universitat Pompeu Fabra, Barcelona GSE and VU Amsterdam

May 23, 2019

Abstract

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, of Euler equations for consumption or output, or of monetary policy rules— have typically relied on using pre-determined 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 (i) substantially under-estimate the slope of the Phillips curve and (ii) over-estimate the role of forward-looking inflation expectations.

JEL classification: C14, C32, E32, E52.

Keywords: Structural equations, instrumental variables, impulse responses, robust inference.

*First draft: May 2018. We thank Christian Brownlees, Luca Fornaro, Patrik Guggenberger, Oscar Jorda, Frank Kleibergen, Sylvain Leduc, Karel Mertens, Barbara Rossi, Adam Shapiro, Jim Stock, Mark Watson and seminar participants at the San Francisco 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 paper 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 acknowledge 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).

1 Introduction

The estimation of structural forward-looking macroeconomic equations is a central tenet of the macroeconomic research agenda. Prominent examples include the estimation of aggregate supply equations like the New-Keynesian Phillips curve (e.g. Gali and Gertler, 1999) and the estimation of aggregate demand equations based on an Euler equation for output—the intertemporal IS curve— (e.g. Fuhrer and Rudebusch, 2004) and a monetary policy rule—the LM curve— (e.g. Clarida, Galí and Gertler, 2000). Additional important examples include the estimation of consumption Euler equations (e.g. Deaton, 1992) and consumption-based asset pricing equations (e.g. Campbell, 2003).

Obtaining reliable estimates for the structural coefficients of forward looking equations has been shown challenging because of 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, i.e. 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; Mavroeidis, 2010; Kleibergen and Mavroeidis, 2009; Mavroeidis, Plagborg-Møller and Stock, 2014). An oft-cited reason is that pre-determined variables are weak instruments.

In this work, we propose a new approach to estimate forward-looking macro equations.

Our approach consists in 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

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, such as oil price shocks.

Using sequences of structural shocks as instruments has an intuitive interpretation as a “regression in impulse response space”. By projecting the structural 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 of interest are replaced by their impulse responses (IRs) to the structural shock. Identification then comes from variation across the horizons of the impulse responses.

Because structural shocks are not necessarily strong instruments,² we rely on weak instrument robust methods for conducting inference, see Andrews, Stock and Sun (2019) 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

¹In 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, i.e., 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.

²Stated differently, the forecast-error variance contribution of the shocks to the macro variables of interest can be small (Gorodnichenko and Lee, 2017; Plagborg-Møller and Wolf, 2018).

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 IR of the residual to a monetary shock should be not be different from zero. But for values away from the truth, the IR of the residual should be a combination of the IRs of future inflation and unemployment (the right-hand side variables of the Phillips curve) and be non-zero.³

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 IR 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 parametrize the residual IR as a quadratic polynomial function which 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, 2019).⁴

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 or high-frequency identified monetary surprises (e.g., Kuttner, 2001) as instruments to identify the structural coefficients. 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 much larger than estimated by OLS or 2SLS using predetermined variables as instruments. In contrast, the role of forward-looking inflation expectations is smaller than

³For instance, when setting the parameters of the Phillips curve to zero, the IR of the residual will correspond to the IR of inflation, the left-hand variable of the Phillips curve.

⁴In the appendix we show the results from a simulation study where we evaluate the finite sample properties of the approach. We simulate data from a typical forward-looking macro equation and obtain excellent performance even with weak structural shocks, i.e., even with shocks that account for a small share of the variance of the macro variables.

estimated with earlier methods. We then re-examine recent evidence of a flattening of the Phillips curve (e.g., Ball and Mazumder (2011)) and in particular whether that flattening is spurious and the by-product of endogeneity biases. Our evidence points to clear changes in the main determinants of inflation: since 1990, forward-looking inflation expectations play a larger role while slack plays a smaller 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 IRs (e.g., Ramey, 2016; Stock and Watson, 2016).⁵

The use of structural shocks as instruments considerably broadens the scope of identification schemes when compared to using predetermined variables, i.e., lags of macro variables, as instruments. While structural shocks are generally not observable, the literature has produced a variety of proxies for structural shocks, which are sufficient for conducting instrumental variable based inference (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 series can potentially be exploited for identifying different structural equations.⁶

While we propose a simple and efficient method to use shocks as instruments in single equation regressions, other estimation methods are possible. Drawing on our insight that a regression with shocks as instruments boils down to a regression in impulse response space, we

⁵Alternative to the limited information approach is the full-information approach which specifies a system of structural equations, typically a dynamic stochastic general equilibrium (DSGE) model. By imposing a theoretical model on all the variables in the system, full-information methods have the potential to improve estimator precision, but they also introduce the risk of misspecification in other equations, inducing bias or inconsistency of the parameters of interest. The method we propose preserves the limited-information nature of the exercise, as it allows researchers to focus on a single macro equation of interest, without having to take a stand on the theoretical model underlying all the endogenous variables.

⁶In 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.

note that forward-looking macro equations could also be estimated from a two-step approach: (i) estimate the structural IRs using SVAR-IV or LP-IV,⁷ and (ii) regress the estimated IRs on each other—. The main advantages of our AR-based method relative to these alternatives are twofold. First, using SVAR-IV or LP-IV involves a two-stage estimation procedure, which is less efficient than our one-stage method.⁸ Second, our one-stage method is a limited-information approach that intends to minimize the number of modeling assumptions beyond those on the single equation of interest. In contrast, SVAR-IV relies on specifying a system of reduced-form equations and on choosing the set of endogenous variables and the number of lags, while LP-IV relies on the same additional assumptions to decide on the set of control variables (e.g. Plagborg-Møller and Wolf, 2019).⁹

The remainder of this paper is organized as follows. In Section 2 we review the empirical issues faced by limited-information methods and we discuss the traditional solution that is based on lagged instruments. Section 3 outlines the use of independently identified structural shocks for identification. The estimation methodology is developed in Section 4 and the empirical findings for the Phillips curve are presented in Section 5. Section 6 concludes.

2 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 discuss 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

⁷LP-IV and SVAR-IV were developed in Jordà (2005), Stock and Watson (2018), Mertens and Ravn (2013) and Olea, Stock and Watson (2018).

⁸An additional complication is that it is unclear how weak instrument robust inference should be implemented for the second stage of such procedures.

⁹Note that these two-stage SVAR-IV and LP-IV methods are different from impulse response function matching estimators, which rely on a fully specified DSGE model (e.g., Christiano, Eichenbaum and Evans, 2005). If the researcher is willing to postulate a fully-specified DSGE model, SVAR-IV or LP-IV could directly (i.e., in one step) identify the parameters of the equation of interest through impulse response matching (e.g., Canova and Sala, 2009). However, this comes at the cost of an even higher risk of misspecification than with a reduced-form VAR. Our contribution fits in the tradition of using robust limited-information methods to estimate single equations.

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

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

where y_t is the variable of interest that depends on its own lag, its expected value $E_t(y_{t+1})$, the forcing variable x_t and the disturbance e_t . The expectation $E_t(\cdot)$ 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 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

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

where \hat{x}_t is an observable proxy for the forcing variable.¹⁰ 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.

1. **Confounding with the error term:** Since the error term may simultaneously affect y_t and \hat{x}_t we have $E(\hat{x}_t u_t) \neq 0$.
2. **Measurement error in the forcing variable:** Since the forcing variable is unobserved and thus subject to measurement error we have $E(\hat{x}_t u_t) \neq 0$.
3. **Unobserved inflation expectations:** Since $E_t(y_{t+1})$ is unobserved and thus subject to measurement error we have $E(y_{t+1} u_t) \neq 0$.

¹⁰Other observable proxies for the expectation term, such as expectation measures from surveys, can equally well be considered.

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

The traditional approach for handling the endogeneity problems is to treat y_{t-1} as pre-determined and to 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.

1. $E(e_t z_t^l) = 0$ since $E_{t-1}(e_t) = 0$ *provided that* the error term e_t has no serial correlation.
2. $E((y_{t+1} - E_t(y_{t+1}))z_t^l) = 0$ since $E_t(y_{t+1} - E_t(y_{t+1})) = 0$ under rational expectations and by applying the law of iterated expectations.
3. $E((\hat{x}_t - x_t)z_t^l) = 0$ *provided that* the measurement error $\hat{x}_t - x_t$ has no serial correlation

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

Unfortunately, this approach faces difficult 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.¹¹ A potential way of avoiding this concern

¹¹This can happen if the disturbance e_t is auto-correlated, 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) shows 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).

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, p163). The informativeness of the instruments can be improved by exploiting instabilities in the moment conditions, see Magnusson and Mavroeidis (2014) and Antoine and Boldea (2018).

3 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 equations like (2). Let ε_t^i denote the structural shock of type i for time period t .¹² 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 (2). To this extent let $\varepsilon_{t:t-H}^i = (\varepsilon_t^i, \dots, \varepsilon_{t-H}^i)'$.

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

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

$$\mathbb{E}(\varepsilon_{t:t-H}^i (y_{t-1}, y_{t+1}, \hat{x}_t)) \quad \text{full column rank} \quad (\text{Relevance})$$

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

¹²We refer to Ramey (2016), Blanchard and Watson (1986) and Bernanke (1986) for more discussion regarding the definition of a structural shock.

exogeneity and relevance conditions, and the identification of impulse response functions.

