

# Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve<sup>†</sup>

SOPHOCLES MAVROEIDIS, MIKKEL PLAGBORG-MØLLER,  
AND JAMES H. STOCK\*

*We review the main identification strategies and empirical evidence on the role of expectations in the New Keynesian Phillips curve, paying particular attention to the issue of weak identification. Our goal is to provide a clear understanding of the role of expectations that integrates across the different papers and specifications in the literature. We discuss the properties of the various limited-information econometric methods used in the literature and provide explanations of why they produce conflicting results. Using a common dataset and a flexible empirical approach, we find that researchers are faced with substantial specification uncertainty, as different combinations of various a priori reasonable specification choices give rise to a vast set of point estimates. Moreover, given a specification, estimation is subject to considerable sampling uncertainty due to weak identification. We highlight the assumptions that seem to matter most for identification and the configuration of point estimates. We conclude that the literature has reached a limit on how much can be learned about the New Keynesian Phillips curve from aggregate macroeconomic time series. New identification approaches and new datasets are needed to reach an empirical consensus. (JEL C51, D84, E12, E24, E31)*

## 1. Introduction

The idea that there is a trade-off between the rates of inflation and unemployment

(or related measures of real economic activity), at least in the short run, is widely accepted in the economics profession and guides monetary policy making by major

\*Mavroeidis: University of Oxford and INET at Oxford. Plagborg-Møller: Harvard University. Stock: Harvard University. We benefited from comments by Guido Ascari, Gary Chamberlain, Janet Currie, Roger Farmer, Jeff Fuhrer, Max Kasy, Greg Mankiw, Kristoffer Nimark, Adrian Pagan, Barbara Rossi, Francisco Ruge-Murcia, Mark Watson, Mike Woodford, Francesco Zanetti, five anonymous referees, and seminar participants at Bath, Birkbeck, the

ECB, Harvard, Oxford, UPF, the 2013 AEA Meetings, the 2013 DAEiNA meeting, and the 2013 BMRC-QASS conference at Brunel University. Mavroeidis acknowledges financial support from the Leverhulme Trust through a Philip Leverhulme Prize, and the European Commission through Marie Curie Fellowship CIG 293675.

<sup>†</sup>Go to <http://dx.doi.org/10.1257/jel.52.1.124> to visit the article page and view author disclosure statement(s).

central banks. Phillips (1958) provided the first formal statistical evidence on this trade-off using data on wage inflation in the United Kingdom, though the idea had existed well before that.<sup>1</sup> Samuelson and Solow (1960) extended what they called “Phillips’ curve” to U.S. data and to price inflation. The subsequent study of the nature of the Phillips trade-off and its implications for monetary policy and business cycle fluctuations has been one of the most active research areas in economics over the last fifty years.

The Phillips curve has a fascinating history, marked by landmark contributions and heated policy debates; see Gordon (2011) for an insightful recent survey. The literature is so large that it is impossible to address all major contributions to it in any single survey article. Instead, we focus on what is currently the most widely used model of this kind, the New Keynesian Phillips curve (NKPC), which has gained its popularity from its appealing theoretical microfoundations and what appeared to be early empirical success. The theory of the NKPC was laid out mostly in the 1980s and 1990s, and it is, by now, a standard feature of modern macroeconomic textbooks; see, e.g., Woodford (2003) and Galí (2008). The key property of the NKPC is that inflation is primarily a forward-looking process, driven by expectations of future real economic activity, rather than past shocks. From a policy perspective, this severely limits the scope for actively exploiting the Phillips trade-off. Instead, the forward-looking behavior provides a central role for monetary policy rules and opens the door to expectations management and

communications as tools of monetary policy.<sup>2</sup> Indeed, the early empirical success of the NKPC, along with its rigorous microfoundations, has led to its widespread adoption as the key price determination equation in policy models used at central banks around the world.

This survey focuses on the empirics of the NKPC, and more specifically on the recent literature on the estimation of the NKPC using limited-information methods such as the generalized method of moments (GMM). The literature, which dates back to the seminal papers by Roberts (1995), Fuhrer and Moore (1995), Galí and Gertler (1999), and Sbordone (2002), is vast. By focusing on limited-information methods, we exclude estimation of the NKPC using full-information (system) methods, in which the NKPC is one of multiple structural equations within a simultaneous system, 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 NKPC parameters of interest (Lindé 2005; Beyer et al. 2008; Fukač and Pagan 2010). In contrast, because they do not impose economic structure elsewhere in the model, the limited-information methods we consider here are robust to extraneous model misspecification. Full-information methods have been the subject of recent reviews by An and Schorfheide (2007) and Schorfheide (2011). Another promising strand of the literature uses detailed microdata to study price dynamics; see the review by Nakamura and Steinsson (2013). These papers are outside the scope of our review,

<sup>1</sup>Mankiw (2001), p. C46 offers a prescient quote from the work of Hume (1987[1752]). Irving Fisher demonstrated a strong correlation between U.S. inflation and unemployment already in 1926—cf. Fisher (1973)—but he interpreted the causality as running from prices to real economic activity.

<sup>2</sup>The policy relevance of this research agenda is highlighted in two speeches by Federal Reserve Chairman Bernanke (2007, 2008).

which focuses only on studies that use macro data.

Despite the apparent early empirical success of the NKPC, the literature that we survey is full of puzzles. What should be relatively innocuous changes in instruments used, in vintages of data and in model specification all seem to matter significantly for the results. For example, simply reestimating the benchmark NKPC in Galí and Gertler (1999), using the same data series, method, and time period, but with revised data, reduces the estimate on the activity variable (real marginal cost) by half and makes the coefficient no longer statistically significant (Rudd and Whelan 2007). This is but a single example of a high degree of sensitivity in this literature to minor econometric changes. Our goals in this review are to understand the reasons for this sensitivity and, more specifically, to provide a clear understanding of the role of expectations that integrates across the different papers and specifications. We do so first by reviewing the papers in the literature and the econometric theory underlying their approaches, then by estimating multiple specifications using a common dataset. Since the first empirical work on the NKPC, there have been significant methodological developments in the area of estimation with weak instruments, and our analysis draws heavily on these methods to help explain the puzzles in the literature. Earlier surveys on the NKPC include Henry and Pagan (2004), Ólafsson (2006), Rudd and Whelan (2007), Nason and Smith (2008b), and Tsoukis, Kapetanios, and Pearlman (2011). We extend these surveys by emphasizing the many econometric issues raised by estimation of the NKPC, and our empirical analysis spans a much wider range of estimation approaches and specifications than previous individual papers have considered (in fact, we suspect we estimated more NKPC specifications than the entire preceding literature combined).

The outline of the paper is as follows. Section 2 briefly reviews the derivation of the NKPC under the Calvo (1983) assumption on price setting, followed by a description of the main extensions and empirical specifications. We emphasize that uncertainty about the NKPC parameters translates into significant uncertainty about the New Keynesian model's policy implications. Departing from the rational expectations assumption has nontrivial consequences for the model. Due to space constraints, we do not discuss recent movements away from the NKPC, such as state dependent pricing (Dotsey, King, and Wolman 1999) or imperfect information (Mankiw and Reis 2011).

Section 3 reviews the various limited-information econometric methods that have been used in the study of the NKPC, including instrumental variables (IV), minimum distance (MD), and maximum likelihood (ML), and the use of survey data on inflation expectations. Furthermore, we propose a new idea for identification using data revisions as external instruments that obviates the need to impose ad hoc exclusion restrictions on the dynamics. We compare the different methods under both strong and weak instruments. In the case of strong instruments, we provide results, not previously explicitly available in the literature, that permit comparison of the various estimators, and we highlight the trade-off between efficiency and robustness. We pay particular attention to methods that are robust to weak instruments. The expectation of future inflation is the key endogenous covariate in the NKPC. Because inflation is notoriously hard to forecast, it is difficult to find exogenous (i.e., lagged) economic variables that correlate strongly with expected future inflation; in other words, potential instruments that satisfy the exclusion restriction will likely be weak. We show that when this is the case, even estimators that do not explicitly rely on instrumental variable techniques can be

severely biased. Hence, weak instrument issues provide a unifying explanation of the sensitivity of NKPC estimates and of the puzzling disagreement between analyses based on standard inference procedures. We also discuss complications arising from misspecification of the NKPC.

Section 4 surveys the vast empirical literature on the NKPC, covering over one hundred papers we have found on this topic. The initial success of the rational expectations specification (with the labor share as the proxy for firm marginal cost) around the turn of the millennium was quickly followed by doubts about robustness to data choices and estimation methods. A plethora of extensions of the basic model have been pursued with no clear universal consequences. Approaches that exploit the nonrationality of expectations or positive trend inflation have recently been gaining traction, while a parallel strand of the literature has emphasized the weak identification issues inherent in estimating the NKPC. Due to differences across papers in datasets, instrument choices, specifications, estimators, and attention to the weak identification problem, no consensus has been reached on parameter values or the reasons for the variability in estimates, and policy implications are entirely unclear. To date, few papers have sought to compare or integrate more than a couple of the empirical strategies, leaving the research program in considerable disarray.

Section 5 provides a new set of empirical results based on a common dataset and a flexible empirical strategy that spans multiple popular approaches in the literature. Like most papers, we focus on quarterly post-war U.S. data. Apart from the standard data series, we have assembled a unique real-time dataset on the labor share. By computing point estimates of the NKPC parameters from a comprehensive set of specifications that combine data and model choices from the literature, we find that the specification

uncertainty is vast: almost any parameter combination that is even remotely close to the range considered in the literature can be generated by some a priori unobjectionable specification. Furthermore, given a particular specification, the sampling uncertainty is large. We show this by computing weak identification robust confidence sets for several benchmark specifications. One type of specification that appears to be typically better identified uses survey forecasts as proxies for inflation expectations; however, such specifications are only microfounded if survey forecasts are rational, which does not seem to be the case empirically. Survey specifications are also less suitable for counterfactual policy analysis and forecasting.

Section 6 concludes by summarizing the main lessons from the literature and our empirical exercise. We recommend that future research pursues substantially new types of datasets, as well as estimation approaches that are tailored to handle the identification problem.

## 2. *Economic Theory, Specifications, and Policy Implications*

The origins of the NKPC can be traced back to the late seventies, in the work of Fischer (1977) and Taylor (1979). The NKPC is a forward-looking model of inflation, according to which current inflation is determined by expected future inflation and marginal costs. It implies that monetary policy can affect inflation through the management of inflation expectations. This contrasts sharply with the traditional “old” Phillips curve, which yields a strongly path-dependent inflation process, so that disinflation can be slow and costly (Mankiw 2001). The importance of expectations was highlighted early on by Phelps (1967) and Friedman (1968). Their so-called “expectations-augmented” Phillips curves emphasized that the inflation/unemployment trade-off shifts with expected

inflation, a property shared by the Phillips curves of the New Classical literature of the 1970s (e.g., Sargent and Wallace 1975). The key difference of these Phillips curves from the NKPC is that past (and thus predetermined) expectations about current inflation matter, not expectations about the future.

### 2.1 Economic Foundations of the Model

A simple derivation of the NKPC can be obtained as follows.<sup>3</sup> The basic ingredient for the derivation of the NKPC is a microeconomic environment with identical monopolistically competitive firms facing constraints on price adjustment. We consider here only time-contingent pricing constraints. The details of the constraints do not matter much for the final form of the NKPC (Roberts 1995), so we focus on the assumption of Calvo (1983), which is analytically attractive. Prices are expressed in logs and inflation in percentage points. All variables, except prices and inflation, are expressed in percent deviations from a zero-inflation steady state. We discuss the assumption of zero steady-state inflation below. The Calvo framework assumes that each firm in the economy has a constant probability,  $1 - \theta$ , of optimally adjusting its price in any given period. Because the economy consists of a continuum of identical firms, by the law of large numbers it follows that a fraction,  $\theta$ , of firms cannot change their prices in any given period, and that prices remain fixed on average for  $1/(1 - \theta)$  periods. Therefore, the parameter  $\theta \in (0, 1)$  is an index of price rigidity. Assume that each firm produces a differentiated product and faces a constant price elasticity of demand  $\varepsilon > 1$  for its product. Let  $p_{i,t}^*$  denote the optimal price chosen by a firm  $i \in [0, 1]$  if it gets to reoptimize in period  $t$ . By the law of large numbers, the aggregate price level  $p_t$  evolves as a convex combination of last period's price

and the cross-sectional average of the current reset prices:

$$(1) \quad p_t = \theta p_{t-1} + (1 - \theta) \int_0^1 p_{i,t}^* di.$$

In the absence of price rigidity, monopolistically competitive firms set prices as a markup over their nominal marginal costs. With price rigidity, maximization over all expected discounted future profits induces firms to take into account the probability that they will not be able to reset their prices optimally in the future. Let  $\beta$  denote the common subjective discount factor. The optimal reset price can, in a first-order log-linearization, be expressed as a markup over a weighted average of current and expected future marginal costs:

$$(2) \quad p_{i,t}^* = (1 - \theta\beta) \sum_{j=0}^{\infty} (\theta\beta)^j E_{i,t}(mc_{i,t+j}^n),$$

where  $mc_{i,t+j}^n$  is the nominal marginal cost faced at time  $t + j$  for a firm  $i$  that was last able to reset its price optimally at time  $t$ , and  $E_{i,t}$  denotes the expectation with respect to the beliefs of firm  $i$ . Relating this to the aggregate marginal cost  $mc_t^n$  requires a specification of the production function. Suppose the production function is Cobb–Douglas with labor elasticity  $1 - \alpha$ . Then it can be shown that, to a first-order approximation,

$$(3) \quad mc_{i,t+j}^n = mc_{t+j}^n + \frac{\alpha\varepsilon}{1 - \alpha} (p_{i,t}^* - p_{t+j}).$$

Substituting (3) into (2) and rearranging yields

$$(4) \quad p_{i,t}^* - p_{t-1} = (1 - \theta\beta) \sum_{j=0}^{\infty} (\theta\beta)^j \times [\kappa E_{i,t}(mc_{t+j}) + E_{i,t}(\pi_{t+j})],$$

where  $\kappa = \frac{1 - \alpha}{1 - \alpha + \alpha\varepsilon} \leq 1$ , inflation is given by  $\pi_t = p_t - p_{t-1}$ , and  $mc_t = mc_t^n - p_t$

<sup>3</sup> For detailed derivations, see chapter 3 of Woodford (2003) or chapter 3 of Galí (2008).



denotes aggregate real marginal costs. Inserting (4) into (1), we find

$$(5) \quad \pi_t = (1 - \theta)(1 - \theta\beta) \sum_{j=0}^{\infty} (\theta\beta)^j \\ \times [\kappa \hat{E}_t(mc_{t+j}) + \hat{E}_t(\pi_{t+j})],$$

where  $\hat{E}_t = \int_0^1 E_{i,t} di$  is the *cross-sectional average expectation operator*.

Until this point, we have not imposed any restrictions on the nature of firms' beliefs about future economic conditions. Assume now that firms have identical, rational expectations, i.e.,  $E_{i,t} \equiv E_t$ . Then the cross-sectional expectation equals the rational expectation,  $\hat{E}_t = E_t$ . If we shift equation (5) forward by one period and take time- $t$  expectations on both sides, we get

$$(6) \quad E_t(\pi_{t+1}) = (1 - \theta)(1 - \theta\beta) \sum_{j=0}^{\infty} (\theta\beta)^j \\ \times [\kappa E_t(mc_{t+j+1}) + E_t(\pi_{t+j+1})],$$

where we have, importantly, used the law of iterated expectations,  $E_t[E_{t+1}(\cdot)] = E_t(\cdot)$ . The infinite sum on the right-hand side of (6) is closely related to the infinite sum on the right-hand side of (5) (with  $\hat{E}_t = E_t$ ). Combining these two equations gives rise to an expectational difference equation for inflation:

$$(7) \quad \pi_t = \beta E_t(\pi_{t+1}) + \frac{(1 - \theta)(1 - \theta\beta)}{\theta} \kappa mc_t.$$

Empirical analyses of the NKPC use proxies for the real marginal cost measure  $mc_t$ . Under the already exploited assumption of Cobb–Douglas production technology,  $mc_t$  is proportional both to the labor share of income (nominal labor compensation divided by nominal output) and the output gap (the deviation of real output from the level that would obtain if prices were fully flexible). Letting  $x_t$  denote a candidate proxy

for real marginal cost and adding an unrestricted unobserved disturbance term,  $u_t$ , we can rewrite the model as

$$(8) \quad \pi_t = \beta E_t(\pi_{t+1}) + \lambda x_t + u_t.$$

This is the baseline purely forward-looking NKPC that we will refer to in subsequent sections. The disturbance term  $u_t$  can be interpreted as measurement error or any other combination of unobserved cost-push shocks, such as shocks to the markup or to input (e.g., oil) prices.

### 2.1.1 Nonrational Expectations

While we did not need to impose assumptions on firms' individual beliefs to arrive at expression (5) for inflation, the derivation of equation (6)—and thus the difference equation (7)—crucially relied on the fact that the rational expectation operator  $E_t$  satisfies the law of iterated expectations. However, this law does not hold in general for the cross-sectional average expectation operator  $\hat{E}_t$ , even if individual firm expectations do satisfy the law. Hence, under general nonrational expectation formation, the difference equation specification of the NKPC in equation (7) is not consistent with the above microeconomic foundations that constitute the standard New Keynesian modeling framework (a similar argument is given by Preston 2005).<sup>4</sup> This will be the case, for instance, if firms' expectations are not based on the same information set or if they are not perfectly model-consistent. As we discuss in section 3.1, this has implications for empirical tests of the NKPC that use survey forecasts to proxy for the expectation term. Preston (2005), Angeletos and La'O (2009),

<sup>4</sup> It is possible for nonrational expectations models to result in a difference equation NKPC. Adam and Padula (2011) derive a difference equation under a set of assumptions on the possibly nonrational cross-sectional expectation operator; their "Condition 1" essentially invokes the law of iterated expectations at the aggregate level.

and Kurz (2011) have derived microfounded inflation equations in certain models with nonrational or heterogeneous expectations.

## 2.2 Extensions

It was recognized early on that the purely forward-looking NKPC (8) has difficulty fitting aggregate U.S. inflation dynamics; see Fuhrer and Moore (1995) and Galí and Gertler (1999). This led to specifications that included lagged inflation terms in the model. This is often called “intrinsic inflation persistence.” Galí and Gertler (1999) introduce lagged terms by assuming that a fraction of firms update their prices using some backward-looking rule of thumb, while Fuhrer and Moore (1995) generate persistence through staggered relative wage contracts. Another popular device is to assume that the fraction  $\theta$  of firms that are unable to reoptimize their prices in the Calvo model instead index prices to past inflation; see Christiano, Eichenbaum, and Evans (2005) and Sbordone (2005, 2006).<sup>5</sup>

Because it could be thought of as a combination of new and old Phillips curves, such a specification is referred to as a “hybrid NKPC.” In principle, if the objective is to nest traditional Phillips curves, one could allow for any number of lagged inflation terms in the model. An appropriate baseline hybrid specification that nests most traditional Phillips curve specifications would take the form

$$(9) \quad \gamma(L) \pi_t = \gamma_f E_t(\pi_{t+1}) + \lambda x_t \\ + \eta' w_t + u_t,$$

where  $\gamma(L) = 1 - \gamma_1 L - \gamma_2 L^2 - \dots - \gamma_l L^l$  is a lag polynomial,  $x_t$  is the main forcing

variable,  $w_t$  denotes additional controls, and  $u_t$  is an unobserved shock. If the lag polynomial only features one lag, the coefficient  $\gamma_1$  is often denoted  $\gamma_b$ . Equation (9) nests the pure NKPC (8) with  $\gamma(L) = 1$  and  $\eta = 0$ . With  $\gamma_f = 0$ , it also nests the backwards-looking “old” Phillips curve and, in particular, Gordon’s (1990) “triangle” model. It is more general than the typical hybrid NKPC specifications that only include one lag of inflation, such as Galí and Gertler (1999), Sbordone (2005, 2006), and Christiano, Eichenbaum, and Evans (2005). The latter are based on indexation to last quarter’s inflation, but this is clearly arbitrary and can be easily generalized to include more general indexation schemes, such as a weighted average of inflation over the previous four quarters (this includes, as a special case, indexation to last *year’s* inflation, which nests the Atkeson and Ohanian 2001 specification), or richer rule-of-thumb behavior by backward-looking firms; see Zhang and Clovis (2010). Restrictions on  $\gamma(L)$  are exclusion restrictions on the dynamics of inflation, which are typically used to provide instruments for identification. Therefore, such exclusion restrictions are not innocuous. A popular restriction, which is seldom rejected by the data, is that the inflation coefficients sum to 1, i.e.,  $\gamma_f = 1 - (\gamma_1 + \gamma_2 + \dots + \gamma_l) = \gamma(1)$ .

The parameters of equation (9) are often referred to as “reduced-form” or “semistructural” because they are functions of the deeper structural parameters of the microfounded model. For example, when the discount factor is one, the Galí and Gertler (1999) specification has  $\gamma_f = \theta/(\theta + \omega)$  and  $\gamma_1 = 1 - \gamma_f$ , where  $\omega$  is the fraction of price setters who are backward-looking. Restrictions on the admissible range of the deep parameters can affect the range of the semistructural ones and thus have nontrivial implications for inference.

<sup>5</sup>Sheedy (2010) and Yao (2011) observe that intrinsic persistence arises when the hazard rate of price resetting (the Calvo parameter) is not constant. See also Kozicki and Tinsley (2002) for further discussion on the sources of lagged inflation dynamics in the NKPC.

### 2.2.1 Trend Inflation

The derivation of equation (8) follows from log-linearizing firms' optimizing conditions around a zero-inflation steady state. Allowing for nonzero steady state inflation, often referred to as "trend inflation," has important implications for the specification of the NKPC, as established by Kozicki and Tinsley (2002), Ascari (2004), and Cogley and Sbordone (2008). Trend inflation  $\bar{\pi}_t$  corresponds to long-run inflation expectations, i.e.,  $\bar{\pi}_t = \lim_{T \rightarrow \infty} E_t(\pi_{t+T})$ . With nonzero trend inflation, the NKPC cannot, in general, be written in the difference equation form (9), as extra forward-looking terms enter on the right-hand side and the semi-structural parameters are functions of trend inflation. However, Cogley and Sbordone (2008) show that if nonresetting firms' prices are indexed to a mixture of past inflation  $\pi_{t-1}$  (weight  $\rho$ ) and current trend inflation  $\bar{\pi}_t$  (weight  $1 - \rho$ ), then

$$(10) \quad \hat{\pi}_t = \gamma_f E_t \hat{\pi}_{t+1} + \gamma_b \hat{\pi}_{t-1} + \lambda x_t \\ + \gamma_b (\beta E_t \Delta \bar{\pi}_{t+1} - \Delta \bar{\pi}_t),$$

where  $\gamma_f = \beta/(1 + \beta\rho)$ ,  $\gamma_b = \rho/(1 + \beta\rho)$ , and we define the *inflation gap*  $\hat{\pi}_t = \pi_t - \bar{\pi}_t$ . If trend inflation is constant,  $\bar{\pi}_t \equiv \bar{\pi}$ , the last term above drops out, and we are left with a standard NKPC in which the inflation gap replaces raw inflation.<sup>6</sup> If, furthermore,  $\beta = 1$ , then  $\gamma_f + \gamma_b = 1$ , so the NKPC can be expressed in terms of the change in inflation  $\Delta \pi_t = \Delta \hat{\pi}_t$ , causing the constant trend inflation to drop out of the relation altogether. Suppose instead that trend inflation is time-varying. If  $\beta = 1$  and  $E_t \Delta \bar{\pi}_{t+1} = \Delta \bar{\pi}_t$  (the change in trend inflation is unforecastable), the last term on the right-hand side of

(10) vanishes and we are left with an NKPC relation in terms of the inflation gap. Such a specification also obtains if nonreset prices are fully indexed to trend inflation ( $\rho = 0$ ).

In the rest of this paper we will focus on inflation gap specifications of the NKPC, i.e., relations of the form (10) without the last term on the right-hand side. We do this to keep the exposition simple, but acknowledge that we do not give the trend inflation issue as much attention as it deserves. Interested readers are referred to the review by Ascari and Sbordone (2013).

### 2.3 Policy Implications

Ideally, estimation uncertainty in the NKPC parameters would only translate into limited ambiguity about the effects of shocks and policy interventions on the broader economy. Unfortunately, this is not the case for the range of NKPC parameter estimates reported in the literature. Figure 1 displays impulse responses of inflation and the output gap to a twenty-five basis point monetary policy shock in the canonical three-equation New Keynesian model (Galí 2008). The model and calibration are described in section A.1 in the Appendix, and they are based on a hybrid NKPC with one lag of inflation whose coefficient  $\gamma_1$  is equal to  $1 - \gamma_f$ . By a monetary policy shock, we mean a shock to the innovation in the AR(1) process for the Taylor rule disturbance. We treat the semi-structural parameters  $\gamma_f$  and  $\lambda$  as being variation-free from the remaining structural parameters (which are calibrated as in chapter 3.4 of Galí 2008) and vary them over a set of values that is consistent with the spread of estimates reported in the literature, see section 4 below, namely  $\gamma_f = 0.3, 0.4, \dots, 0.8$  and  $\lambda = 0.01, 0.03, 0.05$ . As figure 1 shows, this leads to a wide range of possible impulse responses, with substantially different short-run dynamics and steady state return times. For given  $\lambda$ , lower values of  $\gamma_f$  imply more sluggish adjustment and more

<sup>6</sup>This is equivalent to an NKPC written in terms of raw inflation  $\pi_t$  with an intercept that depends on  $\bar{\pi}$ , as established by Yun (1996) for the special case  $\rho = 0$ .



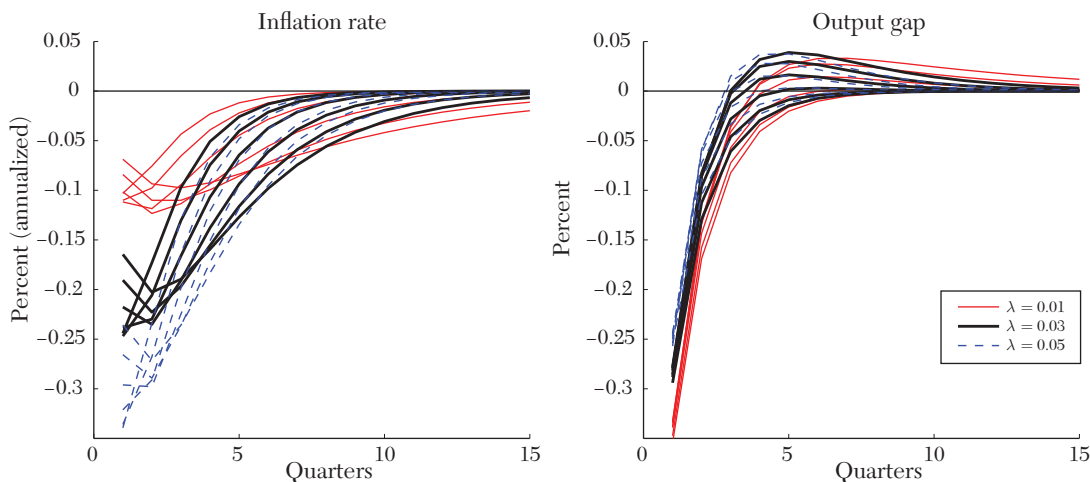


Figure 1. Impulse Responses to Monetary Policy Shock

Notes: Impulse responses of the inflation and the output gap to a 25 basis point monetary policy shock in a standard three-equation New Keynesian model with  $\gamma_f = 0.3, 0.4, \dots, 0.8$  and  $\lambda = 0.01, 0.03, 0.05$ . The other parameters are calibrated to the benchmark values listed in Appendix A.1. More sluggish responses correspond to lower values of  $\gamma_f$ . The figure was generated using Dynare (Adjemian et al. 2011).

hump-shaped dynamics in inflation. The disparity between the high- $\gamma_f$  and low- $\gamma_f$  dynamics increases the lower that  $\lambda$  is. For  $\lambda = 0.03$  (the thick curves in the figure), the effect of the monetary policy shock is felt for about twice as long for  $\gamma_f = 0.3$  than for  $\gamma_f = 0.8$ . The most negative cumulative fifteen-quarter inflation impulse response in figure 1 is five times larger in magnitude than the least negative cumulative inflation response; the ratio between the most and least negative cumulative output gap responses is 5.7. This sensitivity of key economic measures to the NKPC parameters extends beyond the simple model considered here, as demonstrated by Fuhrer (1997), Mankiw (2001), and Estrella and Fuhrer (2002).<sup>7</sup> If indeed the NKPC is a good approximation to actual price setting, it is therefore highly desirable

<sup>7</sup>Adding estimation uncertainty in the non-NKPC equations to the mix will, of course, generate even greater uncertainty about appropriate impulse responses.

from a policy perspective to obtain precise estimates of the NKPC coefficients.<sup>8</sup>

### 3. Econometric Methods

In this section we describe the main estimators that have been used in the literature and discuss their properties under strong and weak identification. We focus on the “semi-structural” parametrization of the NKPC, as opposed to the underlying structural parameters, to facilitate comparison across different specifications.<sup>9</sup> For ease of exposition, we focus in this section on estimation of

<sup>8</sup>The degree of inflation indexation also influences the relative optimality of policies that flexibly target the price level or inflation rate (Woodford 2003, chapter 8.2.1).

<sup>9</sup>Moreover, estimation of the structural parameters raises some additional issues if the mapping from the semi-structural to structural parameters is not injective, so that the latter are not globally identified even when the former are (Ma 2002).

the pure NKPC (8), but all our points generalize to the hybrid specification (9).

### 3.1 Estimators

A glance at the NKPC (8) reveals two immediate estimation issues. First, as noted by Roberts (1995), the forcing variable  $x_t$  may be correlated with the structural error term  $u_t$  (e.g., they may both be driven in part by cost-push shocks). Second, the inflation expectation term  $E_t(\pi_{t+1})$  is certainly endogenous and—even worse—it is unobservable. Different empirical approaches in the literature differ mainly in the way they deal with inflation expectations. They can be usefully categorized as follows:

1. Replace expectations by realizations and use appropriate instruments (generalized instrumental variables (GIV)).
2. Derive expectations from a particular reduced-form model (vector autoregression (VAR)).
3. Use direct measures of expectations (Survey).

We also propose a new strategy based on the use of data revisions as instruments. This can be thought of as a variant of the first approach. All of these approaches can be implemented using GMM (Hansen 1982). We use GMM as a common unifying framework in our discussion and empirical work, which is convenient because weak identification robust methods are readily available for GMM. GMM estimation is briefly described in section A.2.1 in the Appendix.

#### 3.1.1 GIV

This approach was originally proposed for the estimation of rational expectation models by McCallum (1976). Hansen and Singleton (1982), who studied estimation of Euler equation models, called it GIV

estimation. It has been popularized in the estimation of the NKPC by the seminal contributions of Roberts (1995) and Galí and Gertler (1999). It is the most common approach in the literature because it is simple to implement and more robust than the alternatives. Identification is obtained via exclusion restrictions, i.e., excluding lags of variables from the model and using them as instruments.

The simplest and most common implementation is to replace the rational expectation  $E_t(\pi_{t+1})$  in the difference equation (8) by the realization  $\pi_{t+1}$ . This yields

$$(11) \quad \pi_t = \beta\pi_{t+1} + \lambda x_t + \underbrace{u_t - \beta[\pi_{t+1} - E_t(\pi_{t+1})]}_{\tilde{u}_t},$$

where the residual  $\tilde{u}_t$  differs from  $u_t$  because it includes the future one-step-ahead inflation forecast error. Let  $\vartheta = (\beta, \lambda)$  and define the “residual” function

$$(12) \quad h_t(\vartheta) = \pi_t - \beta\pi_{t+1} - \lambda x_t.$$

Suppose  $Z_t$  is a vector of valid instruments, such that

$$(13) \quad E[Z_t h_t(\vartheta)] = 0$$

holds at  $\vartheta = \vartheta_0$ , the true parameter value. Efficient GMM estimation (see section A.2.1 in the Appendix) is based on the sample moments  $f_T(\vartheta) = T^{-1} \sum_{t=1}^T Z_t h_t(\vartheta)$  and a heteroskedasticity and autocorrelation consistent (HAC) estimator of their variance, because  $h_t(\vartheta_0)$  is generally autocorrelated due to the presence of the inflation forecast error.

The most common identifying assumption in the literature, which is the one used by Galí and Gertler (1999), is that the cost-push shock  $u_t$  satisfies  $E_{t-1}(u_t) = 0$ .

Under the rational expectations assumption, this implies  $E_{t-1}(\tilde{u}_t) = 0$  by the law of iterated expectations. This yields unconditional moment restrictions of the form (13) with  $Z_t = Y_{t-1}$ , for any vector of predetermined (i.e., known at time  $t-1$ ) variables  $Y_{t-1}$ . Any predetermined variables can be used as instruments, and different implementations of GIV estimation differ mainly in the choice of instruments. For example, Rudd and Whelan (2005) obtain alternative GMM moment conditions by iterating the NKPC forward. As we show in section A.2.2 in the Appendix, this is equivalent to estimating the original NKPC difference equation (8) with transformed instruments.

### 3.1.2 VAR

Almost all of the papers that use the second approach rely on the assumption that the reduced-form dynamics of the variables can be represented by a finite-order VAR. This is why we refer to it as the VAR approach.

Suppose the information set consists of current and lagged values of some  $n$ -dimensional vector  $z_t$  (this includes at least  $\pi_t, x_t$ , and any other variables that are to be used as instruments), and that  $z_t$  admits a finite-order VAR representation of order  $l$ , which can be written in companion form as

$$(14) \quad Y_t = AY_{t-1} + V_t,$$

where  $Y_t, V_t$  are  $nl \times 1$  and  $A$  is  $nl \times nl$ . If  $E_t(V_{t+1}) = 0$ , equation (14) implies that  $E_t(\pi_{t+1}) = \zeta'Y_t$ , where  $\zeta$  is the row of  $A$  that corresponds to the inflation equation in the VAR. Substituting this in the NKPC (8), yields the moment conditions  $E[(\pi_t - \beta Y_t' \zeta - \lambda x_t) Y_{t-1}] = 0$ , which determine the structural parameters  $\vartheta = (\beta, \lambda)$  given  $\zeta$ , and  $\zeta$  is identified by the reduced-form equation  $E[(\pi_t - Y_{t-1}' \zeta) Y_{t-1}] = 0$ . The VAR assumption suggests using  $Y_{t-1}$  as instruments, so the model can be

estimated by GMM with the  $2nl$  moment conditions

$$(15) \quad E[\hat{h}_t(\vartheta_0, \zeta_0) \otimes Y_{t-1}] = 0, \quad \text{where}$$

$$(16) \quad \hat{h}_t(\vartheta, \zeta) = (\pi_t - \beta Y_t' \zeta - \lambda x_t; \pi_t - Y_{t-1}' \zeta)'$$

Here  $\zeta_0$  is the true value of  $\zeta$ . The seminal papers in this strand of the literature are Fuhrer and Moore (1995) and Sbordone (2002), and they use two different econometric implementations: maximum likelihood (VAR-ML) and minimum distance (VAR-MD), respectively. We describe these methods in section A.2.3 in the Appendix. The realization that the VAR assumption to identification can be easily imposed in GMM, which we refer to as VAR-GMM, appears to be new in the literature.<sup>10</sup> VAR-GMM, VAR-ML, and VAR-MD are numerically identical if the model is just-identified, i.e., if the dimension of  $Y_t$  is the same as the dimension of the parameter vector  $\vartheta$ .

Bayesian inference is mostly used for full-information analysis of DSGE models (e.g., Lubik and Schorfheide 2004), which is beyond the scope of the present survey. A few notable limited-information Bayesian studies of the NKPC are reviewed in section 4. These all rely on the VAR assumption.

### 3.1.3 Surveys

This approach was introduced by Roberts (1995) and, after a lag, has seen increasing popularity. Under the survey approach, direct measures of expectations from surveys, e.g., the Survey of Professional Forecasters (SPF) or the Federal Reserve's "Greenbooks," are used as proxies for inflation expectations in the NKPC. Let  $\pi_{t+j|t}^s$  denote the  $j$ -step-ahead

<sup>10</sup>Guerrieri, Gust, and López-Salido (2010) use a VAR-GMM procedure to estimate an open-economy NKPC, but they do not compare their estimates to non-VAR specifications. See also Sbordone (2005) for a discussion of the relation between VAR-MD and GIV.

survey forecast of inflation at time  $t$ . The most common implementation substitutes the one-quarter-ahead forecast  $\pi_{t+1|t}^s$  for  $E_t(\pi_{t+1})$  in the NKPC (8) to get

$$(17) \quad \pi_t = \beta\pi_{t+1|t}^s + \lambda x_t + u_t + \beta\varepsilon_t,$$

where

$$(18) \quad \varepsilon_t \equiv E_t(\pi_{t+1}) - \pi_{t+1|t}^s.$$

The survey error  $\varepsilon_t$  can be a combination of measurement error and news shocks, the latter arising when survey responses are based on a smaller information set than the one the agents in the model use. Identification depends on the properties of this survey error  $\varepsilon_t$ , as well as on the correlation between  $\pi_{t+1|t}^s$  and the cost-push shock  $u_t$ . Some authors treat  $\pi_{t+1|t}^s$  as exogenous, but we argue below that the assumptions underlying this are too strong. Alternatively, one may use  $\pi_{t+1|t-1}^s$ , which is certainly predetermined, instead of  $\pi_{t+1|t}^s$ , which is typically measured within the quarter. Another possibility is to treat survey forecasts as endogenous and use predetermined variables as instruments, as in the GIV approach.

Because *estimation* of the NKPC using survey forecast data obviates the need for modeling inflation expectations, some authors interpret the procedure as allowing for nonrational price setting. When the NKPC is to be used for policy purposes, such as in forecasting, the lack of a dynamic model for expectations becomes a disadvantage. To date, few papers have attempted to model nonrational survey expectations formation (see Fuhrer 2012). As argued in section 2.1, the difference equation NKPC (8) is, in general, inconsistent with the standard New Keynesian framework if firm expectations are not model-consistent or if they are based on dispersed information. Consequently, survey specifications of the form (17) are only microfounded if the inflation forecasts

firms rely on are rational and identical across firms. However, common empirical findings are that survey forecasts of inflation violate testable implications of rationality and the dispersion of expectations across individual forecasters is large; see, for example, Thomas (1999) and Mankiw, Reis, and Wolfers (2004). This point does not seem to have been taken to heart by the empirical NKPC literature. While the proper microfoundations for price setting under nonrational expectation formation are lacking, the survey forecast specification may still be taken as a primitive, which nicely summarizes our intuition about price setting being partially forward-looking, partially backward-looking as well as being responsive to aggregate demand conditions.

Nunes (2010), Fuhrer and Olivei (2010), and Fuhrer (2012) estimate versions of the NKPC where inflation expectations are specified as a combination of rational and survey expectations, replacing  $E_t(\pi_{t+1})$  with  $\varphi E_t(\pi_{t+1}) + (1 - \varphi)\pi_{t+1}^s$  in the NKPC (8).<sup>11</sup> The parameter  $\varphi$  is not identified if survey expectations are rational.

### 3.1.4 External Instruments

Many of the time series typically used to estimate the NKPC, such as GDP deflator inflation, the labor share, and the output gap, undergo large revisions over time. Because firms' expectations of current and future economic conditions feature prominently in the New Keynesian model, it is crucial to keep in mind the availability of data at different points in time when estimating the NKPC. With access to real-time data, i.e., datasets of different vintages, one possible identification

<sup>11</sup>Nunes (2010) argues that this nesting can be motivated by a variant of the argument of Galí and Gertler (1999), where a fraction of rule-of-thumb firms set their prices using published professional inflation forecasts as opposed to lagged inflation. This justification cannot apply to the Federal Reserve's Greenbook forecasts, as they are only released to the public with lags of several years.

strategy is to use past data vintages of inflation and the forcing variable (and perhaps other time series) as instruments.

We appear to be the first to consider estimation of the NKPC using such external instruments. An advantage of the approach is that it avoids placing exclusion restrictions on the NKPC. This is discussed more formally in section A.2.4 of the Appendix. Regardless of the number and types of right-hand side variables in the hybrid specification (9), past-vintage instruments should not affect current inflation if we control for the latest-vintage data, and thus are plausibly uncorrelated with the GIV error  $\tilde{u}_t$  in (11). Hence, real-time instruments are plausibly exogenous, although they could potentially be very weak. In section 5 we empirically evaluate the success of the external instruments approach.

### 3.2 Comparison of Estimators under Conventional Asymptotics

To compare the properties of the sampling distributions of the various estimators, we start out by outlining the trade-offs between efficiency and robustness under the conventional asymptotic approximations. The conventional asymptotic theory, which is the main analytical tool in graduate econometrics textbooks, implies that GMM estimators of the parameters  $\vartheta$  are consistent and asymptotically normal under certain regularity conditions; see Newey and McFadden (1994). In linear IV models, this asymptotic theory has the first-stage  $F$  statistic, which measures the explanatory power of the instruments for the endogenous variable, tending to infinity at the same rate as the sample size, so we refer to the theory as *strong-instrument asymptotics*. In section 3.3, we will argue that the strong instrument approximation is not empirically relevant in the present context, as it does not capture the kind of near-irrelevance of the instruments that is prevalent in the estimation of

the NKPC. Still, it is useful to first establish the properties of the estimators in the most familiar analytical framework.

Most of the estimators are obtained from conditional moment restrictions of the form  $E_{t-1}[h_t(\vartheta_0)] = 0$ , for which the theory of optimal instruments of Chamberlain (1987) provides an efficiency bound.  $E_{t-1}[h_t(\vartheta_0)] = 0$  implies that all predetermined variables  $Y_{t-j}$ ,  $j \geq 1$ , are admissible instruments. The optimal choice of instruments is derived in section A.2.1 of the Appendix, and we summarize the findings below.

#### 3.2.1 GIV

For GIV estimators, the residual function  $h_t(\vartheta_0) = u_t - \beta[\pi_{t+1} - E_t(\pi_{t+1})]$  is generally autocorrelated, so optimal instruments are given by an infinite-order moving average of  $Y_{t-1}$ . No papers have attempted to use optimal instruments because their derivation requires modeling the conditional mean and variance of the data.<sup>12</sup> Therefore, we cannot rank GIV estimators reviewed in this paper in terms of efficiency. Indeed, claims about their relative efficiency are not formally justified.

#### 3.2.2 VAR

In contrast, the VAR-GMM estimator is based on the moment conditions (15). Let  $v_{\pi t}$  denote the VAR error term in the reduced-form inflation equation in (14). When the VAR assumption holds, the VAR residuals  $\hat{h}_t(\vartheta_0, \zeta_0) = (u_t, v_{\pi t})'$  are serially uncorrelated, unlike the GIV residuals  $h_t(\vartheta_0)$  in (12). If the VAR residuals are also conditionally homoskedastic, then the VAR-GMM estimator does indeed use optimal

<sup>12</sup>Fuhrer and Olivei (2005) propose a GMM implementation of a restricted VAR approach, which they refer to as producing optimal instruments. Because these instruments are linear combinations of  $Y_{t-1}$ , they are not optimal in Chamberlain's sense; cf. section A.2.1.



instruments, and is therefore asymptotically more efficient than the corresponding GIV estimators that do not impose the VAR assumption. But note that the key conditions for this, strong instruments and conditional homoskedasticity, are arguably too strong in this application.

Fuhrer and Moore (1995), and several later papers, estimate the NKPC by ML. To evaluate the likelihood, they combine the NKPC with reduced-form equations for all variables other than inflation to form a complete “limited-information” system of equations, and use an algorithm by Anderson and Moore (1985) that finds a rational expectations solution for any given value of the parameters. For certain parameter combinations, there may be multiple stable rational expectations solutions, a situation known as indeterminacy (see, e.g., Lubik and Schorfheide 2004), and so the likelihood is not uniquely determined by the NKPC and remaining reduced-form parameters. When the solution is unique (determinacy), the reduced-form is a (restricted) finite-order VAR. The Fuhrer–Moore approach restricts the parameter space to the region in which indeterminacy does not occur. The determinacy assumption is standard in full-information estimation of DSGE models—for example, it is imposed by Dynare (Adjemian et al. 2011)—but it can be restrictive.<sup>13</sup> Kurmann (2007) proposes a simple method for evaluating the limited-information likelihood under the assumption that the reduced form is a finite-order VAR without imposing determinacy; see section A.2.3 in the Appendix for details. He demonstrates in an empirical application that his method can give very different

results from the Fuhrer–Moore approach. This finding does not suffice to infer that the determinacy assumption is incorrect, as the estimates can differ due to sampling uncertainty.<sup>14</sup>

The other prominent approach to imposing the VAR assumption on the reduced-form dynamics is VAR-MD; see Sbordone (2005). As explained in section A.2.3 of the Appendix, the difference between VAR-ML and VAR-MD is that the former uses a restricted and the latter an unrestricted estimator for the reduced-form VAR parameters. Therefore, the relationship between VAR-ML and VAR-MD is analogous to the relationship between limited information maximum likelihood and two-stage least squares (2SLS), respectively, in the linear IV regression model (Fukač and Pagan 2010).

The assumption that the dynamics of the data can be represented as a finite-order VAR is restrictive. One well-known case when this assumption fails is when there is indeterminacy and sunspots, and the reduced form has moving average components. When the moving average roots are large (i.e., nearly noninvertible), a finite-order VAR may produce an inaccurate representation of the dynamics. Infinite-order VARs may also arise for other reasons, e.g., by omitting relevant variables from a finite-order VAR.<sup>15</sup> Therefore, the VAR approach to identification is less robust than GIV. To gain intuition about the restrictiveness of the VAR assumption, it is useful to think of the

<sup>13</sup> Kurmann (2007) shows that even if the equilibrium of the underlying structural model of the economy is determinate, a limited-information system of equations, where some of the structural equations have been replaced by their reduced form, may have an indeterminate solution. See section 3.2 of Kurmann (2007) for an example.

<sup>14</sup> Kurmann (2007) does not formally test the hypothesis that his approach fits the data significantly better than the Fuhrer–Moore approach. It is not trivial to develop a test for this hypothesis, especially when the structural parameters may be weakly identified. Note that if the model restrictions hold and the true parameters imply determinacy, the Fuhrer–Moore estimator (which imposes a correct restriction) will be more efficient than the Kurmann estimator under conventional asymptotics.

<sup>15</sup> See also Fernández-Villaverde et al. (2007). But note that omission of variables from a VAR does not necessarily cause misspecification (Fukač and Pagan 2007).

analogy to iterated versus direct forecasts, a point made in Magnusson and Mavroeidis (2010). The VAR-GMM moment conditions (15) differ from the GIV moment conditions  $E[Y_{t-1}(\pi_t - \beta\pi_{t+1} - \lambda x_t)] = 0$ , in that GIV uses direct projections of future inflation on predetermined variables, while VAR-GMM uses iterated multistep forecasts from a VAR.

### 3.2.3 Surveys

The two alternative survey estimators that we consider here are those that use one-step-ahead forecasts of inflation and those that use lagged two-steps-ahead forecasts. The former substitutes  $\pi_{t+1|t}^s$  for  $E_t(\pi_{t+1})$  in (8), yielding equations (17)–(18). The second possibility is to substitute  $\pi_{t+1|t-1}^s$  for  $E_t(\pi_{t+1})$  in (8) to get

$$(19) \quad \pi_t = \beta\pi_{t+1|t-1}^s + \lambda x_t + u_t + \beta\tilde{\varepsilon}_t,$$

$$(20) \quad \tilde{\varepsilon}_t = [E_t(\pi_{t+1}) - E_{t-1}(\pi_{t+1})] \\ + [E_{t-1}(\pi_{t+1}) - \pi_{t+1|t-1}^s].$$

The first component of  $\tilde{\varepsilon}_t$  is orthogonal to  $t-1$  information, and the second component has the same interpretation as the one-step survey error in (18). The difference from the previous case is that  $\pi_{t+1|t-1}^s$  is certainly predetermined, so it can be more plausibly treated as exogenous.

Some studies treat survey forecasts as exogenous for the estimation of the NKPC (Rudebusch 2002; Adam and Padula 2011). This can be justified under very specific assumptions about the timing of expectations and the nature of the disturbance term in the model. For example, if there is no cost-push shock, i.e.,  $u_t = 0$  in (17), and the disturbance term  $\varepsilon_t$  defined in (18) is a pure news shock, then  $\pi_{t+1|t}^s$  in equation (17) will be exogenous. This will not be true if  $\varepsilon_t$  is a classical measurement error. When  $u_t \neq 0$ , exogeneity of  $\pi_{t+1|t}^s$  requires that it should be predetermined, i.e., measured *before*  $\pi_t$ ,

but survey data are actually collected within the quarter. Equation (19) overcomes this problem, because  $\pi_{t+1|t-1}^s$  is certainly predetermined, but exogeneity still requires that  $\pi_{t+1|t-1}^s$  must be uncorrelated with  $\tilde{\varepsilon}_t$  in (20). This will hold if expectations are rational and  $[E_{t-1}(\pi_{t+1}) - \pi_{t+1|t-1}^s]$  is a news shock. Of course, even if the survey forecast is exogenous, the forcing variable  $x_t$  may still be endogenous.

If survey forecasts or the forcing variable are endogenous, then we need to find instruments. If measurement errors are unsystematic, in the sense that they are unpredictable from information at time  $t-1$ , and survey forecasts are unbiased, i.e., rational based on their information set, then  $E_{t-1}(\varepsilon_t) = 0$  in (18). Therefore, moment conditions for (17) are the same as for the GIV approach (13) where  $\pi_{t+1}$  has been replaced with  $\pi_{t+1|t}^s$  and the instruments are predetermined variables, including perhaps lags of  $\pi_{t+1|t}^s$ . Our view is that it is more robust to treat survey data as endogenous and use instruments for them. Nevertheless, we study the empirical implications of treating survey forecasts as exogenous in section 5.

### 3.3 Weak Identification

*Identification* of the structural parameter vector  $\vartheta$  requires that the GMM moment conditions are satisfied only at the true value  $\vartheta_0$ . Identification is clearly a necessary condition for obtaining useful estimators of  $\vartheta$ , but it is not sufficient. *Weak identification* arises when  $\vartheta$  is almost unidentified, i.e., when the moment conditions are close to being satisfied for all parameters  $\vartheta$  in a nonvanishing neighborhood of the true value  $\vartheta_0$ . In instrumental variable settings, weak identification arises when the instruments are only weakly correlated with the endogenous regressors. When identification is weak, conventional strong instrument asymptotic theory provides a poor approximation to the sampling distribution of GMM estimators and tests,

even in large samples. Instead, estimators can have a very non-Gaussian distribution and be severely biased toward their ordinary least squares (OLS) or nonlinear least squares counterparts, while conventional confidence sets may drastically undershoot their advertised coverage rates. Kleibergen and Mavroeidis (2009) discuss these issues at length in the context of estimation of the NKPC.

As we pointed out in the introduction, identification of the NKPC is likely to be weak because of the familiar empirical finding that changes in inflation are hard to forecast (Atkeson and Ohanian 2001; Stock and Watson 2007), implying that potential instruments that are plausibly exogenous (i.e., lagged) must necessarily be close to irrelevant. Indeed, we demonstrate empirically in section 5 that weak identification is pervasive in U.S. data, confirming a common finding in the literature. Furthermore, there are good theoretical reasons to expect identification to be weak; see Mavroeidis (2005) and Nason and Smith (2008a). For example, it is straightforward to see that, when the NKPC is flat, i.e.,  $\lambda = 0$  in (8), inflation is driven only by cost-push shocks. If these shocks are unpredictable, then so is inflation, and the coefficient on inflation expectations is unidentified because no relevant predetermined instruments exist. Therefore, the model predicts that identification will become arbitrarily weak as the slope of the NKPC  $\lambda$  gets closer to zero. Another situation in which identification is weak is when monetary policy is very effective in anchoring short-term inflation expectations. If inflation expectations do not vary, their effect on inflation is again unidentified. In other words, effective economic policy is bad for econometric analysis (Mavroeidis 2010; Cochrane 2011).

When identification is weak, we show below that GIV and VAR-based estimators of the NKPC can be biased in different directions. This helps explain some of the

systematic differences in empirical results that we report in section 5. There, we also find that estimates are extremely sensitive to specification choices, which is consistent with the moment conditions being insensitive to the value of the parameter  $\vartheta$  around the true value  $\vartheta_0$ . Recent advances in econometrics have made it possible to do inference that is fully robust to weak identification. Because these weak identification robust methods have been derived using alternative asymptotic approximations that do not assume strong instruments, they are reliable irrespective of the strength of the instruments. Surveys on the consequences of weak identification and methods of inference that are robust to it include Stock, Wright, and Yogo (2002), Dufour (2003), Andrews and Stock (2005), and Mikusheva (2013).

### 3.3.1 Simulation Studies

Mavroeidis (2005) illustrates the points mentioned above in the context of GIV estimation of the NKPC, and Kleibergen and Mavroeidis (2009) conduct an extensive set of simulations demonstrating the performance of several alternative GMM methods. The conclusion of these simulation exercises is that the theoretical consequences of weak identification are clearly borne out in empirically realistic tests of the NKPC.

The finite-sample performance of VAR-based estimation methods under weak identification has not received much attention in the literature. Therefore, we provide simulation results comparing four procedures: GIV estimation (Galí and Gertler 1999), VAR-MD (Sbordone 2005), VAR-ML (Kurmann 2007), and VAR-GMM (introduced above). For simulation purposes, we write the NKPC as

$$(21) \quad \pi_t = c + \gamma_f E_t(\pi_{t+1}) + (1 - \gamma_f)\pi_{t-1} + \lambda x_t + u_t,$$

TABLE 1  
DGPs FOR MONTE CARLO SIMULATIONS

DGP	NKPC		Red. form: forcing var.				Red. form: inflation				Conc.
	$\gamma_f$	$\lambda$	$\xi_{\pi 1}$	$\xi_{x 1}$	$\xi_{\pi 2}$	$\xi_{x 2}$	$\zeta_{\pi 1}$	$\zeta_{x 1}$	$\zeta_{\pi 2}$	$\zeta_{x 2}$	
<i>1a</i>	0.7	0.03	0.20	0.80	-0.10	0.10	1.00	-0.07	0.01	-0.01	1.8
<i>1b</i>	0.7	0.03	0.10	0.70	0.10	-0.10	0.79	0.25	0.05	-0.05	108.4
<i>2a</i>	0.3	-0.03	0.20	0.80	-0.10	0.10	0.98	-0.06	0.01	-0.01	0.7
<i>2b</i>	0.3	-0.03	0.70	1.05	0.08	-0.08	0.91	-0.07	-0.01	0.01	30.1

Notes: List of our DGPs. Columns 2 and 3 list the true values of  $(\gamma_f, \lambda)$ . The following eight columns list the VAR(2) reduced form coefficients in the equations for the forcing variable ( $\xi$ ) and in inflation ( $\zeta$ ). The last column lists the minimum eigenvalue of the population concentration matrix ("Conc."); cf. section A.3 in the Appendix. Numbers in columns 8 through 11 have been rounded off to two decimal points, while numbers in the last column have been rounded off to one decimal point.

with accompanying VAR(2) reduced-form dynamics

$$(22) \quad \pi_t = \zeta' Y_{t-1} + v_{\pi t}, \quad x_t = \xi' Y_{t-1} + v_{x t}.$$

Here,  $Y_t = (\pi_t, x_t, \pi_{t-1}, x_{t-1}, 1)'$ , and the reduced-form coefficients are  $\zeta = (\zeta_{\pi 1}, \zeta_{x 1}, \zeta_{\pi 2}, \zeta_{x 2}, c_{\pi})'$  and  $\xi = (\xi_{\pi 1}, \xi_{x 1}, \xi_{\pi 2}, \xi_{x 2}, c_x)'$ . We consider four data generating processes (DGPs) here, as summarized in table 1. For each DGP, we simulate samples of 200 observations each and calculate point estimates for each of the four estimators.<sup>16</sup> We execute 10,000 Monte Carlo repetitions per DGP. The GIV estimator that we consider estimates  $(\gamma_f, \lambda, c)$  by linear GMM, with  $\Delta\pi_t$  as the dependent variable,  $(\pi_{t+1} - \pi_{t-1})$  and  $x_t$  (and a constant) as regressors, and  $Y_{t-1}$  as instruments.<sup>17</sup> The three VAR-based methods exploit the VAR(2) reduced form for  $(\pi_t, x_t)$  to estimate the same three

parameters, as explained in section 3.1.<sup>18</sup> Details about the specifications and estimation procedures, as well as a measure of the strength of identification, are given in section A.3 of the Appendix.

DGPs *1a* and *1b* have  $\gamma_f = 0.7$ ,  $\lambda = 0.03$ , and  $c = 0$  as true NKPC parameters. Such parameter values represent typical estimates from the literature; cf. section 4. In *DGP 1a*, our choice of reduced-form parameters  $\xi$  for the forcing variable are based on OLS estimates on quarterly U.S. data with the labor share as  $x_t$ . The implied reduced-form parameters  $\zeta$  for inflation feature very limited second-lag dynamics relative to the variances of  $\pi_t$  and  $x_t$ , so inflation is hard to predict with lagged variables and identification is weak.<sup>19</sup> *DGP 1b* has  $\xi$  set to empirically unrealistic values that yield much better predictability of inflation and, thus, much stronger identification. The top panels in figure 2 display

<sup>16</sup>Because the specifications are overidentified, the VAR estimators do not coincide.

<sup>17</sup>Observe that the NKPC (21) may equivalently be written  $\Delta\pi_t = c + \gamma_f E_t(\pi_{t+1} - \pi_{t-1}) + \lambda x_t + u_t$ .

<sup>18</sup>Our VAR-MD implementation differs from that in Sbordone (2005) in the choice of distance function and weight matrix. We focus on Kurmann's (2007) version of VAR-ML for the reasons stated in section 3.2.

<sup>19</sup>Mavroeidis (2005) and Nason and Smith (2008a) discuss the role of lag dynamics in identification of the NKPC.