3.1 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 of macroeconomic variables 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 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 equation residual u_t can be written as linear functions of the structural shocks.¹³ Under these assumptions, the exogeneity and relevance conditions can be restated as

$$\beta_h^u = 0 \quad \forall \quad h = 0, \dots, H \quad (\text{Exogeneity})$$

$$[\beta_{h-1}^y, \beta_{h+1}^y, \beta_h^{\hat{x}}]_{h=0}^H \quad \text{linearly independent} \quad (\text{Relevance})$$

where β_h^j is the IR 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 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 $(\lambda, \gamma_f, \gamma_b)$ are set at their true values, the IR of the residual u_t should be zero (under correct specification).

The relevance condition states that the impulse responses of the observed forcing variable \hat{x} and of past and future y are not linearly dependent, which includes as a special case that the IRs should be non-zero.

The reformulation of the exogeneity and relevance conditions implies that all the informa-

¹³Note that these assumptions are only made to illustrate the approach, they are not required for the properties of our estimators that are discussed below.

tion needed to recover the coefficients of the structural equation are encoded in the impulse response functions of the observables to the structural shocks. To see this, pre-multiply the forward looking equation (2) by ε_{t-h}^i and take the expectation, we immediately obtain

$$\beta_h^y = \gamma_b \beta_{h-1}^y + \gamma_f \beta_{h+1}^y + \lambda \beta_h^{\hat{x}}, \quad \forall h = 0, \dots, H. \quad (3)$$

Expression (3) implies that we can identify the coefficients of the forward looking macro equation from a regression – across h – of the IR of the outcome variable on its own lags and leads, and on the IR of the forcing variable, i.e., from a regression in “impulse response space”.¹⁴ Intuitively, the exogeneity condition implies that (3) holds, while the relevance condition implies that the dynamics of the IRs of $(y_{t-1}, y_{t+1}, \hat{x}_t)$ are rich enough such that there exist a unique parameter vector $(\lambda, \gamma_f, \gamma_b)$ satisfying (3).

3.2 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.

The Phillips curve

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

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

¹⁴Specifically, by minimizing the sum of squared residuals $\sum_{h=0}^H (\beta_h^y - \gamma_b \beta_{h-1}^y - \gamma_f \beta_{h+1}^y - \lambda \beta_h^{\hat{x}})^2$, we can find the structural coefficients that best fit equation (2) for any h . This is an OLS regression in “impulse response space”, i.e., a regression across the horizon h of the IRs. While the “regression in impulse response space” interpretation is helpful to get the intuition behind our instrumental variable 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 (3) after replacing β_h^y and $\beta_h^{\hat{x}}$ by their sample counterparts are equivalent to computing the GMM estimator for the structural equation (1) with instruments $\{\varepsilon_t^i, \dots, \varepsilon_{t-H}^i\}$ and with the GMM weighting matrix equal to the identity matrix. This choice is not efficient and not robust to weak instruments. Our preferred methodology is described in the estimation section.

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 (1).

Re-writing (4) to highlight the endogeneity issues, we have

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

The Phillips curve includes all three sources of endogeneity discussed in section 2: (i) confounding from cost-push factors, which can simultaneously affect inflation and the forcing variable, (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, i.e., that they are both (i) exogenous and (ii) relevant.¹⁵

Exogeneity: The exogeneity condition $E(\varepsilon_{t:t-H}^i u_t) = 0$ 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.

Condition (i) holds since monetary shocks are, by their very nature, orthogonal to exogenous factors.¹⁶ Condition (ii) holds under the assumption that money is neutral under flexible prices, a relatively mild and uncontroversial assumption.¹⁷ Condition (iii) holds

¹⁵We focus on monetary shocks for ease of exposition, but additional instrument candidates for structural shocks include alternative aggregate demand shocks (e.g., government spending shocks) or productivity shocks (e.g., Fernald, 2012).

¹⁶If the cost push factors are not strictly exogenous, for instance, if oil prices (a type of cost-push factor) respond to US monetary policy, one would need to add oil prices in the Phillips curve and instrument that additional variable. Again, 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).

¹⁷The exogeneity condition $E(\varepsilon_{t-j}^m (\hat{x}_t - x_t)) = 0$ is verified, if $E(\varepsilon_{t-j}^m (\hat{g}_t^n - g_t^n)) = 0$, which holds if monetary policy is neutral under flexible prices.

under rational expectation or provided that survey measures of inflation expectations are available and accurate up to some additive (and possibly autocorrelated) measurement error term.¹⁸

Relevance: Monetary shocks are relevant instruments if they affect inflation and the output gap. This implies that (in addition to the Phillips curve (4)), there must exist *an* underlying IS curve, i.e., 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 (4) involves at least two endogenous variables (in our case three: 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 linear dependent. From the intuition in Section 3.1 it follows that the relevance condition holds if and only if the IRs 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 iid process (see appendix A for a formal derivation), so we again 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.

The Euler equation

Consider a linearized Euler equation of the form

$$x_t = \gamma_b x_{t-1} + \gamma_f E_t(x_{t+1}) - \lambda(i_t - E_t(\pi_{t+1}) - r_t^n), \quad (6)$$

¹⁸The exogeneity condition $E(\varepsilon_{t-j}^m(\pi_{t+1} - E_t\pi_{t+1})) = 0$ is satisfied under rational expectations, since the law of iterated expectations implies $E(\varepsilon_{t-j}^m(\pi_{t+1} - E_t\pi_{t+1})) = E(\varepsilon_{t-j}^m E_t(\pi_{t+1} - E_t\pi_{t+1})) = 0$. For departures of 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.

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. 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, 1988; Yogo, 2004; Ascari, Magnusson and Mavroeidis, 2016).

Rewriting the Euler equation to highlight the endogeneity issues gives

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

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 output gap and (iii) unobserved inflation expectations and output gap expectations.¹⁹

Again, monetary shocks are good candidates for valid instruments to identify (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.²⁰ 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, notably oil price shocks (e.g., Hamilton, 2003). Cost-push shocks are relevant instruments as long as there exist some underlying Phillips curve and monetary rule with rich enough

¹⁹The residual u_t satisfies

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

Equation (6) admits the general form discussed in section 2, but with one additional source of endogeneity compared to the Phillips curve: Because the left-hand side variable in (6) is the unobserved variable x_t , serially correlated measurement error in x_t will imply $E(\hat{x}_{t-1}u_t) \neq 0$.

²⁰In the baseline New-Keynesian model, productivity shocks drive the natural real rate of interest (Galí, 2015).

dynamics (that need not be specified), such that the IRs 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.

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

$$i_t = \gamma_b i_{t-1} + \gamma_f E_t(\pi_{t+1}) + \lambda x_t + \varepsilon_t^m, \quad (8)$$

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

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

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

The sources of endogeneity bias in (9) are confounding from monetary shocks, unobserved inflation expectations, and measurement error in the output gap.²¹ In this case, productivity shocks and oil price 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 IRs of inflation and the output gap to those shocks are not linear functions of one another.

4 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

²¹The residual is given by $u_t = \varepsilon_t^m + \gamma_f(E_t(\pi_{t+1}) - \pi_{t+1}) + \lambda(x_t - \hat{x}_t)$

compact model representation

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

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

While structural shocks are typically not observed, the 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). 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 hold when we replace $\varepsilon_{t:t-H}^i$ by $\xi_{t:t-H}^i$.

4.1 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

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

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 for instance Theorem 5.23). Based on such normal limiting approximation we may conduct hypothesis tests and construct confidence intervals.

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, 2017; 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 $H = 8$ and $H = 16$ for quarterly data as this is the horizon for which macroeconomic IRs are typically found to be significantly different from zero.²² When the number of instruments used is large relative to 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).²³

4.2 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. Additionally, we exploit the impulse response intuition from Section 3.1 to reduce the number of effective instruments, thus avoiding the many instruments problem.

To outline our approach, consider testing the hypothesis $H_0 : \delta = \delta_0$. From the exogeneity condition $E(\xi_{t:t-H}^i u_t) = 0$ it follows that we can base such tests on the distributed lag model

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

where θ is the $(H + 1) \times 1$ *impulse response function* of the macro equation residual u_t to the proxies $\xi_{t:t-H}^i$ and η_t is a disturbance term.²⁴ 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

²²For example, when considering the Phillips curve where y_t corresponds to inflation, the inflation response to a monetary policy shock takes approximately 8-12 quarters to reach its peak (see online appendix).

²³The many instruments problem is well known in the cross-section iv literature and alternative estimators have been proposed, see for example Hausman et al. (2012). To the best of our knowledge such alternative estimators have not been developed for time series instrumental variable regressions and in general they rely on the assumption that increasing sequences of instruments are effectively strong instruments which might not be the case in our setting.

²⁴Note that we changed the IR notation from β to θ to highlight that this is the IRF to the proxies for the structural shocks instead of the structural shocks themselves. Regardless, all IR intuition from Section 3.1 continues to apply.