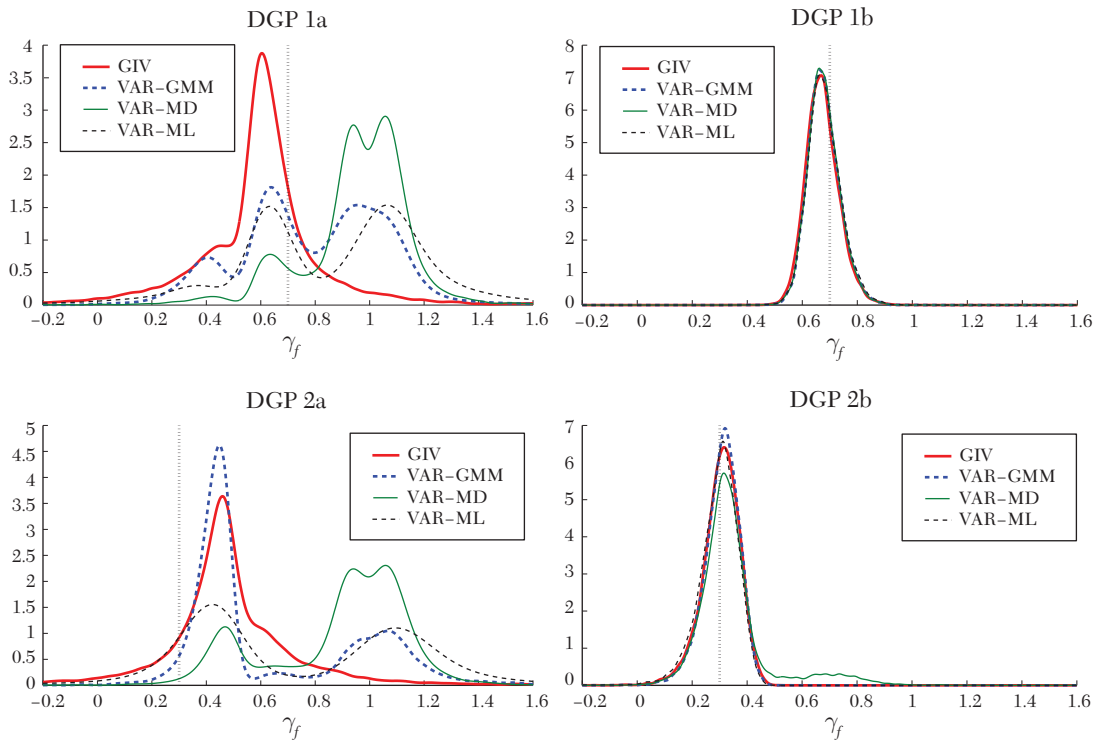


Figure 2. Sampling Distribution of  $\gamma_f$  Estimators

Notes: Kernel-smoothed density estimates of the sampling distribution of  $\gamma_f$  estimators in the hybrid NKPC model (21) for the DGPs listed in table 1. The dotted vertical line marks the true parameter value.

the densities of the sampling distributions of the  $\gamma_f$  estimators under *DGPs 1a* and *1b*. Evidently, the four estimators exhibit quite different behaviors in the weakly identified parametrization, *DGP 1a*. The GIV estimates of  $\gamma_f$  are biased downwards toward the probability limit of the OLS estimator, which is close to 0.5 for all DGPs in this paper.<sup>20</sup> While the sampling distribution density for GIV has rather fat tails, it is single-peaked and not

far from bell-shaped.<sup>21</sup> In contrast, the three VAR-based estimators all exhibit a distinct bimodal behavior, with a large (or even dominant) share of estimates concentrating around  $\gamma_f = 1$ . VAR-MD is particularly problematic in this regard. Due to the biased and decidedly non-Gaussian finite-sample distributions of the VAR methods, conventional strong-instrument inference procedures will give spurious results. In the strongly identified parametrization, *DGP 1b*, the situation is entirely different. Here the sampling densities

<sup>20</sup>By “OLS” we mean a simple regression of  $\Delta\pi_t$  on  $(\pi_{t+1} - \pi_{t-1})$  and  $x_t$ .

<sup>21</sup>Simulations with smaller true values of  $\lambda$  yield a stronger bias of  $\gamma_f$  estimates toward 0.5.



of all four estimators are of the conventional Gaussian shape, and only a slight downward finite-sample bias remains. The VAR estimates do not cluster around  $\gamma_f = 1$ . As the strong-instrument efficiency comparison in section 3.2 predicts, the sampling densities for the three VAR methods are ever so slightly more narrowly concentrated around the true value  $\gamma_f = 0.7$  than the GIV density.

*DGPs 2a* and *2b* set  $\gamma_f = 0.3$ , a value at the lower end of estimates reported in the literature, and  $\lambda = -0.03$ , but are otherwise analogs of *DGPs 1a* and *1b*, respectively. It is striking that the reduced-form parameters for *DGPs 1a* and *2a* are so similar, even though the structural NKPC parameters are completely different; cf. table 1. The sensitivity of the mapping between reduced-form and structural parameters is a key symptom of weak identification. Because structural estimation works by backing out structural parameters from estimates of reduced-form features of the data, it is clear that weak identification will have serious consequences regardless of the estimation method, which is indeed what we find in our simulations. The bottom panels in figure 2 display the sampling distributions of the  $\gamma_f$  estimators under *DGPs 2a* and *2b*. The results are similar to those for *DGPs 1a* and *1b* (the slight bimodality of the VAR-MD estimator under *DGP 2b* disappears if the strength of identification is increased further). In particular, the spurious clustering of VAR estimates around  $\gamma_f = 1$  under weak identification remains even with the true  $\gamma_f$  value set equal to 0.3. This is interesting, as Sbordone (2005) and Kurmann (2007) both report rather large VAR-based estimates of  $\gamma_f$ , and our empirical VAR-GMM estimates in section 5 similarly concentrate around 1.

Section A.3 in the Appendix refers to an online supplement that provides additional simulation results and Matlab code. Among other things, we find—not surprisingly—that misspecification of the VAR can result in

coefficient biases in any direction, irrespective of the strength of identification.

### 3.3.2 VAR Methods

To understand the behavior of VAR methods under weak identification, we set  $c = \lambda = 0$  for ease of exposition, but our results apply more generally. Equation (21) can then be written as

$$(23) \quad \Delta\pi_t = \delta E_t(\Delta\pi_{t+1}) + \bar{u}_t,$$

where  $\delta = \gamma_f/(1 - \gamma_f)$  and  $\bar{u}_t = u_t/(1 - \gamma_f)$ . Suppose the econometrician bases his analysis on a reduced-form AR(1) in  $\Delta\pi_t$ .<sup>22</sup> Let  $\rho_1 = E[\Delta\pi_t \Delta\pi_{t-1}]/E[(\Delta\pi_t)^2]$  denote the first autocorrelation of  $\Delta\pi_t$ , and let  $\hat{\rho}_1$  be its sample analog. Since the specification is just identified (we have one reduced-form parameter to identify  $\gamma_f$ ), all VAR estimators coincide, and it is easy to show that the VAR estimator of  $\delta$  is  $1/\hat{\rho}_1$ . The implied  $\gamma_f$  estimator is then

$$(24) \quad \hat{\gamma}_f^{\text{VAR}} = \frac{1}{1 + \hat{\rho}_1}.$$

Because the primary cause of weak identification is precisely that  $\rho_1 \approx 0$  (inflation changes are nearly unforecastable), we expect to find  $\hat{\gamma}_f^{\text{VAR}} \approx 1$ . This must be the case for any true value of  $\gamma_f$  that leads to  $\rho_1 \approx 0$ , i.e., for any empirically realistic DGP. This is different from the weak instrument behavior of the GIV estimator  $\hat{\gamma}_f^{\text{GIV}}$ . As mentioned earlier, this estimator is biased toward the probability limit of the OLS estimator of  $\gamma_f$  in the regression of  $\Delta\pi_t$  on  $(\pi_{t+1} - \pi_{t-1})$ , which equals  $1/2$ .<sup>23</sup> We are thus able to explain both biases observed in figure 2.

<sup>22</sup>Our weakly identified *DGPs 1a* and *2a* nearly have this reduced form, as their reduced-form coefficient on  $\pi_{t-1}$  in the inflation equation is approximately 1.

<sup>23</sup>Under stationarity, the OLS probability limit is  $E[(\pi_{t+1} - \pi_{t-1})\Delta\pi_t]/E[(\pi_{t+1} - \pi_{t-1})^2] = \{E[(\Delta\pi_t)^2] + E[\Delta\pi_{t+1} \Delta\pi_t]\}/\{2E[(\Delta\pi_t)^2] + 2E[\Delta\pi_{t+1} \Delta\pi_t]\} = 1/2$ .

Another way to view the weak identification VAR bias is as follows. The VAR estimator of  $\delta$  is precisely the inverse of the OLS estimator in the regression suggested by equation (23), i.e., the regression of  $\Delta\pi_{t+1}$  on  $\Delta\pi_t$ . There are only two ways in which the AR(1) assumption can hold. The first possibility is that  $\bar{u}_t \equiv 0$ , i.e., the NKPC is exact. The second possibility is that  $\bar{u}_t \neq 0$ , but the reduced form for inflation changes is  $\Delta\pi_t = \bar{u}_t$ .<sup>24</sup> This is trivially an AR(1) with coefficient 0. Note, however, that if the DGP for inflation really has this form,  $\gamma_f$  is unidentified. This is the sense in which VAR methods can lead to spurious identification: If the solution is  $\Delta\pi_t = \bar{u}_t$ , VAR methods—by being equivalent to OLS estimation of (23)—implicitly select the other possibility  $\bar{u}_t \equiv 0$  and obtain a seemingly very precise estimate of  $\gamma_f$  close to 1, even though the parameter is unidentified.

Our discussion has focused on the tractable limits of  $\lambda = 0$  and exact identification. While we believe that the intuition translates to empirically realistic settings with  $\lambda \approx 0$  and overidentified specifications, there is clearly room for more research on these matters. We stress that it is not obvious from our results or previous analyses in the literature that GIV methods perform either better or worse under weak identification than VAR-based methods. Only recently has the econometric literature begun to analyze the consequences of weak identification for estimators other than GMM. Magnusson and Mavroeidis (2010) introduce a robust MD test and apply it to the NKPC. Robust VAR-ML analysis of the NKPC has not been attempted yet in the literature, but weak identification robust procedures for general

ML analysis are currently under development; cf. Andrews and Mikusheva (2013) and references therein.

### 3.4 Other Issues

A number of other econometric issues have been raised in the literature.

#### 3.4.1 Number of Instruments

Early implementations of the GIV approach to identification of the NKPC used a large number of instruments relative to the sample size; Galí and Gertler (1999) used four lags of six variables on a sample of 160 observations. This practice is subject to the pitfall of “many instrument” biases, and most subsequent studies have used a significantly smaller number of instruments. There is a large econometric literature on this issue; see, e.g., Hansen, Hausman, and Newey (2008). It is well known that use of many instruments biases 2SLS estimators towards OLS. Intuitively, in the limit case where the number of instruments is the same as the sample size, the first-stage yields perfect fit, so 2SLS is identical to OLS. The problem becomes more severe if the instruments are many and weak, which is the relevant framework for the NKPC. A recent contribution by Newey and Windmeijer (2009) demonstrates some robustness properties of the weak identification robust methods to many weak instruments. However, Newey and Windmeijer only cover situations in which the instruments are sufficiently informative for the model to be strongly identified. Therefore, their results exclude cases in which the instruments may be *arbitrarily* weak (e.g., completely irrelevant).<sup>25</sup> They

<sup>24</sup> Such a reduced form can arise in two ways. If  $\gamma_f < 1/2$ , the solution is determinate and equation (23) can be iterated forward to yield  $\Delta\pi_t = \bar{u}_t$ . Alternatively, if  $\gamma_f \geq 1/2$ , the solution is indeterminate, and the so-called minimum state variable solution  $\Delta\pi_t = \bar{u}_t$  satisfies the NKPC.

<sup>25</sup> Andrews and Stock (2007) provide conditions under which the weak identification robust tests remain valid under many instruments that may be arbitrarily weak, but these results are for the homoskedastic linear IV regression model, so they do not apply to the NKPC.

also ignore the complications arising from uncertainty in the estimation of the long-run variance of the moment conditions, which can be substantial when the number of moment conditions is large. Therefore, we recommend against the use of many instruments in the estimation of the NKPC.

### 3.4.2 *Unit Roots*

Empirically, U.S. inflation appears close to nonstationary in certain subsamples. Fanelli (2008), Mikusheva (2009), Boug, Cappelen, and Swensen (2010), and Nymoén, Swensen, and Tveter (2012) discuss the implications of inflation having a unit root. This raises issues about the validity of inference when the unit root is left unaccounted for. When the inflation coefficients in the NKPC add up to 1, the model can be written in terms of changes in inflation. If the instrument set also uses lags of  $\Delta\pi_t$ , inference will be robust to inflation having a unit root. A number of papers have taken that approach (e.g., Kleibergen and Mavroeidis 2009) and we study it further in our empirical section.<sup>26</sup>

### 3.4.3 *Misspecification*

The various methods described so far may be affected differently by misspecification of the NKPC. Jondeau and Le Bihan (2003, 2008) study some special cases of misspecification in the form of omitted lags of inflation in the NKPC that bias the GIV estimator of the coefficient on future inflation upwards and the VAR-ML estimator downwards. This is consistent with the difference among empirical estimates reported by Fuhrer (1997), Galí and Gertler (1999), and Jondeau and Le Bihan (2005), so Jondeau and Le Bihan argue that misspecification could

be the source of disparity in the estimates. However, differences remain when the NKPC is extended to include more lags of inflation.<sup>27</sup> Mavroeidis (2005) explores the implications of omitted dynamics for the bias of GIV estimation of the hybrid NKPC.

Cagliarini, Robinson, and Tran (2011) and Imbs, Jondeau, and Pelgrin (2011) discuss another type of misspecification due to aggregation bias. Building on Carvalho (2006), they show that heterogeneity of (Calvo) price rigidity across economic sectors can bias estimates of average price rigidity upwards and bias estimates of the slope of the aggregate NKPC toward zero. Cagliarini, Robinson, and Tran (2011) trace this bias to the presence of an additional error term in the aggregate NKPC resulting from sectoral heterogeneity.

### 3.4.4 *Autocorrelated Cost-Push Shocks*

Autocorrelation of  $u_t$  in the NKPC (8) violates the common identifying assumption  $E_{t-1}(u_t) = 0$ . Because cost-push shocks induce endogenous movements in observables, in general, autocorrelated cost-push shocks imply that lagged variables will be correlated with  $u_t$ , so that, with the exception of external instruments, all other identifying assumptions listed earlier become invalid. Zhang and Clovis (2010) reiterate this point and perform autocorrelation tests on the residual of the Galí and Gertler (1999) specification. They find evidence of significant residual autocorrelation, which can be removed by including three lags of inflation in the NKPC. Note that the GIV residuals  $\tilde{u}_t$  in (11) include a future inflation forecast error, so they may exhibit moving average (MA) of order 1 autocorrelation even when the structural error  $u_t$  is not autocorrelated (Jondeau

<sup>26</sup> Note that the restriction that inflation coefficients sum to 1 is not sufficient for inflation to have a unit root. This depends on the dynamics of the forcing variable (Bårdsen, Jansen, and Nymoén 2004).

<sup>27</sup> As discussed in section 3.2, Kurmann (2007) offers another possible explanation for why VAR-ML estimates that impose determinacy, such as those reported in Fuhrer (1997) and Jondeau and Le Bihan (2005), may differ from GMM estimates.

and Le Bihan 2005; Mavroeidis 2005; and Eichenbaum and Fisher 2007). Boug, Cappelen, and Swensen (2010) identify the cost-push shock  $u_t$  via VAR-ML and find it to be serially correlated. They also recommend using more lags of inflation in the NKPC. Kuester, Müller, and Stölting (2009) show by simulation that GIV estimates of the slope of the NKPC are biased downwards when  $u_t$  is autocorrelated, and the Hansen (1982)  $J$  test has little power against this misspecification in realistic sample sizes.

### 3.4.5 Subsample Stability

Stability tests of the model parameters can be used to test the immunity of the NKPC to the Lucas (1976) critique, as well as to assess the importance of time varying trend inflation and lack of full indexation to it. The standard stability tests of Andrews (1993), Andrews and Ploberger (1994), and Sowell (1996) require strong instruments, but weak identification robust versions are available (Caner 2007; Magnusson and Mavroeidis 2013). Castle et al. (2010) give an extensive discussion of the consequences of structural breaks in the NKPC.

## 4. Survey of the Empirical Literature

This section surveys the empirical literature on the NKPC. Rather than maintaining a strict chronological order, we have attempted to group the various contributions into the main econometric approaches that were introduced in section 3. Figure 3 and table 2 present a representative set of results from some of the most frequently cited studies; additional papers are referenced below. The major points of controversy in the literature concern the relative importance of forward- and backward-looking price setting behavior, as well as the degree to which real activity influences inflation dynamics. Although several methodological contributions have been proposed since the beginning of the research

program, an empirical consensus is not yet in sight.

### 4.1 Initial Breakthroughs

Limited-information testing of the NKPC was initiated by Fuhrer and Moore (1995) and Roberts (1995). As mentioned by Roberts, previous econometric tests of New Keynesian pricing equations had been based on full-information (system) methods under the assumption of rational expectations. Roberts (1995) shows that three different theoretical frameworks—the staggered contracts model of Taylor (1980), the infinite-horizon staggered pricing model of Calvo (1983), and the quadratic adjustment cost model of Rotemberg (1982)—all lead to (what came to be known as) a difference equation specification of the pure NKPC, with the output gap as forcing variable. He suggests two different limited-information approaches to testing the relationship: first, the use of survey expectations as proxies for the expectation term, and second, McCallum's (1976) technique of subsuming the rational expectations forecast error into the equation's error term and instrumenting for next period's inflation and the output gap with lagged variables (GIV estimation, in the terminology of section 3.1). Using annual U.S. data, Roberts finds a significant role for the output gap.

Seminal contributions by Galí and Gertler (1999) and Sbordone (2002) helped propel the NKPC research agenda into the forefront of empirical macroeconomics. Both papers take the now-standard microfounded rational expectations pricing model to U.S. data and obtain results that are supportive of the model's fit. Furthermore, both sets of authors exploit the model's implication that aggregate marginal cost may be proxied by the labor share. Indeed, Galí and Gertler establish that the NKPC only fits U.S. data if the labor share is used as forcing variable instead of the output gap, which may be mismeasured. Galí

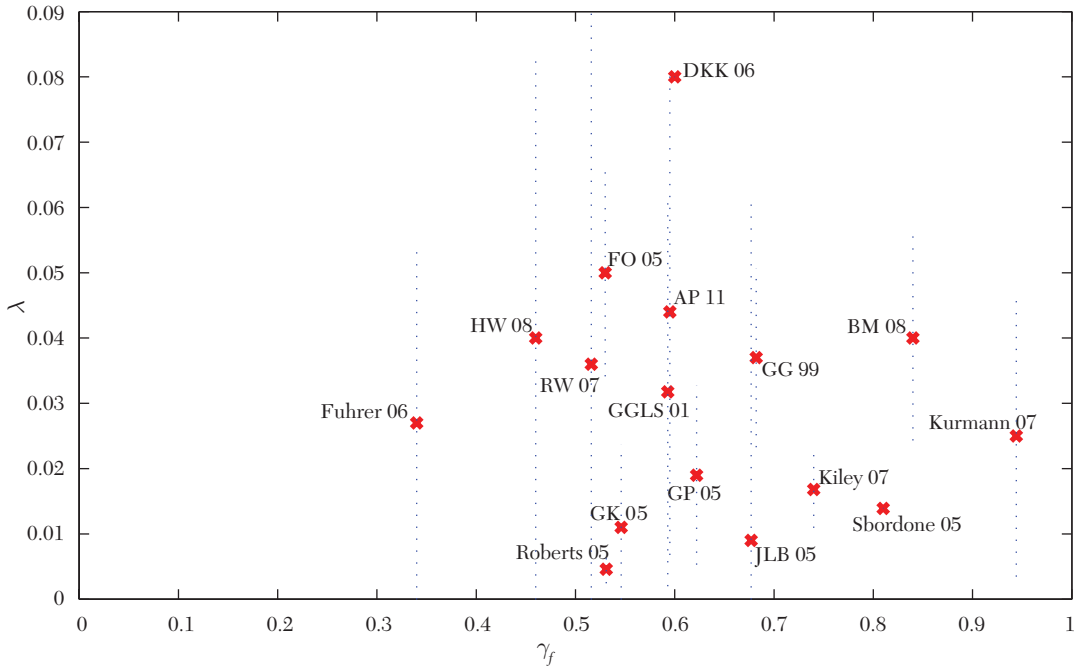


Figure 3. Point Estimates Reported in the Literature

Notes: Point estimates of  $\lambda$  (vertical axis) and  $\gamma_f$  (horizontal axis) reported in the literature. Only estimates that use U.S. data and the labor share as forcing variable are plotted. For some papers the semistructural point estimates have been imputed from point estimates of deeper parameters. The dotted blue lines indicate 95 percent confidence intervals for  $\lambda$  where available. We include papers with readily available estimates and more than twenty-five Google Scholar citations as of mid-September 2012: Galí and Gertler (1999); Galí, Gertler, and López-Salido (2001); Fuhrer and Olivei (2005); Gagnon and Khan (2005); Guay and Pelgrin (2005); Henzel and Wollmershäuser (2008); Jondeau and Le Bihan (2005); Roberts (2005); Sbordone (2005); Dufour, Khalaf, and Kichian (2006); Fuhrer (2006); Kiley (2007); Kurmann (2007); Rudd and Whelan (2007); Brissimis and Magginas (2008); and Adam and Padula (2011).

and Gertler also develop the now-standard *hybrid* NKPC, whose lagged inflation terms introduce intrinsic persistence of the inflation rate on top of the extrinsic persistence imparted by the forcing variable.

#### 4.2 GIV Estimation

Using linear and nonlinear GIV methods, Galí and Gertler (1999) find that, while the backward-looking inflation term is significant, the forward-looking rational expectations term dominates; they also obtain a significant and correctly signed

coefficient on the labor share (unlike the output gap). The NKPC restrictions are not rejected by overidentification tests or by visual inspection of fitted inflation. Galí, Gertler, and López-Salido (2001) take the model to aggregate Eurozone data, largely confirming the U.S. findings. Benigno and López-Salido (2006) find some heterogeneity in estimated coefficients for major Eurozone countries. Eichenbaum and Fisher (2004, 2007) evaluate a variant of the NKPC with price indexation that was developed by Christiano, Eichenbaum, and Evans (2005),



TABLE 2  
OVERVIEW OF ESTIMATION APPROACHES IN THE LITERATURE

Papers	Estimation approach	Expectation versus lags	Slope	Rejection of model?
Galí and Gertler (1999), Galí et al. (2001, 2005)	RE GIV.	Forward-looking behavior dominant, but backward-looking term significant.	Significantly positive for labor share.	No, based on over-ID test and visual fit.
Fuhrer and Moore (1995), Fuhrer (1997, 2006)	RE VAR-ML (AIM).	Price setting not very forward looking; need large intrinsic persistence.	Positive for both labor share and output gap, but significance varies.	Pure NKPC rejected based on LR test and IRFs.
Roberts (1995, 1997, 2005)	GIV, VAR-ML, IRF matching; RE and survey forecasts.	Sluggish survey forecasts impart necessary persistence. For RE, need more than one lag of inflation.	Positive for both labor share and output gap, but significance varies.	No.
Sbordone (2002, 2005)	RE VAR-MD.	Forward-looking behavior clearly dominant, but lag is significant.	Positive but marginally insignificant in hybrid model.	No, based on over-ID test and visual fit.
Rudd and Whelan (2005, 2006, 2007)	RE GIV (iterated).	Lagged inflation very significant.	Neither labor share nor output gap adds explanatory power.	Yes, forcing variable doesn't help explain inflation.
Rudebusch (2002)	OLS; survey forecasts.	Four-quarter MA of lagged inflation receives larger weight than forecast.	Output gap coefficient positive and significant.	No.
Ravenna and Walsh (2006)	RE GIV, interest rate added to NKPC.	(Pure NKPC.)	(Not directly estimated.)	No, based on over-ID test.
Cogley and Sbordone (2008)	Bayesian estimation using VAR with drifting parameters and stochastic volatility.	Backward-looking term insignificant once trend inflation is accounted for.	(Not directly estimated.)	No, based on visual fit and magnitude of forecast errors.

*Notes:* Overview of well-cited papers in the limited-information empirical NKPC literature. Papers have been grouped according to authorship. Each series of papers has been ranked by the number of Google Scholar citations (as of mid-September 2012) for the single most-cited paper within the series.

and they also introduce variable elasticity of demand, capital adjustment costs, and pricing implementation lags. Blanchard and Galí (2007) consider a model with real wage rigidity, which leads to an NKPC featuring the unemployment rate as forcing variable;

GIV estimation on U.S. data yields intuitively reasonable coefficients with significant forward-looking behavior. Krause, López-Salido, and Lubik (2008), Ravenna and Walsh (2008), and Blanchard and Galí (2010) explore NKPCs with explicit labor market

frictions that lead to alternative expressions for marginal cost. Chowdhury, Hoffmann, and Schabert (2006) and Ravenna and Walsh (2006) assume that firms must borrow to pay their wage bill up front each period, which leads to the so-called cost channel of interest rates, i.e., marginal cost is directly influenced by the interest rate. These papers find that, for most countries, the Treasury bill rate enters significantly into an extended NKPC when estimated by GIV, but the coefficients on the forward- and backward-looking inflation terms are not affected much relative to the baseline. Neiss and Nelson (2005) compute the output gap that is implied by a standard New Keynesian model; this theoretically consistent measure turns out to be essentially uncorrelated with quadratically detrended output (which is used by Roberts 1995, and Galí and Gertler 1999), and GIV estimates of the slope of the NKPC are even more significant than when using the labor share.<sup>28</sup> Gagnon and Khan (2005) extend the NKPC to a more general constant elasticity of substitution (CES) production function and find that structural GIV estimates imply less price stickiness than under the usual Cobb–Douglas specification. In addition to a CES production function, McAdam and Willman (2012) add varying capacity utilization, which decreases the estimated coefficient on the inflation expectation term. Batini, Jackson, and Nickell (2005) and Rumler (2007) estimate open-economy NKPCs by GIV on data from European countries, finding a significant role for international variables. Gwin and VanHoose (2008) and Shapiro (2008) construct alternative measures of firm marginal costs from microlevel and sectoral data.

While the previously mentioned papers find a significant, and often dominant, role for the forward-looking rational expectations

term, a number of papers that use the GIV framework have raised issues with the mainstream analysis (see also the discussion of weak identification below). Bårdsen, Jansen, and Nymoen (2004) point out that the literature has mostly not rejected the homogeneity restriction (i.e., that the coefficients on last and next period's inflation sum to 1), which, under strict exogeneity of the forcing variable, implies that inflation is non-stationary. They show empirically that GIV estimates are quite sensitive to the choice of instrument set and estimator (see also Guay and Pelgrin 2005), and the Galí, Gertler, and López-Salido (2001) hybrid NKPC is rejected in favor of alternative, encompassing models of inflation. Fuhrer and Olivei (2005) employ a reduced-form VAR to compute expectations of next period's inflation and the output gap; they then use the computed expectations as instruments. Their estimate of the coefficient on the forward-looking expectation term is much smaller than the traditional GIV estimate. A series of papers by Rudd and Whelan (2005, 2006, 2007) contend that the Galí and Gertler (1999) estimation approach yields spurious results. Rudd and Whelan criticize the use of the labor share as a proxy for marginal cost due to its countercyclicality.<sup>29</sup> They demonstrate that, provided the NKPC leaves out explanatory variables, the use of instruments outside of the model (such as interest rates or wage and commodity price inflation) may bias the estimates in the direction of establishing a high degree of forward-looking behavior. Furthermore, Rudd and Whelan conduct several tests of the incremental explanatory power of the labor share and conclude that it adds essentially no information to inflation forecasting. Estimating the model by GIV in iterated form (cf. section A.2.2 in the Appendix) yields a smaller coefficient

<sup>28</sup>See also Fukač and Pagan (2010) for a critique of “off-model” output detrending.

<sup>29</sup>This issue is further pursued by Mazumder (2010).

on the forward-looking term. Finally, data revisions since 1999 have eroded the significance of the labor share, even in the original Galí and Gertler setup. Galí, Gertler, and López-Salido (2005) counter that Rudd and Whelan (2005) use a parametrization of the model that does not correspond to the structural parameters in Galí and Gertler (1999) and Galí, Gertler, and López-Salido (2001). Galí, Gertler, and López-Salido (2005) show that if the same parametrization is used, iterated GIV estimation confirms the results in Galí and Gertler (1999) and Galí, Gertler, and López-Salido (2001).

As witnessed by the myriad of parallel sub-models and methods, the literature is still far from producing a consensus set of specifications or empirical conclusions, even within the relatively narrow rational expectations GIV framework.

#### 4.3 VAR Estimation

In their seminal paper, Fuhrer and Moore (1995) augment a Taylor (1980) pricing equation with reduced-form VAR equations for the output gap and Treasury bill rate and estimate the resulting system by ML, using the AIM routine from Anderson and Moore (1985) to solve for a rational expectations solution given the parameters. Fuhrer and Moore reject the restrictions implied by the standard pricing model based on a formal likelihood ratio test and inspection of the implied impulse responses, which display too little inflation persistence. The data is more favorable to an alternative real wage contracting model that implies sticky *inflation* instead of just sticky prices. Fuhrer (1997) uses a similar approach to test for the significance of forward-looking rational inflation expectations relative to backward-looking (adaptive) expectations; he finds that the rational expectations component is insignificant. Subsequent papers have applied the AIM-based VAR-ML method to the Galí and Gertler (1999) hybrid NKPC. Fuhrer and

Olivei (2005) and Fuhrer (2006) find a small coefficient on the forward-looking term relative to that on lagged inflation. Roberts (2005), who also considers GIV and impulse response matching, estimates a hybrid NKPC with four lags of inflation, obtaining about 50 percent weight on forward-looking behavior. Jondeau and Le Bihan (2005) estimate the NKPC on data from the United States and major European countries. They find that GIV estimates of the coefficient on forward-looking expectations tend to be high, while VAR-ML estimates tend to be lower. Kiley (2007) uses VAR-ML to estimate an NKPC specification with four lags of inflation and expectations of next-period inflation taken with respect to *previous*-period (rather than current-period) information. Here the forward-looking term is dominant, and the Bayesian Information Criterion indicates that the structural model provides as good a fit to U.S. data as a reduced-form VAR.

Kurmann (2007) criticizes the AIM-based approach to ML estimation, as it imposes the extraneous assumption that the rational expectations solution must be unique (determinate); cf. section 3.2. Using an ML method that does not impose uniqueness, he finds evidence of a large share of forward-looking behavior in the Galí and Gertler (1999) U.S. dataset, which contrasts with estimates obtained under the additional uniqueness assumption. It is an open question whether imposing the determinacy assumption matters empirically across other datasets and NKPC specifications. Korenok, Radchenko, and Swanson (2010) also eschew the AIM algorithm and instead write their model in a form that is amenable to Kalman filtering.<sup>30</sup>

Sbordone (2002) tests the pure NKPC on U.S. data using a two-step approach akin to

<sup>30</sup>Other papers that use filtering techniques include Lee and Nelson (2007), Kim and Kim (2008), and Kim, Manopimoke, and Nelson (forthcoming).

that of Campbell and Shiller (1987, 1988). In a first step, she fits a reduced-form VAR to the data. When iterated forward, the pure NKPC implies that inflation is given by an expected present value of future marginal costs, and this quantity may be evaluated using the fitted VAR. The structural parameters of the NKPC can then be estimated by minimizing the squared distance between model-implied and actual inflation. The estimated Calvo (1983) parameter is in line with microestimates of price stickiness. Sbordone (2005) refines the estimation approach by interpreting it as MD estimation and accounting for sampling uncertainty of the first-step estimated VAR (see also Kurmann 2005). She provides estimates of the hybrid NKPC, broadly confirming the conclusions in Galí and Gertler (1999). Tillmann (2008) uses a related MD approach to assess the importance of the cost channel of monetary policy. Sbordone (2006) develops a model of joint price and wage determination, which is estimated by MD. Coenen, Levin, and Christoffel (2007) construct a model with a general, nonconstant hazard rate of price resetting, which they estimate using an indirect inference procedure that matches the model-implied dynamics to the estimated reduced-form VAR. Carriero (2008) rejects the cross-equation restrictions that the one-lag NKPC places on a reduced-form VAR in inflation and the labor share. Guerrieri, Gust, and López-Salido (2010) develop a microfounded open-economy NKPC in which the relative price of foreign goods enters. To estimate it they use a multi-equation GMM approach that adds reduced-form VAR equations for the labor share and relative foreign goods prices. Their preferred specification yields an insignificant coefficient on lagged inflation. Cornea et al. (2013) use VAR methods to estimate an NKPC with evolutionary switching between forward-looking and backward-looking inflation expectations; they find substantial

time-variation and heterogeneity in the type of expectations formation.

Fanelli (2008) and Boug, Cappelen, and Swensen (2010) conduct likelihood-based estimation of the hybrid NKPC, taking into account the possibility that the variables are cointegrated. The likelihood is derived conditional on a reduced-form vector error-correction model for inflation and the output gap, using the Kurmann (2007) approach that does not impose uniqueness of the rational expectations solution. Both papers find that the NKPC restrictions are rejected for the Eurozone. Boug et al. (2010) find some support for the hybrid NKPC in U.S. data, although the residuals are significantly autocorrelated, violating an assumption of the model. The ML estimate of the coefficient on the inflation expectations term is much larger than that on the lagged inflation term.

While popular in the DSGE literature, Bayesian methods have only been used in a few limited-information analyses of the NKPC. Fuhrer and Olivei (2010) and Fukač and Pagan (2010) compute posteriors for the parameters in versions of the NKPC where the expectation of next period's inflation is determined by a reduced-form VAR. Cogley and Sbordone (2008) introduce drifting trend inflation into the standard New Keynesian model, which changes the form of the NKPC. They estimate the model using quasi-Bayesian methods, conditional on a reduced-form VAR with drifting parameters and stochastic volatility. Their imputed inflation gap (i.e., the difference between inflation and its trend) is much less persistent than raw inflation, and the quasiposteriors indicate that once the trend is accounted for, there is no need to allow for backward-looking price indexation (see also Sahuc 2006; Hornstein 2007). Barnes et al. (2011) and Gumbau-Brisa, Lie, and Olivei (2011) argue, however, that this conclusion is sensitive to how the NKPC restrictions

are imposed in the estimation. Despite the reservations, the trend inflation research agenda is rapidly becoming one of the most well-cited branches of the NKPC literature.

#### 4.4 Estimation Using Survey Expectations

As mentioned, Roberts (1995) uses survey measures of inflation expectations from the Michigan and Livingston surveys as an alternative to the rational expectations GIV approach.<sup>31</sup> Roberts (1997) finds that the apparent sluggishness and nonrationality of these survey forecasts generate sufficient inflation persistence in the U.S. NKPC, and the data favors such a specification to the Fuhrer and Moore (1995) sticky inflation model. Rudebusch (2002) estimates the hybrid NKPC on U.S. data by OLS, with Michigan survey data proxying for inflation expectations. He finds a relatively small coefficient on the expectations term, but a significantly positive coefficient on the output gap. Adam and Padula (2011) use SPF inflation forecasts and also find the forcing variable to be significant, regardless of whether they use the labor share or output gap, but their OLS estimate of the coefficient on the expectation term is slightly larger than that on lagged inflation. Kozicki and Tinsley (2002) estimate various pricing equations for the United States and Canada using survey forecasts and allowing for nonzero trend inflation. Gerberding (2001), Paloviita and Mayes (2005), Paloviita (2006, 2008), Henzel and Wollmershäuser (2008), and Koop and Onorante (2012) estimate NKPCs for European countries using various measures of survey expectations of inflation and various estimation procedures. The estimated extent of forward-looking pricing behavior varies greatly between studies and specifications. Brissimis and Magginas (2008) use SPF forecasts and the Federal Reserve's Greenbook

projections to estimate the U.S. NKPC by GMM. They find a dominant role for forward-looking expectations and a significantly positive coefficient on the labor share. Zhang, Osborn, and Kim (2008, 2009) consider SPF, Greenbook, and Michigan survey forecasts. Unlike the rational expectations specification, the survey forecast NKPC gets a positive and significant coefficient on the output gap, but its estimates appear more unstable over subsamples (see also Kim and Kim 2008). Mazumder (2011) uses SPF, Michigan, and Greenbook forecasts to test the NKPC with a procyclical measure of marginal costs developed in Mazumder (2010).

Nunes (2010) simultaneously includes rational expectations and SPF forecasts in an NKPC. The GMM estimates point to nonrational expectations only playing a minor role in explaining U.S. inflation dynamics. This conclusion is disputed by Fuhrer and Olivei (2010) and Fuhrer (2012), who proxy for the rational expectations term with expectations from a reduced-form VAR and estimate the NKPC by Bayesian and ML methods. Smith (2009) gives conditions under which it is advantageous to include data on survey forecasts for *statistical* reasons, even if the researcher has a purely rational NKPC in mind.

Survey forecast methods have established a commanding presence in the NKPC literature. So far, the literature has only scratched the surface in terms of providing full-fledged microfoundations, and a detailed understanding of the interplay between nonrational expectation formation and price setting remains elusive.

#### 4.5 Identification Issues and Robust Inference

The literature's awareness of the problems associated with weak identification has grown over time. Galí, Gertler, and López-Salido (2001) guide the choice of their instrument set by the first-stage  $F$  statistic. Mavroidis (2004, 2005) provides

<sup>31</sup> Roberts still instruments for the output gap as it may be correlated with the inflation error term.



analytical and simulation evidence that explains why weak identification is likely to be an issue for NKPC estimation. Ma (2002) is the first paper to compute weak identification robust confidence sets (specifically, the Stock and Wright (2000)  $S$  set) for the NKPC, finding the data to be completely uninformative about the structural parameters. Dufour, Khalaf, and Kichian (2006) compute Anderson and Rubin (1949) and Kleibergen (2002) confidence sets for both GIV and survey forecast specifications. The U.S. GIV confidence region is fairly large, while the survey forecast one is empty; no NKPC specification seems to fit Canadian data. Nason and Smith (2008a) reject the hybrid NKPC for Canada, the United Kingdom, and the United States using the Anderson and Rubin (1949) test and the Guggenberger and Smith (2008) generalized empirical likelihood (GEL) test. In contrast, Martins and Gabriel (2009) find very wide robust GEL confidence sets. Using a variety of GMM-based robust tests, Kleibergen and Mavroeidis (2009) conclude that inflation appears to be significantly forward-looking, but the confidence regions are wide. Dufour, Khalaf, and Kichian (2010a, 2010b) carry out robust inference on certain extensions of the NKPC with real wage rigidities and labor market frictions. Magnusson and Mavroeidis (2010) develop a weak identification robust version of Sbordone's (2005) MD test, finding somewhat smaller confidence regions than when using a robust GIV approach. Kleibergen and Mavroeidis (forthcoming) demonstrate the consequences of ignoring weak identification in Bayesian analyses of the NKPC and propose ways of circumventing the problems.

Some papers have devised methods for improving the strength of identification in GIV estimation of the NKPC. Dees et al. (2009) obtain instruments for individual-country NKPCs by estimating a multicountry cointegrating VAR. Building on Beyer et al. (2008), Kapetanios and

Marcellino (2010) and Kapetanios, Khalaf, and Marcellino (2011) develop identification robust theory for GMM testing using instruments that have been estimated by principal components from a large set of candidate variables. While this seems to improve identification of the slope of the NKPC, the relative shares of forward- and backward-looking behavior remain very weakly identified. Motivated by the Lucas critique, Magnusson and Mavroeidis (2013) suggest using robust parameter instability tests to improve inference about the NKPC.

The lessons from weak identification analyses have so far only had limited impact on the broader NKPC literature. Papers that do mention the identification issue often either treat it as merely another robustness check or incorrectly dismiss it as a strictly GMM-specific problem. The consequence is that comparison of results across papers is difficult.

## 5. Empirical Synthesis

In this section, we generate estimates of the NKPC corresponding to a wide selection of empirical approaches from the literature.<sup>32</sup> Because we use a common dataset for all estimates, we are able to highlight the sensitivity of the inference to choices of specification and econometric strategy. While our results largely confirm several isolated results in recent strands of the literature, they also convey the strong message that the *specification* uncertainty surrounding estimation of the NKPC is vast. We then show, using a number of benchmark specifications, that even given a model, the *sampling* uncertainty of the estimates tends to be large. Both of these conclusions can be explained by the weakness of identification. We also demonstrate that the potential for

<sup>32</sup> Estimation results are obtained using Ox (Doornik 2007).

the data to distinguish between rational and nonrational price setting is limited.

### 5.1 Data

As in most of the literature, our dataset features U.S. aggregate time series at a quarterly frequency, with the largest possible sample extending from 1947q1 to 2011q4. Most series have been downloaded from the St. Louis Fed's FRED database. The data consists of alternative series for price and wage inflation, the labor share, output, interest rates, and survey measures of inflation expectations. We use the abbreviation "NFB" for the nonfarm business sector. See section A.4 in the Appendix for a detailed description of the data and transformations.

A few of our data series deserve mention here. Survey forecasts of inflation are taken from the SPF and the Federal Reserve's Greenbooks. We consider both one-quarter-ahead inflation forecasts made at time  $t$ ,  $\pi_{t+1|t}^s$ , and two-quarters-ahead inflation forecasts made at time  $t-1$ ,  $\pi_{t+1|t-1}^s$ .

Inflation gaps are calculated as the raw inflation rate minus a measure of trend inflation. Our two model-based measures of trend inflation are the smoothed (two-sided) and filtered (one-sided) permanent components of inflation from the unobserved components stochastic volatility model of Stock and Watson (2007, 2010). For consumer price index (CPI) inflation, ten-year CPI inflation forecasts serve as an additional measure of trend inflation (this series starts in 1991).

*Real-time* data on inflation and output is obtained from the Philadelphia Fed's website. We have compiled a unique dataset on real-time changes in the labor share (real unit labor cost), for use as instruments, by combining internal records from the Bureau of Labor Statistics with figures from the bureau's historical news releases.<sup>33</sup>

<sup>33</sup>We are grateful to Shawn Sprague for assisting us in obtaining the real-time labor share data.

Our output gaps include the official estimate from the Congressional Budget Office (CBO) as well as various detrended output series. We also compute labor share gaps. This is done to remove trends such as the recent dramatic decline in the labor share, which may arguably be attributed to secular changes outside of the New Keynesian model.<sup>34</sup> In addition to full-sample gaps, we use real-time output data or current-vintage labor share data to compute *one-sided* gaps, for which the trend is determined using only data points up to time  $t$ . Because such series do not estimate the trend from future data, they (or their lags) can more plausibly be treated as exogenous for estimation purposes.<sup>35</sup> Another stationary analog of the labor share is the cointegrating relationship between real wages and labor productivity found by Sbordone (2005). Like most of the literature, we consider non-detrended labor share series as well.

In our empirical analysis we ignore measurement error in the estimates of the trends of inflation and forcing variables.

### 5.2 Specification Sensitivity

We take the specification of Galí and Gertler (1999) and Galí, Gertler, and López-Salido (2001) to be our benchmark: a hybrid NKPC (9) with one lag of inflation and the labor share as forcing variable, estimated by GIV under the rational expectations assumption. As discussed in section 4, GIV analyses typically find point estimates of the coefficient on expectations  $\gamma_f$  in the 0.5–0.7 range, and the coefficient on lagged inflation  $\gamma_b$ , the measure of intrinsic persistence, is often significantly positive and

<sup>34</sup>See Gwin and VanHoose (2008) for a discussion of the need to detrend marginal cost measures.

<sup>35</sup>Unfortunately, we cannot use our real-time labor share data to construct actual real-time labor share gaps. Because the Bureau of Labor Statistics (BLS) base year changes over time, we can only compute real-time *changes* in the (log) labor share, not levels. Our one-sided labor share gaps therefore rely on current-vintage data.

TABLE 3  
 BASELINE GIV ESTIMATES USING DIFFERENT DATA VINTAGES

Data vintage	Const.	$\lambda$	$\gamma_f$	$\gamma_b$	Hansen test
1998	0.041 (0.030)	0.026 (0.013)	0.615 (0.057)	0.340 (0.058)	5.263 [0.628]
2012	-0.049 (0.040)	0.018 (0.012)	0.719 (0.099)	0.240 (0.095)	9.816 [0.199]

*Notes:* Comparison of GIV estimates of the hybrid NKPC based on 1998 and 2012 vintages of data. The estimation sample is 1970q1 to 1998q1. Inflation: GDP deflator. Labor share: NFB. Instruments: four lags of inflation and two lags of the labor share, wage inflation, and quadratically-detrended output. Estimation method: CUE GMM. Weight matrix: Newey and West (1987) with automatic lag truncation (4 lags). Standard errors in parentheses and  $p$ -values in square brackets.

not significantly different from  $1 - \gamma_f$ . The coefficient on the labor share  $\lambda$  is generally estimated to be positive, but borderline significant (using the usual strong-instrument inference). In table 3 we replicate these findings using data of the same vintage as Galí and Gertler (1999), but with the Galí, Gertler, and López-Salido (2001) instrument set.<sup>36</sup> Later papers have mostly obtained insignificant  $\lambda$  estimates, and like Rudd and Whelan (2007) we find that this is even true on the Galí and Gertler (1999) sample if revised data (as of 2012) is used. Using the output gap as forcing variable also typically yields an insignificant estimate of  $\lambda$ , and early papers in the literature tended to find negative point estimates.

The estimation results reported in the literature differ in terms of the choice of data series, estimation sample, and various other aspects of the specification, such as the number of inflation lags, any additional

regressors, the measurement of inflation expectations, and the identification assumptions, including the set of instruments and other identifying restrictions. As we showed in figure 3, estimates of  $\lambda$  and  $\gamma_f$  reported in various papers differ markedly, but the key message is that all highly cited papers obtain a positive slope coefficient ( $\lambda > 0$ ), and, with the exception of Fuhrer (2006) and Henzel and Wollmershäuser (2008), generally find forward-looking behavior to be dominant ( $\gamma_f > 0.5$ ). The results presented in figure 3 are a tiny subset of possible specifications. Table 4 presents various dimensions of the specification choice that have been considered in the literature.<sup>37</sup> These combinations of choices produce a very large number of specifications that are not objectionable on a priori grounds.

<sup>36</sup>We obtained the 1998 vintage data from Adrian Pagan. We use the continuous updating estimator (CUE) rather than 2-step GMM (cf. section A.2.1 in the Appendix) because the former is invariant to reparametrization of the moment conditions. The results are comparable to the bottom two rows of table 2 in Galí, Gertler, and López-Salido (2001).

<sup>37</sup>The only components of the table that have not been explored extensively in the literature are some of the real-time data series (but see Paloviita and Mayes (2005), Dufour, Khalaf, and Kichian (2006), and Wright (2009)) and the use of survey expectations as instruments (but see Wright (2009) and Nunes (2010)). The latter is motivated by evidence that surveys typically forecast inflation better than most alternatives; see Ang, Bekaert, and Wei (2007).

TABLE 4  
 NKPC SPECIFICATION COMBINATIONS

Specification settings	Options
Inflation ( $\pi_t$ )	GDP deflator, CPI, chained GDP def., GNP def., chained GNP def., NFB GDP def., PCE, core PCE, core CPI, filtered GDP def. gap, smoothed GDP def. gap, filt. CPI gap, sm. CPI gap, SPF-based CPI gap, filt. core CPI gap, sm. core CPI gap, filt. PCE gap, sm. PCE gap, filt. core PCE gap, sm. core PCE gap
Labor share (ls)	NFB, NFB coint. relation, HP filtered NFB gap, Baxter-King filt. NFB gap, linearly detrended NFB gap, quadratically detrended NFB gap, real-time NFB HP gap, real-time NFB BK gap, real-time NFB lin. detr. gap, real-time NFB quadr. detr. gap
Output gap (ygap)	CBO, HP filt., BK filt., lin. detr., quadr. detr., real-time HP filt., real-time BK filt., real-time lin. detr., real-time quadr. detr.
Reduced form	Unrestricted, VAR
Survey forecasts ( $\pi_{t \tau}^s$ )	SPF CPI, SPF GDP def., GB GDP def.
Expectations	$\pi_{t+1}$ (endogenous), $\pi_{t+1 t}^s$ (endog.), $\pi_{t+1 t}^s$ (exogenous) $\pi_{t+1 t-1}^s$ (endog.), $\pi_{t+1 t-1}^s$ (exog.)
Instruments	GG: 4 lags of $\pi_t$ , ls, ygap, 10y–90d yield spread, wage infl., commodity price infl. GGLS: 4 lags of $\pi_t$ and 2 lags of ls, ygap, wage infl. small: 4 lags of $\pi_t$ and 3 lags of forcing variable exact: 1 extra lag of each endog. regr. (just-identified) RT: 2 real-time lags of GDP def. inflation, $\Delta$ ls, ygap survey: 2 lags of 1-quarter SPF/GB forecasts, forcing variable Extra regressors (e.g., oil) added to instruments (if endog., use 2 lags)
Inflation lags	0 lags (pure NKPC), 1 lag, 4 lags
Parameter restrictions	No restrictions, $\gamma(1) = \gamma_f$ (inflation coefficients sum to 1) With $\gamma(1) = \gamma_f$ , use lags of $\Delta\pi_t$ instead of $\pi_t$ as instruments
Oil shocks	None, log change of WTI spot price divided by GDP def.
Interest rate	None, 90-day Treasury rate
Sample	Full available, 1960–1997, 1968–2005, 1968–2008, 1971–2008, 1981–2008, 1984–end of sample
GMM estimator	2-step, CUE

*Notes:* List of the specification options that we consider when estimating the NKPC (9). The efficient GMM weight matrix is computed using the Newey and West (1987) heteroskedasticity and autocorrelation consistent estimator with automatic lag truncation, except for VAR specifications, which use the White (1980) heteroskedasticity consistent estimator.

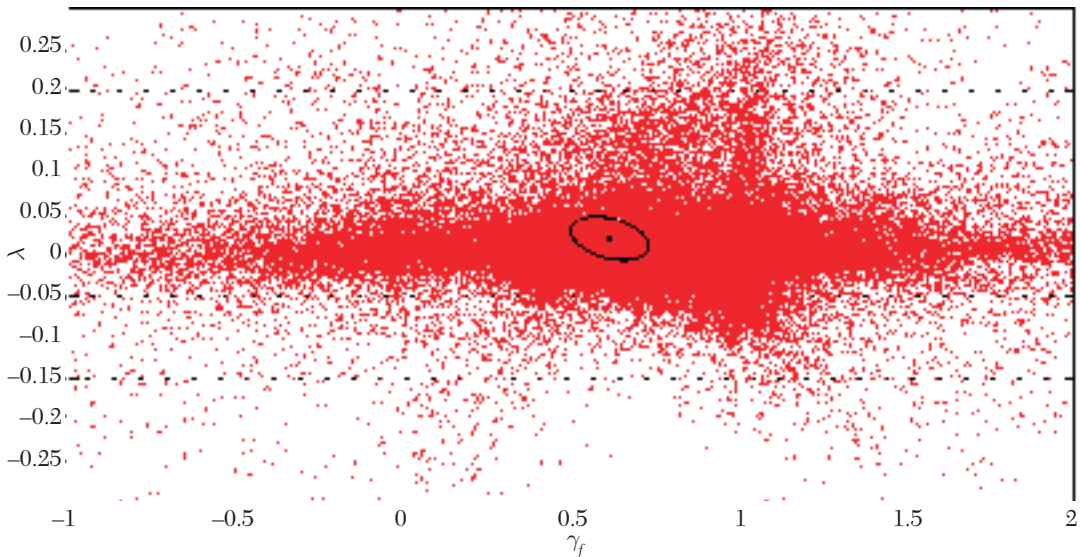


Figure 4. Point Estimates: Labor Share Specifications

Notes: Point estimates of  $\lambda$ ,  $\gamma_f$  from the various specifications listed in table 4 that use the labor share as forcing variable, excluding real-time and survey instrument sets. The black dot and ellipse represent the point estimate and 90 percent joint Wald confidence set from the 1998 vintage results in table 3.

To gauge the sensitivity of the results about the importance of forward-looking behavior to variations in data, sample, and identification assumptions, we obtain estimates of the coefficients ( $\lambda$ ,  $\gamma_f$ ) in the baseline NKPC (9) for various combinations of the specification choices listed in table 4. We then plot the point estimates in  $(\gamma_f, \lambda)$ -space. These plots do not convey any information about sampling uncertainty, i.e., they are not confidence sets. Confidence sets for a subset of those specifications are analyzed in section 5.3 below. However, these plots, which we refer to as “clouds,” do give a useful visual impression of the specification uncertainty. We study the specifications with the labor share and output gap as forcing variable separately, because the coefficient  $\lambda$  on the forcing variable is not comparable across these cases. As we are only able to report a limited number of results here, we invite interested

readers to explore the myriad of possible clouds using our interactive Matlab plotting tool, available in the online supplement.<sup>38</sup>

We first look at the specification settings that have been used in the literature (i.e., not using real-time data or survey expectations as instruments). Figures 4 and 5 report the results with the labor share and output gap as forcing variable, respectively. Figure 4 also contains the Galí and Gertler (1999) vintage point estimate and associated Wald confidence ellipse from table 3 for comparison. These plots contain more than 600,000 estimates combined. Observe that the plotted parameter space  $(\gamma_f, \lambda) \in [-1, 2] \times [-0.3, 0.3]$  is much larger than that of figure 3. Table 5 reports

<sup>38</sup>[http://www.aeaweb.org/jel/app/mar14\\_Mav\\_doc.zip](http://www.aeaweb.org/jel/app/mar14_Mav_doc.zip).



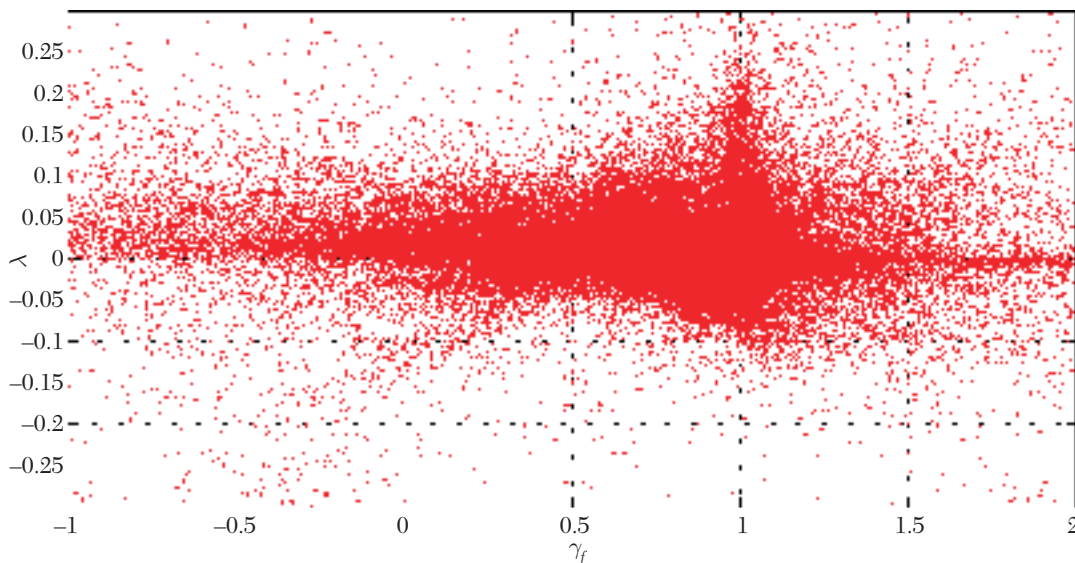


Figure 5. Point Estimates: Output Gap Specifications

Notes: Point estimates of  $\lambda$ ,  $\gamma_f$  from the various specifications listed in table 4 that use the output gap as forcing variable, excluding real-time and survey instrument sets.

TABLE 5  
SUMMARY STATISTICS FOR ESTIMATION RESULTS

Parameter	Labor share		Output gap	
	$\lambda$	$\gamma_f$	$\lambda$	$\gamma_f$
Median	0.004	0.753	0.004	0.760
Fifth percentile	-0.068	-0.648	-0.070	-0.771
Ninety-fifth percentile	0.135	1.814	0.133	1.831
Fraction both positive		0.525		0.505
... and signif. (5% one-sided $t$ test)		0.102		0.179
... and $\gamma_f > 0.5$		0.087		0.155
Median $F_{HAR}$	63.73	3.079	166.46	4.154
Fraction rejections by 5 percent Hansen test		0.033		0.032

Notes: Summary statistics of estimation results across specifications listed in table 4, excluding real-time and survey instrument sets. Hansen test is evaluated at the CUE using larger critical values that are robust to weak identification, cf. section A.2.6 of the Appendix; results for this statistic do not include VAR-GMM specifications.

summary statistics for the point estimates in figures 4 and 5.

The main messages from the figures are that (i) estimates of the coefficient on

the forcing variable are symmetrically dispersed around zero, and (ii) estimates of the coefficient on expectations are on average around 3/4 and very dispersed, though the vast majority (around 90 percent) of those are positive. Importantly, only about half of the estimates lie in the positive orthant  $\lambda > 0, \gamma_f > 0$ . Moreover, the fraction of cases in which  $\lambda$  and  $\gamma_f$  both appear statistically significantly positive using (one-sided) 5 percent-level individual  $t$  tests is quite small, while most of the reported estimates in the literature appear to fall into that category. It is interesting that the frequency of significantly positive coefficients for the output gap specifications is almost double the frequency for the labor share ones. This is not in line with the view that NKPC specifications with the output gap as forcing variable more frequently have estimates with the “wrong sign” than do specifications using the labor share as forcing variable (Galí and Gertler 1999). It is important to stress that the results based on  $t$  tests are reported for the comparison with the literature, and they do not yield reliable evidence on the significance of the coefficients. In the next subsection we report results that are robust to weak identification.

To shed some light on the issue of weak identification, the penultimate row of table 5 reports the median value of the heteroskedasticity and autocorrelation robust first-stage  $F$  statistic of Montiel Olea and Pflueger (2013), denoted  $F_{HAR}$ . A low value of this statistic can be thought of as a warning sign for weak instruments.<sup>39</sup> We see that instruments are quite strong for forecasting the forcing variable (the median  $F$  statistic is 63.7 for labor share specifications

and 166.5 for output gap specifications) but rather weak for forecasting the inflation expectation proxy (median  $F$  is 3.1 and 4.2, respectively).<sup>40</sup> Even though this is not a formal test of weak instruments and we do not recommend the use of pretests in place of weak identification robust inference, these results reinforce the intuition that changes in inflation are hard to forecast and we should, therefore, worry about weak identification. Figure 6 displays smoothed density estimates of our first-stage  $F$  statistics for forecasting the expectations proxy, treating rational expectations GIV separately from time- $t$  dated SPF/Greenbook forecasts, and using all instrument sets in table 4. As one might expect, time- $t$  dated survey forecasts are much better predicted by the various instrument sets than is next period’s realized inflation. The median  $F$  for forecasting next period’s inflation is 2.7 across all labor share specifications (3.6 across output gap specifications). In comparison, the median is 12.8 (12.5) for SPF/Greenbook forecasts, and if we restrict attention to the instrument set that includes lagged survey forecasts, the median  $F$  is even higher at 42.1 (43.9). This suggests that survey forecast specifications of the NKPC may be more strongly identified than their GIV counterparts, and the evidence reported in section 5.3 corroborates this conjecture.

The final row of table 5 reports the rejection frequencies of a weak identification robust version of Hansen’s (1982)  $J$  test of overidentifying restrictions; see section A.2.6 in the Appendix for details. The rejection frequencies are just over 3 percent at the 5 percent level, so there is no systematic evidence

<sup>39</sup>The well-known rule of thumb of  $F > 10$  is a commonly used benchmark (Stock and Yogo 2002), although Montiel Olea and Pflueger (2013) show that this condition is neither necessary nor sufficient for instruments to be strong in the presence of heteroskedasticity and autocorrelation.

<sup>40</sup>Low values of the  $F$  statistic for forecasting the forcing variable do arise in specifications that use the real-time instrument set. For labor share specifications, the median value of the  $F$  statistic is 9.5 for the real-time regressions, whereas the median is 63.1 for all other instrument sets in table 4. For output gap specifications, the corresponding medians are 69.3 and 170.6, respectively.