$\theta = 0$. Intuitively, for values of the macro parameters close to their true values, the IR of the residual $u_t = y_t - w_t'\delta_0$ to the structural shock proxies should be not be different from zero. Conversely, for values away from the truth the IR of the residual should be a combination of the IRs of \hat{x}_t , future and past y_t (the right-hand side variables of the macro equation) and thus be non-zero.²⁵

Testing $H_0 : \delta = \delta_0$ is thus easily implemented by testing $\theta = 0$ using an AR (Anderson and Rubin, 1949) type statistic. The important feature of such AR-type statistic is that its limiting distribution does not depend on the strength of the instruments (e.g. Staiger and Stock, 1997).²⁶

The baseline AR-statistic is given by

$$AR(\delta_0) = n\hat{\theta}'\hat{\Sigma}_\theta^{-1}\hat{\theta}, \quad (13)$$

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 asymptotic variance of $\hat{\theta}$.

Unfortunately, hypothesis tests based on the standard AR-statistic have poor power and size properties when the number of instruments is large relative to the sample size, see Andrews, Stock and Sun (2019).²⁷ To reduce the dimension of the problem, we go back to Almon (1965) and re-parameterize the elements of the impulse response θ as a polynomial function

$$\theta_h = a + bh + ch^2, \quad \text{for } h = 0, \dots, H, \quad (14)$$

where a, b and c are the polynomial coefficients. Alternative basis functions for θ_h can also

²⁵For instance, when setting the parameters of the macro equations to zero, the IR of the residual will correspond to the IR of y_t , the left-hand variable.

²⁶In 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 $\xi_{t:t-H}^i$. 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).

²⁷Note that this also holds for other conventional weak instrument robust statistic such as the LR and LM statistic. Further, our simulation study below shows that the $AR(\delta_0)$ statistic is over-sized when $H = 16$ and $n = 200$. This corresponds with the theoretical derivations in Andrews and Stock (2007) who require that $H^3/n \rightarrow 0$ as $H, n \rightarrow \infty$ for the consistency of the test based on the AR-statistic.

be considered, but the polynomial one is attractive in our setting as the resulting estimation problem remains linear. Intuitively, this approach will allow us to reduce the noise in the AR-statistic by exploiting the fact that the IRs of macro variables are typically believed to be smooth functions.²⁸

With this parameterization in place we obtain

$$y_t - w_t' \delta_0 = \theta_a' z_t^i + \eta_t , \quad (15)$$

where the Almon-polynomial coefficients are captured by $\theta_a = (a, b, c)'$ and

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

Notice that z_t^i is merely a deterministic linear function of the exogenous structural shocks and hence z_t^i inherits the exogeneity properties of $\xi_{t:t-H}^i$, i.e. we have $E(z_t^i (y_t - w_t' \delta_0)) = 0$ under H_0 .

The imposed Almon restriction implies that the number of instruments reduces to three, the number of endogenous variables. 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.

For these reasons, we propose the Almon (1965) restricted AR statistic that we define as

$$AR_a(\delta_0) = n \hat{\theta}_a' \hat{\Sigma}_{\theta_a}^{-1} \hat{\theta}_a , \quad (16)$$

where $\hat{\theta}_a$ is the OLS estimate of equation (15) and $\hat{\Sigma}_{\theta_a}$ is any consistent estimate for the asymptotic variance of $\hat{\theta}_a$. In practice, we compute the variance matrix using the approach

²⁸Note that our approach remains valid (our confidence sets have correct size) even if the true IRs are not smoothed functions and a quadratic polynomial provides a poor approximation. The Almon-restriction will however impose a cost in terms of lower power.

outlined in Newey and West (1994). Under mild assumptions, which do not involve the strength of the instruments, we have that under H_0 the AR_a statistic converges to a chi-squared distribution with three degrees of freedom (e.g., Stock and Wright, 2000; Chernozhukov and Hansen, 2008). Confidence sets for δ are computed by inverting the AR_a statistic for different values of $\delta_0 \in \mathcal{D} \subset \mathbb{R}$.

Finally, it is worth mentioning that the Almon restriction can also be used to reduce the number of instruments when considering standard moment estimators. In particular, we may consider the Almon restricted moment estimator

$$\hat{\delta}_a^{IV} = \left(S'_{zw} \hat{\Omega}_z S_{zw} \right)^{-1} S'_{zw} \hat{\Omega}_z s_{zy} , \quad (17)$$

where $S_{zw} = \frac{1}{n} \sum_{t=1}^n z_t^i w_t'$, $s_{zy} = \frac{1}{n} \sum_{t=1}^n z_t^i y_t$ and $\hat{\Omega}_z$ is some positive definite weight matrix. This moment estimator does not suffer from the many instrument problem, thanks to the Almon-restriction, but 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.

4.3 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. The full description of the study is presented in Appendix B.

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., Gorodnichenko and Lee, 2017; Plagborg-Møller and Wolf, 2018), and notably the fact that monetary shocks may account for a relatively small share of the variance of macro variables. Based on this data generating process we compared the standard Wald test (based on the 2SLS moment estimator with $H = 16$ structural shocks), the Almon restricted Wald test (based on the 2SLS moment

estimator with Almon restriction), the standard AR test (based on using $H = 16$ structural shocks) and the Almon-restricted AR_a test.

We compared the empirical rejection frequencies of these tests and found that only the AR_a test has correct size. All other tests severely over-reject. 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.

5 The US Phillips curve

In this section we discuss estimation results for the New Keynesian Phillips curve for the United States. We consider a standard hybrid Phillips curve of the form

$$\pi_t = \gamma_b \pi_{t-1}^4 + \gamma_f E_t(\pi_{t+4}^4) + \lambda x_t + \varepsilon_t^s, \quad (18)$$

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 3.2 we showed that one can identify the parameters of (18) by using monetary policy shocks as instrumental variables. To operationalize the use of monetary shocks for identification we rely on several 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 will also rely on the recent high-frequency identification (HFI) approach pioneered by Kuttner (2001) and Gürkaynak, Sack and Swanson (2005) and use surprises in fed funds futures prices around FOMC announcement as proxies for monetary shocks.

Inflation is measured from changes in the PCE price level and we take the unemployment

rate as our baseline real activity variable. An advantage of our approach is that, unlike earlier methods, we can directly proxy the unemployment gap with the unemployment rate, since projecting on monetary shocks *projects out* the natural rate of unemployment from the Phillips curve equation.²⁹ As an alternative, we will also report results based on using the output gap as the forcing variable.

5.1 Full sample estimates

For our full sample estimates, we first treat lagged inflation as exogenous to focus on the effects of expected future inflation and economic slack. We then relax this assumption and also instrument lagged inflation with monetary shocks.

Treating lagged inflation as exogenous

Figure 1 presents our AR_a -based confidence sets for the Phillips curve coefficients λ and γ_f estimated using the Romer and Romer narrative shocks as instruments with $H = 16$ over 1969-2007. In addition, the blue dot reports the OLS estimate, the green dot reports the 2SLS estimate with lagged macro variables as instruments, and the red dot reports the Almon-restricted 2SLS estimate that is obtained from (17) with $H = 16$.³⁰

The main results are as follows:

- The slope of the Phillips curve (λ) implied by our Almon-restricted 2SLS point estimate and by our confidence sets is significantly different from zero and substantially larger than suggested by the OLS estimate, in line with what one would expect from confounding from supply shocks (since supply shocks lead to a positive correlation between inflation and the unemployment gap) and measurement error in the unemployment gap (leading to an attenuation bias).

²⁹The unobservability of the natural level of output has been a severe limitation in the Phillips curve literature, and our ability to project out the natural level of output is a key advantage of our approach. We do not rely on the labor share (Gali and Gertler, 1999), which suffers from a number of issues, most notably non-stationarity, which can substantially affect point estimates (e.g., Mavroeidis, Plagborg-Møller and Stock, 2014).

³⁰For all these estimates we first project out the lag of inflation and a constant from equation (18). This effectively treats π_{t-1}^4 as pre-determined which is an assumption that we relax below.

- The confidence sets point to a modest role for forward-looking inflation expectations (γ_f), in fact close to zero. Interestingly, note that the width of the confidence sets in the γ_f direction is relatively small. This implies that the role of inflation expectations is well identified, but that their influence is modest and much smaller than suggested by OLS.
- Figure 4 presents results similar to Figure 1 but based on using as forcing variable the output gap, defined as the log deviation of real GDP from its natural level, as estimated from an HP filter with $\lambda^{hp} = 1600$. The results are similar: (i) the slope of the Phillips curve (implied by our Almon-restricted 2SLS point estimate or implied by our confidence sets) is substantially larger than estimated by OLS, more than twice as large in the case of 2SLS, and (ii) our approach points to a small role for inflation expectations, smaller than suggested by OLS.

To provide some intuition on how we construct our confidence sets from the impulse responses of the residual, Figure 2 displays the heatmap of the AR_a statistic, which can be seen an F-test of overall significance for the IR of the Phillips curve residual to a monetary shock. Darker (bluer) values indicate values of the AR_a statistic close to zero —IRs of the residual close to zero— and thus more “likely” parameter values.