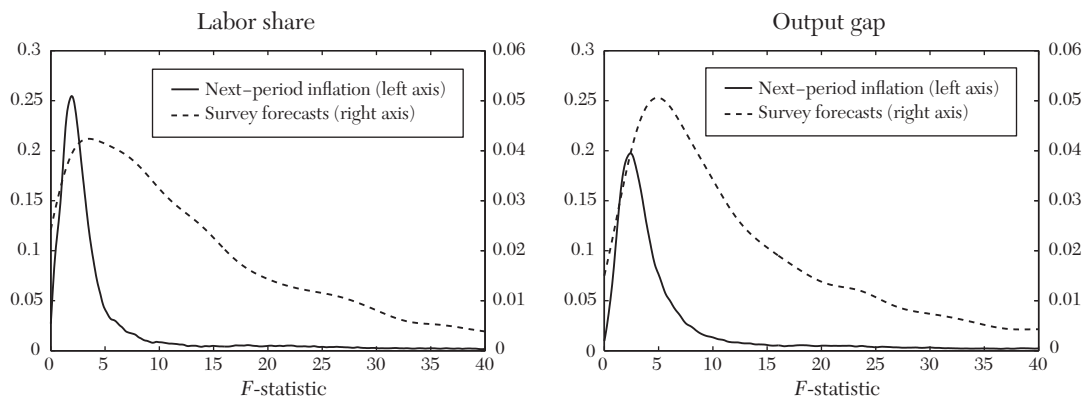


Figure 6. First-Stage  $F$ -Statistics

Notes: Smoothed density estimates of robust first-stage  $F$  statistics for forecasting the expectation proxy, using realized next-period inflation (solid line, left axis) or time- $t$  dated endogenous SPF/GB survey forecasts of inflation (dotted line, right axis). The left and right panels show results for all labor share and output gap specifications, respectively, that are listed in table 4.

against the validity of the overidentifying restrictions. Notice, however, that this test is less powerful than the standard  $J$  test because it uses larger critical values.

We now take a closer look at the different dimensions of the specification choice. In the following we do not exclude estimates that use the real-time or survey instrument sets. In the remainder of this subsection, the discussion is organized in self-contained paragraphs that can be skipped without affecting the readability of the rest of the article. Additional details are provided in section A.5 of the Appendix.

### 5.2.1 CUE versus 2-Step GMM

We generate GMM estimates using both the efficient two-step estimator and the continuously updated estimator (CUE) of Hansen, Heaton, and Yaron (1996). These are described in section A.2.1 in the Appendix. Table 6 compares summary statistics of point estimates based on two-step and CUE GMM for the various specifications

listed in table 4.<sup>41</sup> The two-step and CUE are very similar for  $\lambda$ , and CUE is typically larger than two-step for  $\gamma_f$ . Moreover, the two-step estimates are closer to the corresponding OLS estimates than CUE. This finding is consistent with the well-known bias of GMM estimators towards the OLS probability limit, which is stronger for two-step than for CUE (Stock, Wright, and Yogo 2002). The relatively better bias properties of the CUE come at the cost of greater dispersion, which is confirmed by the 90 percent interquartile ranges: the ones for CUE are more than double the corresponding ones for two-step. Bårdsen, Jansen, and Nymoen (2004) and Guay and Pelgrin (2005) also report large sensitivity of NKPC estimates to the choice of GMM estimator, as well as to the set of instruments.

<sup>41</sup>In a small number of cases, the CUE either failed to converge or produced estimates that were very large in absolute value. This is consistent with the well-known property that the finite sample distribution of the CUE has fat tails (Hansen, Heaton, and Yaron 1996).

TABLE 6  
PAIRWISE COMPARISON OF POINT ESTIMATES: 2-STEP VERSUS CUE GMM

Parameter	Labor share				Output gap			
	$\lambda$		$\gamma_f$		$\lambda$		$\gamma_f$	
	2S	CUE	2S	CUE	2S	CUE	2S	CUE
Median	0.003	0.004	0.676	0.798	0.004	0.002	0.670	0.784
90 percent IQR	0.141	0.273	1.213	2.855	0.123	0.264	1.250	3.272
Med. diff 2S-CUE	-0.000		-0.030		0.000		-0.024	
Med. diff from OLS	0.000	0.002	0.168	0.239	-0.002	-0.003	0.156	0.213
Med. diff 2S-CUE (GG)	-0.001		-0.057		0.002		-0.041	
Med. diff from OLS (GG)	0.001	0.002	0.124	0.208	-0.002	-0.005	0.129	0.195

*Notes:* Comparison of 2-step and CUE GMM estimates for the specifications in Table 4 (excluding VAR specifications, for which we only computed 2-step GMM). “90 percent IQR” is the difference between the ninety-fifth and fifth percentiles. Rows labeled “GG” focus on results for the GG instrument set.

TABLE 7  
PAIRWISE COMPARISON OF POINT ESTIMATES: IMPACT OF VAR ASSUMPTION

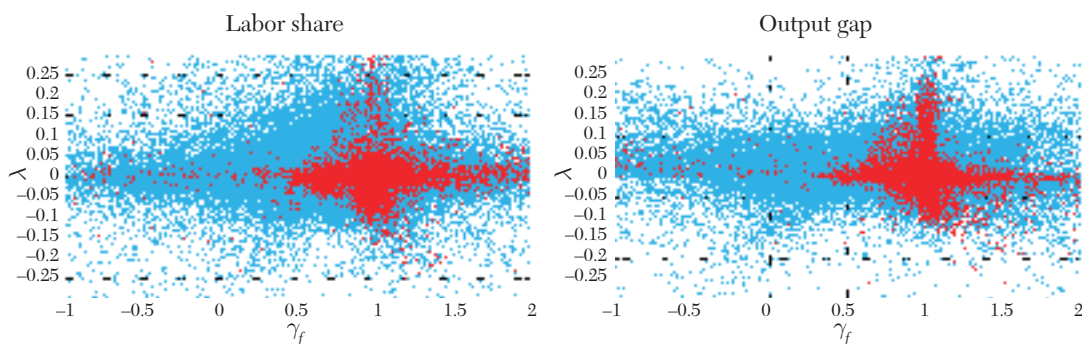
Parameter	Labor share				Output gap			
	$\lambda$		$\gamma_f$		$\lambda$		$\gamma_f$	
	VAR	GIV	VAR	GIV	VAR	GIV	VAR	GIV
Median	0.002	0.002	0.997	0.747	0.002	0.001	0.997	0.750
90 percent IQR	0.137	0.087	1.912	1.124	0.198	0.106	3.207	1.116
Median diff	-0.001		0.178		0.000		0.176	
Fraction positive diff	0.481		0.844		0.513		0.797	

*Note:* Effect of imposing the VAR identification assumption.

### 5.2.2 VAR Assumption

Our VAR-GMM estimates are based on the moment condition (15). The reduced form evolution of inflation is thus restricted to be a linear function of the variables in the instrument set. Table 7 reports summary statistics comparing GIV and VAR-GMM estimates, while figure 7 plots clouds for estimates that impose the VAR assumption and those that do not. There is no noticeable difference in the estimates of  $\lambda$  between

the VAR and GIV methods, but there is a substantial difference in  $\gamma_f$ —in the vast majority of cases (about 80 percent), imposing the VAR assumption increases the estimate of  $\gamma_f$ , and the median estimate is actually one. This is consistent with the results reported in Sbordone (2005) that use VAR-MD and find no role for intrinsic persistence, as well as with the VAR-ML results in Kurmann (2007). It is inconsistent with Fuhrer (2006, 2012), who additionally imposes determinacy; cf.



*Figure 7.* Point Estimates: VAR (Dark Gray, Red Online) versus Unrestricted (Light Gray, Blue Online)  
*Notes:* Point estimates of the coefficient on the forcing variable and inflation expectations for the specification combinations listed in table 4. The dark gray (red online) points correspond to estimates that impose the VAR assumption, while the light gray (blue online) points do not impose the assumption. The left and right panels plot specifications with the labor share and output gap as forcing variable, respectively.

section 3.2. As we pointed out in section 3.3, weak identification can cause VAR estimates of  $\gamma_f$  to be biased toward 1. Imposing the additional restrictions that coefficients on inflation in the NKPC sum to one (i.e.,  $\gamma(1) = \gamma_f$  in equation (9)) and that inflation enters the VAR in first differences (thus using lags of  $\Delta\pi$  as instruments) causes  $\gamma_f$  estimates to concentrate even more tightly around one, but also increases the dispersion of the  $\lambda$  estimates.

### 5.2.3 Survey Forecasts

There are large and systematic differences in the effect of using survey inflation forecasts relative to rational expectations GIV across labor share and output gap specifications, sample period, inflation series (GDP deflator versus CPI), and forecast source (SPF versus Greenbook). Survey forecasts typically increase the estimate of  $\lambda$  across most specifications and sample periods, especially when the output gap is used as forcing variable. The estimate of  $\gamma_f$  moves in different directions across specifications:

it is typically much lower than GIV in labor share specifications and either the same or higher in output gap specifications. This is illustrated in figure 8, which plots the post-1984 cloud for rational expectations GIV estimates against that for time- $(t-1)$  dated exogenous SPF forecasts (results are similar for other choices of survey forecasts), treating GDP deflator and CPI specifications separately. Further details are given in table 9 in section A.5 of the Appendix.

Subsample variation is also quite striking. The reduction in  $\gamma_f$  relative to GIV is much more evident in the post-1981 sample (SPF CPI forecasts are only available from 1981q3). Survey specifications with CPI inflation typically yield much larger estimates of  $\gamma_f$  than those with GDP deflator inflation. SPF and Greenbook forecasts do not yield systematically different full-sample estimates, though there are some systematic difference to the estimates of  $\gamma_f$  before and after 1984. Treating surveys as endogenous or exogenous does not seem to make much difference to the central tendency of the



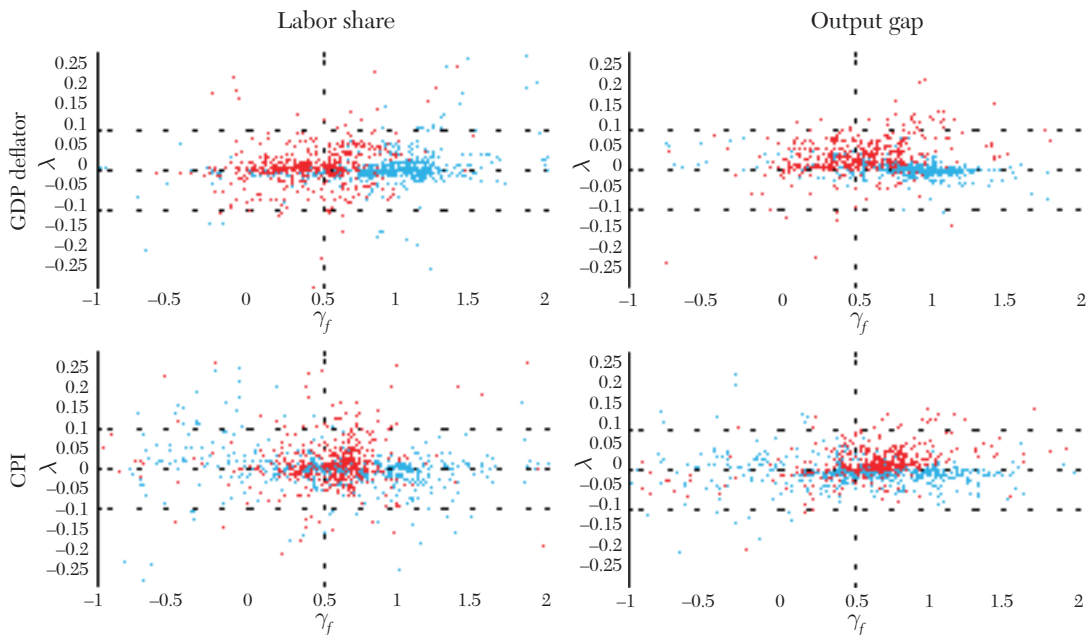


Figure 8. Point Estimates: Survey Forecasts (Dark Gray, Red Online) versus RE GIV (Light Gray, Blue Online)  
 Notes: Point estimates of the coefficient on the forcing variable and inflation expectations using time- $(t-1)$  dated exogenous SPF forecasts (dark gray, red online) versus RE GIV (light gray, blue online), for the various specification settings listed in table 4. The top row of figures use GDP deflator inflation, while the bottom row uses CPI inflation. In each row, the left and right panels correspond to labor share and output gap specifications, respectively. Sample: 1984q1–2011q4.

estimates, though it does make a difference to dispersion (the latter estimates are a lot less dispersed, as expected).

#### 5.2.4 Instruments

The last two rows of table 6 give median differences for specifications using the Galí and Gertler (1999) instrument set (GG), which is considerably larger than the rest. This instrument set produces estimates for  $\gamma_f$  that are typically lower than average. The GG estimates are also less dispersed and more concentrated around the OLS estimates. Other than GG, the choice of instrument set does not substantially change the central tendency of the estimates.

#### 5.2.5 Number of Inflation Lags in the NKPC

Estimates of  $\gamma_f$  are very sensitive to the number of inflation lags included in the model, while estimates of  $\lambda$  seem to be unaffected, on average. Specifically, adding lags of inflation to the NKPC tends to reduce the estimate of  $\gamma_f$  by about 0.25 when we add one lag to the pure NKPC, and by a similar amount when we add three more lags. This corroborates results reported by Rudd and Whelan (2005), but it need not be due to misspecification of the more restrictive NKPC specifications, as was suggested by Rudd and Whelan (2005) and Mavroeidis

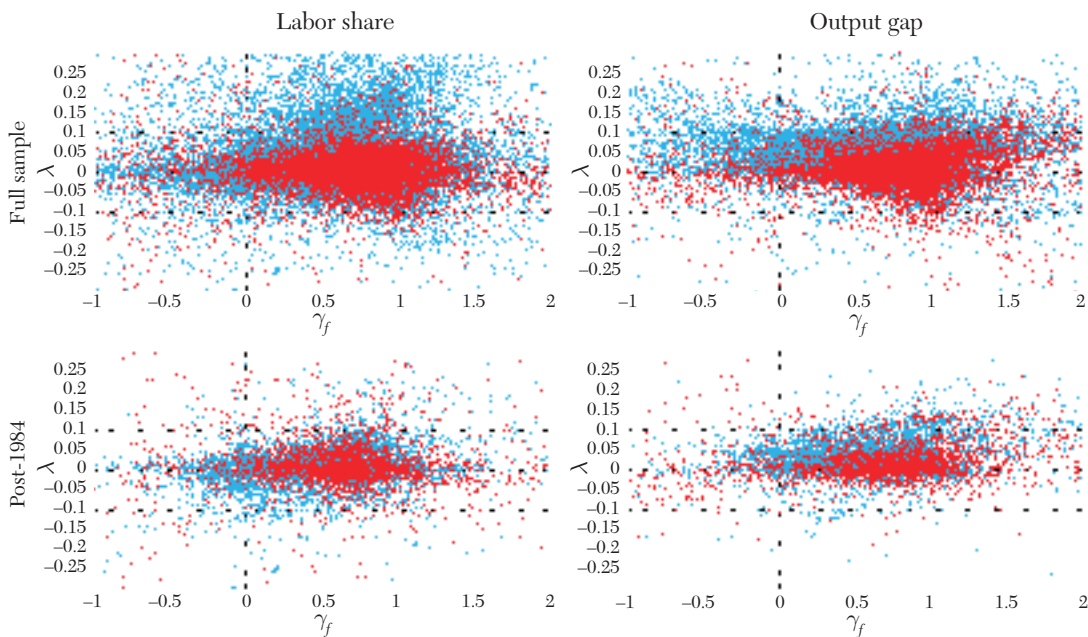


Figure 9. Point Estimates: CPI (Dark Gray, Red Online) versus GDP Deflator (Light Gray, Blue Online) Inflation  
 Notes: Point estimates of the coefficient on the forcing variable and inflation expectations using CPI (dark gray, red online) versus GDP deflator (light gray, blue online), for the various specification settings listed in table 4. The top row gives results for all available samples, while the bottom row selects only the post-1984q1 sample. The left and right panels plot specifications with the labor share and output gap as forcing variable, respectively.

(2005). The direction of the movement in the point estimates is entirely consistent with the possibility that specifications with more inflation lags are more weakly identified, in which case estimates of  $\gamma_f$  would exhibit a larger bias towards  $\gamma_f = 1/2$ .<sup>42</sup> Indeed, the median first-stage  $F_{HAR}$  statistic for inflation expectations are 24.5, 3.2, and 2.3, for the 0, 1, and 4-lag NKPC models, respectively, across all specifications that use the labor share as the forcing variable.<sup>43</sup> This is further

corroborated by the size of the robust confidence regions reported in the next subsection: they get progressively larger as we move from zero lags to four lags.

### 5.2.6 Inflation Series

Figure 8 indicates that survey forecast specifications are more sensitive to the choice of inflation series than GIV estimates are. Figure 9 compares GDP deflator and CPI estimates, pooling across all GIV and survey forecast specifications and all subsamples in table 4. Estimates of both parameters are considerably more dispersed in GDP deflator specifications, but the median difference across these inflation series is very small. This is partly a result of the general

<sup>42</sup>As explained in section 3.3, the probability limit of the OLS  $\gamma_f$  estimator is often close to 0.5 for empirically realistic NKPC parametrizations.

<sup>43</sup>The corresponding numbers for output gap specifications are 25.2, 4.1, and 3.1.

decrease in the dispersion of estimates from the pre-1984 to the post-1984 sample, since CPI specifications are under-represented in samples that contain data before 1981 due to the lack of CPI survey forecasts. The bottom row of plots in figure 8 compares CPI versus GDP deflator estimates for a common post-1984 sample, and it is apparent that the dispersion of the estimates is generally smaller and not substantially different across inflation series.

Using inflation gaps to account for trend inflation tends to produce somewhat lower estimates of  $\gamma_f$ , irrespective of whether the labor share or output gap is used as the forcing variable. For  $\lambda$ , there is a small positive difference only in output gap specifications. The reason these differences are not large may be that the sum of the coefficients on inflation are close to one, thus mitigating the impact of any trend inflation, as discussed in section 2.2. The inflation gap point estimates do, however, cluster much tighter around  $\lambda = 0$ .

The remaining inflation series yield results that are similar to either GDP deflator or CPI estimates. Using the chain-type GDP price index gives very similar results to those for GDP deflator inflation. For GIV, personal consumption expenditure (PCE) inflation estimates are similar to CPI results, although in output gap specifications  $\lambda$  tends to be estimated higher with CPI compared to PCE. There is little difference between using CPI/PCE inflation and their core inflation equivalents, except that the core estimates are less dispersed.

### 5.2.7 Output Gap and Labor Share Series

There is generally very little systematic difference in the results based on alternative labor share and output gap series, except that use of detrended labor share series (labor share “gaps”), using either pseudo-real-time or full-sample trends, increases the dispersion of the estimates of  $\lambda$  without much change in central tendency. This could be due to the

fact that the detrended series are harder to forecast, thus making identification somewhat weaker.<sup>44</sup> A striking conclusion is that the addition of a good decade’s worth of data (and data revisions) since Galí and Gertler (1999) completely overturns their conclusion that labor share specifications yield markedly different results from output gap specifications.

### 5.2.8 Sample

There is little systematic difference in the central tendency of estimates before and after 1984, cf. table 10 in the Appendix. As figure 8 suggests, survey estimates are, however, sensitive to sample choice. This is consistent with Zhang, Osborn, and Kim (2008). Figure 10 reports pre- and post-1984 estimates in the survey specifications with GDP deflator inflation. For rational expectations GIV specifications, the central tendency of estimates does not depend much on the choice of sample, but post-1984 estimates are more tightly concentrated around  $\lambda = 0$ .

### 5.2.9 Other Specification Choices

The restriction that coefficients sum to one does not matter much except for VAR specifications, as discussed above. Use of oil prices or interest rates in the NKPC does not affect the central tendency of the point estimates. This is consistent with Chowdhury, Hoffmann, and Schabert (2006) and Ravenna and Walsh (2006).

## 5.3 Sampling Uncertainty

The previous subsection focused on specification sensitivity, characterized by the variation in point estimates across specifications. We now turn to sampling uncertainty, which we measure conventionally using confidence

<sup>44</sup>The median of the first-stage  $F_{HAR}$  statistic for the labor share is 71 for the levels and 50 for the gap specifications, respectively.

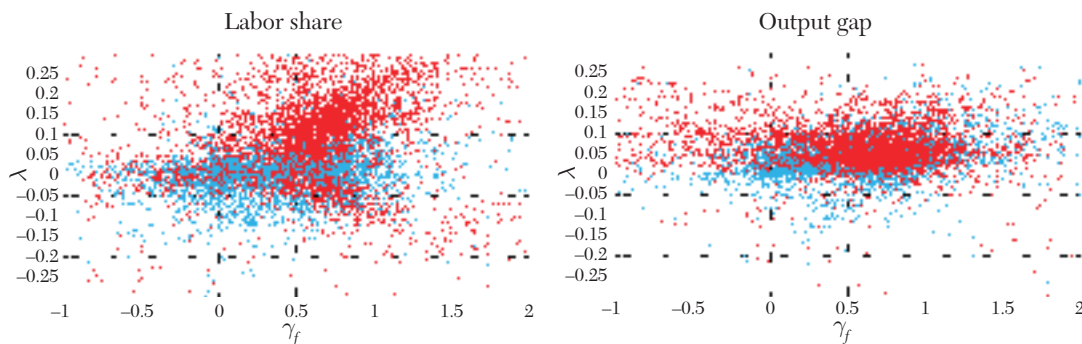


Figure 10. Survey Forecasts in Pre-1984 (Dark Gray, Red Online) versus Post-1984 (Light Gray, Blue Online) Samples

Notes: Point estimates of the coefficient on the forcing variable and inflation expectations in survey forecast specifications (SPF and GB) on pre-1983q4 (dark gray, red online) and post-1984q1 (light gray, blue online) samples. The inflation series is GDP deflator inflation. The left and right panels plot specifications with the labor share and output gap as forcing variable, respectively.

sets for selected specifications based on methods that are robust to weak identification. Our robust confidence sets, called  $S$  sets, are based on the  $S$  test of Stock and Wright (2000), described in section A.2.5 of the Appendix. This is a test of the validity of the model's identifying restrictions at a hypothesized value of the structural parameters. Other weak identification robust methods, such as conditional likelihood ratio or score tests (Moreira 2003; Kleibergen 2005) are more powerful than the  $S$  test under strong identification, but they are technically more involved and computationally more demanding. We do not report results based on those tests because, in all the cases that we considered, they gave similar results to the  $S$  test.

$S$  sets are obtained by inverting the  $S$  test, i.e., by performing an  $S$  test for each candidate value of the parameters in the parameter region and collecting all the points that are not rejected at the given significance level. Unlike Wald sets, which are elliptical and can be computed analytically,  $S$  sets need to be computed by grid search over the parameter space and they can be disjoint. In this exercise, we use the same parameter

region as the one that was used for the cloud plots (which includes over 90 percent of all point estimates), namely,  $\lambda \in [-0.3, 0.3]$  and  $\gamma_f \in [-1, 2]$ . For each specification, we evaluate the test at over 1,000 grid points. Because this procedure is computationally intensive, we consider only a subset of all the specifications listed in table 4, consisting of about 1,400 specifications; see table 8.<sup>45</sup> The cloud of point estimates for the specifications in table 8 is qualitatively similar to that for the full set of specifications in table 4. Perhaps not surprisingly, the union of the joint 90 percent  $S$  sets for all specifications in table 8 covers the entire parameter region in our plots.<sup>46</sup> These findings are detailed in section A.5 of the Appendix.

To get a sense of the impact of different specification choices on sampling uncertainty, we compare the average size of

<sup>45</sup>Computation of the  $S$  sets for VAR-GMM takes about one hundred times longer than using the other single-equation methods. Therefore, we only consider sixteen specifications that impose the VAR assumption.

<sup>46</sup>The union of the  $S$  sets may be formally interpreted as a projection-based grand  $S$  set that projects over a latent hyperparameter which indexes the different specifications.

TABLE 8  
NKPC SPECIFICATION COMBINATIONS USED FOR ROBUST CONFIDENCE REGIONS

Specification settings	Options
Inflation ( $\pi_t$ )	GDP deflator, CPI
Labor share (ls)	NFB, real-time NFB BK gap
Output gap (ygap)	CBO, real-time BK filter
Reduced form	Unrestricted, VAR
Survey forecasts ( $\pi_{t \tau}^s$ )	SPF CPI, SPF GDP def., GB GDP def.
Expectations	$\pi_{t+1}$ (endogenous), $\pi_{t+1 t}^s$ (endogenous)
Instruments	GGLS, small, exact, RT, survey
Inflation lags	0, 1, 4
Parameter restrictions	No restrictions, $\gamma(1) = \gamma_f$ (inflation coefficients sum to 1)
Sample	Full available, pre-1984, 1984–end of sample

Note: Different specifications of the NKPC for which we compute robust confidence sets.

90 percent and 95 percent  $S$  sets across different specification choices (see section A.5 in the Appendix for details). The  $S$  sets are generally quite large, covering on average between 1/3 (90 percent level) and 1/2 (95 percent level) of the parameter space for both labor share and output gap specifications.<sup>47</sup> However, there is systematic variation in size across specification choices. With regards to the impact of adding lags of inflation to the NKPC, the size of the  $S$  sets becomes progressively larger as we move from zero to four lags. The difference between the pure and 1-lag hybrid NKPC is small, but adding three more lags of inflation to the hybrid NKPC roughly doubles the  $S$  sets, on average. The size of the  $S$  sets is smaller over the full sample than over pre- and post-1984 subsamples, as expected, but pre-1984  $S$  sets are smaller than post-1984  $S$  sets. More striking differences arise when we compare rational expectations GIV to survey inflation expectations and when we compare different instrument sets.  $S$  sets

for rational expectations GIV are on average much larger than for surveys, as anticipated in the discussion of first-stage  $F$  statistics above, and it looks like most of the difference arises from using Greenbook forecasts. With regards to the different instruments, real-time (external) instruments yield the largest  $S$  sets covering 50–80 percent of the parameter space. These are almost double the size of the  $S$  sets for exactly identified models, which are the smallest. Use of lagged survey forecasts as instruments produces on average smaller  $S$  sets than using lags of realized inflation, as conjectured by Wright (2009).

It is interesting to compute how often the  $S$  sets for  $(\lambda, \gamma_f)$  lie entirely in the positive orthant, as would be required to find significant evidence of forward-looking behavior. First, recall that  $S$  sets can be empty, which would indicate violation of the model's overidentifying restrictions, but the frequency of empty  $S$  sets (for the overidentified specifications) is considerably below the nominal significance level, so there is no systematic evidence against the validity of the identifying restrictions. Jointly significantly positive coefficients  $\lambda$  and  $\gamma_f$  occur in a very

<sup>47</sup> Additionally, the  $S$  sets are, on average, three to seven times larger than the corresponding Wald ellipses.



small fraction of the specifications considered (less than 5 percent at the 10 percent level).<sup>48</sup> This happens more frequently when the output gap rather than the labor share is used as the forcing variable. Interestingly, when the forcing variable is the output gap, we obtain significantly positive coefficients only when survey forecasts are used to proxy for inflation expectations, whereas when it is the labor share, the occurrence of positive  $S$  sets is equally (un)likely for survey and rational expectations GIV specifications. Significantly positive coefficients almost never arise when 4 lags of inflation are included in the NKPC, or when real-time instruments are used. Detailed results are provided in section A.5 in the Appendix.

Next, we draw 90 percent  $S$  sets and Wald confidence ellipses (based on the CUE) for  $(\lambda, \gamma_f)$  in the NKPC for a number of different specifications. The complete collection of robust confidence sets can be accessed using our interactive Matlab plotting tool in the online supplement (cf. footnote 38). Figure 11 reports the results for specifications based on GDP deflator inflation using either the labor share (NFB) or output gap (CBO) as forcing variables, imposing the restriction that inflation coefficients sum to 1, and using the “small” instrument set (three lags of  $\Delta\pi_t$  and  $x_t$ ). Three samples are considered: the full available sample, and the pre-1984 and post-1984 subsamples. The confidence sets are not completely uninformative, and they are particularly tight along the  $\lambda$  axis over the full sample, but rather wide across the  $\gamma_f$  axis. All  $S$  sets (and most Wald ellipses) contain  $\lambda = 0$ . For most specifications, identification is sufficiently weak for the results to be consistent both with the view that there is no forward-looking behavior, i.e., no role for expectations in price setting, as well as

with the view that expectations matter a lot. Martins and Gabriel (2009) and Kleibergen and Mavroeidis (2009) reach similar conclusions. Regarding subsample variation, even though the point estimates differ considerably across the pre- and post-1984 samples, the sampling uncertainty is so large that we cannot infer that the coefficients have changed over time.<sup>49</sup>

Figure 12 reports confidence sets for the full sample based on the assumption that the reduced form is a VAR(3) in the change in inflation and the forcing variable. The point estimate of  $\gamma_f$  is larger than one when either the labor share or the output gap are used as forcing variables, and the confidence sets are considerably tighter than for the corresponding GIV specification—compare with the top row of figure 11. Hence, the VAR assumption appears to be informative in these specifications, consistent with the results in Magnusson and Mavroeidis (2010). However, we should stress that, due to computational limitations, we have only looked at very few VAR specifications, so this result should be viewed as tentative. A more thorough investigation is needed in order to assess the validity of the VAR assumption.