Recall that for values away from the truth, the IR of the residual should be a combination of the IRs of future inflation and unemployment and thus be non-zero. Thus, a simple sanity check can be conducted by plotting the IR of the residual for nine different values of the parameter pair (λ, γ_f) , see figure 3. The small red dots in the top-left quadrant of Figure 2 denote the different parameter values corresponding to the nine impulse responses. For the Almon-restricted 2SLS estimate for $\hat{\lambda}$ and $\hat{\gamma}_f$ (center dot of figure 2), the IR of the residual (center of figure 3) is close to zero, consistent with the idea that the point estimates are close to the true values. As we move away from these values however, the IRs are driven by the IR of inflation and/or unemployment. For instance, setting γ_f to 1 and λ to 0 (top-right panel), the residual u_t is equal to $-(\pi_{t+1} - \pi_t)$, and the IR displays a positive hump because

an increase in monetary policy lowers inflation (Romer and Romer, 2004), i.e., raises the negative of first-differenced inflation. Keeping $\gamma_f = 1$ but lowering λ to $\hat{\lambda} - 1$ (top-left panel), the IR is even stronger because the residual u_t is now equal to $u_t - (\pi_{t+1} - \pi_t)$, and a contractionary monetary shock raises unemployment and lowers inflation. A similar intuition holds throughout the λ - γ_f parameter grid.

Interestingly, Figure 3 highlights why the confidence sets for λ are much larger than the confidence sets for γ_f . Moving sideways from the center (lowering λ but holding γ_f at $\hat{\gamma}_f$) the IR is not particularly strong, but moving up or down, the IR becomes quickly very large. The reason is that the IR of unemployment is less well estimated than the IR of inflation, as shown in the online appendix. Thus, as we move away from the 2SLS point estimate, we are able to quickly reject values of γ_f , but not values of λ .

Treating lagged inflation as endogenous

While the previous results treated π_{t-1} as exogenous, this is not valid if the Phillips curve residual displays auto-correlation coming from e.g., auto-correlated cost-push shocks or from auto-correlated measurement error in the inflation expectation proxy. An important advantage of our approach compared to using pre-determined variables as instruments is that we can easily accommodate auto-correlation in the Phillips curve residual.³¹

Figures 5 and 6 depict the 68 and 95 confidence volumes for λ , γ_f and γ_b using the unemployment rate as the forcing variable. Similar results holds using the output gap (see the online appendix). A few comments are in order:

- Even with three endogenous variables, the confidence volume remains convex and not much larger than when we treated π_{t-1} as exogenous. In fact, the 68 percent volume is even closed, as shown by Figure 6.
- The results for λ and γ_f are consistent with our previous results: the slope of the

³¹In contrast, with pre-determined variables, autocorrelation in the residual leads to a trade-off between the exogeneity condition and the relevance condition as increasing the lag length dramatically worsens the weak instrument problem.

Phillips curve is significantly different from zero and substantially larger than OLS, while the coefficient on inflation expectations is small and not significantly different from zero. The 68 (95) percent coefficient set for lagged inflation (γ_b) ranges from about .7 to 1.5 (.2 to 1.8) and includes .89 (our Almon-restricted 2SLS estimate) roughly in its center.

- We cannot reject that the sum of the coefficients on lagged and future inflation is statistically different from one, as was the case with OLS estimates.

5.2 The Phillips curve over time

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) and Blanchard (2016)).

In this section, we parallel the Phillips curve literature and study the evolution of the Phillips curve over time. Importantly, unlike earlier Phillips-curve based studies, our approach projects out confounding factors and measurement error, so that it will allow us to evaluate whether the recent flattening of the Phillips curve is spurious and caused by endogeneity issues. For instance, a change in the variance contribution of supply shocks (which varies the magnitude of the endogeneity bias, see Gordon, 2011), a change in the central bank's systematic response to supply shocks,³² or mis-measured movements in the natural rate of unemployment could all give the illusion of a change in the Phillips curve parameters.³³

³²For instance, if the Fed is more aggressive after 1990 in its efforts to stabilize inflation and becomes more successful in stabilizing inflation —as in the inflation targeting era—, the stronger endogenous response of monetary policy can create the illusion of a decline in the slope of the Phillips curve (McLeay and Tenreyro, 2018).

³³For instance, if the natural rate of unemployment was over-estimated during the late 1990s, the unemployment gap would have been under-estimated (not small enough), leading to a downward-biased estimate

To investigate a possible change in the Phillips curve coefficients since 1990, we draw on HFI monetary surprises. Specifically, we follow Gertler and Karadi (2015) and consider changes in Federal Funds futures rates (FF) around FOMC announcement dates as external instruments.³⁴ This monetary surprise is plausibly uncorrelated with other shocks because the surprise is measured over a short announcement window.

Since HFI monetary surprises are only available in the more recent period (1990-2007),³⁵ we report two sets of results to assess the evolution of the Phillips curve over time: first, the Phillips curve estimated using the Romer-Romer narrative shocks over 1969-1989, and second the Phillips curve estimated using the HFI monetary surprises over 1990-2007.³⁶

The results are shown in Figures 7, which displays the AR_a -based confidence sets for the two sample periods. Figure 8 plots the same set of results for the Phillips curve based on the output gap.

Regardless of whether we use unemployment or the output gap as the forcing variable, the distinctively different shapes of the confidence sets highlight changes in the roles that real activity and expectations play in determining inflation.

- For the pre-1990 period, the confidence sets are centered around zero for γ_f —pointing to a negligible role for forward-looking inflation expectations—, but the confidence sets include mostly positive values (in absolute terms) for the slope λ —pointing to a large effect of slack on inflation—. The opposite holds for the post-1990 period: the confidence sets are closer to zero for λ , but distinctively in positive territory for γ_f . This finding is consistent with a change in the main determinants of inflation, with forward-looking inflation expectations playing a larger role, and slack playing a smaller role.³⁷

of the slope of the Phillips curve.

³⁴Similarly to Gertler and Karadi (2015), we use “FF4”, the three month ahead monthly Fed Funds futures and we present results based on these instruments.

³⁵We intentionally exclude the post-2007 period (the zero lower-bound period) during which forward guidance played a more prominent role and the Fed started employing unconventional monetary tools that could have different effects than conventional monetary tools, see Swanson (2017).

³⁶As an alternative, we can also use the Romer and Romer shocks over 1990-2007. The results (shown in the online appendix) are similar.

³⁷The same conclusions hold if we use core inflation instead of headline inflation, or whether we use the

- In terms of point estimates, the OLS estimates show a substantial decline in λ the slope of the Phillips curve —from -0.24 to -.07 for the unemployment rate—, consistent with previous studies. Interestingly, our Almon-restricted 2SLS point estimates confirm that decline with a comparable *proportional* decline —from -1.01 to -0.27—. That being said, even after 1990, the slope of the Phillips curve is substantially larger than suggested by OLS estimates. Similar results hold for the output gap with λ declining from 0.55 to 0.22 according to our 2SLS estimates (versus a decline of 0.11 to 0.06 according to OLS).
- Our Almon-restricted 2SLS estimates also suggest a large increase in γ_f after 1990; from 0.03 to 0.84 for the Phillips curve with the unemployment rate and from 0.22 to 0.95 for the Phillips curve with the output gap.

6 Conclusion

In this paper, 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 paper, 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 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 Romer and Romer shocks post-1990 instead of HFI surprises (see online appendix).

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 paper focuses on estimating linear equations, using shocks as instruments instead of pre-determined variables can be also used to estimate non-linear forward-looking equations, which is of high relevance for the asset pricing literature (Hansen and Singleton, 1982).

References

- Almon, Shirley.** 1965. “The Distributed Lag Between Capital Appropriations and Expenditures.” *Econometrica*, 33(1): 178–196.
- Anderson, T. W., and Herman Rubin.** 1949. “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations.” *Ann. Math. Statist.*, 20(1): 46–63.
- Andrews, Donald W.K., and James H. Stock.** 2007. “Testing with many weak instruments.” *Journal of Econometrics*, 138(1): 24 – 46.
- Andrews, Isaiah, James Stock, and Liyang Sun.** 2019. “Weak Instruments in IV Regression: Theory and Practice.” *Annual Review of Economics*. forthcoming.
- Antoine, B., and O. Boldea.** 2018. “Efficient Inference with Time-Varying Information and the New Keynesian Phillips Curve.” *Journal of Econometrics*, 204: 268–300.
- Ascari, Guido, Leandro M. Magnusson, and Sophocles Mavroidis.** 2016. “Empirical evidence on the Euler equation for consumption and output in the US.” Working Paper.
- Ball, Laurence, and Sandeep Mazumder.** 2011. “Inflation Dynamics and the Great Recession.” *Brookings Papers on Economic Activity*, 337–405.
- Bekker, Paul A.** 1994. “Alternative Approximations to the Distributions of Instrumental Variable Estimators.” *Econometrica*, 62(3): 657–681.
- Bernanke, Ben S.** 1986. “Alternative explanations of the money-income correlation.” *Carnegie-Rochester Conference Series on Public Policy*, 25: 49 – 99.
- Blanchard, Olivier.** 2016. “The Phillips Curve: Back to the ’60s?” *American Economic Review*, 106(5): 31–34.
- Blanchard, Olivier, and Mark W. Watson.** 1986. “Are Business Cycles All Alike?” In *The American Business Cycle: Continuity and Change*. 123–180. National Bureau of Economic Research, Inc.
- Caldara, Dario, and Edward Herbst.** 2018. “Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs.” *American Economic Journal: Macroeconomics*. forthcoming.
- Campbell, John Y.** 2003. “Consumption-based asset pricing.” *Handbook of the Economics of Finance*, 1: 803–887.
- Canova, Fabio, and Luca Sala.** 2009. “Back to square one: Identification issues in DSGE models.” *Journal of Monetary Economics*, 56(4): 431–449.
- Chernozhukov, Victor, and Christian Hansen.** 2008. “The Reduced Form: A Simple Approach to Inference with Weak Instruments.” *Economics Letters*, 100: 68 – 71.