Figure 13 reports confidence sets based on survey specifications. Results are reported for SPF and Greenbook GDP deflator inflation forecasts, as well as SPF CPI inflation forecasts. For SPF GDP deflator forecasts, the results are quite similar to the corresponding rational expectations GIV specifications, given in the bottom row of figure 11 (post-1984 sample). However, when we use the Greenbook forecasts, confidence sets become considerably smaller.  $S$  sets

<sup>48</sup>Specifically, this is the fraction of the  $S$  sets that are nonempty and lie entirely in the area  $(\lambda, \gamma_f) \in (0, .3] \times (0, 2]$ .

<sup>49</sup>Using weak identification robust stability tests, Kleibergen and Mavroeidis (2009) find some evidence of instability before 1984, when they use a shorter pre-1984 sample, but no evidence of instability using their full 1960–2008 sample, which is consistent with Magnusson and Mavroeidis (2013).

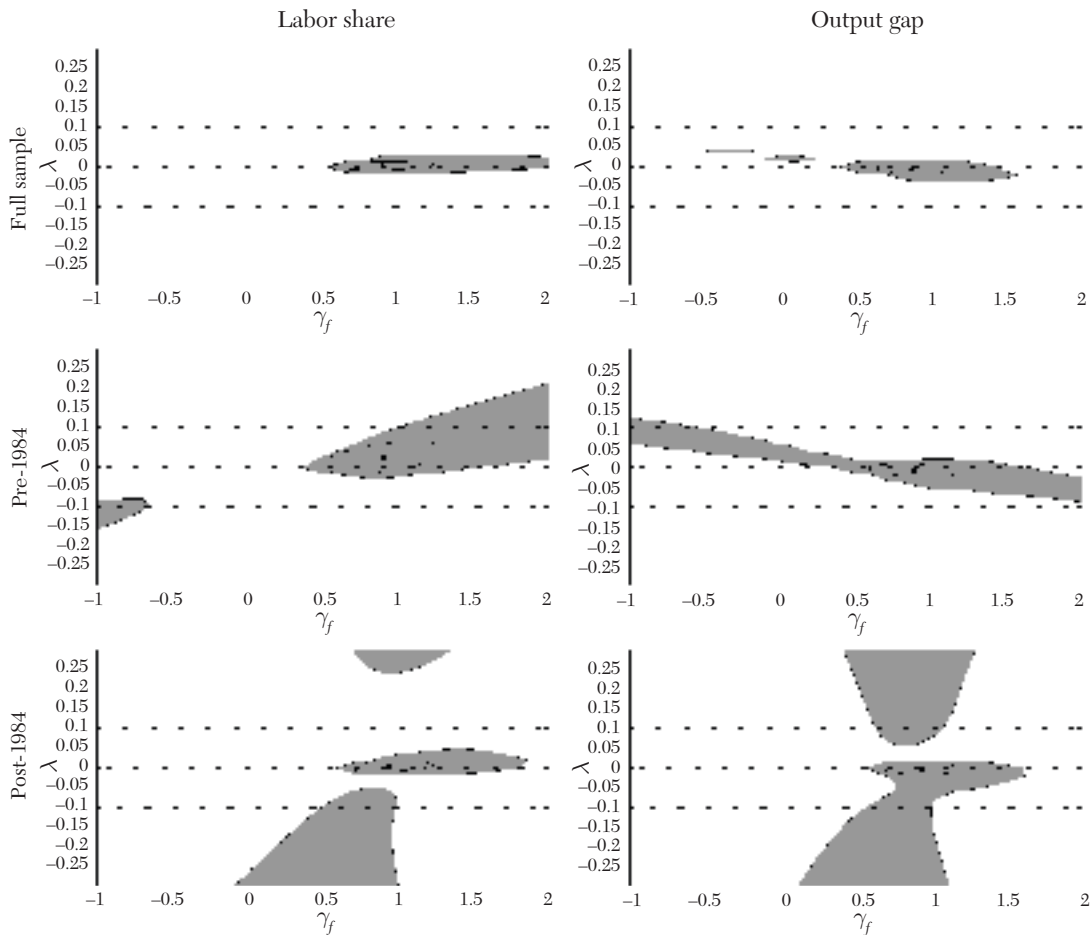


Figure 11. Robust Confidence Regions: RE Specifications

Notes: 90 percent  $S$  set (gray), 90 percent Wald ellipse, and CUE GMM point estimate (bullet) of the coefficients of the labor share and future inflation in the hybrid NKPC specification with one lag of inflation, where inflation coefficients sum to one. Inflation: GDP deflator. Forcing variable: NFB labor share (left panels), CBO output gap (right panels). Instruments: three lags of  $\Delta\pi_t$  and the forcing variable. Sample: starts 1948q2 (labor share), 1949q4 (output gap), ends 2011q3; full sample (top row), pre-1983q4 (middle row), post-1984q1 (bottom row). Weight matrix: Newey–West with automatic lag truncation.

based on SPF CPI forecasts are comparable in size and have considerable overlap with those based on SPF GDP deflator forecasts. Results for SPF GDP deflator inflation specifications over the pre- and post-1984 samples look very different. In particular, the 90 percent Wald ellipses do not overlap, which,

if identification were strong, would suggest time variation in the coefficients of the NKPC, as suggested by Zhang, Osborn, and Kim (2008). However, the  $S$  sets do overlap considerably over the two subsamples, so it is not clear whether the survey-based NKPC is unstable.

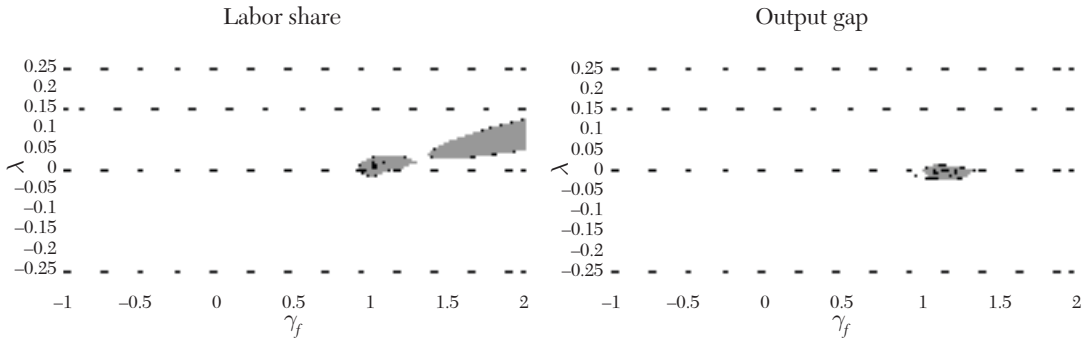


Figure 12. Robust Confidence Regions: VAR Specifications

Notes: 90 percent  $S$  set (gray), 90 percent Wald ellipse, and CUE VAR-GMM point estimate (bullet) of the coefficients of the labor share and future inflation in the hybrid NKPC specification with one lag of inflation, where inflation coefficients sum to one. Inflation: GDP deflator. Forcing variable: NFB labor share (left panel), CBO output gap (right panel). Instruments: three lags of  $\Delta\pi_t$  and the forcing variable, implying VAR(3) reduced form. Sample: starts 1948q2 (labor share), 1949q4 (output gap), ends 2011q3. Weight matrix: White (1980) heteroskedasticity-consistent.

To assess the empirical success of the external instruments approach, we plot robust confidence sets for GDP deflator inflation using the real-time instrument set in figure 14. These figures are comparable to the top row in figure 11, although the sample starts in 1971 due to data availability. Figure 14 demonstrates the unfortunate fact that the most plausibly exogenous instrument set also results in very weak identification, as the 90 percent robust confidence sets contain all reasonable  $\gamma_f$  values. Post-1984 confidence sets (not reported) are even larger.

#### 5.4 Nesting Rational Expectations and Survey Expectations

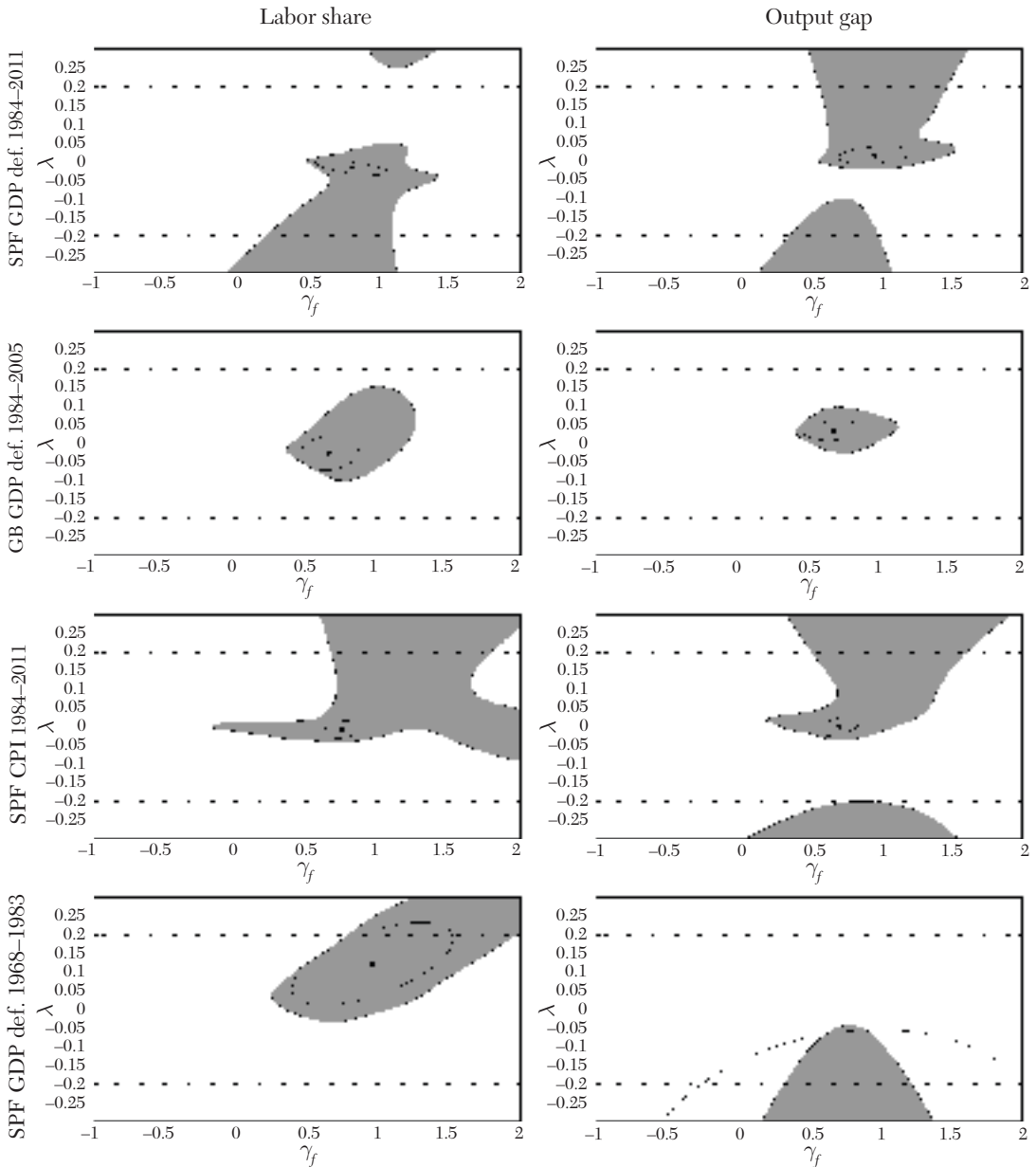
Finally, we assess the relative importance of rational and survey expectations in the NKPC, as studied by Nunes (2010), Fuhrer and Olivei (2010), and Fuhrer (2012); cf. section 3.1. Figure 15 reports CUE estimates and 90 percent  $S$  sets for the coefficients of future inflation rational expectations and

time- $t$  dated one-quarter-ahead forecasts of inflation in the model

$$(25) \quad \pi_t = \lambda x_t + \gamma_{RE} \pi_{t+1} + \gamma_s \pi_{t+1|t} + \gamma_b \pi_{t-1} + \tilde{u}_t.$$

The coefficient  $\lambda$  here is treated as well-identified, and it is concentrated out.<sup>50</sup> We consider both SPF and Greenbook GDP deflator inflation forecasts over the full available samples as well as a sample that starts in 1984q1. The instrument set is the same as in Nunes (2010), i.e., GGLS (see table 4), plus two lags of survey inflation forecasts. The point estimates generally indicate a dominant role for rational expectations, consistent with the evidence in Nunes (2010), and different from the preferred estimates

<sup>50</sup>This is a reasonable assumption since in all cases the first-stage  $F_{HAR}$  statistic for  $x_t$  is over 100.



*Figure 13.* Robust Confidence Regions: Survey Forecast Specifications

*Notes:* 90 percent S set (gray), 90 percent Wald ellipse, and CUE GMM point estimate (bullet) of the coefficients of the labor share and future inflation in the hybrid NKPC specification with one lag of inflation, where inflation coefficients sum to one. Forcing variable: NFB labor share (left panels), CBO output gap (right panels). Instruments: three lags of  $\Delta\pi_t$  and the forcing variable. Specifications: SPF GDP deflator 1984q1-2011q4 (top row), GB GDP deflator 1984q1-2005q4 (second row), SPF CPI 1984q1-2011q4 (third row), SPF GDP deflator 1968q3-1983q4 (bottom row), all time- $t$  dated and endogenous. Weight matrix: Newey–West with automatic lag truncation.

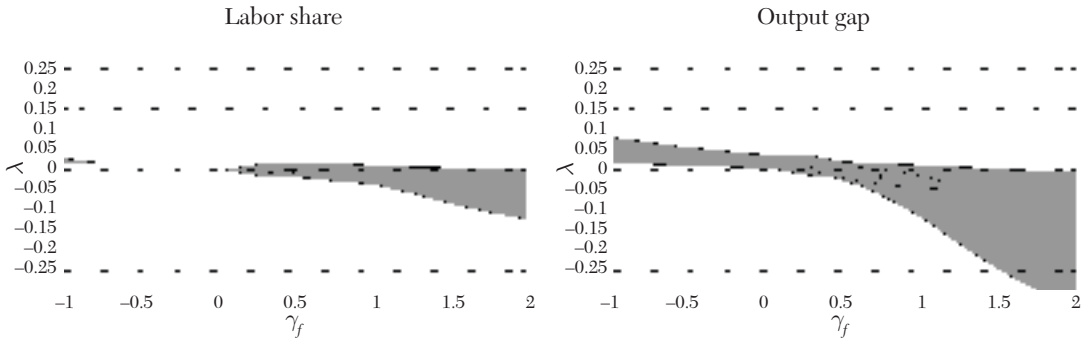


Figure 14. Robust Confidence Regions: Real-Time Instruments

Notes: 90 percent  $S$  set (gray), 90 percent Wald ellipse, and CUE GMM point estimate (bullet) of the coefficients of the labor share and future inflation in the hybrid NKPC specification with one lag of inflation, where inflation coefficients sum to one. Inflation: GDP deflator. Forcing variable: NFB labor share (left panel), CBO output gap (right panel). Instruments: two lags of GDP deflator inflation, the output gap, and the change in the labor share, all measured in real time. Sample: 1971q1-2011q2. Weight matrix: Newey–West with automatic lag truncation.

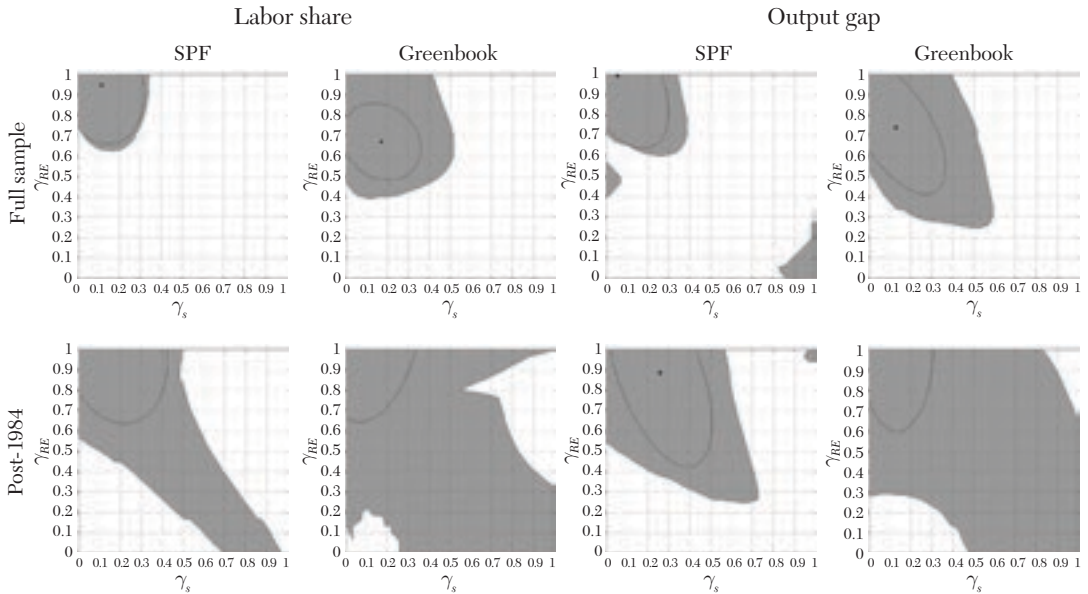


Figure 15. Robust Confidence Regions: Specifications with Both Rational and Survey Expectations

Notes: 90 percent  $S$  set (gray), 90 percent Wald ellipse, and CUE GMM point estimate (bullet) of the coefficients of future inflation ( $\gamma_{RE}$ , vertical axis) and single-quarter forecasts ( $\gamma_s$ , horizontal axis) in the nesting NKPC specification (25). Inflation: GDP deflator. Forcing variable: NFB labor share (left two columns), CBO output gap (right two columns). Instruments: GGLS plus two lags of survey forecasts. Sample: full available (top row), 1984q1–end of sample (bottom row). Survey forecasts: SPF (first and third columns), Greenbook (second and fourth columns), all time- $t$  dated and endogenous. Weight matrix: Newey–West with automatic lag truncation.



in Fuhrer and Olivei (2010) and Fuhrer (2012). However, as acknowledged by Nunes (2010), sampling uncertainty is very large, and there is considerable sensitivity to data and estimation sample. Only when we use the labor share as forcing variable and the full available sample can we conclude that the rational expectations term is dominant. Interestingly, all the confidence sets exclude  $\gamma_{RE} = \gamma_s = 0$ .<sup>51</sup>

## 6. Conclusion

Based on the foregoing comparison of more than one hundred papers from the literature with our analysis of thousands of a priori reasonable NKPC specifications estimated on U.S. data, we reach six main conclusions.

First, estimation of the NKPC using macro data is subject to a severe weak instruments problem. Consequently, seemingly innocuous specification changes lead to big differences in point estimates. The specification sensitivity is even larger than what has been reported in the literature. Moreover, given a choice of specification, sampling uncertainty is typically large, as weak identification robust confidence sets often cover a substantial part of the parameter space. While these findings are purely empirical, there are good theoretical explanations for why identification of the NKPC is weak.

Second, we do not reject the NKPC—far from it. However, we are unable to pin down the role of expectations in the inflation process sufficiently accurately for the results to be useful for policy analysis. The evidence is consistent both with the view that

expectations matter a lot, as well as with the opposite view that they matter very little.

Third, because standard inference methods and efficiency comparisons are unreliable, weak identification robust methods should be used when possible. Weak identification is not a GMM-specific problem.

Fourth, estimation methods that rely on the assumption that inflation expectations can be proxied by a reduced-form VAR typically point toward a much greater role for forward-looking expectations in price determination than do less restrictive estimators. We demonstrate that VAR-based inference can be spurious when identification is weak. Because the VAR assumption is not innocuous, we recommend that VAR estimates be compared to non-VAR estimates when possible.

Fifth, it is hard to interpret the empirical results from specifications that use survey forecasts to proxy for inflation expectations. They often appear to be more strongly identified than other types of specifications, but they are particularly sensitive to the choice of forecast source, sample, and inflation series. Moreover, the survey forecast specification of the NKPC is not microfounded unless the forecasts are rational, which does not seem to hold empirically. It is an interesting topic for future research to develop an internally consistent framework for analyzing inflation dynamics under nonrational expectation formation.

Sixth, researchers should be aware of the large and frequent revisions to NKPC data. We have proposed an estimation method that uses revisions as external instruments. While its assumptions are appealingly unrestrictive, it does not yield informative empirical results.

The evidence we present in this paper leads us to conclude that identification of the NKPC is too weak to warrant research on conceptually minor extensions. Issues related to the choice of explanatory variables,

<sup>51</sup> A number of authors have argued that nonrational expectations can account for the intrinsic persistence in the NKPC (Roberts 1997; Brissimis and Magginas 2008; Nunes 2010). Estimation results that impose  $\gamma_b = 0$  (not reported) are quite similar to the ones reported.

instruments, alternate data constructions, and small modifications of the model are likely to be dwarfed by identification problems. Instead, we think it will be more fruitful to explore fundamentally new sources of identification, such as micro/sectoral data, cross-country models, information from large datasets, and stability restrictions. Some recent papers have taken up this challenge, and we hope more will follow. The onus is not purely on applied researchers; theoretical macroeconomists can help by developing models that can be taken to the data in ways that directly address the identification issue.

## Appendix

### A.1 Calibration of Impulse Responses

The model used to generate the impulse responses in figure 1 is based on the canonical three-equation New Keynesian framework as described by Galí (2008):

$$\begin{aligned}
 (26) \quad \pi_t &= \gamma_f E_t(\pi_{t+1}) + (1 - \gamma_f)\pi_{t-1} \\
 &+ \lambda mc_t, \\
 x_t &= -\frac{1}{\sigma}(i_t - E_t(\pi_{t+1}) - \rho) + E_t(x_{t+1}), \\
 mc_t &= \left(\sigma + \frac{\phi + \alpha}{1 - \alpha}\right)x_t, \\
 i_t &= \rho + \varphi_\pi \pi_t + \varphi_x x_t + v_t^i, \\
 v_t^i &= \rho_v v_{t-1}^i + \varepsilon_t^v.
 \end{aligned}$$

The first equation is a hybrid NKPC, the second is the dynamic IS curve, the third relates log real marginal cost (in deviation from the zero inflation steady state)  $mc_t$  to the log output gap  $x_t$ , the fourth is a Taylor rule for the nominal interest rate  $i_t$ , and the fifth equation specifies that the Taylor rule

disturbance  $v_t^i$  follows an AR(1) process. We call  $\varepsilon_t^v$  the monetary policy shock.

When calibrating the structural parameters of the model, we use the benchmark values on page 52 of Galí (2008). The rate of time preference is  $\rho = -\log(0.99)$ , the elasticity of intertemporal substitution is  $1/\sigma = 1$ , the Frisch elasticity of labor supply is  $\phi = 1$ , the labor exponent in the production function is  $1 - \alpha = 2/3$ , the Taylor rule coefficients are  $(\varphi_\pi, \varphi_x) = (1.5, 0.5/4)$ , and the AR(1) coefficient for the Taylor rule disturbance is  $\rho_v = 0.5$ .

### A.2 Econometric Methods

#### A.2.1 GMM Estimation and Optimal Instruments

GMM estimation can be briefly described as follows. Let  $f_T(\vartheta)$  denote sample moments, whose expectation vanishes at the true value of the parameters. For example, for the moment conditions (13) we set  $f_T(\vartheta) = T^{-1} \sum_{t=1}^T Z_t h_t(\vartheta)$ . Define the GMM objective function

$$(27) \quad S_T(\vartheta, \bar{\vartheta}) = f_T(\vartheta)' W_T(\bar{\vartheta}) f_T(\vartheta),$$

where  $\bar{\vartheta}$  is some preliminary estimator of  $\vartheta$ , and  $W_T$  is a weighting matrix that may depend on the data and on  $\vartheta$ . A GMM estimator is the minimizer of  $S_T(\vartheta, \bar{\vartheta})$  with respect to  $\vartheta$ , if it exists. Given the particular choice of moments  $f_T(\vartheta)$ , efficient GMM estimation requires  $W_T(\bar{\vartheta})$  to be a consistent estimator of the inverse of the variance of  $\sqrt{T}f_T(\vartheta)$ —the long-run variance of the moment conditions. The most commonly used GMM estimator is a 2-step estimator, where the preliminary estimator  $\bar{\vartheta}$  is obtained using some weight matrix that does not depend on  $\vartheta$ . When the moment conditions are linear,  $\bar{\vartheta}$  may be obtained using two-stage least squares.

Setting  $\bar{\vartheta} = \vartheta$ , so the efficient weight matrix estimator  $W_T(\vartheta)$  is evaluated at the same parameters as the sample moments  $f_T(\vartheta)$ , yields the CUE, which was proposed by Hansen, Heaton, and Yaron (1996). Two-step GMM and CUE are asymptotically equivalent under strong identification, but the latter has certain advantages under weak identification (see, e.g., Stock, Wright, and Yogo 2002).

### A.2.1.1 Optimal Instruments

When identification is given by conditional moment restrictions of the form  $E_{t-1}[h_t(\vartheta_0)] = 0$ , where  $h_t(\cdot)$  is an  $s \times 1$  vector-valued function, there is an infinite number of predetermined variables  $Z_t$  that can be used as instruments to form unconditional moment restrictions  $E[Z_t h_t(\vartheta_0)] = 0$ . Efficiency (under strong identification) in the class of all GMM estimators amounts to choosing the instruments in a way that minimizes the asymptotic variance of the GMM estimator among all possible instruments  $Z_t \in I_{t-1}$ , where  $I_{t-1}$  denotes the information set at time  $t - 1$ . If the residual function  $h_t(\vartheta_0)$  is a martingale difference sequence (MDS), the optimal instruments are given by

$$(28) \quad Z_t^o = \left[ E_{t-1} \left( \frac{\partial h_t(\vartheta_0)}{\partial \vartheta'} \right) \right]' \\ \times \{ E_{t-1} [h_t(\vartheta_0) h_t(\vartheta_0)'] \}^{-1},$$

see Chamberlain (1987).

When estimating the NKPC by VAR-GMM, the two-dimensional residual vector (16) satisfies the conditional moment restriction  $E_{t-1} \widehat{h}_t(\vartheta_0) = 0$ . The residual vector is an MDS because it is adapted to the information set at time  $t$ . Moreover, the VAR assumption implies that  $E_{t-1}[\partial \widehat{h}_t(\vartheta_0) / \partial \vartheta']$  is spanned by  $Y_{t-1}$ . So, applying the formula

for the optimal instruments (28), we see that under conditional homoskedasticity, the optimal instruments are spanned by  $Y_{t-1}$ .

In the case of GIV estimation of the NKPC, the residuals are not adapted to  $I_t$  since  $h_t(\vartheta_0) = \tilde{u}_t \in I_{t+1}$ , see equation (12). Under the assumption  $E_{t-1}(u_t) = 0$ ,  $h_t(\vartheta_0)$  can be represented as a moving average of order 1, e.g.,  $h_t(\vartheta_0) = v_t - \phi v_{t+1} = \phi(L^{-1}) v_t$ , say, where  $v_t$  is an MDS with  $E_{t-1}(v_t) = 0$ . Following Hayashi and Sims (1983), the optimal instruments can be obtained as follows. First, forward-filter  $h_t(\vartheta_0)$  to get  $v_t = \phi(L^{-1})^{-1} h_t(\vartheta_0)$ . Then compute the optimal instruments  $Z_t^o$  for  $E_{t-1}(v_t) = 0$  using (28). Finally, transform these instruments to the optimal instruments  $\tilde{Z}_t^o$  for  $E_{t-1}[h_t(\vartheta_0)] = 0$ , which are given by  $\tilde{Z}_t^o = \phi(L)^{-1} Z_t^o = \sum_{j=0}^{\infty} \phi^j Z_{t-j}^o$ .

### A.2.2 GIV Estimation with Iterated Instruments

Rudd and Whelan (2005) suggested the following alternative to the Galí and Gertler (1999) approach. Iterating equation (8)  $q$  periods forward using  $E_t(u_{t+j}) = 0$ ,  $j > 0$ , and the law of iterated expectations, we get

$$(29) \quad \pi_t = \beta^{q+1} E_t(\pi_{t+q+1}) \\ + \lambda \sum_{j=0}^q \beta^j E_t(x_{t+j}) + u_t.$$

Rudd and Whelan (2005) use the GIV approach to estimate the above relation. We now point out how the iterated method relates to the previously described Galí and Gertler (1999) procedure. Using the definition of the residual  $h_t(\vartheta)$  in (12), equation (29) can be equivalently written as

$$E_t \left[ \sum_{j=0}^q \beta^j h_{t+j}(\vartheta) \right] = u_t.$$

The identifying restriction  $E_{t-1}(u_t) = 0$  then implies the unconditional moment restrictions

$$(30) \quad E \left[ \sum_{j=0}^q \beta^j Y_{t-1} h_{t+j}(\vartheta) \right] = 0,$$

where  $Y_{t-1}$  is a vector of lags of  $\pi_t$ ,  $x_t$ , and any other variables used in the analysis. If we further assume that the distribution of the data is stationary, so that  $E[Y_{t-1} h_{t+j}(\vartheta)] = E[Y_{t-j-1} h_t(\vartheta)]$ , then equation (30) is equivalent to

$$(31) \quad E \left[ \left( \sum_{j=0}^q \beta^j Y_{t-j-1} \right) h_t(\vartheta) \right] = 0.$$

This makes it clear that the only difference between the iterated moment conditions (30) and the difference equation moment conditions (13) is in the choice of instruments. That is, the underlying identifying assumption  $E_{t-1}(u_t) = 0$  is the same, but each method uses a different subset of all admissible instruments.

### A.2.3 Alternative VAR Estimators

#### A.2.3.1 VAR-MD

This approach was introduced by Campbell and Shiller (1987) for the estimation of asset pricing models and was popularized in the NKPC literature by the work of Sbordone (2002, 2005, and 2006). It can be described briefly as follows. The structural model (8) implies restrictions on the reduced-form VAR coefficients  $A$  in (14). These restrictions can be written as  $g(A, \vartheta) = 0$ , where  $g$  is a vector-valued “distance” function. Typically, the number of restrictions exceeds the number of structural parameters, so the MD estimator is defined as the minimizer of the objective function

$$g(\hat{A}, \vartheta)' W g(\hat{A}, \vartheta),$$

where  $\hat{A}$  is a consistent first-step estimator of the reduced-form parameters (such as the OLS estimator), and  $W$  is a possibly random weight matrix. The optimal choice of  $W$  is a consistent estimator of the inverse of the asymptotic variance of  $g(\hat{A}, \vartheta)$ .

Define  $e_\pi$  and  $e_x$  to be the unit vectors with 1 in the position of  $\pi_t$  and  $x_t$  in  $Y_t$ , respectively. If we take time- $(t-1)$  expectations on both sides of the difference equation specification (8) and use the VAR implication  $E_t(Y_{t+1}) = AY_t$ , we obtain the parameter restrictions

$$(32) \quad g_1(A, \vartheta)' \equiv e'_\pi A - \beta e'_\pi A^2 - \lambda e'_x A = 0.$$

Note that  $A = E(Y_t Y'_{t-1}) E(Y_{t-1} Y'_{t-1})^{-1} = E(Y_{t+1} Y'_t) E(Y_t Y'_t)^{-1}$ , so that (32) can be equivalently written as

$$(33) \quad E[Y_{t-1}(\pi_t - \beta Y'_t \zeta - \lambda x_t)] = 0,$$

where  $\zeta = A'e_\pi$  are the coefficients of the projection of  $\pi_{t+1}$  on  $Y_t$ . These are exactly the moment conditions for VAR-GMM given in (15).<sup>52</sup>

The VAR-MD distance function is not unique. If we iterate the pure NKPC (8) forward an infinite number of times, we obtain the so-called “closed-form” solution

$$(34) \quad \pi_t = \lambda \sum_{j=0}^{\infty} \beta^j E_t(x_{t+j}) + u_t,$$

provided the series converges and the terminal condition  $\lim_{\tau \rightarrow \infty} E_t(\beta^\tau \pi_{t+\tau}) = 0$  holds. Using  $E_t(Y_{t+j}) = A^j Y_t$ , we can write (34) as

$$\pi_t = \lambda(I - \beta A)^{-1} x_t + u_t.$$

<sup>52</sup>The VAR-GMM estimator reduces to the VAR-MD estimator for a particular (inefficient) choice of block diagonal GMM weight matrix. Thus, VAR-MD can be viewed as a variant of VAR-GMM, which imposes that the OLS moment conditions  $E[(Y_t - AY_{t-1})Y'_{t-1}] = 0$  for the VAR companion matrix  $A$  hold exactly.

The assumption  $E_{t-1}(u_t) = 0$  implies the restrictions

$$(35) \quad g_2(A, \vartheta)' = e'_\pi A - \lambda e'_x A(I - \beta A)^{-1} \\ = 0.$$

The distance function  $g_1(A, \vartheta)$  defined in (32) satisfies  $g_1(A, \vartheta) = (I - \beta A)'g_2(A, \vartheta)$ . Because  $(I - \beta A)$  is nonsingular, the two sets of restrictions are equivalent. However, the MD estimator is not invariant to nonlinear transformations of the distance function, so in finite samples the choice of distance function matters.<sup>53</sup>

#### A.2.3.2 VAR-ML

Suppose

$$(36) \quad z_t = \sum_{j=1}^l A_j z_{t-j} + v_t$$

denotes the  $l$ -th order VAR of  $z_t$ , whose companion representation was given in equation (14) above, where  $A_j$  are  $n \times n$  coefficient matrices, and  $v_t$  is an  $n \times 1$  vector of reduced-form errors. We have omitted deterministic terms for simplicity. An alternative to MD estimation is to maximize the likelihood function of the finite-order VAR (36) subject to the cross-equation restrictions (32) implied by the structural NKPC (8).

ML estimation of the constrained VAR is typically implemented by solving out the equality constraints (32) to express some of the reduced-form parameters in the likelihood in terms of the structural parameters,  $\vartheta$ , and the remaining reduced-form parameters. Denoting the latter as  $\psi$ , the restricted reduced-form coefficients can be expressed

as  $A_j(\vartheta, \psi)$ ,  $j = 1, \dots, l$ . Assume, as in most of the literature, that the VAR errors are i.i.d. Gaussian and homoskedastic, i.e.,  $v_t \sim N(0, \Omega)$ , where  $\Omega$  is an  $l \times l$  positive definite variance matrix. After concentrating with respect to  $\Omega$ , the log-likelihood function can be written as

$$(37) \quad L(\vartheta, \psi) = \text{constant} \\ - \frac{T}{2} \log \det [\hat{\Omega}(\vartheta, \psi)],$$

where  $\hat{\Omega}(\vartheta, \psi) = T^{-1} \sum_{t=l+1}^T \hat{v}_t(\vartheta, \psi) \hat{v}_t(\vartheta, \psi)'$  and  $\hat{v}_t(\vartheta, \psi) = z_t - \sum_{j=1}^l A_j(\vartheta, \psi) z_{t-j}$ . The choice of  $\psi$  is not unique, i.e., there are several ways of imposing the restrictions (32) on the likelihood. The pioneering approach by Fuhrer and Moore (1995) chooses  $\psi$  to be all the coefficients in the VAR except those corresponding to the equation for inflation. Computation of  $A_j(\vartheta, \psi)$  then requires solving for the reduced-form coefficients in the inflation equation as functions of all other structural and reduced-form parameters. There are generically multiple solutions to this problem, so this mapping is not unique, and evaluating the likelihood (37) at all of the possible VAR solutions can be impractical; see Kurmann (2007). Fuhrer and Moore (1995) circumvent this issue by restricting the parameter space to the *determinacy region*, which by definition contains the parameter combinations for which there is a unique stable VAR solution.

An alternative approach, proposed by Kurmann (2007), is to set  $\psi$  equal to all the reduced-form VAR coefficients except those corresponding to the equation for the forcing variable  $x_t$ . He shows that the mapping  $A_j(\vartheta, \psi)$  is then unique, except on a set of measure zero, and so evaluation of the likelihood (37) is straightforward on the entire parameter space, also outside the determinacy region. Inside the determinacy region the

<sup>53</sup>This was briefly discussed in Sbordone (2005), and in more detail in Barnes et al. (2011). An exception occurs when the model is just-identified and the equations  $g_1(\hat{A}, \vartheta) = 0$  and  $g_2(\hat{A}, \vartheta) = 0$  can be solved for  $\vartheta$  as a function of  $\hat{A}$ , in which case the VAR-MD estimator does not depend on the choice of distance function.



method gives the same likelihood as the Fuhrer–Moore approach. The following example from Kurmann (2007) illustrates. Suppose the reduced form is a VAR(1) in  $(\pi_t, x_t)'$ :

$$(38) \quad \begin{aligned} \pi_t &= a_{\pi\pi} \pi_{t-1} + a_{\pi x} x_{t-1} + v_{\pi t}, \\ x_t &= a_{x\pi} \pi_{t-1} + a_{xx} x_{t-1} + v_{xt}, \end{aligned}$$

where  $v_{\pi t}$  and  $v_{xt}$  are i.i.d. reduced-form shocks. The restrictions (32) can be expressed as

$$\begin{aligned} a_{x\pi}(\lambda + \beta a_{\pi\pi}) &= (1 - \beta a_{\pi\pi}) a_{\pi\pi}, \\ a_{xx}(\lambda + \beta a_{\pi\pi}) &= (1 - \beta a_{\pi\pi}) a_{\pi x}. \end{aligned}$$

If we solve these equations for  $a_{\pi\pi}, a_{\pi x}$  as functions of  $\vartheta = (\beta, \lambda)'$  and  $\psi = (a_{x\pi}, a_{xx})'$ , then it can be shown that there are generically three solutions; see Kurmann (2007, sec. 2).<sup>54</sup> If we instead set  $\psi = (a_{\pi\pi}, a_{\pi x})'$  and solve for the reduced-form parameters  $a_{x\pi}, a_{xx}$ , then there is a unique solution unless  $\lambda + \beta a_{\pi\pi} = 0$ .

### A.2.3.3 Relationship between VAR Methods

Let  $\hat{A}$  be the OLS estimator of the VAR coefficients. VAR-ML can be thought of as minimizing the distance between  $\hat{A}$  and  $A(\vartheta, \psi)$  with respect to  $\vartheta$  and  $\psi$ . VAR-MD instead sets  $\psi$  equal to its OLS estimator and only minimizes this distance with respect to  $\vartheta$ . Thus, the relationship between VAR-MD and VAR-ML is analogous to the relationship between 2SLS and limited information maximum likelihood (LIML), respectively, in the textbook linear IV model (Fukač and Pagan 2010). The analogy suggests that VAR-ML and VAR-MD should be asymptotically equivalent under strong identification, but not so under weak identification.

<sup>54</sup>The multiplicity of solutions increases with the dimension of the VAR.

Moreover, computation of VAR-MD (like 2SLS) is easier than VAR-ML (like LIML). Another difference is that VAR-ML is invariant to nonlinear transformations, i.e., it gives the same results in finite samples whether we specify the model as a difference equation (8) or in closed form (34).

An advantage of the VAR-GMM estimator relative to VAR-MD and VAR-ML is that it is easy to add zero restrictions to the coefficients of the reduced-form VAR so as to avoid many instrument issues. For instance, if you want to use four lags of inflation but only two lags of  $x_t$  and other variables in the VAR, as in Galí, Gertler, and López-Salido (2001), you just need to include only those variables in  $Y_{t-1}$  in the moment conditions (15)–(16). Hence, it is straightforward to check the implications of imposing the VAR assumption given any choice of instruments, as we do in our empirical section.

### A.2.4 External Instruments

Consider a generalized version of the model (9) that does not place any exclusion restrictions on the lags (we assume  $w_t$  in equation (9) is part of  $Y_{t-1}$ ):

$$(39) \quad \pi_t = \lambda x_t + \gamma_f \pi_{t+1}^e + \delta' Y_{t-1} + u_t.$$

We use  $\pi_{t+1}^e$  to denote inflation expectations so as to allow for the possibility that these may not be rational. Define  $Y_t^s$  to be the vintage- $s$  observation, i.e., the statistical agency's estimate of  $Y_t$  published at time  $s$ . Variables without superscripts denote the latent true values of the series. Rearrange (39):

$$\begin{aligned} \pi_t &= \lambda x_t + \gamma_f \pi_{t+1} + \delta' Y_{t-1} + \tilde{u}_t, \\ \tilde{u}_t &= u_t + \gamma_f (\pi_{t+1}^e - \pi_{t+1}). \end{aligned}$$

Suppose that

$$(40) \quad E(u_t | Y_{t-1}, \{Y_{t-1}^r, \text{all } r\}) = 0.$$



This assumption can be interpreted as saying that the only way that data revisions enter the model is through their use in forming expectations. For  $Y_{t-1}^r$  to be valid instruments it must be the case that

$$(41) \quad E(\tilde{u}_t | Y_{t-1}, Y_{t-1}^r) = 0.$$

If  $\gamma_f = 0$ , then  $\tilde{u}_t = u_t$  and data revisions are exogenous by (40). Whether they are relevant is an empirical issue, and depends on the extent to which expectations are formed using published data.

If instead  $\gamma_f \neq 0$ , things are more complicated. Let  $\pi_{t+1}^*$  denote the rational expectation of  $\pi_{t+1}$ , and suppose that

$$(42) \quad \pi_{t+1}^e = \pi_{t+1}^* + \zeta_t,$$

where  $E(\zeta_t | Y_{t-1}^{t-1}) = 0$ , and

$$(43) \quad E(\pi_{t+1} - \pi_{t+1}^* | Y_{t-1}^{t-1}) = 0.$$

Condition (42) holds if agents have rational expectations, in which case  $\zeta_t = 0$ , but it also holds under departures from rational expectations, in which case  $\zeta_t$  is some “opinion” that is orthogonal to observable vintage- $(t-1)$  data. Condition (43) says that the information set used to compute  $\pi_{t+1}^*$  contains  $Y_{t-1}^{t-1}$ . Under these conditions, it can be shown that  $E(\tilde{u}_t | Y_{t-1}^{t-1}) = 0$ . In other words,  $Y_{t-1}^{t-1}$  is an exogenous instrument in the model (39). But this treats  $Y_{t-1}$  as endogenous, which leaves the model underidentified (there are two more endogenous variables,  $\pi_{t+1}$  and  $x_t$ , than instruments). This identification problem can be “solved” by imposing some exclusion restrictions on elements of  $Y_{t-1}$ , though this goes against the idea of external instruments. Alternatively, if we replace  $Y_{t-1}^{t-1}$  in conditions (42) and (43) with  $\{Y_{t-1}^r\}_{r \leq t-1}$ , then we could use several vintages of  $Y_{t-1}$  as instruments. This would satisfy the order condition for identification, but those instruments are likely to be weak.

### A.2.5 S Test

The  $S$  statistic for testing the null hypothesis  $H_0: \vartheta = \vartheta_0$  is given by  $T$  times the value of the continuous updated GMM objective function (27) at  $\vartheta_0$ , i.e.,  $TS_T(\vartheta_0, \vartheta_0)$ . Under  $H_0$  and some regularity conditions, this statistic is asymptotically  $\chi^2(k)$  with degrees of freedom equal to the number of moment restrictions (or instruments), irrespective of whether the model is identified or not. A  $(1-a)$  percent level  $S$  set is obtained by collecting all points  $\vartheta_0$  for which  $S_T(\vartheta_0, \vartheta_0)$  does not exceed the  $(1-a)$  percentile of  $\chi^2(k)$ . When the model also contains exogenous and predetermined variables, e.g.,  $w_t$  and  $\pi_{t-j}$  in (9), their coefficients are concentrated out in order to improve the power of the test; see theorem 3, Stock and Wright (2000).

### A.2.6 Weak Identification Robust Hansen Test

The minimum value of the  $S$  statistic,  $\min_{\vartheta} TS_T(\vartheta, \vartheta)$ , coincides with Hansen’s  $J$  test of overidentifying restrictions that is based on the continuous updated GMM objective function; see Hansen, Heaton, and Yaron (1996). Its strong-instruments asymptotic distribution under the null of correct specification is the usual  $\chi^2(k-p)$ , where  $k$  is the number of identifying restrictions and  $p$  is the total number of estimated parameters. Under weak instruments, the asymptotic distribution of this statistic is bounded by  $\chi^2(k-q)$ , where  $q$  is the number of coefficients on exogenous regressors (cf. Stock and Wright 2000, theorem 3). Hence, since  $q < p$ , a robust version of the test can be obtained using the larger critical values associated with quantiles of  $\chi^2(k-q)$ . The robust test is less powerful than the standard one, because it uses the same test statistic but larger critical values.

### A.3 Simulation Study

Here we give details on the simulations presented in section 3.3. The main parameter choices for our four DGPs are listed in table 1. All DGPs set the intercepts in the reduced-form VAR equal to 0. The innovations  $v_t = (v_{\pi_t}, v_{x_t})'$  are distributed i.i.d. Gaussian with mean zero and covariance matrix  $\Omega = ((0.07, 0.03)', (0.03, 0.70)')$ , a typical reduced-form estimate on quarterly U.S. data from 1960–2011. The last column in table 1 lists the smallest eigenvalue of the population concentration matrix for the GIV specification. This is a measure of the strength of identification, for which higher values mean stronger identification. It can loosely be thought of as an analog of 2 times the smallest first-stage  $F$  statistic in homoskedastic linear IV (there are two endogenous regressors).

For DGPs 1a and 2a, the reduced-form coefficients  $\xi$  are set to values that are close to the OLS estimates on the above-mentioned sample, with  $x_t$  equal to the labor share. DGPs 1a–b are indeterminate, since none of the Blanchard and Kahn (1980) generalized eigenvalues are outside the unit circle. DGPs 2a–b are determinate.

The four estimators we consider are implemented as follows. GIV estimation uses efficient two-step linear GMM with instruments  $Y_{t-1}$  and the Newey and West (1987) HAC long-run variance estimator. VAR-GMM is based on efficient two-step GMM with heteroskedasticity robust weight matrix. Because the moment conditions (15) are not linear, we resort to numerical optimization, although we only have to optimize over the scalar parameter  $\gamma_f$ . VAR-MD uses a distance function of the difference equation type (32) and a two-step efficient procedure. The estimator is available in closed form. VAR-ML uses the Kurmann (2007) approach, described in subsection A.2.3, to optimize over the parameters  $(\gamma_f, \lambda, c, \zeta)$ .

This requires numerical optimization, which we carry out using Matlab's `fmincon` routine. The optimizer is provided with analytical first derivatives of the log likelihood, and we consider eight different initial values per estimation.

Matlab code and a full documentation of our approach are available in the online supplement (cf. footnote 38). The documentation also provides a more comprehensive set of results, including additional DGPs (some with vector autoregressive moving average reduced form such that the VAR assumption does not hold), the behavior of  $\lambda$  estimators, and rejection frequencies for  $t$  tests, overidentification tests, and the  $S$ -test.

### A.4 Data Description

Most series mentioned in table 4 are either self-explanatory or described in section 5.1. Unless otherwise noted, the series are from the St. Louis Fed FRED database. All growth rates are logarithmic and quarterly. Here we give details on some of the more involved constructions. Complete data and transformation files are available in the online supplement (cf. footnote 38).

Wage and commodity price inflation, which we use as instruments, refer to the growth in business sector compensation per hour and commodity producer price index inflation from the BLS, respectively (the latter series is not seasonally adjusted). Interest rates are U.S. Treasury rates.

All forcing variables are in logs. “Output” refers to real GDP per capita. We estimate trends by fitting linear or quadratic polynomials in time, or by the Hodrick–Prescott (HP) or Baxter–King filters. The Baxter–King filtered gaps retain cycles of duration between 6 and 32 quarters. For the HP filter, we use a smoothing parameter (commonly referred to as  $\lambda$ ) of 1,600 for output and 10,000 for the labor share. Our computation of *real-time* (one-sided) output gaps proceeds as

TABLE 9  
PAIRWISE COMPARISON OF POINT ESTIMATES: RE GIV VERSUS SURVEY FORECASTS

Parameter	Labor share		Output gap	
	$\lambda$	$\gamma_f$	$\lambda$	$\gamma_f$
SPF versus GIV: post-81	0.002	-0.200	0.042	0.092
SPF versus GIV: CPI post-81	-0.002	0.030	0.049	0.286
SPF versus GIV: GDP def. post-81	0.004	-0.326	0.038	-0.043
SPF versus GIV: GDP def. post-68	0.002	-0.200	0.042	0.092
GB versus GIV: post-68	0.039	-0.093	0.047	-0.000
Endogenous versus exogenous	-0.000	-0.000	-0.001	-0.045
GB versus SPF	0.003	-0.016	-0.007	-0.063
GB versus SPF: pre-84	-0.010	0.069	-0.017	0.007
GB versus SPF: post-84	0.009	-0.076	0.003	-0.103

*Notes:* Effect of using observed inflation forecasts (SPF or GB) to proxy for inflation expectations in the NKPC. Numbers are median pairwise differences in estimates across specifications that differ by one characteristic, keeping all other specification aspects constant. For example, “SPF versus GIV” is the median difference of coefficient estimates in SPF specifications from the corresponding RE GIV specifications.

follows. For every quarter, the series of real-time (i.e., then-current estimate of) output per capita is loaded. An AR(6) in changes is fitted to this series and used to generate forecasts several quarters ahead. The detrending routines are then applied to the concatenation of the real-time series and the generated forecasts. Pseudo-real-time labor share gaps are calculated somewhat differently from the output gaps. First, the data used is not actually real-time but is instead based on the latest vintage, as explained in the main text. Second, to better capture the marked decrease in the labor share in the latter half of the sample, the forecasting regression is an AR(15) in second differences.<sup>55</sup>

The real-time labor share dataset is gleaned from the BLS’s Productivity and Cost “Preliminary” news releases on nominal unit labor costs and the implicit price deflator. The

data corresponds approximately to what was known around the middle of each quarter, like in the Philadelphia Fed’s real-time dataset. Data vintages from 1971q2 to 1993q4 have been manually typed in from scanned PDFs of the BLS news releases, available in the St. Louis Fed’s FRASER document database. Vintages from 1994q1 to 2001q2 are parsed from electronic news releases available on the BLS website. Finally, vintages from 2001q3 and onward are parsed from vintages of the BLS internal “edit 60” flat text file.

#### A.5 Additional Empirical Results

Table 9 lists median differences in estimates of  $\lambda$  and  $\gamma_f$  across different survey specifications and over different subsamples. Table 10 reports a number of similar pairwise comparisons across other specification choices. The results are discussed in section 5.2.

<sup>55</sup>For the business sector log labor share, the Akaike information criterion selects 15 lags on the full sample.

TABLE 10  
 PAIRWISE COMPARISONS OF POINT ESTIMATES: OTHER SPECIFICATION CHOICES

Parameter	Labor share		Output gap	
	$\lambda$	$\gamma_f$	$\lambda$	$\gamma_f$
RT versus small instr.	-0.003	0.008	-0.003	-0.012
Survey versus small instr.	0.001	0.001	-0.001	-0.018
1 versus 0 lags	-0.001	-0.256	-0.000	-0.256
4 versus 1 lags	-0.004	-0.223	0.009	-0.277
Post- versus pre-84 sample	-0.008	-0.024	-0.002	0.000
GDP def. versus CPI inflation	0.001	0.011	0.004	0.018
Gap versus level of inflation	-0.001	-0.054	0.013	-0.072
Oil versus no controls	0.000	0.001	-0.001	-0.001
Treas. rate versus no controls	-0.000	0.005	0.001	0.000

Notes: Median pairwise differences in estimates across specifications that differ by one characteristic, keeping all other specification aspects constant. For example, row “RT versus small instr.” compares the specifications that use “small” and “RT” instrument sets; row “1 versus 0 lags” gives the median difference of estimates in the hybrid NKPC with 1 lag from the corresponding estimates in the pure NKPC.

 TABLE 11  
 SIZE OF ROBUST CONFIDENCE REGIONS

Confidence level	Spec.	Sample	Labor share		Output gap	
			90 percent	95 percent	90 percent	95 percent
	All	All	0.33	0.47	0.30	0.47
	0 lags	All	0.19	0.34	0.18	0.34
	1 lag	All	0.24	0.38	0.19	0.34
	4 lags	All	0.49	0.63	0.48	0.67
	All	Full	0.25	0.36	0.24	0.39
	All	Pre-1984	0.35	0.50	0.25	0.41
	All	1984q1–end of sample	0.41	0.56	0.41	0.59
	All RE	1984q1–2011q3	0.49	0.65	0.46	0.65
	all survey	1984q1–end of sample	0.35	0.50	0.38	0.55
	SPF only	1984q1–2011q4	0.37	0.53	0.44	0.62
	GB only	1984q1–2005q4	0.31	0.45	0.26	0.42
	GGLS instr.	1984q1–end of sample	0.33	0.48	0.43	0.60
	Small instr.	1984q1–end of sample	0.46	0.66	0.43	0.59
	Exact instr.	1984q1–end of sample	0.29	0.42	0.32	0.52
	RT instr.	1984q1–end of sample	0.64	0.78	0.54	0.73
	Survey instr.	1984q1–end of sample	0.32	0.47	0.35	0.52

Notes: Size of 90 percent and 95 percent  $S$  sets corresponding to the different specifications of the NKPC in table 8, as a fraction of the parameter space shown in figure 16 (using 1025 grid points). “RE” refers to all RE GIV and VAR-GMM specifications.

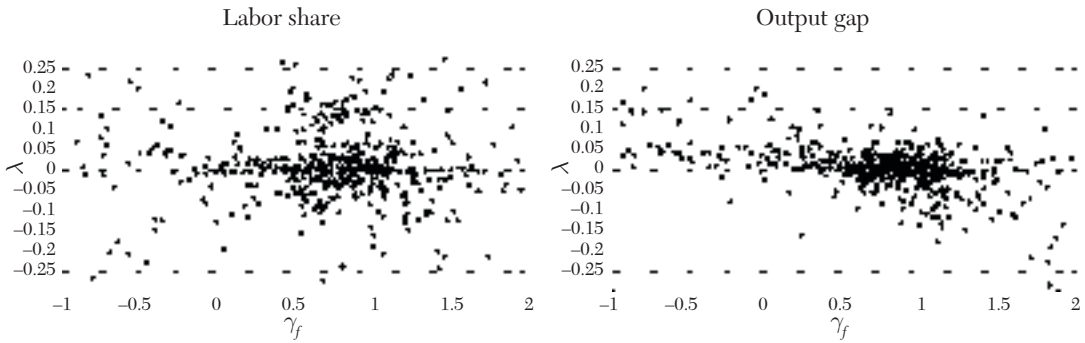


Figure 16. Point Estimates for Specifications in Table 8

Notes: CUE point estimates of the coefficient on the forcing variable and inflation expectations for the various specification settings listed in table 8. The left and right panels plot specifications with the labor share and output gap as forcing variable, respectively.

TABLE 12  
ANATOMY OF ROBUST CONFIDENCE REGIONS

Level	Labor share		Output gap	
	90 percent	95 percent	90 percent	95 percent
No. of specs	698	698	698	698
Fraction empty $S$ set	0.038	0.013	0.016	0.004
Fraction nonempty and positive	0.034	0.010	0.049	0.020
All RE	0.034	0.010	0.000	0.000
All survey	0.035	0.010	0.085	0.035
0 lags	0.072	0.014	0.094	0.036
1 lag	0.046	0.018	0.074	0.032
4 lags	0.004	0.000	0.000	0.000
GGLS instr.	0.062	0.021	0.042	0.000
Small instr.	0.035	0.014	0.035	0.021
Exact instr.	0.000	0.000	0.079	0.043
RT instr.	0.007	0.007	0.000	0.000
Survey instr.	0.069	0.008	0.092	0.038

Notes: Anatomy of 90 percent and 95 percent  $S$  sets for specifications listed in table 8.  $S$  sets are computed on the parameter space shown in figure 16 using 1,025 grid points. "Fraction empty" lists fraction of cases with empty  $S$  set. "Fraction nonempty and positive" lists fraction of cases for which  $S$  set is nonempty and lies entirely in positive orthant; rows "all RE" to "survey instr." give the latter fraction for different subcases. "RE" refers to all RE GIV and VAR-GMM specifications.

Figure 16 displays point estimates for all the specifications listed in table 8. For the collection of labor share specifications, the union of the joint 90 percent  $S$  sets (not shown) covers the entire plotted parameter space; the same is true for the collection of output gap specifications.

Table 11 reports the average size of 90 percent and 95 percent  $S$  sets as a fraction of the plotted parameter space  $(\gamma_f, \lambda) \in [-2, 1] \times [-0.3, 0.3]$  for various specification choices and samples. Here “all” refers to the different options listed in table 8. The sample end date varies by series: for Greenbook data the sample ends in 2005q4, while for all other series it stretches to 2011q4.

Table 12 reports some additional statistics associated with the  $S$  sets corresponding to the specifications of table 8. Row “percent empty  $S$  set” gives the fraction of the overidentified specifications for which the  $S$  sets are empty at the specified significance level. The rest of the rows give the frequency of nonempty and positive  $S$  sets for all specifications and for various subcategories.