- Chernozhukov, Victor, Christian Hansen, and Michael Jansson.** 2009. “Admissible Invariant Similar Tests for Instrumental Variable Regression.” *Econometric Theory*, 25(3): 806–818.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans.** 2005. “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy.” *Journal of Political Economy*, 113(1): 1–45.
- Clarida, Richard, Jordi Galí, and Mark Gertler.** 2000. “Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory.” *The Quarterly Journal of Economics*, 115(1): 147–180.
- Cochrane, John.** 2004. “Comments on “A new measure of monetary shocks: Derivation and implications” by Romer and Romer.” NBER EFG Meeting.
- Coibion, Olivier, Yuriy Gorodnichenko, and Mauricio Ulate.** 2017. “The cyclical sensitivity in estimates of potential output.” National Bureau of Economic Research.
- Deaton, Angus.** 1992. *Understanding consumption*. Oxford University Press.
- Fernald, John G.** 2012. “A quarterly, utilization-adjusted series on total factor productivity.” Federal Reserve Bank of San Francisco Working Paper Series 2012-19.
- Frisch, Ragnar.** 1934. “Statistical Confluence Analysis by Means of Complete Regression Systems.” Publication No. 5 of the University Institute of Economics, Oslo.
- Fuhrer, Jeffrey C., and Glenn D. Rudebusch.** 2004. “Estimating the Euler equation for output.” *Journal of Monetary Economics*, 51(6): 1133–1153.
- Galí, Jordi.** 2015. *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press.
- Gali, Jordi, and Mark Gertler.** 1999. “Inflation dynamics: A structural econometric analysis.” *Journal of Monetary Economics*, 44: 195–222.
- Gertler, M., and P. Karadi.** 2015. “Monetary Policy Surprises, Credit Costs, and Economic Activity.” *AEJ: Macroeconomics*, 7: 44–76.
- Gordon, Robert J.** 2011. “The History of the Phillips Curve: Consensus and Bifurcation.” *Economica*, 78: 10–50.
- Gorodnichenko, Y., and B. Lee.** 2017. “A Note on Variance Decomposition with Local Projections.” Working paper.
- Gürkaynak, Refet, S., Brian Sack, and Eric Swanson.** 2005. “The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models.” *American Economic Review*, 95: 425–436.
- Hall, Robert E.** 1988. “Intertemporal substitution in consumption.” *Journal of political economy*, 96(2): 339–357.

- Hamilton, James D.** 2003. “What is an oil shock?” *Journal of Econometrics*, 113(2): 363–398.
- Hansen, Lars Peter.** 1982. “Large sample properties of generalized method of moments estimators.” *Econometrica*, 50: 1029–1054.
- Hansen, Lars Peter, and Kenneth J Singleton.** 1982. “Generalized instrumental variables estimation of nonlinear rational expectations models.” *Econometrica*, 1269–1286.
- Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson.** 2012. “Instrumental variable estimation with heteroskedasticity and many instruments.” *Quantitative Economics*, 3: 211–255.
- Jordà, Oscar.** 2005. “Estimation and Inference of Impulse Responses by Local Projections.” *The American Economic Review*, 95: 161–182.
- Kilian, Lutz.** 2008. “Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?” *The Review of Economics and Statistics*, 90(2): 216–240.
- King, Robert G., and Mark W. Watson.** 1994. “The post-war U.S. phillips curve: a revisionist econometric history.” *Carnegie-Rochester Conference Series on Public Policy*, 41: 157–219.
- Kleibergen, F., and S. Mavroeidis.** 2009. “Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve.” *Journal of Business and Economic Statistics*, 27: 293–311.
- Kleibergen, Frank.** 2002. “Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression.” *Econometrica*, 70(5): 1781–1803.
- Kuttner, Kenneth N.** 2001. “Monetary policy surprises and interest rates: Evidence from the Fed funds futures market.” *Journal of Monetary Economics*, 47(3): 523–544.
- Magnusson, Leandro, and Sophocles Mavroeidis.** 2014. “Identification using stability restrictions.” *Econometrica*, 82: 1799–1851.
- Mavroeidis, Sophocles.** 2005. “Identification Issues in Forward-Looking Models Estimated by GMM, with an Application to the Phillips Curve.” *Journal of Money, Credit and Banking*, 37(3): 421–448.
- Mavroeidis, Sophocles.** 2010. “Monetary Policy Rules and Macroeconomic Stability: Some New Evidence.” *American Economic Review*, 100(1): 491–503.
- Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H. Stock.** 2014. “Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve.” *Journal of Economic Literature*, 52: 124–188.

- McLeay, Michael, and Silvana Tenreyro.** 2018. “Optimal Inflation and the Identification of the Phillips Curve.” Centre for Macroeconomics (CFM) Discussion Papers 1815.
- Mertens, Karel, and Morten O. Ravn.** 2013. “The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States.” *American Economic Review*, 103(4): 1212–1247.
- Moreira, Marcelo J.** 2003. “A Conditional Likelihood Ratio Test for Structural Models.” *Econometrica*, 71(4): 1027–1048.
- Moreira, Marcelo J.** 2009. “Tests with correct size when instruments can be arbitrarily weak.” *Journal of Econometrics*, 152(2): 131 – 140.
- Newey, Whitney K., and Kenneth D. West.** 1994. “Automatic Lag Selection in Covariance Matrix Estimation.” *The Review of Economic Studies*, 61(4): 631–653.
- Olea, Jose L.Montiel, James H. Stock, and Mark W. Watson.** 2018. “Inference in Structural Vector Autoregressions Identified With an External Instrument.” Working paper.
- Plagborg-Møller, Mikkel, and Christian K. Wolf.** 2018. “Instrumental Variable Identification of Dynamic Variance Decompositions.” working paper.
- Plagborg-Møller, Mikkel, and Christian K Wolf.** 2019. “Local projections and VARs estimate the same impulse responses.” working paper.
- Ramey, Valerie.** 2016. “Macroeconomic Shocks and Their Propagation.” In *Handbook of Macroeconomics*, , ed. J. B. Taylor and H. Uhlig. Amsterdam, North Holland:Elsevier.
- Ramey, Valerie A., and Sarah Zubairy.** 2018. “Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data.” *Journal of Political Economy*, 126.
- Reiersol, Olav.** 1941. “Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis.” *Econometrica*, 9(1): 1–24.
- Romer, Christina D., and David H. Romer.** 2004. “A New Measure of Monetary Shocks: Derivation and Implications.” *American Economic Review*, 94: 1055–1084.
- Staiger, Douglas, and James H. Stock.** 1997. “Instrumental variables regression with weak instruments.” *Econometrica*, 65: 557–586.
- Stock, James H., and Jonathan H. Wright.** 2000. “GMM with Weak Identification.” *Econometrica*, 68(5): 1055–1096.
- Stock, James H., and Mark W. Watson.** 2016. “Chapter 8 - Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics.” In . Vol. 2 of *Handbook of Macroeconomics*, , ed. John B. Taylor and Harald Uhlig, 415 – 525. Elsevier.

- Stock, James H., and Mark W. Watson.** 2018. "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments." *The Economic Journal*, 128(610): 917–948.
- Swanson, Eric T.** 2017. "Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets." National Bureau of Economic Research Working Paper 23311.
- Tenreyro, Silvana, and Gregory Thwaites.** 2016. "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions." *American Economic Journal: Macroeconomics*, 8(4): 43–74.
- White, H.** 2000. *Asymptotic Theory for Econometricians - 2nd edition*. San Diego, California:Academic Press.
- Yogo, Motohiro.** 2004. "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak." *The Review of Economics and Statistics*, 86(3): 797–810.
- Zhang, Chengsi, and Joel Clovis.** 2010. "The New Keynesian Phillips Curve of rational expectations: A serial correlation extension." *Journal of Applied Economics*, 13(1): 159–179.

Appendix A: The rank condition for a forward looking structural equation

Consider the general forward looking structural equation

$$y_t = \gamma_b y_{t-1} + \gamma_f E_t y_{t+1} + \lambda x_t + e_t \quad (19)$$

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

$$x_t = \rho x_{t-1} + \varepsilon_t + \alpha e_t. \quad (20)$$

with e_t and ε_t some iid shocks, and γ_b , γ_f , λ , ρ and α parameters of the model.

Proposition 1. *The model characterized by (19) and (20) 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 (19).*

Proof. Solving for x_t and y_t , we get

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

with α some no-zero parameter and where δ_1 and δ_2 are the stable and unstable roots of the second order-difference equation given by (19).³⁸

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

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

with $\kappa = E(\pi_t \varepsilon_t) = \frac{\lambda}{\delta_2 \gamma_f (1 - \rho / \delta_2)} \neq 0$.³⁹ $\det \Gamma = \kappa \delta_1^2 (\rho + \delta_1 + \rho(\rho + 1))$, so that the rank condition is satisfied if $\delta_1 \neq 0$, i.e., if $\gamma_b \neq 0$. \square

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, i.e., that inflation cannot be strictly forward-looking ($\gamma_b \neq 0$). We can relax this assumption at the expense of assuming more 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: 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 $H_0 : \delta = \delta_0$ using different methods based on using structural shocks as instruments.