#### REFERENCES

- Adam, Klaus, and Mario Padula. 2011. “Inflation Dynamics and Subjective Expectations in the United States.” *Economic Inquiry* 49 (1): 13–25.
- Adjemian, Stéphane, et al. 2011. “Dynare: Reference Manual Version 4.” Dynare Working Paper 1.
- An, Sungbae, and Frank Schorfheide. 2007. “Bayesian Analysis of DSGE Models.” *Econometric Reviews* 26 (2–4): 113–72.
- Anderson, Gary, and George Moore. 1985. “A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models.” *Economics Letters* 17 (3): 247–52.
- Anderson, T. W. and Herman Rubin. 1949. “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations.” *Annals of Mathematical Statistics* 20 (1): 46–63.
- Andrews, Donald W. K. 1993. “Tests for Parameter Instability and Structural Change with Unknown Change Point.” *Econometrica* 61 (4): 821–56.
- Andrews, Donald W. K., and Werner Ploberger. 1994. “Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative.” *Econometrica* 62 (6): 1383–1414.
- Andrews, Donald W. K., and James H. Stock. 2005. “Inference with Weak Instruments.” National Bureau of Economic Research Technical Working Paper 313.
- Andrews, Donald W. K., and James H. Stock. 2007. “Testing with Many Weak Instruments.” *Journal of Econometrics* 138 (1): 24–46.
- Andrews, Isaiah, and Anna Mikusheva. 2013. “Maximum Likelihood Inference in Weakly Identified DSGE Models.” Unpublished.
- Ang, Andrew, Geert Bekaert, and Min Wei. 2007. “Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?” *Journal of Monetary Economics* 54 (4): 1163–1212.
- Angeletos, George-Marios, and Jennifer La’O. 2009. “Incomplete Information, Higher-Order Beliefs and Price Inertia.” *Journal of Monetary Economics* 56 (Supplement): S19–37.
- Ascari, Guido. 2004. “Staggered Prices and Trend Inflation: Some Nuisances.” *Review of Economic Dynamics* 7 (3): 642–67.
- Ascari, Guido, and Argia M. Sbordone. 2013. “The Macroeconomics of Trend Inflation.” Unpublished.
- Atkeson, Andrew, and Lee E. Ohanian. 2001. “Are Phillips Curves Useful for Forecasting Inflation?” *Federal Reserve Bank of Minneapolis Quarterly Review* 25 (1): 2–11.
- Bårdsen, Gunnar, Eivlev S. Jansen, and Ragnar Nymoen. 2004. “Econometric Evaluation of the New Keynesian Phillips Curve.” *Oxford Bulletin of Economics and Statistics* 66: 671–86.
- Barnes, Michelle L., Fabia Gumbau-Brisa, Denny Lie, and Giovanni P. Olivei. 2011. “Estimation of Forward-Looking Relationships in Closed Form: An Application to the New Keynesian Phillips Curve.” Federal Reserve Bank of Boston Working Paper 11-3.
- Batini, Nicoletta, Brian Jackson, and Stephen Nickell. 2005. “An Open-Economy New Keynesian Phillips Curve for the U.K.” *Journal of Monetary Economics* 52 (6): 1061–71.
- Benigno, Pierpaolo, and J. David López-Salido. 2006. “Inflation Persistence and Optimal Monetary Policy in the Euro Area.” *Journal of Money, Credit, and Banking* 38 (3): 587–614.
- Bernanke, Ben S. 2007. “Inflation Expectations and Inflation Forecasting.” Speech at the Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute, July 10.
- Bernanke, Ben S. 2008. “Outstanding Issues in the Analysis of Inflation.” Speech at the Federal Reserve Bank of Boston’s 53rd Annual Economic Conference, June 9.
- Beyer, Andreas, Roger E. A. Farmer, Jerome Henry, and Massimiliano Marcellino. 2008. “Factor Analysis in a Model with Rational Expectations.” *Econometrics Journal* 11 (2): 271–86.
- Blanchard, Olivier, and Jordi Galí. 2007. “Real Wage Rigidities and the New Keynesian Model.” *Journal of Money, Credit, and Banking* 39: 35–65.
- Blanchard, Olivier Jean, and Jordi Galí. 2010. “Labor Markets and Monetary Policy: A New Keynesian



- Model with Unemployment." *American Economic Journal: Macroeconomics* 2 (2): 1–30.
- Blanchard, Olivier Jean, and Charles M. Kahn. 1980. "The Solution of Linear Difference Models under Rational Expectations." *Econometrica* 48 (5): 1305–11.
- Boug, Pal, Adne Cappelen, and Anders Rygh Swensen. 2010. "The New Keynesian Phillips Curve Revisited." *Journal of Economic Dynamics and Control* 34 (5): 858–74.
- Brissimis, Sophocles N., and Nicholas S. Magginas. 2008. "Inflation Forecasts and the New Keynesian Phillips Curve." *International Journal of Central Banking* 4 (2): 1–22.
- Cagliarini, Adam, Tim Robinson, and Allen Tran. 2011. "Reconciling Microeconomic and Macroeconomic Estimates of Price Stickiness." *Journal of Macroeconomics* 33 (1): 102–20.
- Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3): 383–98.
- Campbell, John Y., and Robert J. Shiller. 1987. "Cointegration and Tests of Present Value Models." *Journal of Political Economy* 95 (5): 1062–88.
- Campbell, John Y., and Robert J. Shiller. 1988. "The Dividend–Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies* 1 (3): 195–228.
- Caner, Mehmet. 2007. "Boundedly Pivotal Structural Change Tests in Continuous Updating GMM with Strong, Weak Identification and Completely Unidentified Cases." *Journal of Econometrics* 137 (1): 28–67.
- Carriero, Andrea. 2008. "A Simple Test of the New Keynesian Phillips Curve." *Economics Letters* 100 (2): 241–44.
- Carvalho, Carlos. 2006. "Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks." *B.E. Journal of Macroeconomics: Frontiers of Macroeconomics* 2 (1).
- Castle, Jennifer L., Jurgen A. Doornik, David F. Hendry, and Ragnar Nymoen. 2010. "Testing the Invariance of Expectations Models of Inflation." University of Oxford Department of Economics Discussion Paper 510.
- Chamberlain, Gary. 1987. "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions." *Journal of Econometrics* 34 (3): 305–34.
- Chowdhury, Ibrahim, Mathias Hoffmann, and Andreas Schabert. 2006. "Inflation Dynamics and the Cost Channel of Monetary Transmission." *European Economic Review* 50 (4): 995–1016.
- 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.
- Cochrane, John H. 2011. "Determinacy and Identification with Taylor Rules." *Journal of Political Economy* 119 (3): 565–615.
- Coenen, Gunter, Andrew T. Levin, and Kai Christoffel. 2007. "Identifying the Influences of Nominal and Real Rigidities in Aggregate Price-Setting Behavior." *Journal of Monetary Economics* 54 (8): 2439–66.
- Cogley, Timothy, and Argia M. Sbordone. 2008. "Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve." *American Economic Review* 98 (5): 2101–26.
- Cornea, Adriana, Cars Hommes, and Domenico Masaro. 2013. "Behavioral Heterogeneity in U.S. Inflation Dynamics." Unpublished.
- Dees, Stephane, M. Hashem Pesaran, L. Vanessa Smith, and Ron P. Smith. 2009. "Identification of New Keynesian Phillips Curves from a Global Perspective." *Journal of Money, Credit, and Banking* 41 (7): 1481–1502.
- Doornik, Jurgen. 2007. *Object-Oriented Matrix Programming Using Ox*, Third edition. London: Timberlake Consultants Press.
- Dotsey, Michael, Robert G. King, and Alexander L. Wolman. 1999. "State-Dependent Pricing and the General Equilibrium Dynamics of Money and Output." *Quarterly Journal of Economics* 114 (2): 655–90.
- Dufour, Jean-Marie. 2003. "Identification, Weak Instruments, and Statistical Inference in Econometrics." *Canadian Journal of Economics* 36 (4): 767–808.
- Dufour, Jean-Marie, Lynda Khalaf, and Maral Kichian. 2006. "Inflation Dynamics and the New Keynesian Phillips Curve: An Identification Robust Econometric Analysis." *Journal of Economic Dynamics and Control* 30 (9–10): 1707–27.
- Dufour, Jean-Marie, Lynda Khalaf, and Maral Kichian. 2010a. "Estimation Uncertainty in Structural Inflation Models with Real Wage Rigidities." *Computational Statistics and Data Analysis* 54 (11): 2554–61.
- Dufour, Jean-Marie, Lynda Khalaf, and Maral Kichian. 2010b. "On the Precision of Calvo Parameter Estimates in Structural NKPC Models." *Journal of Economic Dynamics and Control* 34 (9): 1582–95.
- Eichenbaum, Martin, and Jonas D. M. Fisher. 2004. "Evaluating the Calvo Model of Sticky Prices." National Bureau of Economic Research Working Paper 10617.
- Eichenbaum, Martin, and Jonas D. M. Fisher. 2007. "Estimating the Frequency of Price Re-optimization in Calvo-Style Models." *Journal of Monetary Economics* 54 (7): 2032–47.
- Estrella, Arturo, and Jeffrey C. Fuhrer. 2002. "Dynamic Inconsistencies: Counterfactual Implications of a Class of Rational-Expectations Models." *American Economic Review* 92 (4): 1013–28.
- Fanelli, Luca. 2008. "Testing the New Keynesian Phillips Curve through Vector Autoregressive Models: Results from the Euro Area." *Oxford Bulletin of Economics and Statistics* 70 (1): 53–66.
- Fernández-Villaverde, Jesus, Juan F. Rubio-Ramírez, Thomas J. Sargent, and Mark W. Watson. 2007. "ABCs (and Ds) of Understanding VARs." *American Economic Review* 97 (3): 1021–26.
- Fischer, Stanley. 1977. "Wage Indexation and Macroeconomic Stability." *Carnegie-Rochester Conference Series on Public Policy* 5: 107–47.

- Fisher, Irving. 1973. "I Discovered the Phillips Curve: A Statistical Relation between Unemployment and Price Changes." *Journal of Political Economy* 81 (2 Part 1): 496–502.
- Friedman, Milton. 1968. "The Role of Monetary Policy." *American Economic Review* 58: 1–17.
- Fuhrer, Jeffrey C. 1997. "The (Un)Importance of Forward-Looking Behavior in Price Specifications." *Journal of Money, Credit, and Banking* 29 (3): 338–50.
- Fuhrer, Jeffrey C. 2006. "Intrinsic and Inherited Inflation Persistence." *International Journal of Central Banking* 2 (3): 49–86.
- Fuhrer, Jeffrey C. 2012. "The Role of Expectations in Inflation Dynamics." *International Journal of Central Banking* 8: 137–65.
- Fuhrer, Jeffrey C., and George Moore. 1995. "Inflation Persistence." *Quarterly Journal of Economics* 110 (1): 127–59.
- Fuhrer, Jeffrey C., and Giovanni P. Olivei. 2005. "Estimating Forward-Looking Euler Equations with GMM and Maximum Likelihood Estimators: An Optimal-Instruments Approach." In *Models and Monetary Policy: Research in the Tradition of Dale Henderson, Richard Porter, and Peter Tinsley*. Board of Governors of the Federal Reserve System, 87–114. Washington, D.C.
- Fuhrer, Jeffrey C., and Giovanni P. Olivei. 2010. "The Role of Expectations and Output in the Inflation Process: An Empirical Assessment." Federal Reserve Bank of Boston Public Policy Brief 10-2.
- Fukač, Martin, and Adrian Pagan. 2007. "An Estimated DSGE Model for the United Kingdom: Commentary." *Federal Reserve Bank of St. Louis Review* 89 (4): 233–40.
- Fukač, Martin, and Adrian Pagan. 2010. "Limited Information Estimation and Evaluation of DSGE Models." *Journal of Applied Econometrics* 25 (1): 55–70.
- Gagnon, Edith, and Hashmat Khan. 2005. "New Phillips Curve under Alternative Production Technologies for Canada, the United States, and the Euro Area." *European Economic Review* 49 (6): 1571–1602.
- Galí, Jordi. 2008. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton and Oxford: Princeton University Press.
- Galí, Jordi, and Mark Gertler. 1999. "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44 (2): 195–222.
- Galí, Jordi, Mark Gertler, and J. David López-Salido. 2001. "European Inflation Dynamics." *European Economic Review* 45 (7): 1237–70.
- Galí, Jordi, Mark Gertler, and J. David López-Salido. 2005. "Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve." *Journal of Monetary Economics* 52 (6): 1107–18.
- Gerberding, Christina. 2001. "The Information Content of Survey Data on Expected Price Developments for Monetary Policy." Deutsche Bundesbank Research Centre Discussion Paper 09/01.
- Gordon, Robert J. 1990. "U.S. Inflation, Labor's Share, and the Natural Rate of Unemployment." In *Economics of Wage Determination*, edited by Heinz König, 1–34. New York: Springer-Verlag.
- Gordon, Robert J. 2011. "The History of the Phillips Curve: Consensus and Bifurcation." *Economica* 78 (309): 10–50.
- Guay, Alain, and Florian Pelgrin. 2005. "The U.S. New Keynesian Phillips Curve: An Empirical Assessment." Unpublished.
- Guerrieri, Luca, Christopher Gust, and J. David López-Salido. 2010. "International Competition and Inflation: A New Keynesian Perspective." *American Economic Journal: Macroeconomics* 2 (4): 247–80.
- Guggenberger, Patrik, and Richard J. Smith. 2008. "Generalized Empirical Likelihood Tests in Time Series Models with Potential Identification Failure." *Journal of Econometrics* 142 (1): 134–61.
- Gumbau-Brisa, Fabia, Denny Lie, and Giovanni P. Olivei. 2011. "A Response to Cogley and Sbordone's Comment on 'Closed-Form Estimates of the New Keynesian Phillips Curve with Time-Varying Trend Inflation'." Federal Reserve Bank of Boston Working Paper 11-4.
- Gwin, Carl R., and David D. VanHoose. 2008. "Alternative Measures of Marginal Cost and Inflation in Estimations of New Keynesian Inflation Dynamics." *Journal of Macroeconomics* 30 (3): 928–40.
- Hansen, Christian, Jerry Hausman, and Whitney Newey. 2008. "Estimation with Many Instrumental Variables." *Journal of Business and Economic Statistics* 26 (4): 398–422.
- Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (4): 1029–54.
- Hansen, Lars Peter, John Heaton, and Amir Yaron. 1996. "Finite-Sample Properties of Some Alternative GMM Estimators." *Journal of Business and Economic Statistics* 14 (3): 262–80.
- Hansen, Lars Peter, and Kenneth J. Singleton. 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica* 50 (5): 1269–86.
- Hayashi, Fumio, and Christopher A. Sims. 1983. "Nearly Efficient Estimation of Time Series Models with Predetermined, but Not Exogenous, Instruments." *Econometrica* 51 (3): 783–98.
- Henry, S. G. B., and A. R. Pagan. 2004. "The Econometrics of the New Keynesian Policy Model: Introduction." *Oxford Bulletin of Economics and Statistics* 66: 581–607.
- Henzel, Steffen, and Timo Wollmershäuser. 2008. "The New Keynesian Phillips Curve and the Role of Expectations: Evidence from the CESifo World Economic Survey." *Economic Modelling* 25 (5): 811–32.
- Hornstein, Andreas. 2007. "Evolving Inflation Dynamics and the New Keynesian Phillips Curve." *Federal Reserve Bank of Richmond Economic Quarterly* 93 (4): 317–39.
- Hume, David. 1987. "Of Money." In *Essays, Moral,*

- Political, and Literary*, edited by Eugene F. Miller, 281–94. Indianapolis: Liberty Fund, 1752.
- Imbs, Jean, Eric Jondeau, and Florian Pelgrin. 2011. “Sectoral Phillips Curves and the Aggregate Phillips Curve.” *Journal of Monetary Economics* 58 (4): 328–44.
- Jondeau, Eric, and Hervé Le Bihan. 2003. “ML vs. GMM Estimates of Hybrid Macroeconomic Models (with an Application to the ‘New Phillips Curve’).” Banque de France notes d’études et de recherche 103.
- Jondeau, Eric, and Hervé Le Bihan. 2005. “Testing for the New Keynesian Phillips Curve: Additional International Evidence.” *Economic Modelling* 22 (3): 521–50.
- Jondeau, Eric, and Hervé Le Bihan. 2008. “Examining Bias in Estimators of Linear Rational Expectations Models under Misspecification.” *Journal of Econometrics* 143 (2): 375–95.
- Kapetanios, George, Lynda Khalaf, and Massimiliano Marcellino. 2011. “Factor Based Identification-Robust Inference in IV Regressions.” Unpublished.
- Kapetanios, George, and Massimiliano Marcellino. 2010. “Factor-GMM Estimation with Large Sets of Possibly Weak Instruments.” *Computational Statistics and Data Analysis* 54 (11): 2655–75.
- Kiley, Michael T. 2007. “A Quantitative Comparison of Sticky-Price and Sticky-Information Models of Price Setting.” *Journal of Money, Credit, and Banking* 39 (1): 101–25.
- Kim, Chang-Jin, and Yunmi Kim. 2008. “Is the Backward-Looking Component Important in a New Keynesian Phillips Curve?” *Studies in Nonlinear Dynamics and Econometrics* 12 (3).
- Kim, Chang-Jin, Pym Manopimoke, and Charles Nelson. Forthcoming. “Trend Inflation and the Nature of Structural Breaks in the New Keynesian Phillips Curve.” *Journal of Money, Credit, and Banking*.
- Kleibergen, Frank. 2002. “Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression.” *Econometrica* 70 (5): 1781–1803.
- Kleibergen, Frank. 2005. “Testing Parameters in GMM without Assuming That They Are Identified.” *Econometrica* 73 (4): 1103–23.
- Kleibergen, Frank, and Sophocles Mavroeidis. 2009. “Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve.” *Journal of Business and Economic Statistics* 27 (3): 293–311.
- Kleibergen, Frank, and Sophocles Mavroeidis. Forthcoming. “Identification Issues in Limited-Information Bayesian Analysis of Structural Macroeconomic Models.” *Journal of Applied Econometrics*.
- Koop, Gary, and Luca Onorante. 2012. “Estimating Phillips Curves in Turbulent Times Using the ECB’s Survey of Professional Forecasters.” European Central Bank Working Paper 1422.
- Korenok, Oleg, Stanislav Radchenko, and Norman R. Swanson. 2010. “International Evidence on the Efficacy of New-Keynesian Models of Inflation Persistence.” *Journal of Applied Econometrics* 25 (1): 31–54.
- Kozicki, Sharon, and P. A. Tinsley. 2002. “Alternative Sources of the Lag Dynamics of Inflation.” Federal Reserve Bank of Kansas City Research Working Paper 02-12.
- Krause, Michael U., J. David López-Salido, and Thomas A. Lubik. 2008. “Inflation Dynamics with Search Frictions: A Structural Econometric Analysis.” *Journal of Monetary Economics* 55 (5): 892–916.
- Kuester, Keith, Gernot J. Müller, and Sarah Stöltgen. 2009. “Is the New Keynesian Phillips Curve Flat?” *Economics Letters* 103 (1): 39–41.
- Kurmann, Andre. 2005. “Quantifying the Uncertainty about the Fit of a New Keynesian Pricing Model.” *Journal of Monetary Economics* 52 (6): 1119–34.
- Kurmann, Andre. 2007. “VAR-Based Estimation of Euler Equations with an Application to New Keynesian Pricing.” *Journal of Economic Dynamics and Control* 31 (3): 767–96.
- Kurz, Mordecai. 2011. “A New Keynesian Model with Diverse Beliefs.” Unpublished.
- Lee, Jaejoon, and Charles R. Nelson. 2007. “Expectation Horizon and the Phillips Curve: The Solution to an Empirical Puzzle.” *Journal of Applied Econometrics* 22 (1): 161–78.
- Lindé, Jesper. 2005. “Estimating New-Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach.” *Journal of Monetary Economics* 52 (6): 1135–49.
- Lubik, Thomas A., and Frank Schorfheide. 2004. “Testing for Indeterminacy: An Application to U.S. Monetary Policy.” *American Economic Review* 94 (1): 190–217.
- Lucas, Robert E., Jr. 1976. “Econometric Policy Evaluation: A Critique.” *Carnegie-Rochester Conference Series on Public Policy* 1: 19–46.
- Ma, Adrian. 2002. “GMM Estimation of the New Phillips Curve.” *Economics Letters* 76 (3): 411–17.
- Magnusson, Leandro M., and Sophocles Mavroeidis. 2010. “Identification-Robust Minimum Distance Estimation of the New Keynesian Phillips Curve.” *Journal of Money, Credit, and Banking* 42 (2–3): 465–81.
- Magnusson, Leandro M., and Sophocles Mavroeidis. 2013. “Identification Using Stability Restrictions.” Unpublished.
- Mankiw, N. Gregory. 2001. “The Inexorable and Mysterious Tradeoff between Inflation and Unemployment.” *Economic Journal* 111 (471): C45–61.
- Mankiw, N. Gregory, and Ricardo Reis. 2011. “Imperfect Information and Aggregate Supply.” In *Handbook of Monetary Economics, Volume 3A*, edited by Benjamin M. Friedman and Michael Woodford, 183–229. Amsterdam and Boston: Elsevier, North-Holland.
- Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers. 2004. “Disagreement about Inflation Expectations.” In *NBER Macroeconomics Annual 2003*, edited by Mark Gertler and Kenneth Rogoff, 209–70. Cambridge, Mass. and London: MIT Press.
- Martins, Luis F., and Vasco J. Gabriel. 2009. “New

- Keynesian Phillips Curves and Potential Identification Failures: A Generalized Empirical Likelihood Analysis." *Journal of Macroeconomics* 31 (4): 561–71.
- Mavroeidis, Sophocles. 2004. "Weak Identification of Forward-Looking Models in Monetary Economics." *Oxford Bulletin of Economics and Statistics* 66 (Supplement): 609–35.
- 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): 421–48.
- Mavroeidis, Sophocles. 2010. "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence." *American Economic Review* 100 (1): 491–503.
- Mazumder, Sandeep. 2010. "The New Keynesian Phillips Curve and the Cyclicalities of Marginal Cost." *Journal of Macroeconomics* 32 (3): 747–65.
- Mazumder, Sandeep. 2011. "The Empirical Validity of the New Keynesian Phillips Curve Using Survey Forecasts of Inflation." *Economic Modelling* 28 (6): 2439–50.
- McAdam, Peter, and Alpo Willman. 2012. "Technology, Utilization and Inflation: What Drives the New Keynesian Phillips Curve?" University of Surrey Discussion Paper 09/12.
- McCallum, Bennett T. 1976. "Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates." *Econometrica* 44 (1): 43–52.
- Mikusheva, Anna. 2009. "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve: Comment." *Journal of Business and Economic Statistics* 27 (3): 322–23.
- Mikusheva, Anna. 2013. "Survey on Statistical Inferences in Weakly-Identified Instrumental Variable Models." *Applied Econometrics* 29: 117–31.
- Montiel Olea, José Luis, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business and Economic Statistics* 31 (3): 358–69.
- Moreira, Marcelo J. 2003. "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica* 71 (4): 1027–48.
- Nakamura, Emi, and Jon Steinsson. 2013. "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications." National Bureau of Economic Research Working Paper 18705.
- Nason, James M., and Gregor W. Smith. 2008a. "Identifying the New Keynesian Phillips Curve." *Journal of Applied Econometrics* 23 (5): 525–51.
- Nason, James M., and Gregor W. Smith. 2008b. "The New Keynesian Phillips Curve: Lessons from Single-Equation Econometric Estimation." *Federal Reserve Bank of Richmond Economic Quarterly* 94 (4): 361–95.
- Neiss, Katharine S., and Edward Nelson. 2005. "Inflation Dynamics, Marginal Cost, and the Output Gap: Evidence from Three Countries." *Journal of Money, Credit, and Banking* 37 (6): 1019–45.
- Newey, Whitney K., and Daniel McFadden. 1994. "Large Sample Estimation and Hypothesis Testing." In *Handbook of Econometrics, Volume 4*, edited by Robert F. Engle and Daniel L. McFadden, 2111–2245. Amsterdam; London and New York: Elsevier, North-Holland.
- Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–08.
- Newey, Whitney K., and Frank Windmeijer. 2009. "Generalized Method of Moments with Many Weak Moment Conditions." *Econometrica* 77 (3): 687–719.
- Nunes, Ricardo. 2010. "Inflation Dynamics: The Role of Expectations." *Journal of Money, Credit, and Banking* 42 (6): 1161–72.
- Nymoen, Ragnar, Anders Rygh Swensen, and Eivind Tveter. 2012. "Interpreting the Evidence for New Keynesian Models of Inflation Dynamics." *Journal of Macroeconomics* 34 (2): 253–63.
- Ólafsson, Thorvarður Tjörvi. 2006. "The New Keynesian Phillips Curve: In Search of Improvements and Adaptation to the Open Economy." Central Bank of Iceland Working Paper 31.
- Paloviita, Maritta. 2006. "Inflation Dynamics in the Euro Area and the Role of Expectations." *Empirical Economics* 31 (4): 847–60.
- Paloviita, Maritta. 2008. "Comparing Alternative Phillips Curve Specifications: European Results with Survey-Based Expectations." *Applied Economics* 40 (16–18): 2259–70.
- Paloviita, Maritta, and David Mayes. 2005. "The Use of Real-Time Information in Phillips-Curve Relationships for the Euro Area." *North American Journal of Economics and Finance* 16 (3): 415–34.
- Phelps, E. S. 1967. "Phillips Curves, Expectations of Inflation and Optimal Unemployment over Time." *Economica N.S.* 34 (135): 254–81.
- Phillips, A. W. 1958. "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957." *Economica N.S.* 25 (100): 283–99.
- Preston, Bruce. 2005. "Learning about Monetary Policy Rules When Long-Horizon Expectations Matter." *International Journal of Central Banking* 1 (2): 81–126.
- Ravenna, Federico, and Carl E. Walsh. 2006. "Optimal Monetary Policy with the Cost Channel." *Journal of Monetary Economics* 53 (2): 199–216.
- Ravenna, Federico, and Carl E. Walsh. 2008. "Vacancies, Unemployment, and the Phillips Curve." *European Economic Review* 52 (8): 1494–1521.
- Roberts, John M. 1995. "New Keynesian Economics and the Phillips Curve." *Journal of Money, Credit, and Banking* 27 (4): 975–84.
- Roberts, John M. 1997. "Is Inflation Sticky?" *Journal of Monetary Economics* 39 (2): 173–96.
- Roberts, John M. 2005. "How Well Does the New Keynesian Sticky-Price Model Fit the Data?" *B.E. Journal of Macroeconomics: Contributions to Macroeconomics* 5 (1).
- Rotemberg, Julio J. 1982. "Sticky Prices in the United States." *Journal of Political Economy* 90 (6): 1187–1211.
- Rudd, Jeremy, and Karl Whelan. 2005. "New Tests of



- the New-Keynesian Phillips Curve." *Journal of Monetary Economics* 52 (6): 1167–81.
- Rudd, Jeremy, and Karl Whelan. 2006. "Can Rational Expectations Sticky-Price Models Explain Inflation Dynamics?" *American Economic Review* 96 (1): 303–20.
- Rudd, Jeremy, and Karl Whelan. 2007. "Modeling Inflation Dynamics: A Critical Review of Recent Research." *Journal of Money, Credit, and Banking* 39: 155–70.
- Rudebusch, Glenn D. 2002. "Assessing Nominal Income Rules for Monetary Policy with Model and Data Uncertainty." *Economic Journal* 112 (479): 402–32.
- Rumler, Fabio. 2007. "Estimates of the Open Economy New Keynesian Phillips Curve for Euro Area Countries." *Open Economies Review* 18 (4): 427–51.
- Sahuc, Jean-Guillaume. 2006. "Partial Indexation, Trend Inflation, and the Hybrid Phillips Curve." *Economics Letters* 90 (1): 42–50.
- Samuelson, P. A., and R. M. Solow. 1960. "Analytical Aspects of Anti-inflation Policy." *American Economic Review* 50 (2): 177–94.
- Sargent, Thomas J., and Neil Wallace. 1975. "Rational Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule." *Journal of Political Economy* 83 (2): 241–54.
- Sbordone, Argia M. 2002. "Prices and Unit Labor Costs: A New Test of Price Stickiness." *Journal of Monetary Economics* 49 (2): 265–92.
- Sbordone, Argia M. 2005. "Do Expected Future Marginal Costs Drive Inflation Dynamics?" *Journal of Monetary Economics* 52 (6): 1183–97.
- Sbordone, Argia M. 2006. "U.S. Wage and Price Dynamics: A Limited-Information Approach." *International Journal of Central Banking* 2 (3): 155–91.
- Schorfheide, Frank. 2011. "Estimation and Evaluation of DSGE Models: Progress and Challenges." National Bureau of Economic Research Working Paper 16781.
- Shapiro, Adam Hale. 2008. "Estimating the New Keynesian Phillips Curve: A Vertical Production Chain Approach." *Journal of Money, Credit, and Banking* 40 (4): 627–66.
- Sheedy, Kevin D. 2010. "Intrinsic Inflation Persistence." *Journal of Monetary Economics* 57 (8): 1049–61.
- Smith, Gregor W. 2009. "Pooling Forecasts in Linear Rational Expectations Models." *Journal of Economic Dynamics and Control* 33 (11): 1858–66.
- Sowell, Fallaw. 1996. "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework." *Econometrica* 64 (5): 1085–1107.
- Stock, James H., and Mark W. Watson. 2007. "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit, and Banking* 39 (Supplement 1): 3–33.
- Stock, James H., and Mark W. Watson. 2010. "Modeling Inflation After the Crisis." National Bureau of Economic Research Working Paper 16488.
- Stock, James H., and Jonathan H. Wright. 2000. "GMM with Weak Identification." *Econometrica* 68 (5): 1055–96.
- Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business and Economic Statistics* 20 (4): 518–29.
- Stock, James H., and Motohiro Yogo. 2002. "Testing for Weak Instruments in Linear IV Regression." National Bureau of Economic Research Technical Working Paper 284.
- Taylor, John B. 1979. "Staggered Wage Setting in a Macro Model." *American Economic Review* 69 (2): 108–13.
- Taylor, John B. 1980. "Aggregate Dynamics and Staggered Contracts." *Journal of Political Economy* 88 (1): 1–23.
- Thomas, Lloyd B., Jr. 1999. "Survey Measures of Expected U.S. Inflation." *Journal of Economic Perspectives* 13 (4): 125–44.
- Tillmann, Peter. 2008. "Do Interest Rates Drive Inflation Dynamics? An Analysis of the Cost Channel of Monetary Transmission." *Journal of Economic Dynamics and Control* 32 (9): 2723–44.
- Tsoukis, Christopher, George Kapetanios, and Joseph Pearlman. 2011. "Elusive Persistence: Wage and Price Rigidities, the New Keynesian Phillips Curve and Inflation Dynamics." *Journal of Economic Surveys* 25 (4): 737–68.
- White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4): 817–38.
- Woodford, Mike. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton and Oxford: Princeton University Press.
- Wright, Jonathan H. 2009. "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve: Comment." *Journal of Business and Economic Statistics* 27 (3): 323–26.
- Yao, Fang. 2011. "Monetary Policy, Trend Inflation and Inflation Persistence." Sonderforschungsbereich 649 Discussion Paper 2011-008.
- Yun, Tack. 1996. "Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles." *Journal of Monetary Economics* 37 (2): 345–70.
- Zhang, Chengsi, and Joel Clovis. 2010. "The New Keynesian Phillips Curve of Rational Expectations: A Serial Correlation Extension." *Journal of Applied Economics* 13 (1): 159–79.
- Zhang, Chengsi, Denise R. Osborn, and Dong Heon Kim. 2008. "The New Keynesian Phillips Curve: From Sticky Inflation to Sticky Prices." *Journal of Money, Credit, and Banking* 40 (4): 667–99.
- Zhang, Chengsi, Denise R. Osborn, and Dong Heon Kim. 2009. "Observed Inflation Forecasts and the New Keynesian Phillips Curve." *Oxford Bulletin of Economics and Statistics* 71 (3): 375–98.