³⁸We have $\delta_1 = \frac{1 - \sqrt{1 - 4\gamma_b \gamma_f}}{2\gamma_f}$ and $\delta_2 = \frac{1 + \sqrt{1 - 4\gamma_b \gamma_f}}{2\gamma_f}$.

³⁹This 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 $j > 0$.

Simulation design

We consider the following data generating process

$$\begin{aligned} y_t &= \gamma_b y_{t-1} + \gamma_f E_t(y_{t+1}) + \lambda x_t + e_t \\ x_t &= \rho_1 x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t^i + \alpha e_t, \end{aligned} \tag{21}$$

where the forcing variable x_t follows an AR(2) process. Model (21) has two structural shocks: e_t and ε_t^i . We assume, without loss of generality, that ξ_t^i , our instrument for ε_t^i satisfies $\varepsilon_t^i = \xi_t^i$. Further, we emphasize that although model (21) is highly stylized it includes all the elements that are required to evaluate our methodology. The choice for the AR(2) processes is motivated by empirical evidence for the output gap and the unemployment rate which are typical choices for the forcing variable in the Phillips curve.

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 rate: $\rho_1 = 1.2$ and $\rho_2 = -0.4$. We fix $\alpha = -1$ 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 (2017), Plagborg-Møller and Wolf (2018) and Caldara and Herbst (2018). 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 (2017) 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.⁴⁰ 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 $e_t \sim N(0, 1)$. 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% in the forcing variable x_t . These percentages increase when we increase σ_i . In Table 1 we provide the details. The last scenario where $\sigma_i = 1$ is perhaps over optimistic 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 for monetary policy shocks, e.g. Gorodnichenko and Lee (2017), Plagborg-Møller and Wolf (2018) and Caldara and Herbst (2018).

For each combination of parameter values and sample sizes $n = 200, 500$ we simulate 5.000 datasets and for each dataset we test the hypothesis $H_0 : \delta = \delta_0$ using the methodology outlined in Section 4. In particular, we consider the standard Wald test based on the two stage least squares estimator⁴¹, the standard Wald test based on the Almon-restricted two-stage least squared estimator (17), the standard *AR* test given in equation (13) and

⁴⁰When using local projection methods they find substantially larger influences of the monetary shocks.

⁴¹That 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.

our preferred Almon (1965) restricted AR_a test as defined in equation (16). All tests are implemented using $H = 16$ shocks-as-instruments which corresponds to our empirical work and the intuition that the influence of most macro-economic shocks should be negligible after four years. Note that for the Almon restricted Wald test and the AR_a test the effective number of instruments remains 3 regardless of the value of H .

Results

We report the average rejection frequencies ($\alpha = 0.05$ level) for the different test statistics in Table 2. We find the following patterns. First, the standard Wald statistic based on the normal limiting distribution of the two stage least squares estimator is severely over-sized when the strength of the instruments is small. This holds for both the Almon-restricted Wald test and the unrestricted version that uses $H = 16$ instruments. The empirical rejection frequency is much larger when compared to the nominal size when the variance of the structural shocks is relatively small, e.g. $\sigma_i = 0.1, 0.25, 0.5$. The Almon-restricted version performs slightly better as it only suffers from the weak instruments problem and not from the many instruments problem. The unrestricted Wald test is unreliable across all specifications.

Further, the conventional AR statistic (denoted by AR) based on $H = 16$ structural shocks is severely oversized as well. This corresponds to the theoretical derivations of Andrews and Stock (2007) who show that the AR test is only correct when $H^3/n \rightarrow 0$, this is clearly not the case in the current setting where $H = 16$ and $n = 200, 500$.

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

Table 1: 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%

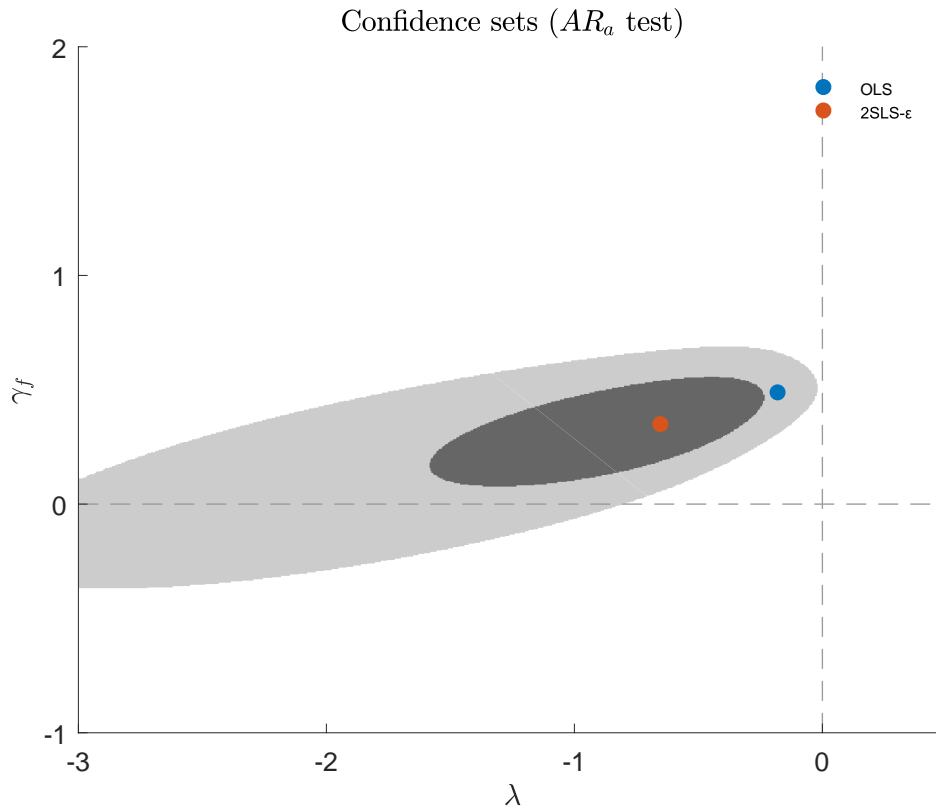
Notes: The table reports the details for the different simulation designs considered. We show the average percentage of variance explained by the structural shock in the variables y_t and x_t , respectively. The remainder of the variance is explained by the – non-structural – shock ε_t^o .

Table 2: Simulation results: Rejection frequencies

n	H	σ_i^2	IV- ε	IV $_{a-\varepsilon}$	AR	AR_a
200	16	0.10	0.470	0.295	0.209	0.047
200	16	0.25	0.338	0.296	0.205	0.048
200	16	0.50	0.093	0.201	0.198	0.047
200	16	1.00	0.010	0.059	0.199	0.046
500	16	0.10	0.426	0.334	0.098	0.051
500	16	0.25	0.282	0.287	0.102	0.047
500	16	0.50	0.053	0.192	0.097	0.049
500	16	1.00	0.024	0.060	0.094	0.048

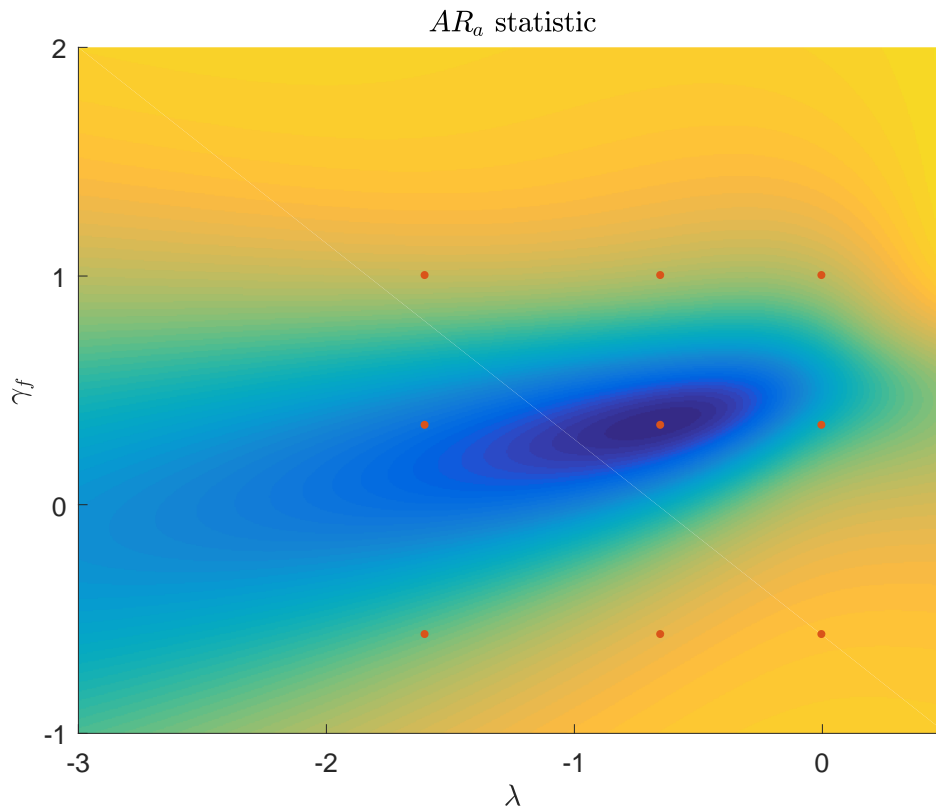
Notes: The table reports the empirical rejection frequencies for $H_0 : \delta = \delta_0$ with level $\alpha = 0.05$. For the IV- ε estimator these correspond to the Wald statistic based on the limiting distribution of the 2SLS estimator with $H = 16$ instruments. The IV $_{a-\varepsilon}$ corresponds to the Wald statistic based on the limiting distribution of the Almon-restricted 2SLS estimator. The AR column corresponds to the test based on the Anderson-Rubin statistic that was computed using $H = 16$ 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 (16). For all estimators the variance covariance matrices are computed using Newey and West (1994) with automatic bandwidth selection.

Figure 1: The Phillips curve — 1969-2007, RR identification



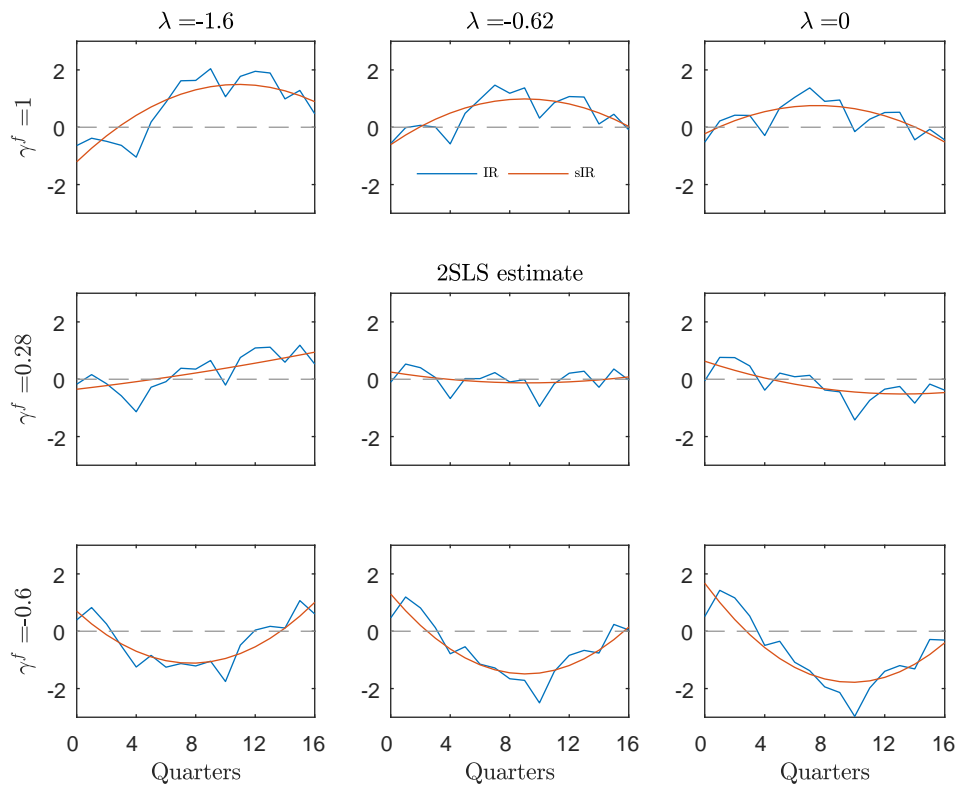
Notes: 95 and 68 percent robust confidence sets for the Phillips curve coefficients obtained by inverting the AR_a test over the parameter space of λ (the slope of the Phillips curve) and γ_f (the loading on inflation expectations). Estimation based on using the Romer-Romer (RR) monetary shocks as instruments for 1969-2007. The red dot (2SLS- ε) is the Almon-restricted 2SLS estimate using the RR monetary shocks as instruments, the green dot (2SLS-GIV) is the 2SLS estimate using 4 lags of inflation and unemployment as instruments, the blue dot is the OLS estimate.

Figure 2: The Phillips curve — 1969-2007, RR identification



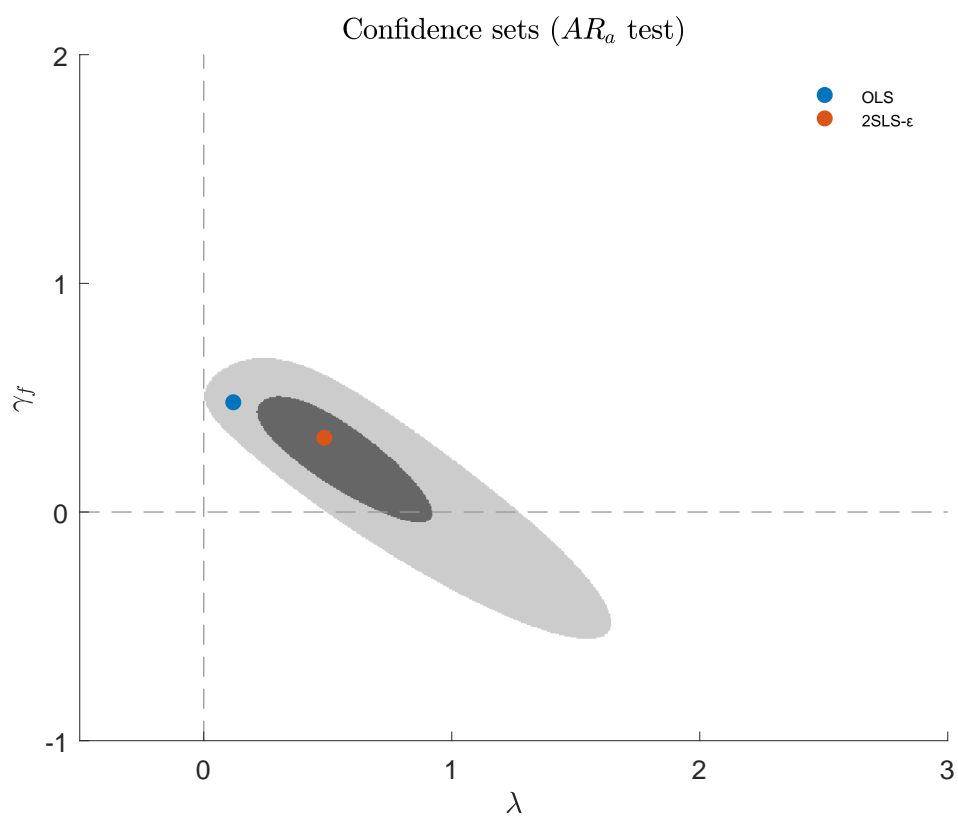
Notes: Heatmap of the Almon AR statistic (AR_a) across the parameter space of λ (the slope of the Phillips curve) and γ_f (the loading on inflation expectations). Estimation based on using the Romer-Romer (RR) monetary shocks as instruments for 1969-2007. The red dots denote the parameter values corresponding to the nine impulse response plotted in figure 3, with the center dot corresponding to the Almon-restricted 2SLS estimate using the RR monetary shocks as instruments.

Figure 3: Impulse Responses of the Phillips curve residual



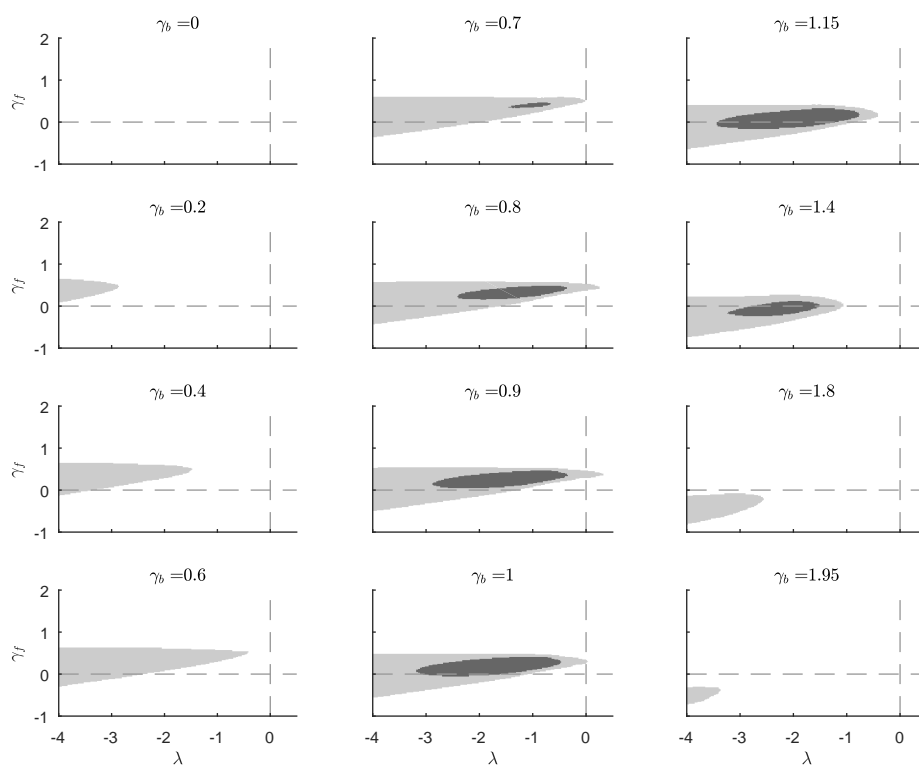
Notes: Impulse responses (IR, blue color) of the Phillips curve residual for different values of λ and γ_f . The small red dots in the top-left quadrant of Figure 2 denote the parameter values corresponding to these nine impulse response, with the center dot corresponding to our Almon-restricted 2SLS estimate. The corresponding impulses responses smoothed with an Almon restriction (sIR) are reported in red.

Figure 4: The output gap Phillips curve — 1969-2007, RR identification



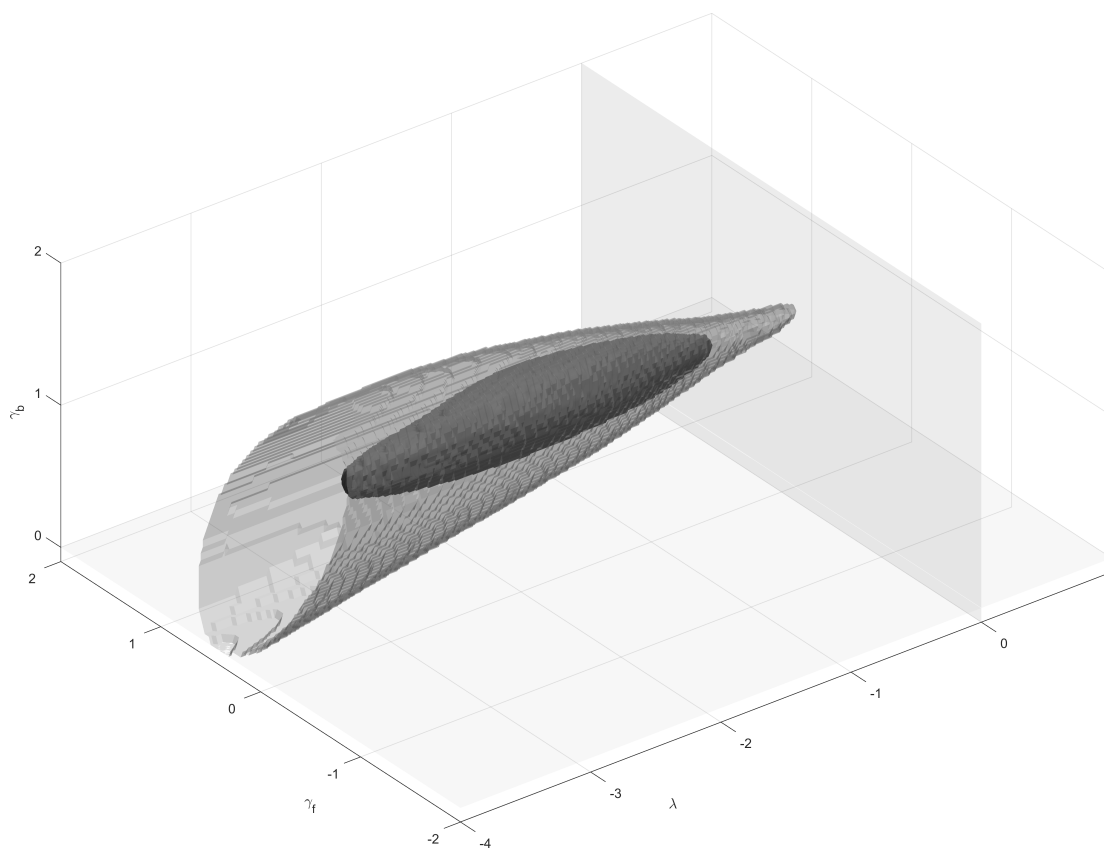
Notes: 95 and 68 percent robust confidence sets for the Phillips curve (output gap) coefficients obtained by inverting the AR_a tests over the parameter space of λ and γ_f . Estimation based on using the Romer-Romer (RR) monetary shocks as instruments for 1969-2007. The red dot (2SLS- ε) is the Almon-restricted 2SLS estimate using lags of the Romer and Romer monetary shocks as instruments, the green dot (2SLS-GIV) is the 2SLS estimate using 4 lags of inflation and output gap as instruments, the blue dot is the OLS estimate.

Figure 5: The Phillips curve — 1969-2007, RR id., π_{t-1} endogenous



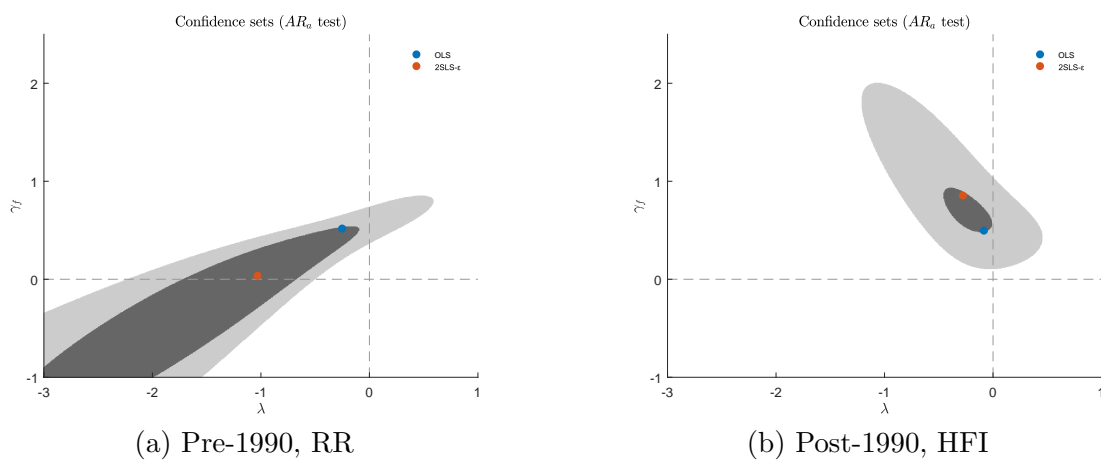
Notes: Robust confidence sets for the Phillips curve coefficients on unemployment (λ), expected future inflation (γ_f) and past inflation (γ_b). Estimation based on using the Romer-Romer (RR) monetary shocks as instruments for 1969-2007. 68 percent (dark grey) and 95 percent (light grey) confidence sets obtained by inverting the AR_a tests over the parameter space of λ , γ_f and γ_b .

Figure 6: The Phillips curve — 1969-2007, RR id., π_{t-1} endogenous



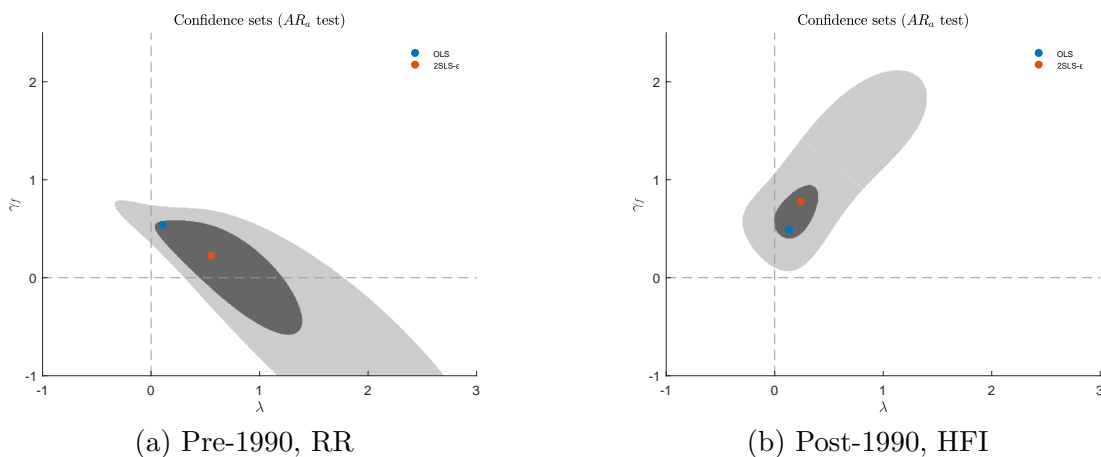
Notes: Robust confidence sets for the Phillips curve coefficients on unemployment (λ), expected future inflation (γ_f) and past inflation (γ_b). Estimation based on using the Romer-Romer (RR) monetary shocks as instruments for 1969-2007. 68 percent (dark grey) and 95 percent (light grey) confidence sets obtained by inverting the AR_a tests over the parameter space of λ , γ_f and γ_b .

Figure 7: The Phillips curve over time — RR/HFI



Notes: Left column: Confidence sets from the AR_a statistic based on using the Romer-Romer (RR) monetary shocks over 1969q1–1989q4. Right column: Confidence sets from the AR_a statistic based on using the High-Frequency Identified (HFI) monetary surprises (“FF4”, the three month ahead monthly Fed Funds futures) over 1990q1-2007q4. The inner set is the 68% confidence set and the outer-set is the 95% confidence set. The red dot is the 2SLS estimate using lags of the RR or HFI monetary shocks as instruments, and the blue dot is the OLS estimate.

Figure 8: The output gap Phillips curve over time — RR/HFI



Notes: Left column: Confidence sets from the AR_a statistic based on using the Romer-Romer (RR) monetary shocks over 1969q1–1989q4. Right column: Confidence sets from the AR_a statistic based on using the High-Frequency Identified (HFI) monetary surprises (“FF4”, the three month ahead monthly Fed Funds futures) over 1990q1-2007q4. The inner set is the 68% confidence set and the outer-set is the 95% confidence set. The red dot is the 2SLS estimate using lags of the RR or HFI monetary shocks as instruments, and the blue dot is the OLS estimate.